



COLEGIO DE POSTGRADUADOS

**INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN
EN CIENCIAS AGRÍCOLAS**

CAMPUS MONTECILLO

**POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA**

Uso de los Modelos Credit Scoring en Microfinanzas

Arturo Escalona Cortés

T E S I S

**PRESENTADA COMO REQUISITO PARCIAL PARA
OBTENER EL GRADO DE:**

MAESTRO EN CIENCIAS

**MONTECILLO, TEXCOCO, EDO. DE MÉXICO
2011**

La presente tesis titulada: **Uso de los Modelos Credit Scoring en Microfinanzas**, realizada por el alumno: **Arturo Escalona Cortés**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA ESTADÍSTICA

CONSEJO PARTICULAR

CONSEJERO _____
Dr. Humberto Vaquera Huerta

ASESOR _____
Dr. Saturnino Mora Flores

ASESOR _____
Dr. David H. del Valle Panigua

ASESOR _____
Dr. Javier Suarez Espinosa

Montecillo, Texcoco, Estado de México, Enero 2011.

Uso de los Modelos Credit Scoring en Microfinanzas

Arturo Escalona Cortés

Resumen

En el presente trabajo se presenta un modelo *credit scoring* en microfinanzas el cual permita calificar los solicitantes de un microcrédito. El modelo propuesto se basa en el modelo de regresión logística. La evaluación del modelo incluye el ajuste medido por la estadística Hosmer-Lemeshow, poder predictivo medido por la R^2 , poder discriminatorio medido por la curva ROC y área bajo la curva ROC. Además, se determinó un punto de corte para validar el modelo. La información de clientes durante el período de mayo de 2008 a abril de 2009 de la empresa microfinanciera **MásKapital** se utilizó como base para desarrollar y calibrar el modelo. Se obtuvo un buen ajuste de este, reflejado en los resultados de validación que, presentan una clasificación total correcta del 60%.

Palabras clave: riesgo crediticio, regresión logística, microcréditos, sensibilidad, especificidad.

Using Credit Scoring Models to Microfinance

Arturo Escalona Cortés

Abstract

In this paper we develop a *credit scoring* model for microfinance which can qualify applicants for a microcredit. The proposed model is based on a logistic regression model. The evaluation model includes adjustment measured by the Hosmer-Lemeshow statistic, predictive power as measured by the R^2 , discriminatory power as measured by the ROC curve and area under the ROC curve. In addition, we determined a cutoff point to validate the model. The information provided by the Microfinance company **MásKapital** from may 2008 to april 2009 was used to develop and calibrate the model. The model fits well and the results show a right total clasification of 60 %.

Key words: Credit risk, logistic regression, microcredit, sensitivity, specificity.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado durante la realización de mis estudios de postgrado.

Al Colegio de Postgraduados, por haberme brindado la oportunidad de seguir mi formación académica en sus aulas.

A los integrantes de mi Consejo Particular:

Dr. Humberto Vaquera Huerta, por su excelente dirección, sin la cual me hubiera sido más difícil la culminación del presente trabajo.

Dr. Javier Suarez Espinosa, por sus observaciones y ayuda desinteresada en la realización del trabajo.

Dr. David H. del Valle Paniagua, por sus comentarios y revisión detallada del trabajo.

Dr. Saturnino Mora Flores, por sus comentarios tan acertados y revisión del trabajo.

A la empresa MásKapital por permitirme utilizar información real y por el apoyo brindado en la elaboración de este trabajo.

A mis profesores, compañeros de clases y todos aquellos que de alguna u otra manera fueron copartícipes de esta tarea.

DEDICATORIA

A mis **padres:** Elmer Escalona Barrios y María del Rosario Cortés León, por sus consejos y exhortación desde siempre, para continuar mi formación académica.

A mi **esposa:** Cynthia y mis **hijos:** Arturo y Sebastián, por su paciencia y cariño.

A mis **hermanos:** Omar, Elmer, Yuridia y Uriel.

Índice

1. Introducción	1
2. Objetivos	3
3. Las Microfinanzas	4
3.1. Microfinanzas en el Mundo	5
3.2. Microfinanzas en México	6
4. Los Modelos <i>Credit Scoring</i>	11
4.1. <i>Credit Scoring</i>	11
4.2. <i>Credit Scoring</i> en Microfinanzas	12
5. Metodologías aplicadas en los modelos <i>Credit Scoring</i>	15
5.1. Análisis discriminante	15
5.2. Redes Neuronales	16
5.3. Scoring basado en el juicio	16
5.4. Regresión Logística	16
6. <i>Credit Scoring</i> usando Regresión Logística: Caso MásKapital	20
6.1. Antecedentes generales	20

Índice

6.2. Conformación de la Base de Datos	23
6.3. Definición de la variable respuesta y variables explicatorias	24
6.4. Selección de variables	28
6.5. Ajuste del modelo	29
6.6. Poder predictivo	30
6.7. Poder discriminatorio	31
6.8. Elección del Punto de corte	33
6.9. Validación	34
7. Conclusiones	36
Referencias	37
Apéndices	41
Apéndice A:	41

Índice de tablas

6.1. Variables analizadas	25
6.2. Variables categóricas	27
6.3. Type III Analysis of Effects	28
6.4. Summary of Backward Elimination	29
6.5. Partition for the Hosmer and Lemeshow Test	30
6.6. Comparación del valor observado con el resultado del modelo	31
6.7. Regla general del poder discriminatorio	33
6.8. Resultados para validación	35

Índice de figuras

3.1. Activos, préstamos y cantidad de deudores de las IMF en una muestra de países (Delfiner M, Pailhé C y Silvana Perón; 2006)	7
3.2. Diez principales Instituciones Microfinancieras en México, (Benchmarking de las microfinanzas en México:Un informe del sector 2008, Pro Desarrollo AC)	8
3.3. Número de sucursales de instituciones microfinancieras por estado (Benchmarking de las microfinanzas en México:Un informe del sector 2008, Pro Desarrollo AC)	10
4.1. Credit Scoring en Microfinanzas	14
6.1. Calificación GIRAFE -MásKapital México- Noviembre 2009. Planet Rating	21
6.2. Curva ROC	32
6.3. Sensibilidad, especificidad y total correcto	34

Capítulo 1

Introducción

En sus actividades cotidianas, empresas del área financiera y comercial deben tomar decisiones vinculadas con otorgar/rechazar una solicitud de préstamo, aumentar/disminuir la línea de crédito y definir las tasas de interés a aplicar a los clientes, implica una gran volumen de trabajo, por lo que es necesario el contar con mecanismos automatizados de apoyo.

En la década de 1960, en los Estados Unidos comenzaron a desarrollarse y aplicarse las técnicas de *Credit Scoring*. Este término es utilizado para describir formalmente los métodos estadísticos utilizados para la clasificación de los solicitantes de crédito en las clases de riesgo bueno y malo. Tales métodos, han tomado mucha importancia con el espectacular crecimiento del crédito al consumo en los últimos años [Hand y Henley (1997)]

El modelo *Credit Scoring* utiliza medidas cuantitativas de los resultados y características de los préstamos pasados para predecir el rendimiento futuro de los préstamos con características similares. Para las organizaciones de crédito en los países ricos el *credit scoring* ha sido una de las herramientas más importantes en el logro de una mayor eficiencia. La información usada en estos modelos, en países ricos, se basa en la proporcionada por el historial crediticio (buró de crédito) y en la experiencia laboral (salario) del cliente. En microfinanzas, sin embargo, la mayoría de los acreditados son pobres y trabajan por cuenta propia [Schreiner (2000)].

El *Credit Scoring* ha sido aplicado satisfactoriamente en el campo de préstamos hipotecarios y tarjetas de crédito en la banca convencional. En 1993 inició la aplicación de los modelos mencionados en las instituciones de microfinanzas, y desde entonces se han publicado artículos que tratan sobre la adaptación del scoring estadístico al microcrédito (y viceversa).

En el presente trabajo se implementa un modelo *credit scoring* en el sector micro-

1. Introducción

financiero de México, tomando el caso de la empresa microfinanciera *MásKapital*, cuya área de cobertura se ubica en el centro y sur del país. Para el desarrollo de este modelo se usó la herramienta estadística de regresión logística.

En el capítulo III se presenta información general acerca de las microfinanzas en México y en el mundo. En el capítulo IV se describen de manera general los modelos *credit scoring* y su aplicación en microfinanzas. En el capítulo V se mencionan los diferentes enfoques con los que se han desarrollado estos modelos, y se detalla el enfoque de regresión logística. El capítulo VI presenta un ejemplo con datos reales y los resultados obtenidos con el paquete estadístico SAS.

Capítulo 2

Objetivos

- Desarrollar e implementar un modelo *Credit Scoring* adecuado al sector de microfinanzas, basado en el modelo de regresión logística.
- Evaluar el modelo *Credit Scoring* propuesto utilizando una base de datos real: Caso Maskapital.

Capítulo 3

Las Microfinanzas

Las microfinanzas se refieren a los servicios financieros otorgados a personas de bajos ingresos de poblaciones que normalmente no tienen acceso al ahorro, crédito, información financiera, y los seguros. Históricamente, a los pobres se les ha negado el acceso al crédito, y por lo tanto no pueden darse el lujo de gastar sus ingresos y consumo en la actividad empresarial.

Las Microfinanzas tienen como objetivo fundamental impulsar la creación y el desarrollo de pequeñas actividades productivas. El tema de microfinanzas toma relevancia en 1974 cuando el economista bengalí Muhammad Yunus inicia con su proyecto de prestar dinero con fondos propios a gente pobre para microemprendimientos sin exigirles ninguna de las garantías usuales de la banca tradicional. Actualmente el proyecto de Yunus ha crecido al punto de tener ya el Grameen Bank (o Banco de los Pobres), institución financiera que se aboca a la prestación de servicios microfinancieros en Bangladesh. A la fecha Yunus ha ganado el Premio Príncipe de Asturias de la Concordia en 1998, el Premio Internacional Simón Bolívar en 1996, el Premio Nobel de la Paz en 2006, y la medalla de la Libertad en 2009.

En términos generales se concibe a las microfinanzas como un tipo de financiamiento a pequeña escala, pero con la característica de que es para familias pobres. Así, las microfinanzas es el desarrollo de las finanzas al servicio de una población excluida del sistema tradicional capitalista [[Hernández Romero \(2005\)](#)].

Durante mucho tiempo se había pensado que las personas de escasos recursos no podían, por su condición, obtener servicios financieros de calidad, porque ha persistido el mito de que la gente pobre representa altos riesgos crediticios pues son personas que no pueden pagar o son incapaces de ahorrar, debido a que tienen demasiadas carencias para hacerlo [[Conde Bonfil \(2000\)](#), [Mansell Carstens \(1995\)](#)].

Sin embargo, las familias de bajos ingresos quieren y pueden ahorrar, y lo hacen cuan-

3.1. Microfinanzas en el Mundo

do tienen a su alcance instituciones e instrumentos apropiados a sus peculiaridades. Tales innovaciones, que se llevan a cabo en las microfinanzas, han contribuido a generar oportunidades de desarrollo y bienestar a las familias que participan en los proyectos de microfinanzas.

En este trabajo nos referimos específicamente al instrumento de financiación denominado *microcrédito*, del tipo Grupos Solidarios.

3.1. Microfinanzas en el Mundo

Muchos grupos de microfinanzas han estado operando en varios países en años recientes, por ejemplo, los "fondos chit" (India), los "tontines" (África Occidental), "susus" (Ghana), "pasanaku" (Bolivia), etc. Además, varias entidades de influencia global de ahorro formal y de crédito están incursionado en el mundo de las microfinanzas.

En Europa durante la década de 1990 se formaron diversas instituciones de crédito, organizadas por personas pobres, tanto de las zonas rurales como urbanas. Estas instituciones se les llamo Uniones de Crédito, Banco Popular, etc. El objetivo principal de estas instituciones fue el facilitar el acceso al crédito para los pobres frecuentemente ignorados por las grandes instituciones financieras y bancos.

Algunas de las principales instituciones microfinancieras que tienen presencia en varios países se enumeran a continuación:

- **ACCION International:** Creada por un estudiante de derecho de América Latina para ayudar a los pobres que residen en las zonas rurales y urbanas de los países de América Latina. Hoy en día es una de las instituciones microfinancieras más importantes del mundo. Es una red de socios de préstamos que comprende no sólo América Latina sino también EE.UU. y África.
- **Banco SEWA:** En 1973, la Asociación de Trabajadoras Autónomas (SEWA) de Gujarat (India) forman un banco, nombrado como el Banco Cooperativo Mahila SEWA. Hoy en día el número de clientes activos del Banco SEWA lo constituyen más de 30.000.
- **El Grameen Bank (Bangladesh):** Fundado por el Dr. Muhammad Yunus en 1983 (Premio Nobel de la Paz (2006)). Este banco está atendiendo casi 400,000 personas pobres de Bangladesh.
- **Opportunity International:** Ofrece productos financieros a más de dos millones de personas en el mundo en desarrollo. Tiene clientes en más de 20 países.

3.2. Microfinanzas en México

- **Finca Internacional:** Fundada en el año de 1984 por John Hatch, es una Institución con presencia en 22 países en América Latina, África, Asia y Europa del Este, su sede está en la Ciudad de Washington, E.U.A.
- **ProCredit:** Es una red mundial de Bancos y Financieras, se encuentra en 21 países (Latinoamérica, Europa del Este y África).

La figura (3.1) generada por [Delfiner *et al.* (2006)] sobre los activos, préstamos y cantidad de deudores de las instituciones microfinancieras por países en el 2006, nos da una idea de la participación a nivel mundial de las empresas microfinancieras.

A nivel del continente americano [Renso Martínez María (2010)] publica una lista de las mejores empresas microfinancieras de Latinoamérica basado en la información del *Microfinance Information exchange, Inc.* (MIX), el mismo autor de acuerdo con su análisis considera que las cinco mejores empresas microfinancieras son:

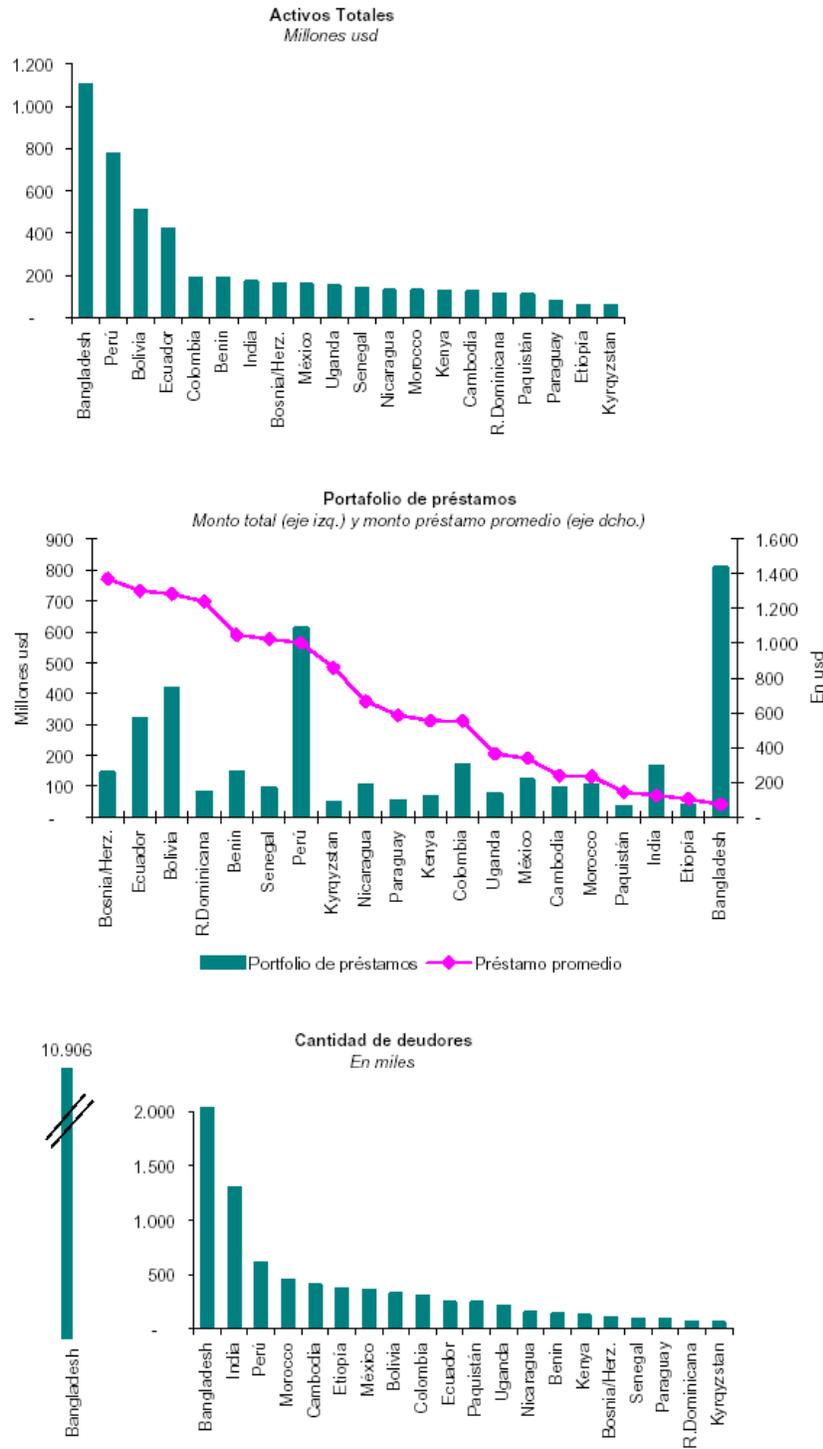
- **CrediAmigo**, Programa de microcrédito productivo orientado del Banco do Nordeste de Brasil, es la institución con mayor alcance de su mercado, además de ser un referente importante en la utilización de metodologías de grupos solidarios.
- **Fundación para el Desarrollo Integral Espoir de Ecuador**, Aplica de manera efectiva la metodología de bancos comunales, logrando por intermedio de ella un destacado control de la cartera en riesgo mayor a 30 días.
- **Pro Mujer Bolivia**, se destaca por su continuo compromiso con la transparencia de la información.
- **Banco ADOPEM** de la República Dominicana, logró incrementar sus operaciones tanto de crédito como de depósitos en un número de depositantes superior al número de prestatarios.
- **Banco Caja Social de Colombia**, este pone énfasis en el crédito de consumo.

3.2. Microfinanzas en México

En México habitan millones de personas que sufren graves carencias de empleo, ingresos, vivienda, servicios públicos, alimentación, oportunidades de educación, etc. En este contexto, señalar que también carecen de los servicios bancarios mínimos parece irrelevante y alejado de las prioridades y demandas más urgentes.

3.2. Microfinanzas en México

Figura 3.1: Activos, préstamos y cantidad de deudores de las IMF en una muestra de países (Delfiner M, Pailhé C y Silvana Perón; 2006)



Sin embargo, la experiencia de proyectos alternativos cuyo principal objetivo ha sido

3.2. Microfinanzas en México

el financiamiento a personas en situación de pobreza extrema, ha demostrado que desplegar esfuerzos contra la desigualdad y la pobreza, es una tarea prioritaria y posible, que exige ser atendida, no sólo por razones económicas y políticas, sino éticas.

Ante el terrible panorama de pobreza y desigualdad social, las microfinanzas deben considerarse una solución, una alternativa propia, creativa, que pueda reconocer las potencialidades reales y hacer una diferencia.

Uno de los instrumentos más importantes de la política social es el acceso a los servicios financieros. Una inmensa mayoría de la población del país no tiene acceso a estos servicios por no contar con un aval. Se trata, desde luego, de un círculo vicioso en el que muchas veces el individuo no puede comenzar a ser productivo porque no puede pedir prestado para impulsar un pequeño esfuerzo empresarial porque no tiene pertenencias ni aval para garantizar el préstamo. Es precisamente en este punto donde las microfinanzas inciden para romper el círculo vicioso.

Las actividades microfinancieras no son un fin en sí mismas, sino que son un instrumento que apoya otras actividades o metas básicas de las organizaciones; se centra en la mayoría de los casos en fortalecer la economía y oportunidades de desarrollo de las personas que viven en situación de pobreza.

Al igual que en otros países del mundo, en México, las microfinanzas están adquiriendo un lugar importante en el financiamiento hacia las clases más desprotegidas, ello se nota en la gran cantidad de actores involucrados en diversos proyectos, programas y eventos dirigidos hacia el microfinanciamiento.

La figura (3.2) muestra las diez principales instituciones microfinancieras en México.

Figura 3.2: Diez principales Instituciones Microfinancieras en México, (Benchmarking de las microfinanzas en México: Un informe del sector 2008, Pro Desarrollo AC)

Instituciones con mayor alcance en México

Posición	Nombre de la IMF	Composición general de cartera*	Número de préstamos vigentes	Cartera bruta de préstamos (Mill. MXN)
1	Compartamos Banco	MIC	869,153	4,120.9
2	Financiera Independencia	CNS	833,902	3,156.1
3	Caja Popular Mexicana	CNS	778,808	12,617.7
4	Caja Libertad	CNS	343,706	6,352.6
5	CAME	Solo MIC	105,778	204.3
6	FINCA México	Solo MIC	87,428	241.7
7	FinComún	MIC	57,535	471.4
8	Apoyo Económico	CNS	23,347	189.8
9	Credi-Capital	Solo MIC	22,745	88.9
10	Soluciones Financieras	Solo MIC	22,095	66.2
	Otras 34 IMF's	•	282,444	1,741.6
Total 2007 (44 IMF's)			3,426,941	29,251.1

3.2. Microfinanzas en México

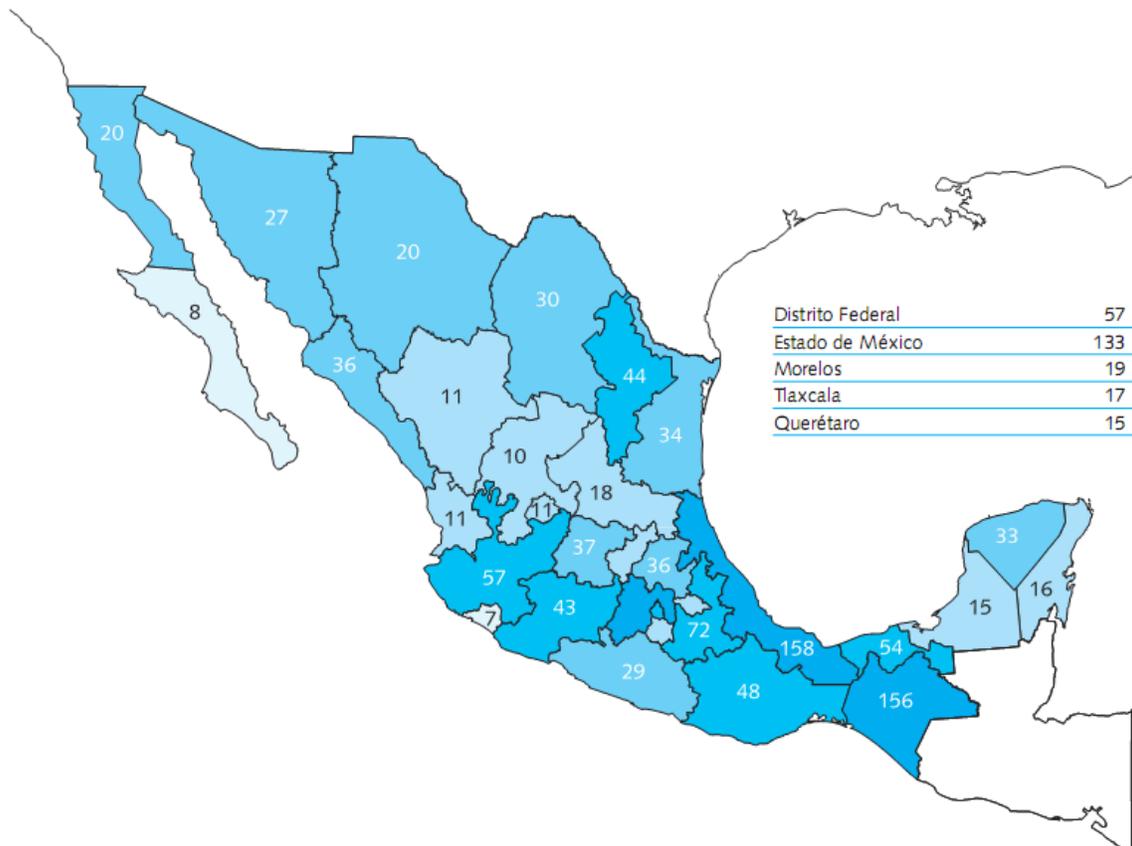
En el informe de Prodesarrollo [[ProDesarrollo Finanzas y Microempresas \(2008\)](#)] se menciona que durante el año 2008 aumentó de manera exponencial el número de instituciones dedicadas a los microcréditos y a los créditos personales. La figura jurídica Sociedad Financiera de Objeto múltiple (SOFOM) ha permitido la rápida creación de nuevas empresas.

De acuerdo con información de la Comisión Nacional para la Protección y Defensa de los Usuarios de Servicios Financieros (CONDUCEF), a dos años de la creación de esta figura jurídica existen seiscientos ochenta y dos SOFOMES, de las cuales alrededor de doscientas cincuenta están relacionadas con los microcréditos y los créditos personales a la población de escasos recursos. La entrada de nuevos actores a contribuido a lograr una mayor cobertura en los estados y a llegar a lugares que estaban casi desatendidos como Baja California Sur.

ProDesarrollo estima que el 80% del total de las sucursales existentes en el país está representado por 79 instituciones microfinancieras. La figura (3.3) muestra el número de sucursales de estas instituciones por estado; donde se observa una mayor concentración en los estados del centro sur y poca cobertura en los estados del norte.

3.2. Microfinanzas en México

Figura 3.3: Número de sucursales de instituciones microfinancieras por estado (Benchmarking de las microfinanzas en México: Un informe del sector 2008, Pro Desarrollo AC)



Capítulo 4

Los Modelos *Credit Scoring*

4.1. *Credit Scoring*

Un prestamista puede enfrentarse a dos tipos de concesión de crédito. Primero, si concede un crédito a un solicitante nuevo y, segundo, como tratar con clientes existentes. El conjunto de modelos de decisión que ayudan en el primer tipo de concesión es llamado *Credit Scoring*; mientras que, las técnicas que ayudan en el segundo se denominan *Behavioral Scoring* [Lyn C. Thomas (2002)]. En este trabajo se trata, particularmente, sobre los modelos *Credit Scoring*.

En una primera aproximación a los mismos, se les puede definir como "métodos estadísticos utilizados para clasificar a los solicitantes de crédito" [Hand y Henley (1997)]. Aunque originalmente se basaban en técnicas estadísticas (en particular, el análisis discriminante), en la actualidad también están basados en técnicas matemáticas, econometricas y de inteligencia artificial.

Credit Scoring es esencialmente una manera de identificar diferentes grupos en una población. La primera propuesta para resolver este problema fue introducida en estadística por Fisher (1936) al buscar diferenciar dos variedades de plantas por las medidas físicas de estas. En 1941, Durand (1941) fue el primero en reconocer que se pueden usar las mismas técnicas para optimizar la discriminación entre préstamos buenos y malos. En esta discriminación, es indispensable el uso de conocimiento sobre el comportamiento y características de préstamos en el pasado para pronosticar el desempeño de préstamos en el futuro. Tal información está contenida en las solicitudes de crédito y en fuentes internas y externas de información.

Así, cuando un asesor de crédito valora el riesgo comparando mentalmente una solicitud de crédito en el presente con la experiencia que este mismo asesor ha acumulado con otros clientes con solicitudes parecidas, está aplicando scoring, aunque sea un

4.2. *Credit Scoring* en Microfinanzas

scoring implícito y subjetivo. Por eso, aunque el nombre scoring puede sonar nuevo es una práctica antigua.

Por otra parte, el scoring estadístico trata el uso de conocimiento cuantitativo acerca del desempeño y características de préstamos pasados registrados en una base de datos electrónica para pronósticar el desempeño de préstamos futuros.

Las principales ventajas que presenta el scoring estadístico cuando se compara con el scoring implícito o subjetivo son: cuantifica el riesgo como una probabilidad, mejora la administración del riesgo, trata exactamente igual a las solicitudes idénticas, puede comunicarse el proceso exacto usado para pronosticar el riesgo, considera una amplia gama de factores simultáneamente, puede probarse antes de usarlo y revela las relaciones entre el riesgo y las características del prestatario, préstamo y prestamista [Schreiner (2002)].

El resultado de la evaluación se refleja en la asignación de alguna medida que permita comparar y ordenar a los evaluados en función de su riesgo, a la vez que cuantifica ese riesgo. Por lo general los modelos *credit scoring* le asignan al evaluado un puntaje o *score*, o una calificación, clasificación o *rating*. Algunos métodos los asignan a grupos en donde cada grupo tiene un perfil de riesgo distinto; sin embargo, en la práctica esto equivale a una calificación. A su vez, estos ordenamientos de los deudores permiten obtener estimaciones más concretas del riesgo; en general se busca obtener alguna estimación de la probabilidad de incumplimiento del deudor asociada a su *score*, *rating* o calificación [Gutiérrez (2007)].

Si bien la escala del *score* puede oscilar entre 0 y 1.000, la misma es arbitraria y depende en última instancia de la construcción del modelo. Generalmente los modelos presentan una relación inversa entre el *score* y el riesgo; es decir, a medida que mejora el *score* o calificación la caída marginal en el riesgo es cada vez menor.

La traducción, usada comúnmente, para el término en inglés *credit scoring* es, calificación de crédito. En este texto se utiliza el término en inglés debido a su amplio uso en la literatura.

4.2. *Credit Scoring* en Microfinanzas

La banca tradicionalmente ha contado con modelos adecuados de *credit scoring* para analizar el riesgo de incumplimiento, pero esto no ha sido así en las instituciones de microfinanzas, donde historiales de crédito poco desarrollados y faltos de información impiden predecir la probabilidad del no pago de deudas mediante el uso de los sistemas tradicionales de *credit scoring* desarrollados para la banca.

4.2. *Credit Scoring* en Microfinanzas

El primer modelo de *credit scoring* para microfinanzas que se conoce fue desarrollado por [Vigano \(1993\)](#) para una institución de microfinanzas de Burkina Faso. Sobre una muestra de 100 microcréditos, y contando con 53 variables iniciales, Vigano utilizó el análisis discriminante para la elaboración del modelo.

Los modelos de *Credit Scoring* en microfinanzas publicados hasta la actualidad generalmente están diseñados en las regiones de América Latina y del Sur de África, como lo evidencian [Vogelgesang \(2003\)](#) y [Kleimeier y Dinh \(2007\)](#) en sus investigaciones.

Experimentos en Bolivia y Colombia sugieren que la calificación estadística puede en efecto mejorar la valoración del riesgo y, por tanto, disminuir los costos de las microfinanzas [[Schreiner \(2000\)](#)].

Entre otros trabajos se encuentran los de [Miller y Rojas \(2005\)](#), que formularon un *credit scoring* de Pymes de México y Colombia respectivamente, mientras que [Milena et al. \(2005\)](#) hicieron lo mismo para microfinancieras de Nicaragua. En Mali, [Diallo \(2006\)](#) empleó la regresión logística para una muestra de 269 créditos de una entidad microbancaria del país. Diallo solo obtuvo 5 variables significativas en su modelo.

Las limitaciones e inconvenientes en la elaboración de sistemas de calificación estadística del cliente potencial plantean dificultades a la hora de construirlos, hecho que se refleja en la escasa literatura existente hasta la fecha sobre modelos de *credit scoring* para microfinanzas. Hay autores que discuten sobre la conveniencia o no y sobre la posibilidad de éxito de los modelos *credit scoring* para las microfinanzas. Ese es el caso de [Dennis \(1995\)](#), [Kulkosky \(1996\)](#) y [Schreiner \(2002\)](#), quienes aportan las limitaciones, ventajas e inconvenientes de los modelos de evaluación del riesgo de crédito en las microfinanzas.

Si lo que se pretende es analizar el comportamiento de pago de un cliente de microcrédito, se requiere una base de datos amplia que recoja la historia de los préstamos que han resultado impagados en algunas de sus cuotas desde su concesión hasta su fecha de vencimiento. Por otro lado, las instituciones de microfinanzas normalmente no incluyen en sus bases de datos la información referente a los clientes a los que se les denegó el crédito porque, en su momento, no pasaron la evaluación estándar del analista. En consecuencia, solo se podrá contar con la información de aquellas solicitudes de crédito que están en la fase de aprobación.

Las microfinanzas requieren la intervención de un asesor de crédito en la recopilación y captación de información para el historial de crédito. Sin embargo, este proceso podría estar muy influenciado por la opinión subjetiva de dicho funcionario.

En el proceso de crédito en microfinanzas los asesores de crédito visitan a cada uno de los integrantes del grupo, para verificar el domicilio y el funcionamiento del negocio al cual se dedica cada prestatario. El proceso dura generalmente una semana. Los datos son registrados en el formulario de la solicitud de crédito. Y una vez que se

4.2. Credit Scoring en Microfinanzas

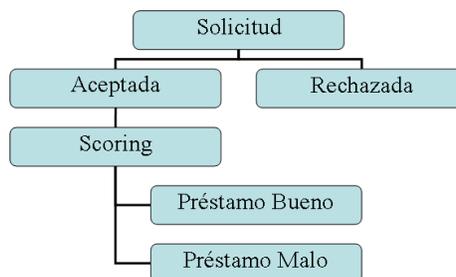
cumple con las políticas de crédito vigentes, el asesor de crédito juzga si el solicitante podrá pagar según lo prometido, haciendo uso del conocimiento personal del cliente y la información que el Grupo Solidario proporcione sobre este.

El incumplimiento de pago debe definirse con cautela, por lo que es necesario identificar todo atraso que conlleve un costo para la organización. Para ello, se ha de verificar que el atraso considerado ha de suponer un incremento en el costo, para la entidad, mayor al habitual. Generalmente, estos incrementos suelen darse en términos de costos administrativos debido al incremento monetario que supone realizar un seguimiento y gestionar el pago de un crédito cuyo reembolso mantiene un retraso considerable.

En microfinanzas, *credit scoring* no reemplaza el tradicional proceso de crédito. No puede reemplazar el conocimiento personal del cliente, que tienen los asesores de crédito y los grupos solidarios [Schreiner (2000)]. Complementa, pero no sustituye, las tecnologías usadas actualmente. Sin embargo, puede ser introducido como una herramienta de refinamiento del proceso de crédito, para combinar mejores prácticas estadísticas y humanas.

En microfinanzas el scoring estadístico ayuda al asesor de crédito a tomar decisiones sobre casos que, sin el scoring, serían aprobados. Esto implica que, el scoring estadístico no da la aprobación a los solicitantes que, sin él, son rechazados [Schreiner y Dellien (2005)].

Figura 4.1: Credit Scoring en Microfinanzas



En el campo, los analistas de crédito deben seguir libres de usar su intuición para prever las evidencias de riesgos que observen, pero de regreso en la oficina, para tomar una decisión, deben auxiliarse de la calificación estadística.

Este trabajo muestra los resultados de un modelo scoring para microfinanzas, desarrollado con datos de préstamos en el estado de Oaxaca, México.

Capítulo 5

Metodologías aplicadas en los modelos *Credit Scoring*

Los principales métodos estadísticos y no estadísticos usados en la construcción de modelos *credit scoring* son: Análisis discriminante, regresión logística [Quintana *et al.* (2005)], redes neuronales, scoring basado en el juicio, regresión lineal, métodos de programación matemática, arboles de decisión, sistemas expertos y métodos no paramétricos [Hand y Henley (1997)]. En este capítulo se describen algunos de estos métodos y se explica a detalle el modelo de regresión logística.

Se ha observado que, *credit scoring* típicamente emplea aproximaciones pragmáticas. Generalmente las aproximaciones dependen de el lugar y tiempo donde se recogen los datos, el contexto socio-cultural y la calidad de los datos. Por ejemplo Desai *et al.* (1996) reporta que el desarrollo de redes neuronales es significativamente mejor que el análisis discriminante, mientras Yobas M. y P. (2000) reporta resultados inversos.

5.1. Análisis discriminante

El análisis discriminante tiene como objetivo analizar la relación entre una variable dependiente categórica y un conjunto de variables independientes cuantitativas, a partir de una serie de funciones discriminantes, que son combinaciones lineales de las variables independientes que mejor discriminan o separan los grupos.

Si alguna de las variables independientes es categórica, es preferible utilizar la regresión logística.

Entre los inconvenientes que presenta el análisis discriminante está la incapacidad para

5.2. Redes Neuronales

el cálculo de las probabilidades de impago. Además, se debe cumplir la condición de que la distribución de las variables independientes sea normal.

Para el caso de diferenciar entre dos grupos se aplica el análisis discriminante simple, en el que la variable dependiente es dicotómica y, donde se obtiene una única función discriminante [Quintana *et al.* (2005)].

5.2. Redes Neuronales

Las redes neuronales se consideran dentro de los modelos de inteligencia artificial. Tratan de imitar al sistema nervioso, de modo que construyen sistemas con cierto grado de inteligencia. La red está formada por una serie de procesadores simples denominados nodos, que se encuentran interconectados entre sí. La finalidad de cada nodo consiste en dar respuesta a una determinada señal de entrada, para obtener una salida. El enfoque de *credit scoring* consiste en considerar como nodos de entrada las características o variables de la operación de crédito. Y el nodo de salida sería la variable respuesta definida como la probabilidad de ser préstamo malo [Cantón *et al.* (2010)].

5.3. Scoring basado en el juicio

Esta técnica consiste en una selección de indicadores y asignación de puntos tomando como base la experiencia de los asesores. El puntaje se asigna de acuerdo a un criterio subjetivo. Estos modelos tienen menos capacidad de pronóstico pero requieren menos datos (o no requieren datos), son más fáciles de elaborar, suelen tener capacidad de pronóstico suficiente y son aceptados por los usuarios con mayor facilidad, especialmente si estos participan en la selección de los indicadores y del sistema de puntos [Schreiner (2008)].

5.4. Regresión Logística

Los modelos de regresión logística permiten calcular la probabilidad que tiene un cliente para pertenecer a uno de los grupos establecidos a priori (no pagador o pagador). La clasificación se realiza de acuerdo con el comportamiento de una serie de variables independientes de cada observación o individuo. La principal ventaja del modelo de regresión logística radica en que no es necesario plantear hipótesis de partida, como por ejemplo la normalidad de la distribución de las variables, mejorando

5.4. Regresión Logística

el tratamiento de las variables cualitativas o categóricas. Además este modelo presenta la ventaja de medir la probabilidad de incumplimiento al mantener la variable explicada siempre dentro de un rango de variación entre cero y uno.

Considerando las ventajas mencionadas anteriormente, en este estudio se usa regresión logística, la cual se describe a continuación.

La regresión logística propone la transformación logit definida como:

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

Tomando esta función como variable dependiente, puede plantearse el buscar para ella una ecuación de regresión tradicional.

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (5.1)$$

Esta transformación tiene muchas propiedades deseables. Es lineal en estos parámetros, continua y su rango va de $-\infty$ a $+\infty$.

Dado el valor de las variables independientes, se puede calcular directamente la estimación de la probabilidad de que ocurra el evento de interés de la siguiente forma:

$$\hat{p} = \pi(\mathbf{x}) = \frac{e^{\text{suma}}}{1+e^{\text{suma}}}; \text{ donde } \text{suma} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k$$

La estimación de los parámetros se obtiene por máxima verosimilitud. Tal estimación se describe como sigue: Si Y es codificada como 0 ó 1 entonces $\pi(\mathbf{x})$ proporciona la probabilidad condicional de que Y es igual a 1 dado \mathbf{x} , denotado por $P(Y = 1|\mathbf{x})$. Por lo tanto, $1 - \pi(\mathbf{x})$ es la probabilidad condicional de que Y sea igual a cero dado \mathbf{x} , $P(Y = 0|\mathbf{x})$. Suponiendo que las observaciones son independientes la función de verosimilitud es:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^n \pi(\mathbf{x})^{y_i} [1 - \pi(\mathbf{x})]^{1-y_i}$$

El principio de máxima verosimilitud establece que se usa como estimador de $\boldsymbol{\beta}$ el valor que maximiza la expresión anterior. Sin embargo, es matemáticamente más fácil trabajar con la log-verosimilitud, definida como:

$$L(\boldsymbol{\beta}) = \ln[l(\boldsymbol{\beta})] = \sum_{i=1}^n \{y_i \ln[\pi(\mathbf{x}_i)] + (1 - y_i) \ln[1 - \pi(\mathbf{x}_i)]\}$$

Para encontrar el valor de $\boldsymbol{\beta}$ que maximiza $L(\boldsymbol{\beta})$ se deriva $L(\boldsymbol{\beta})$ con respecto a los $k+1$ coeficientes, y se igualan a cero las expresiones resultantes.

5.4. Regresión Logística

$$\sum_{i=1}^n [y_i - \pi(\mathbf{x}_i)] = 0$$

y

$$\sum_{i=1}^n x_{ij} [y_i - \pi(\mathbf{x}_i)] = 0 \text{ para } j = 1, \dots, k.$$

Estas expresiones son no lineales con respecto a los coeficientes y requieren métodos especiales para su solución [Hosmer (2000)].

Uno de los métodos iterativos más utilizados es el algoritmo Newton-Raphson, el cual se describe como sigue: Sea $U(\boldsymbol{\beta})$ el vector de las primeras derivadas de $L(\boldsymbol{\beta})$ con respecto a $\boldsymbol{\beta}$ y sea $I(\boldsymbol{\beta})$ la matriz de las segundas derivadas de $L(\boldsymbol{\beta})$

El algoritmo de Newton Raphson es entonces:

$$\boldsymbol{\beta}_{j+1} = \boldsymbol{\beta}_j - I^{-1}(\boldsymbol{\beta}_j)U(\boldsymbol{\beta}_j)$$

Es necesario dar valores iniciales $\boldsymbol{\beta}_0$, se puede empezar simplemente con todos los coeficientes igual a 0 [Allison (1999)]. Los valores iniciales son sustituidos en la ecuación anterior de donde se obtiene el resultado de la primera iteración, $\boldsymbol{\beta}_1$. Ahora se sustituyen estos últimos valores para encontrar $\boldsymbol{\beta}_2$. El proceso es repetido hasta que el máximo cambio en cada parámetro, de un paso al siguiente, sea menor a algún criterio.

Odds y Odds Ratio (OR)

El odds contesta a la pregunta ¿cuántas veces es más probable que ocurra un evento respecto a que no ocurra?. Y se calcula por:

$$odds = \frac{P(evento)}{1-P(evento)}$$

Por otra parte, el OR es una razón de odds y cuantifica la magnitud de la asociación entre el riesgo y el factor de interés.

$$OR = \frac{Odds(evento|caracteristica)}{Odds(evento|ausenciadecaracteristica)}$$

Para una covariable binaria, el OR contesta la pregunta ¿por cuánto se multiplica el riesgo en presencia de la característica de la covariable respecto a la ausencia de esa misma característica?.

Para una covariable categórica, el OR contesta la pregunta ¿por cuánto se multiplica el riesgo en presencia de cierta categoría con respecto a la categoría de referencia?

El OR para una variable explicativa cuantitativa cuantifica la magnitud de la aso-

5.4. Regresión Logística

ciación entre el riesgo y el cambio (aumento) en el factor de interés en una unidad. Por lo tanto, contesta la pregunta ¿por cuanto se multiplica el riesgo ante el aumento en una unidad en la covariable?

Capítulo 6

Credit Scoring usando Regresión Logística: Caso MásKapital

Este capítulo muestra el desarrollo de un modelo *Credit Scoring* usando información de MásKapital.

6.1. Antecedentes generales

KapitalMujer S.A. de C.V. SOFOM E.N.R. es una institución financiera no regulada creada en julio de 1993 con la finalidad de dedicarse a la consultoría y a brindar servicios diversos a bancos comerciales y cuya incursión en el microcrédito se produjo en octubre de 2005. KapitalMujer cambió su denominación original -Eurekasol- a finales de 2007 aunque comercialmente es conocida como "MasKapital". La institución ha tenido una rápida expansión de sus actividades crediticias en los últimos años. Tiene una cartera de 88.5 millones de pesos (6.5 millones de dólares) y cuenta con 30, 528 clientes atendidos desde su Oficina Central -en Oaxaca, Oaxaca- y veinte sucursales repartidas en los estados de Oaxaca, Puebla, Veracruz, Tlaxcala, Morelos y Estado de México. La empresa ofrece créditos a mujeres de zonas periurbanas y rurales mediante la metodología crediticia de Grupos Solidarios.

La sucursal "Oaxaca" es la de mayor antigüedad en la empresa. Se ubica en el municipio de Oaxaca de Juárez. Su zona de influencia comprende este municipio, que es capital del estado, y sus zonas aledañas. Sumando un total de 15 municipios atendidos.

MásKapital se encuentra afiliada a la red Prodesarrollo que aglomera más de 85 instituciones microfinancieras locales. También está asociada a Finanzas y Microempresas AC, asociada a la Federación Atlántico Pacífico de Entidades de Ahorro y Crédito

6.1. Antecedentes generales

A.C.; también participa en ASOFOM la Asociación de Sociedades Financieras de Objeto Múltiple [Rating (2009)].

Figura 6.1: Calificación GIRAFE -MásKapital México- Noviembre 2009. Planet Rating

Indicadores de desempeño

USD	dic. 2006	dic. 2007	dic. 2008	sep. 2009
Activos	4,178,562	5,183,339	7,389,822	11,091,265
Cartera	3,027,926	2,620,200	4,945,634	6,555,589
Clientes activos	8,314	10,083	22,763	30,528
Cartera promedio por cliente	364	260	217	215
Ahorros en garantía ⁽¹⁾	661,314	859,588	870,301	1,467,614
Personal	36	111	217	288
ROE	1.5%	1.5%	3.8%	44.3%
ROA (sin donaciones)	0.5%	0.4%	1.0%	11.2%
Deudas / Patrimonio	2.42x	2.64x	2.85x	3.06x
Rendimiento de cartera	77.4%	100.3%	100.8%	101.6%
Ratio de costos operativos	48.7%	64.5%	81.4%	61.8%
Ratio de costos financieros	13.8%	16.5%	17.4%	14.9%
Ratio de costos de provisión	13.4%	11.1%	4.4%	6.0%
Cartera en riesgo 31 - 365 días	2.8%	4.6%	2.9%	4.5%
Cartera en riesgo > 365 días	0.0%	0.0%	0.3%	0.1%
Tasa de cartera castigada	0.0%	10.8%	3.3%	1.9%

⁽¹⁾Ahorros en garantía: fondos de clientes (10% del monto desembolsado). No se utiliza como financiamiento, sólo se resguarda en cuentas bancarias

Un Grupo Solidario (GS) tiene las siguientes características: es un grupo de 10 a 25 mujeres que, viven en la misma comunidad, son mayores de 18 años y tienen confianza entre sí. Poseen negocio propio, el cual, puede ser informal. Tienen necesidad de crédito para capital de trabajo y no para consumo personal. Además, se comprometen a mejorar su actividad productiva e incrementar sus ingresos. Todos los integrantes son personas que están dispuestas a cumplir con los lineamientos, normas y reglamentos definidos por la empresa; así como también a realizar pago solidario cada vez que sea necesario.

Las características sobresalientes del grupo solidario son: **Autoseleccionado**. Es importante que las mismas clientas elijan a las demás integrantes del grupo. **Solidario**. Si una de las clientas no cumple con los compromisos adquiridos las demás tendrán que hacerlo.

La metodología de grupos solidarios aplicada por MasKapital está destinada a mujeres con negocio dedicadas a actividades productivas, comerciales, artesanales o de servicios con escaso o nulo acceso al sistema financiero tradicional. Los grupos solidarios se forman de 10 a 25 mujeres (siendo lo más frecuente de 11 a 12), con necesidades de crédito que varían entre 1,000 y 30,000 por socia, con un máximo de 3,000 en el primer ciclo, y con límites escalonados en los siguientes ciclos hasta los 30,000.

6.1. Antecedentes generales

El proceso crediticio comienza con la *promoción* y la formación de grupos solidarios por los asesores de crédito en las zonas de influencia. Se realiza el proceso de *autorización* del grupo en tres sesiones o reuniones. El objetivo de estas es proporcionar a las clientas potenciales la información sobre los beneficios, características y compromisos del producto Grupo Solidario de Kapitalmujer que permita la adecuada colocación del crédito. El grupo se reúne en el domicilio de una de la clientas, donde el asesor inicia la capacitación. Al término de tres pláticas de capacitación, se llenan los formatos correspondientes, entre ellos el acta constitutiva del grupo solidario y la solicitud individual de microcrédito.

La concesión de créditos está determinada por políticas de la empresa y en gran medida por el criterio del asesor de crédito. Además, el crédito se considera aprobado sólo con el consentimiento de todos los miembros del grupo solidario. La valoración del riesgo de un crédito permanece en la experiencia del asesor; lo cual, se conoce como scoring subjetivo.

Para el *desembolso* se requiere el depósito en garantía del 10 % del importe financiado. Durante la vida del crédito del cliente puede incluir en la cuota un monto extra (ahorro) voluntariamente que determinará el incremento del préstamo al que podrá acceder en el siguiente ciclo (10 veces el monto ahorrado).

Una vez que el crédito es concedido, ¿cómo recupera MásKapital el crédito más intereses?. El proceso de *recuperación* consta de 16 pagos semanales por parte del GS; donde, en cada pago, todas y cada una de las clientas están comprometidas a pagar la parte que les corresponde. Cada semana el Grupo Solidario y el asesor se reúnen en el lugar y horario acordado por las clientas para realizar el pago correspondiente. La tesorera, que pertenece y es nombrada por el GS, solicita a las clientas que pasen en forma ordenada a realizar su pago.

En caso de que alguna clienta no cumpla con el pago requerido, el GS debe hacer un *pago solidario*, definido como una aportación extra que realizan las clientas para cubrir la cantidad que no pagó la clienta incumplida; de tal manera que el pago semanal del GS sea cubierto en su totalidad.

Algunas semanas antes de concluir la recuperación el asesor realiza la invitación al grupo para solicitar un crédito subsecuente y la integración de nuevas clientas, proceso conocido como *renovación*. Se capacita a las clientas entrantes y nuevamente se llevan a cabo las actividades de autorización. En la última semana de recuperación el gerente preferentemente junto con las clientas, tendrá que verificar y autorizar los montos de renovación. Las garantías liquidadas son devueltas y el expediente es archivado adecuadamente.

Actualmente, MásKapital está elaborando una base de datos electrónica que ayude en la construcción de un sistema de calificación de créditos (*Credit Scoring*), para su posterior implementación. Lo cual, hará posible la administración del riesgo y

6.2. Conformación de la Base de Datos

ayudará al asesor a tomar decisiones basadas en dicha calificación.

Cuando se presenta incumplimiento de pago, inicia el proceso de *cobranza* definido como aquella actividad que se realiza fuera de la reunión de recuperación. En este proceso el asesor entrega notificaciones y visita junto con el grupo a la clienta morosa. Si la clienta hace caso omiso de las notificaciones el gerente de la sucursal intervendrá en el proceso de recuperación. Se da seguimiento a las gestiones de la cobranza administrativa hasta agotarlas.

6.2. Conformación de la Base de Datos

La información utilizada en este estudio, corresponde a la sucursal Oaxaca y fue extraída de las fuentes que a continuación se describen:

- El formato *Control de pagos y depósitos* que se encuentra en los expedientes en papel, de cada grupo solidario. Contiene el registro de los pagos realizados por cada una de las clientas cada semana durante el período ordinario de recuperación; el cual, consta de 16 semanas. De este formato se capturó el número de semanas que la clienta no cumplió con el pago requerido.
- El formato *Solicitud de microcrédito* que también se encuentra en los expedientes en papel. Y de este se capturó el ciclo de clienta y resultado de buró.
- Otra fuente utilizada fue la base de datos electrónica de MásKapital. Donde, se encuentra capturada la información del estudio socioeconómico realizado al cliente.

Toda la información encontrada en papel fue capturada en Excel. Mientras que, con la información electrónica, se realizaron las operaciones necesarias para poder manipular los datos.

En el período de mayo de 2008 a abril de 2009, en la sucursal Oaxaca, se ministró un total de 6,429 préstamos a nivel individual. Donde, cada uno de estos, es parte integrante de un préstamo grupal. En este total existen clientes que solicitan un primer préstamo y clientes con alguna antigüedad.

Se decidió usar, para este estudio, la información de clientes en ciclo 1 por las siguientes razones: es el primer análisis de scoring en MásKapital, que puede ser parte de un sistema integral de scoring; se encontró que el número de préstamos consecutivos al cliente anotado en las solicitudes de microcrédito, es un dato impreciso, lo que dificulta el análisis de préstamos de ciclo mayor a 1. Los registros que no contaron

6.3. Definición de la variable respuesta y variables explicatorias

con la información necesaria para este estudio fueron excluidos. Así, de un total de 2,173 registros capturados en ciclo 1, presentaron información completa 1,610.

Por lo tanto, la base de datos queda conformada por 1,610 registros que comprenden fechas de ministración de mayo de 2008 a abril de 2009. Tal información corresponde a clientes en ciclo 1 que ingresan a grupos solidarios nuevos o con experiencia. Cada registro contiene información de una variable dependiente y 20 variables explicatorias. Estas últimas, abarcan características de grupo, préstamo y prestatario. La base de datos puede ser consultada en www.maskapital.com.mx

6.3. Definición de la variable respuesta y variables explicatorias

El riesgo es entendido como la posibilidad de que suceda un evento no deseado. Para el caso de concesión de créditos en microfinanzas, ese evento no deseado es la posibilidad de que un préstamo concedido se convierta en un préstamo malo.

Los modelos *credit scoring* en microfinanzas buscan calcular la probabilidad de que un préstamo se convierta en malo. Ahora, ¿Cuándo se dice que un préstamo es malo?. A continuación se menciona una serie de sucesos que ayudarán a definir un préstamo malo.

Cuando un grupo hace pagos solidarios, la clienta incumplida esta comprometida a devolver al grupo la cantidad correspondiente. No obstante, cuando los pagos solidarios se realizan con demasiada frecuencia se ocasionan problemas al grupo, que son percibidos fácilmente por el asesor y el GS. Se ha observado que la falta de tres pagos por parte de una clienta, genera conflictos en el grupo. Es necesario cuidar que el GS este libre de integrantes morosos. Puesto que, esto puede repercutir en el cumplimiento general del GS, cayendo en problemas de morosidad.

Por otra parte, las políticas de la empresa manifiestan que en caso de que una clienta presente 3 o mas atrasos, durante el período ordinario de recuperación, será dada de baja del grupo solidario, perdiendo el derecho a renovar su crédito. Con base a estas situaciones se optó por definir como un préstamo malo a aquel que presenta 3 o más semanas de atraso durante el período ordinario de recuperación. Así, la variable respuesta binaria, toma el valor 1 si el préstamo es malo, de lo contrario toma el valor 0.

Ahora, se busca calcular la probabilidad de que la variable respuesta tome el valor 1. Usaremos el modelo de regresión logística binaria con una variable dependiente dicotómica y 20 variables explicatorias. De estas últimas, una es cuantitativa, cuatro son variables binarias y las quince restantes son categóricas.

6.3. Definición de la variable respuesta y variables explicatorias

La variable explicatoria CICLO se refiere al número de préstamos consecutivos realizados al grupo solidario. Es una característica de grupo, y mide la experiencia de este por el número de préstamos. La base de datos contiene grupos con experiencia variada, que puede ir de 1 a 13 préstamos.

La variable binaria TEL se refiere a la disposición, que tiene el cliente, de un número telefónico. Toma el valor 1 si el cliente cuenta con teléfono y 0 si no cuenta con este.

La variable binaria ZONAVIV, toma el valor 1 si el cliente vive en zona rural y 0 si es de zona urbana.

La variable binaria NOHIJ se refiere al número de hijos, toma el valor 1 si el cliente tiene mas de dos hijos y 0 de otra forma.

Y la variable binaria HIJEST es el número de hijos estudiando, toma el valor 1 si tiene hijos estudiando y 0 si no tiene.

La tabla 6.1 presenta la descripción de las variables utilizadas en este estudio.

Tabla 6.1: Variables analizadas.

Num	Variable	Descripción	Tipo
0	Y	Variable dependiente	Binaria
1	CICLO	Número de préstamos al grupo	Cuantitativa
2	MONTO	Monto solicitado	Catagórica
3	BURO	Resultado de buró	Catagórica
4	INGRE	Ingresos del cliente	Catagórica
5	ANTINEG	Antigüedad en el negocio	Catagórica
6	TRANS	Transporte	Catagórica
7	TEL	Teléfono	Binaria
8	ELECTRDM	Electrodomésticos	Catagórica
9	ZONAVIV	Zona de vivienda	Binaria
10	TIPOVIV	Tipo de vivienda	Catagórica
11	NOHAB	Número de habitaciones	Catagórica
12	TECHO	Techo	Catagórica
13	PISO	Piso	Catagórica
14	DREN	Drenaje	Catagórica
15	AGUA	Agua	Catagórica
16	EDAD	Edad del cliente	Catagórica
17	EDOCIVIL	Estado civil	Catagórica
18	ESCOLARI	Escolaridad del cliente	Catagórica
19	NOHIJ	Número de hijos	Binaria
20	HIJEST	Número de hijos estudiando	Binaria

Para cada variable categórica fue necesario seleccionar una categoría de referencia. En cada caso se optó por seleccionar aquella categoría que contiene el mayor número de clientes. Este criterio para seleccionar la categoría de referencia indica las preferencias, conscientes o inconscientes, del asesor en la aprobación de créditos. Así como también la aplicación de las políticas de la empresa en la concesión de créditos. Por tanto, es

6.3. Definición de la variable respuesta y variables explicatorias

en esa categoría donde se encuentra la experiencia del asesor y su buen ojo para seleccionar clientes buenos. Así, es razonable esperar que la categoría de referencia en cada variable presente el menor riesgo.

La variable MONTO es una característica del préstamo y es la cantidad en pesos solicitada por el cliente que puede ser entre 1,000 y 30,000 pesos. Fue clasificada en cuatro categorías: 'menor o igual a 3000', 'mayor a 3000 y menor a 5000', y 'mayor a 5000'.

La variable BURO se refiere al resultado de los reportes de las sociedades de información crediticia y tiene tres opciones: 'sin antecedentes', 'antecedentes positivos' y 'antecedentes negativos'.

La variable INGRE es una característica de la situación financiera del cliente y se refiere a sus ingresos semanales en pesos. Fue clasificada en cuatro categorías: 'menor o igual a 1000', 'mayor a 1000 y menor a 2000', 'mayor a 2000 y menor a 3000', y 'mayor a 3000'.

ANTINEG es la antigüedad que tiene el negocio del cliente, medida en años. Clasificada en cuatro categorías: 'menor o igual a 3', 'mayor a 3 y menor o igual a 5', 'mayor a 5 y menor o igual a 10' y 'mayor a 10'.

TRANS es el tipo de transporte que usa el cliente, que puede ser: 'vehículo propio', 'transporte público' ó 'camina'.

ELECTRD se refiere al número de electrodomésticos con los que cuenta el cliente, las respuesta puede ser cualquiera de estas cuatro: ninguno, 'estufa de gas o refrigerador', '2 de ellos' y 'todos'.

La variable TIPOVIV es la situación de la vivienda y tiene tres opciones: 'vivienda propia', 'rentada' y 'prestada'.

La variable NOHAB se refiere al número de habitaciones en la vivienda y tiene cuatro opciones: 'una habitación', '2 habitaciones', '3 habitaciones' y '4 o mas habitaciones'.

La variable TECHO se refiere al tipo de techo de la vivienda y tiene cuatro opciones: 'Todo concreto', 'concreto y otro', 'ásbесто o aluminio' y 'lámina de cartón o paja'.

La variable PISO se refiere al tipo de piso de la vivienda y tiene tres opciones: 'mosaico', 'concreto' y 'tierra'.

La variable DREN se refiere a al tipo de drenaje con que cuenta la vivienda y tiene tres opciones: 'no tiene', 'letrina' y 'WC'.

La variable AGUA se refiere a la disposición de agua entubada en la vivienda y tiene

6.3. Definición de la variable respuesta y variables explicatorias

tres categorías: 'llave en casa', 'llave pública'y 'no tiene'.

La variable EDAD es la edad del cliente y fue clasificada en cuatro categorías: 'menor a 21', 'mayor a 21 y menor o igual a 50', 'mayor a 50 y menor o igual a 60'y 'mayor a 60'.

La variable EDOCIVIL es el estado civil del cliente y tiene cuatro opciones: 'soltera', 'casada', 'viuda'y 'divorciada'.

La variable ESCOLARI es la escolaridad máxima del cliente y su respuesta entra en una de cuatro opciones: 'ánalfabeta', 'primaria', 'secundaria'y 'bachillerato o mayor'.

La tabla 6.2 muestra en negrita la categoría de referencia para cada una de estas variables.

Tabla 6.2: Variables categóricas.

Variable	Categorías	Variable	Categorías
MONTO	3 Menor o igual a 3000	TECHO	4 Todo concreto
	1 Mayor a 3000 y menor o igual a 5000		1 Concreto y otro
	2 Mayor a 5000		2 Asbesto o aluminio
BURO	2 Antecedentes negativos	PISO	3 Lámina, lona o paja
	3 Sin antecedentes		0 Mosaico
INGRE	1 Antecedentes positivos		3 Concreto
	0 Menor o igual a 1000	2 Tierra	
	4 Mayor a 1000 y menor o igual a 2000	3 WC	
ANTINEG	2 Mayor a 2000 y menor o igual a 3000	AGUA	1 Letrina
	3 Mayor a 3000		2 No tiene
	4 Menor o igual a 3		3 Llave en casa
TRANS	1 Mayor a 3 y menor o igual a 5	EDAD	1 Llave pública
	2 Mayor a 5 y menor o igual a 10		2 No tiene
	3 Mayor a 10		0 18-21
ELECTRDM	0 Propio	3 22 a 50	
	3 Público	1 51 a 60	
	2 Camina	2 mayor a 60	
TIPOVIV	4 Todos	EDOCIVIL	4 Casada
	1 2 de ellos		1 Divorciada
	2 Gas o refrigerador		2 Viuda
NOHAB	3 Ninguno	ESCOLARI	3 Soltera
	3 Propia		0 Bachillerato
	1 Rentada		1 Secundaria
NOHAB	2 Prestada	4 Primaria	
	3 Una	3 Analfabeta	
	2 Dos		
	1 Tres		
	4 Mayor o igual a 4		

El modelo se desarrolló en SAS, usando PROC LOGISTIC [ver apéndice A]. En las siguientes secciones se exponen los resultados obtenidos.

6.4. Selección de variables

Se usó el método de selección Backward para encontrar un modelo que tuviera variables con nivel de significancia menor al 5%. Este método comienza incluyendo todas las variables, en este caso 20 variables. Los resultados de la prueba Wald para los parámetros individuales son examinados. Si existen variables con p-value mayor a 0.05 la variable con mayor p-value es removida y queda excluida. El proceso es repetido con las 19 variables restantes hasta que no haya variables que remover [Ayuda de SAS].

La selección backward se obtiene con la opción SELECTION = B de la sentencia MODEL de PROC LOGISTIC. Los resultados muestran cinco variables significativas de un total de 20 variables explicatorias.

Tabla 6.3: Type III Analysis of Effects

Effect	DF	Wald	
		Chi-Square	Pr > ChiSq
ciclo	1	5.6767	0.0172
trans	2	6.3288	0.0422
techo	3	14.1318	0.0027
dren	2	13.8464	0.0010
hijest	1	5.3522	0.0207

La variable CICLO tiene un coeficiente negativo, indicando que cuando la experiencia del grupo aumenta el riesgo disminuye. Tiene un OR menor a 1 (0.927), y por ser una variable cuantitativa, el OR significa que el incremento en un préstamo en la experiencia de grupo, multiplica el riesgo por 0.927.

La variable HIJEST tiene un coeficiente positivo, y por ser una variable binaria este indica que el riesgo aumenta en presencia de hijos estudiando. Tiene un OR de 1.36 (mayor a 1), que indica que, el riesgo en presencia de hijos estudiando es 1.36 veces el riesgo en ausencia de hijos estudiando.

La variable TECHO presenta mayor riesgo en la categoría techo de *asbesto o aluminio*. Tiene un OR de 2.048 el cual, significa que el riesgo de *asbesto o aluminio* es 2.048 el riesgo de la categoría de referencia *todo concreto*.

La variable DREN presenta mayor riesgo en la categoría *no tiene*. Su OR de 3.339, indica que el riesgo en presencia de esta categoría es 3.339 el riesgo de la categoría de referencia *W.C.*. Por otra parte, el riesgo disminuye con la categoría *letrina*, que es 0.517 el riesgo de *W.C.*.

En la variable TRANS, la categoría que representa mayor riesgo es *transporte público*(categoría de referencia). Mientras que la categoría con menor riesgo es *vehículo*

6.5. Ajuste del modelo

propio, que multiplica por 0.6(OR) el riesgo de la categoría de referencia .

Las últimas variables eliminadas fueron BURO e INGRE. La tabla 6.4 muestra un resumen de la eliminación backward:

Tabla 6.4: Summary of Backward Elimination

Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	nohij	1	19	0.0055	0.9406
2	nohab	3	18	1.1837	0.7569
3	piso	2	17	0.7971	0.6713
4	zonaviv	1	16	0.3742	0.5407
5	tipoviv	2	15	1.3715	0.5037
6	escolari	3	14	3.0501	0.3840
7	edocivil	3	13	3.5588	0.3132
8	monto	2	12	2.0489	0.3590
9	edad	3	11	3.6351	0.3037
10	tel	1	10	0.9975	0.3179
11	electrdm	3	9	3.6834	0.2977
12	agua	2	8	2.6652	0.2638
13	antineg	3	7	5.1771	0.1593
14	ingre	3	6	4.5418	0.2086
15	buro	2	5	3.7130	0.1562

6.5. Ajuste del modelo

El ajuste del modelo fue evaluado con la ayuda de la estadística Hosmer-Lemeshow. La prueba consiste en establecer los deciles del riesgo (o probabilidad de presentar el evento) estimado por el modelo, y en cada una de estas diez categorías se comparan los valores observados y los predichos. Si hay una elevada coincidencia entre observados y esperados, la prueba Ji cuadrada que contrastará ambas distribuciones no mostrará significación estadística y se concluye un buen ajuste. En otras palabras, un modelo tiene un buen ajuste cuando esta prueba es no significativa.

La prueba se obtiene fácilmente con la opción LACKFIT de la sentencia MODEL de PROC LOGISTIC. En este caso el p-value fue de 0.7887, el cual representa un buen ajuste.

6.6. Poder predictivo

Tabla 6.5: Partition for the Hosmer and Lemeshow Test

Group	Total	y = 1		y = 0	
		Observed	Expected	Observed	Expected
1	162	6	4.37	156	157.63
2	161	9	7.54	152	153.46
3	162	9	9.80	153	152.20
4	163	8	12.02	155	150.98
5	162	13	14.18	149	147.82
6	162	19	16.93	143	145.07
7	161	21	20.01	140	140.99
8	161	26	23.74	135	137.26
9	161	24	29.48	137	131.52
10	155	46	42.88	109	112.12

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
4.7036	8	0.7887

6.6. Poder predictivo

El poder predictivo es la capacidad que tiene el modelo para predecir la variable dependiente; basado en los valores de las variables independientes. Es posible tener un modelo que predice muy bien la variable dependiente y, sin embargo, que tenga un ajuste terrible. Y no es poco frecuente tener un modelo que ajuste bien y, sin embargo, que tenga un bajo poder predictivo.

Una de las estadísticas que evalúan el poder predictivo es, la R^2 generalizada, reportada por la opción RSQ de la sentencia MODEL de PROC LOGISTIC; la cual, está basada en la razón de verosimilitud chi-square para probar la hipótesis nula de que todos los coeficientes son 0. Si se denota esta razón como L^2 , y sea n el tamaño de muestra, la R^2 es:

$$R^2 = 1 - \exp\left\{-\frac{L^2}{n}\right\}$$

El límite superior de la R^2 generalizada es menor que 1 porque la variable dependiente es discreta. Para ajustar esto, LOGISTIC divide la R^2 por el límite superior.

De esta manera, obtenemos con SAS una R^2 de 0.0945, valor que indica un bajo poder predictivo.

6.7. Poder discriminatorio

A la capacidad que tiene el modelo para clasificar de manera correcta a los préstamos, se le conoce como poder discriminatorio.

Las curvas ROC (Receiver Operating Characteristic) ofrecen una representación gráfica del poder discriminatorio de un sistema de scoring. Cuanto mas próxima es una curva a la esquina superior izquierda, más alta es la exactitud global de la prueba.

Antes de definir la curva ROC se hacen las siguientes consideraciones. La variable respuesta solo tiene dos valores: malo y bueno. El modelo utilizado clasifica los préstamos como malo ó bueno basándose en un punto de corte y el riesgo estimado. Al comparar el resultado observado con la clasificación hecha por el modelo existen cuatro posibilidades, que pueden resumirse en una tabla

Tabla 6.6: Comparación del valor observado con el resultado del modelo

Valor observado	Resultado del modelo	
	Malo	Bueno
Malo	Verdadero positivo (VP)	Falso negativo (FN)
Bueno	Falso positivo (FP)	Verdadero negativo (VN)

La sensibilidad se define como la probabilidad de que el modelo clasifique como malo a un préstamo malo, y es calculada por la proporción de préstamos malos que fueron clasificados correctamente. Mide la capacidad del modelo para detectar un préstamo malo.

$$Sensibilidad = \frac{VP}{VP+FN}$$

Por otra parte, la especificidad es la probabilidad de que el modelo clasifique como bueno a un préstamo bueno y se calcula por la proporción de préstamos buenos que fueron clasificados correctamente. Mide la capacidad del modelo para descartar mora cuando ésta no existe.

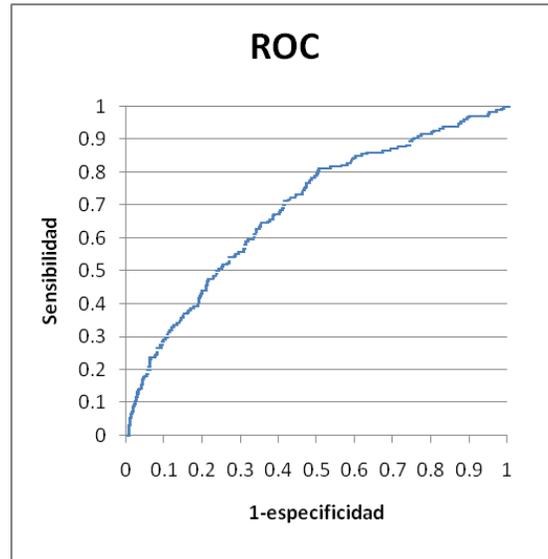
$$Especificidad = \frac{VN}{VN+FP}$$

La curva ROC es un gráfico en el que se observan todos los pares {(1-especificidad), sensibilidad} resultantes de la variación continua de los puntos de corte en todo el rango de resultados observados. En el eje y de coordenadas se sitúa la sensibilidad o proporción de verdaderos positivos, mientras que en el eje x se sitúa $\{1 - especificidad\}$ o proporción de falsos positivos [Burgueño M. J. (1995)].

Los datos necesarios para elaborar la curva ROC se obtienen con la opción OUT-ROC=Option de la sentencia MODEL de PROC LOGISTIC.

6.7. Poder discriminatorio

Figura 6.2: Curva ROC



El área bajo la curva ROC se define como la probabilidad de que para un par de préstamos, uno malo y otro bueno, seleccionados al azar de la población y al aplicarles la prueba *credit scoring*, el préstamo malo presente mayor riesgo que el préstamo bueno.

El área bajo la curva se calcula como sigue: si n_1 denota el número de préstamos con $y = 1$ (malos) y n_0 el número de préstamos con $y = 0$, entonces tenemos $n_1 * n_0$ pares. Cada préstamo con $y = 1$ es pareado con cada préstamo $y = 0$. De esos $n_1 * n_0$ pares, determinamos la proporción de pares donde, $y = 1$ tiene mayor riesgo que $y = 0$ en cada par. Esta proporción es igual al área bajo la curva.

Los resultados de PROC LOGISTIC muestran automáticamente esta proporción con el nombre de *Percent Concordant*, bajo el encabezado *Association of Predicted Probabilities and Observed Responses*. Para el caso en estudio, se tienen 181 préstamos con $y = 1$ y 1429 con $y = 0$, obteniendo un área bajo de la curva igual a 0.69. Esta área significa que para dos préstamos, uno seleccionado aleatoriamente del grupo de malos y otro elegido al azar del grupo de buenos, el préstamo malo presentará un riesgo mayor al bueno el 69% de las veces.

La siguiente tabla presenta una regla general para el poder discriminatorio [Hosmer (2000)]:

6.8. Elección del Punto de corte

Tabla 6.7: Regla general del poder discriminatorio

Si $ROC = 0.5$	Sugiere no discriminación
Si $0.7 \leq ROC < 0.8$	Se considera discriminación aceptable
Si $0.8 \leq ROC < 0.9$	Se considera discriminación excelente
Si $ROC \geq 0.9$	Se considera discriminación extraordinaria

6.8. Elección del Punto de corte

El punto de corte es un valor elegido de la manera mas conveniente para discriminar, en este caso, préstamos buenos y malos. Los préstamos con riesgo por encima de este punto son clasificados como malos y los que están por debajo son clasificados como buenos. Con los datos utilizados en este estudio, las estimaciones del riesgo van de 0.007 a 0.474. Por lo tanto, el punto de corte será elegido entre estas probabilidades.

Existen diferentes criterios para determinar el punto de corte. Una manera es considerar aquel punto con el máximo porcentaje de clasificación correcta total [Rocabado *et al.* (2007)]. Otra propuesta es, elegir el punto que presenta sensibilidad igual a especificidad. No obstante el criterio a seguir está determinado por los objetivos del problema.

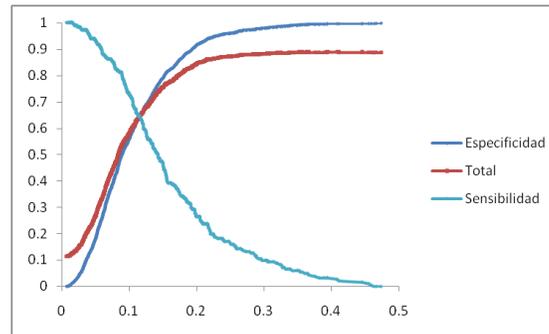
Para el caso *credit scoring* en microfinanzas donde se está tratando con préstamos aprobados por el método tradicional, es muy importante que el modelo clasifique bien a los prestamos malos. Por lo tanto, se elige un punto de corte que tenga una sensibilidad (tasa de verdaderos positivos) mayor a 0.5.

Usando el primer criterio se ve que la máxima clasificación correcta total es de 88.88 %, sin embargo, ahí se observa una sensibilidad de 0.03, que no resulta útil para una aplicación de *credit scoring*. Por lo tanto, este criterio no es conveniente en este caso.

Considerando el segundo criterio, se busca el punto donde sensibilidad sea igual a especificidad. Para esto se toma el punto con la menor diferencia absoluta entre sensibilidad y especificidad. El punto de corte que cumple esta condición es 0.1150. Donde se obtiene una sensibilidad y especificidad de 0.65. La sensibilidad de 0.65 significa que, con este punto de corte, de los 181 malos en la muestra el modelo detectó el 65 %. De la misma manera, la especificidad expresa que de los 1,429 préstamos buenos, el modelo clasificó correctamente al 65 %; con un porcentaje total de clasificación correcta igual a 65 %.

6.9. Validación

Figura 6.3: Sensibilidad, especificidad y total correcto



6.9. Validación

El scoring estadístico tiene la capacidad de ser probado antes de usarse. Este procedimiento revela como hubiera funcionado el scoring si hubiera estado en aplicación al momento de los desembolsos actualmente vigentes [Schreiner (2002)]. La validación se realiza con una muestra no utilizada para construir el modelo.

Para poder validar el modelo se usó una muestra de 402 registros, de los cuales 47 son préstamos malos y 355 son buenos. Tal muestra corresponde al período de mayo de noviembre de 2009, que es inmediato al período que comprende la muestra usada para construir el modelo. Y puede ser consultada en www.maskapital.com.mx

Si esta muestra se une con la original las proporciones son 20 % y 80 % del total, respectivamente; proporciones recomendadas en [Simbaqueba (2004)]. Quien menciona que, aproximadamente el 20 % de la muestra original no debe entrar en el modelo para poder validarlo una vez que este se haya construido. La muestra debe ser aleatoria y es necesario asegurarse que el 20 % de la muestra contenga una proporción similar de buenos y malos como la muestra del 80 %.

Una vez que se estimaron los coeficientes se aplica el modelo a la muestra del 20 % para obtener el riesgo estimado, y usando el punto de corte se hace la discriminación correspondiente. La programación de las operaciones para discriminar se realizó en Excel, de tal manera que llegara a ser suficiente pegar la muestra para obtener las estimaciones deseadas.

Los resultados mostraron una sensibilidad de 0.53 y una especificidad de 0.61, presentando una clasificación total correcta del 60 %. Observe que esta sensibilidad indica que mas del 50 % de los préstamos malos son detectados por el modelo.

6.9. Validación

Tabla 6.8: Resultados para validación

Valor observado	Resultado del modelo	
	Malo	Bueno
Malo	25	22
Bueno	138	217

Capítulo 7

Conclusiones

El trabajo presenta una manera de desarrollar un modelo *credit scoring*. El método de regresión logística fue seleccionado, por tener la ventaja sobre otros métodos de no requerir el supuesto de normalidad y por calcular directamente las probabilidades de impago.

Se conformó una muestra usando información de préstamos de la microfinanciera MásKapital. Dicha base de datos comprende 1,610 registros, correspondiente a préstamos en ciclo 1 en grupos solidarios nuevos o con experiencia.

De las 20 variables explicatorias 5 resultaron significativas usando el criterio de selección Backward (CICLO, HIJEST, TECHO, DREN Y TRANS). La variable más significativa fue DREN. Esta presenta el mayor riesgo en la categoría *no tiene*. Su OR de 3.339 indica que el riesgo en presencia de esta categoría es 3.339 el riesgo de la categoría de referencia *W.C.*. La variable CICLO muestra que la experiencia del grupo es inversamente proporcional al riesgo de impago de los integrantes nuevos.

Las cinco variables que guardan menos relación con el riesgo son: NOHIJ, NOHAB, PISO, ZONAVIV Y TIPOVIV.

El criterio de Hosmer-Lemeshow presenta un p-value de 0.7887, concluyendo un buen ajuste. No obstante, se obtuvo un bajo poder predictivo, evaluado por una R^2 generalizada igual a 0.0945. La literatura menciona que no es poco frecuente encontrarse con este caso de un buen ajuste y un bajo poder predictivo.

El área bajo la curva fue de 0.69, y por regla general del poder discriminatorio, se considera que el modelo tiene una discriminación aceptable.

La manera de definir el punto de corte fue buscando sensibilidad igual a especificidad. Así, fue elegido el punto 0.1150, donde se obtuvo sensibilidad y especificidad de 0.65.

7. Conclusiones

La sensibilidad nos dice que de 181 préstamos malos en la muestra, el modelo detecto el 65 % de ellos.

La validación del modelo fue realizada con una muestra de 402 registros y el punto de corte 0.1150. El modelo detectó el 53 % de los préstamos malos y el 60 % de los registros de esta muestra fue clasificado correctamente. La discriminación es buena y puede mejorarse jugando con los datos, alternando entre construcción del modelo y validación, mejorando la definición de categorías en algunas variables e incluyendo variables que influyan en el riesgo, sugeridas por los expertos.

Referencias

- Allison, P. D. (1999). *Logistic Regression Using the SAS System: Theory and Application*. SAS Institute Inc., Cary, NC, USA.
- Burgueño M. J., G. B. J. L. y G. B. J. M. (1995). Las curvas ROC en la evaluación de las pruebas diagnósticas. *Med Clin (Barc)*, 104, 661–670.
- Cantón, S. R., Rubio, J. L. y Blasco, D. C. (2010). Un Modelo de Crédito Scoring para instituciones de microfinanzas en el marco de Basilea II. *Journal of Economics, Finance and Administrative Science*, Vol. 15 No. 28.
- Conde Bonfil, C. (2000). *¿Pueden ahorrar los pobres? ONG y proyectos gubernamentales en México*. El Colegio Mexiquense y la Colmena Milenaria.
- Delfiner, M., Pailhé, C. y Perón, S. (2006). Microfinanzas: un análisis de experiencias y alternativas de regulación. Documento técnico, El Banco Central de la República Argentina.
- Dennis, W. (1995). Fair Lending and Credit Scoring. *Mortgage Banking*, 56(2), 55–59.
- Desai, V., Crook, J. y Overstreet, G. (1996). A comparison of Neural Networks and Linear Scoring models in the credit union environment. *European Journal of Operational Research*, 95, 24–37.
- Diallo, B. (2006). Un Modele de 'credit scoring' pour une institution de microfinance Africaine: le cas de Nyesigiso au Mali. *Hal Working Papers*.
- Durand, D. (1941). *Risk Elements in Consumer Instalment Financing*. National Bureau of Economic Research.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals Eugen*, 7, 179–188.
- Gutiérrez, G. M. A. (2007). Modelo de Crédito Scoring -Qué, Cómo, Cuando y Para Qué-. *Banco Central de la República Argentina*.
- Hand, D. J. y Henley, W. E. (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. ISSN 1467-985X.
- Hernández Romero, O. (2005). *Las microfinanzas en México, Tendencias y Perspectivas*. Fundación Ayuda en Acción.

Referencias

- Hosmer, D. W. y L. S. (2000). *Applied Logistic Regression*. John Wiley & Sons, Inc., segunda edición.
- Kleimeier, S. y Dinh, T. (2007). A Credit Scoring Model for Vietnam's Retail Banking Market. *International Review of Financial Analysis*, 16(5), 471–495.
- Kulkosky, E. (1996). Credit Scoring Could Have a Down-side, Experts Say. *American Banker*, 161(208), 8.
- Lyn C. Thomas, J. N. C., David B. Edelman (2002). *Credit scoring and its applications..* Siam.
- Mansell Carstens, C. (1995). *Las finanzas populares en México: el redescubrimiento de un sistema financiero olvidado*. Cento de Estudios Monetarios Latinoamericanos, Milenio, Instituto Tecnológico Autónomo de México.
- Milena, E., Miller, M. y Simbaqueba, L. (2005). The Case for Information Sharing by Microfinance Institutions: Empirical Evidence of the Value of Credit Bureau-Type Data in the Nicaraguan Microfinance Sector. *New York: The World Bank, mimeo*.
- Miller, M. y Rojas, D. (2005). Improving Acces to Credit for Smes: An Empirical Analysis of the Viability of Pooled Data SME Credit Scoring Models in Brazil, Colombia & Mexico. *New York: the World Bank*.
- ProDesarrollo Finanzas y Microempresas, A. (2008). Benchmarking de las microfinanzas en méxico:un informe del sector 2008. Reporte tecnico, Microfnance Information Exchange, Inc. (MIX), Mexico.
- Quintana, M. J. M., Gallego, A. G. y Pascual, M. E. V. (2005). Aplicación del Análisis Discriminante y Regresión Logística en el estudio de la morosidad en las entidades financieras. Comparación de resultados. *Pecunia*, 1, 175–199.
- Rating, P. (2009). Calificación girafe -máskapital méxico- noviembre 2009. Inf. téc., Planet Rating. The Glogal Microfinance Rating Agency.
- Renso Martínez María, A. V., Cecilia Rondón (2010). Microfinanzas en las americas: las 100 mejores 2010. Report, Banco Interamericano de Desarrollo, Washington, DC.
- Rocabado, J. R., Flores, C. E., Vargas, J. D. y Peradotto, P. R. (2007). Una aplicacion del modelo de regresion logistica en la predicción del rendimiento estudiantil. *Estudios Pedagógicos XXXIII*, 2, 101–120.
- Schreiner, M. (2000). La Calificacion Estadística en las Microfinanzas: ¿Podra Funcionar? *Microfinance Risk Management and Center for Social Development*.
- Schreiner, M. (2002). Ventajas y Desventajas del Scoring Estadístico para las Microfinanzas. *Microfinance Risk Management and Center for Social Development*.
- Schreiner, M. (2008). ¿Podría el Scoring atraer a inversionistas con fines de lucro al micro-crédito? *Microfinance Risk Management*.

Referencias

- Schreiner, M. y Dellien, H. (2005). El scoring estadístico, los bancos y las microfinanzas: como lograr un balance entre el uso de tecnología y la atención personalizada. *BID. Microempresa. Informe de Avances, diciembre 2005.*, 8. No. 2.
- Simbaqueba, L. (2004). ¿qué es scoring? una visión práctica de la gestión del riesgo de crédito. Inf. téc., Instituto del Riesgo Financiero.
- Vigano, L. (1993). A credit scoring model for development banks: An African case study. *Savings and development*, 17(4), 441–482.
- Vogelgesang, U. (2003). Microfinance in Times of Crisis: The Effects of Competition, Rising Indebtness, and Economic Crisis on Repayment Behaviour. *World Development*, 31(12), 2085–2114.
- Yobas M., C. J. y P., R. (2000). Credit scoring using neural and evolutionary techniques. *IMA Journal of Management Mathematics*, 11, 111–125.

Apéndices

Apéndice A: Programas en SAS

Programa SAS que corre una regresión logística.

```
DATA maskapital;
INPUT ciclo monto buro edad ingre edocivil nohij hijest escolar zonaviv tipoviv nohab
techo piso dren agua trans electrdrm tel antineg y;
DATALINES;
7 3 3 01 0 4 1 1 3 0 3 1 4 3 3 3 3 2 0 2 0
7 3 1 03 4 3 0 0 4 1 2 4 1 3 3 3 3 2 0 1 0
7 3 1 03 4 4 1 1 1 0 3 4 2 3 3 3 0 4 0 2 0
6 3 1 03 4 4 1 1 4 0 3 1 1 3 3 2 3 2 0 2 0
6 3 3 03 4 4 0 0 1 0 3 3 2 3 3 3 0 2 0 4 0
6 3 3 03 0 4 1 1 4 1 3 4 4 3 3 3 3 2 0 3 0
...
;
PROC LOGISTIC DES DATA=maskapital;
CLASS monto buro ingre antineg trans electrdrm tipoviv nohab techo piso dren agua edad
edocivil escolar /PARAM=REF;
MODEL y=ciclo monto buro ingre antineg trans tel electrdrm zonaviv tipoviv nohab techo
piso dren agua edad edocivil escolar nohij hijest;
RUN;
```

Programa en SAS que hace una selección Backward.

```
DATA maskapital;
INPUT ciclo monto buro edad ingre edocivil nohij hijest escolar1 zonaviv tipoviv nohab
techo piso dren agua trans electrdrm tel antineg y;
DATALINES;
7 3 3 01 0 4 1 1 3 0 3 1 4 3 3 3 3 2 0 2 0
7 3 1 03 4 3 0 0 4 1 2 4 1 3 3 3 3 2 0 1 0
7 3 1 03 4 4 1 1 1 0 3 4 2 3 3 3 0 4 0 2 0
6 3 1 03 4 4 1 1 4 0 3 1 1 3 3 2 3 2 0 2 0
6 3 3 03 4 4 0 0 1 0 3 3 2 3 3 3 0 2 0 4 0
6 3 3 03 0 4 1 1 4 1 3 4 4 3 3 3 3 2 0 3 0
...
;
PROC LOGISTIC DES DATA=maskapital;
CLASS monto buro ingre antineg trans electrdrm tipoviv nohab techo piso dren agua edad
edocivil escolar1 /PARAM=REF;
MODEL Y=ciclo monto buro ingre antineg trans tel electrdrm zonaviv tipoviv nohab techo
piso dren agua edad edocivil escolar1 nohij hijest/SELECTION=B;
RUN;
```

Programa SAS que estima la estadística Hosmer-Lemeshow que evalúa el ajuste del modelo

```
DATA maskapital;
INPUT ciclo monto buro edad ingre edocivil nohij hijest escolar_i zonaviv tipoviv nohab
techo piso dren agua trans electrdrm tel antineg y;
DATALINES;
7 3 3 01 0 4 1 1 3 0 3 1 4 3 3 3 3 2 0 2 0
7 3 1 03 4 3 0 0 4 1 2 4 1 3 3 3 3 2 0 1 0
7 3 1 03 4 4 1 1 1 0 3 4 2 3 3 3 0 4 0 2 0
6 3 1 03 4 4 1 1 4 0 3 1 1 3 3 2 3 2 0 2 0
6 3 3 03 4 4 0 0 1 0 3 3 2 3 3 3 0 2 0 4 0
6 3 3 03 0 4 1 1 4 1 3 4 4 3 3 3 3 2 0 3 0
...
;
PROC LOGISTIC DES DATA=maskapital;
CLASS monto buro ingre antineg trans electrdrm tipoviv nohab techo piso dren agua edad
edocivil escolar_i /PARAM=REF;
MODEL Y=ciclo monto buro ingre antineg trans tel electrdrm zonaviv tipoviv nohab techo
piso dren agua edad edocivil escolar_i nohij hijest/LACKFIT;
RUN;
```

Programa SAS que calcula la R^2 generalizada

```
DATA maskapital;
INPUT ciclo monto buro edad ingre edocivil nohij hijest escolar_i zonaviv tipoviv nohab
techo piso dren agua trans electrdrm tel antineg y;
DATALINES;
7 3 3 01 0 4 1 1 3 0 3 1 4 3 3 3 3 2 0 2 0
7 3 1 03 4 3 0 0 4 1 2 4 1 3 3 3 3 2 0 1 0
7 3 1 03 4 4 1 1 1 0 3 4 2 3 3 3 0 4 0 2 0
6 3 1 03 4 4 1 1 4 0 3 1 1 3 3 2 3 2 0 2 0
6 3 3 03 4 4 0 0 1 0 3 3 2 3 3 3 0 2 0 4 0
6 3 3 03 0 4 1 1 4 1 3 4 4 3 3 3 3 2 0 3 0
...
;
PROC LOGISTIC DES DATA=maskapital;
CLASS monto buro ingre antineg trans electrdrm tipoviv nohab techo piso dren agua edad
edocivil escolar_i /PARAM=REF;
MODEL Y=ciclo monto buro ingre antineg trans tel electrdrm zonaviv tipoviv nohab techo
piso dren agua edad edocivil escolar_i nohij hijest/RSQ;
RUN;
```

Programa SAS que calcula sensibilidad y especificidad, y dibuja la curva ROC

```
DATA maskapital;
INPUT ciclo monto buro edad ingre edocivil nohij hijest escolari zonaviv tipoviv nohab
techo piso dren agua trans electrdrm tel antineg y;
DATALINES;
7 3 3 01 0 4 1 1 3 0 3 1 4 3 3 3 3 2 0 2 0
7 3 1 03 4 3 0 0 4 1 2 4 1 3 3 3 3 2 0 1 0
7 3 1 03 4 4 1 1 1 0 3 4 2 3 3 3 0 4 0 2 0
6 3 1 03 4 4 1 1 4 0 3 1 1 3 3 2 3 2 0 2 0
6 3 3 03 4 4 0 0 1 0 3 3 2 3 3 3 0 2 0 4 0
6 3 3 03 0 4 1 1 4 1 3 4 4 3 3 3 3 2 0 3 0
...
;
PROC LOGISTIC DES DATA=maskapital;
CLASS monto buro ingre antineg trans electrdrm tipoviv nohab techo piso dren agua edad
edocivil escolari /PARAM=REF;
MODEL Y=ciclo monto buro ingre antineg trans tel electrdrm zonaviv tipoviv nohab techo
piso dren agua edad edocivil escolari nohij hijest/OUTROC=roc;
PROC PRINT DATA=roc;
PROC PLOT DATA=roc;
PLOT _SENSIT_*_1MSPEC_='*';
RUN;
```