



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA CIVIL INDUSTRIAL

DESARROLLO DE UN MODELO DE COMPORTAMIENTO DE PAGO PARA ESTIMAR EL RIESGO DE CRÉDITO EN MICROEMPRESARIOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

VANNIA SCARLETT ALARCÓN PÉREZ

PROFESOR GUÍA:
CRISTIÁN DANILO BRAVO ROMÁN

MIEMBROS DE LA COMISIÓN:
CARLOS PULGAR ARATA
SOLSIRE GIAVERINI CEA

Este trabajo ha sido parcialmente financiado por el Instituto de Sistemas Complejos de Ingeniería

SANTIAGO DE CHILE

2017

DESARROLLO DE UN MODELO DE COMPORTAMIENTO DE PAGO PARA ESTIMAR EL RIESGO DE CRÉDITO EN MICROEMPRESARIOS

El presente trabajo propone una metodología interna para el desarrollo de un modelo de comportamiento de pago, basado en puntaje, que permita a una institución de inclusión financiera, facilitadora de microcréditos productivos sin captación de ahorros, estimar el riesgo de crédito de su cartera de clientes, cumpliendo los requisitos especificados en el Anexo N°1 del Compendio de Normas Contables de la Superintendencia de Bancos e Instituciones Financieras (SBIF).

Las etapas para la creación del modelo y manejo de la base de datos se ajustan a las fases del proceso CRISP-DM de minería de datos. Como resultado se encuentra un modelo de clasificación basado en regresión logística, estimada bajo el supuesto de independencia de las observaciones con el método de máxima verosimilitud, el que permite mejorar la toma de decisiones de renovación de créditos de la institución en un 34% por sobre discriminación aleatoria, basándose en los resultados del área bajo la curva ROC (también conocido como parámetro AUC), que muestra la proporción de casos verdaderos positivos versus falsos positivos para cada punto de corte del modelo.

A partir de los resultados de la *scorecard* de comportamiento se encontró que, por ejemplo, los clientes más jóvenes son más riesgosos que los adultos mayores, los clientes nuevos son más riesgosos que los antiguos y, que clientes que al cierre de mes han pagado más de la mitad de las cuotas del crédito son menos riesgosos que quienes aún no lo hacen. Así también, se identificaron segmentos accionables de alto riesgo de incumplimiento, a los que la empresa sí o sí debería realizar un seguimiento exhaustivo: cartera de créditos reprogramados, créditos con 30 días de mora o más y créditos con menos de 4 cuotas pagadas a la fecha de observación. Por otro lado, también se identificaron segmentos que representan una buena oportunidad de renovación como, por ejemplo, créditos que tienen un porcentaje de pago acumulado de más del 41% del total pactado a la fecha de observación. Adicionalmente, a partir de la tasa de rechazo de buenos y malos clientes para cada punto de corte del *score*, se encontró que para 625 puntos el costo promedio por renovación es relativamente bajo (\$56.883), y permite filtrar el 62,7% del total de malos pagadores, y que clientes con puntos sobre este valor pueden representar una oportunidad de pre-aprobación de créditos.

En conclusión, como resultado de este trabajo se encuentra que las exigencias que se aplican a instituciones bancarias chilenas para el desarrollo de modelos de comportamiento de pago, también pueden ser aplicadas a instituciones de inclusión financiera, obteniéndose resultados que contribuirían a mejorar la capacidad de estas entidades para discriminar entre buenos y malos pagadores, y reduciendo a su vez los costos por renovar créditos a individuos que con alta probabilidad caerán en incumplimiento. Así, se esperaría que expandir la regulación a este tipo de agentes sea algo considerable de evaluar para garantizar el crecimiento y estabilidad financiera del segmento microempresarial.

Tabla de contenido

Capítulo 1: Introducción.....	1
1.1 Título de la memoria.....	3
1.2 Objetivos.....	3
1.2.1 Objetivo General.....	3
1.2.2 Objetivos Específicos.....	4
1.3 Estructura de la memoria.....	4
1.3.1 Alcances.....	5
Capítulo 2: Marco Teórico.....	6
2.1 La microempresa y su inclusión financiera.....	6
2.2 Relevancia de la microempresa en América Latina.....	7
2.3 Caracterización de la microempresa en Chile.....	9
2.4 Entorno normativo en Riesgo de Crédito.....	11
2.5 Regulación de las microfinanzas en Chile.....	13
2.5.1 Levantamiento de información de otras instituciones.....	15
2.6 Credit Scoring.....	16
2.6.1 Behavioral Scoring.....	17
2.7 Proceso de elaboración de una scorecard.....	19
2.7.1 Medidas estadísticas.....	19
2.7.2 Tendencia Lógica.....	21
2.7.3 Regresión Logística.....	22
2.7.4 Producción final del Scorecard.....	24
2.8 Muestreo de Biron.....	26
2.8.1 Algoritmo.....	27
2.9 Metodología CRISP-DM.....	28
2.9.1 Etapa 1: Comprensión del Negocio.....	29
2.9.2 Etapa 2: Comprensión de los Datos.....	29
2.9.3 Etapa 3: Preparación de los Datos.....	30
2.9.4 Etapa 4: Modelado.....	30
2.9.5 Etapa 5: Evaluación.....	30
2.9.6 Etapa 6: Diseño de Scorecard.....	32
Capítulo 3: Metodología.....	33

3.1	Preliminares: Hipótesis de trabajo	33
3.2	Etapa 1: Comprensión del negocio	35
3.2.1	Caracterización de la organización.....	35
3.2.2	Proceso de otorgación y seguimiento de créditos.....	38
3.2.3	Variables representativas.....	44
3.3	Etapa 2: Comprensión de los datos.....	47
3.3.1	Almacenamiento.....	47
3.3.2	Descripción de los datos	47
3.3.3	Anonimización	48
3.4	Etapa 3: Preparación de los datos	51
3.4.1	Construcción de base de datos tipo panel.....	51
3.4.2	Descripción y construcción de variables de tipo application	53
3.4.3	Definición de parámetros de comportamiento.....	55
3.4.4	Definición de variable objetivo y mora	58
3.4.5	Variables de comportamiento.....	61
3.4.6	Limpieza de los datos	63
3.4.7	Transformación de variables	69
3.4.8	Selección de variables	77
3.5	Etapa 4: Modelado	83
3.6	Etapa 5: Evaluación	85
3.7	Etapa 6: Construcción de Scorecard	87
3.7.1	Definición de parámetros.....	87
Capítulo 4:	Análisis de resultados	92
4.1	Variables que no discriminan	92
4.2	Variables que resultan discriminantes	93
Capítulo 5:	Conclusiones y recomendaciones	97
Capítulo 6:	Bibliografía	100
Capítulo 7:	Anexos	103
7.1	Anexo 1: Almacenamiento	103
7.2	Anexo 2: Descripción de los datos	104
7.3	Anexo 3: Anonimización.....	106
7.4	Anexo 4: Descripción y construcción de variables de tipo application.....	107
7.5	Anexo 5: Selección de variables	112

7.6	Anexo 6: WOE de variables nominales y binarias.	114
7.7	Anexo 7: WOE de variables numéricas.	125

Índice de tablas

Tabla 2.1: Total de instituciones de microfinanzas en América Latina, por tipo de institución, según datos de 2012.....	8
Tabla 2.2: Segmentación de empresas según el nivel de ventas en UF.....	9
Tabla 2.3: Distribución por tamaño de microempresas formales.....	9
Tabla 2.4: Diferencia entre los modelos Application Scoring y Behavioral Scoring.....	16
Tabla 2.5: Poder predictivo de una característica según resultados de IV.	21
Tabla 3.1: Datos Económicos Básicos de Empresa Microfinanzas S.A.....	35
Tabla 3.2: Clientes vigentes de EMF por sucursal, año 2015.	36
Tabla 3.3: Clientes vigentes de EMF por zona, al año 2015.....	37
Tabla 3.4: Características principales de la cartera de clientes de Empresa Microfinanzas, año 2015.	38
Tabla 3.5: Medios de captación de clientes de Empresa Microfinanzas.....	39
Tabla 3.6: Principales variables que se ingresan a la base de datos en cada etapa de la otorgación de créditos.....	40
Tabla 3.7: Ejemplo de registro del plan de pago de un crédito en la base de datos.	41
Tabla 3.8: Ejemplo de registro de gestiones de cobranza en la base de datos.....	42
Tabla 3.9: Ejemplo del registro de un crédito reprogramado en el plan de pago.	42
Tabla 3.10: Código de estado de la cuota en el plan de pago, variable ST.	43
Tabla 3.11: Ejemplo del registro de un crédito refinanciado en el plan de pago.	43
Tabla 3.12: Variables que podrían ser significativas para explicar el comportamiento de pago de un microempresario.	45
Tabla 3.13: Variables relacionados con el comportamiento de pago.	46
Tabla 3.14: Variables creadas como resultado de la anonimización de datos relevantes.	49
Tabla 3.15: Ejemplo de configuración de datos tipo panel.	51
Tabla 3.16: Tabla Maestra N°1.	52
Tabla 3.17: Tabla Maestra N°2.	53
Tabla 3.18: Variables asociadas a características del cliente, clasificadas por tipo.	54
Tabla 3.19: Variables asociadas a características del negocio del cliente, clasificadas por tipo...	54
Tabla 3.20: Variables asociadas con características del crédito otorgado al cliente, clasificadas por tipo.....	55
Tabla 3.21: Ejemplo de cálculo de MoraDefault.....	58
Tabla 3.22: Ejemplo de cálculo de Default90.	59
Tabla 3.23: Ejemplo de cálculo de mora para construcción de variables de comportamiento.....	60

Tabla 3.24: Ejemplo de cálculo de mora para construcción de variables de comportamiento (2).	60
Tabla 3.25: Variables de comportamiento relacionadas con la morosidad.	61
Tabla 3.26: Variables de comportamiento relacionadas con la cobranza.	62
Tabla 3.27: Variables de comportamiento relacionadas con los pagos del crédito.	62
Tabla 3.28: Variables de comportamiento relacionadas con el crédito anterior.	63
Tabla 3.29: Descripción de variable EdadCierre, antes de realizar modificaciones.	65
Tabla 3.30: Descripción de variable EdadCierre, después de las modificaciones.	66
Tabla 3.31: Distribución en cuartiles de la variable MesesEmp.	67
Tabla 3.32: Caracterización numérica de la tabla maestra para el desarrollo del modelo.	67
Tabla 3.33: Muestreo de la base de datos. Representa el 10,5% del total.	68
Tabla 3.34: Proporción de buenos y malos clientes de la muestra.	68
Tabla 3.35: Muestra de entrenamiento y Muestra de testeo	71
Tabla 3.36: Muestra luego de incorporar los datos que el Muestreo de Biron dejó fuera.	71
Tabla 3.37: Cálculo de WOE para los atributos de la variable EdadCierre.	72
Tabla 3.38: Cálculo de WOE para los atributos de la variable CodEscolaridad.	75
Tabla 3.39: Cálculo de WOE para los atributos de la variable CodEscolaridad, agrupación de segmentos.	75
Tabla 3.40: Cálculo de WOE para los atributos de la variable CodEscolaridad, agrupación de segmentos (2).	76
Tabla 3.41: Regla sobre poder predictivo de IV en una variable.	77
Tabla 3.42: Ejemplo de modificación de segmentación de variable para mejorar el IV.	77
Tabla 3.43: Ejemplo de modificación de segmentación de variable para mejorar el IV (2).	78
Tabla 3.44: Correlación de variables muy correlacionadas con respecto a la variable objetivo.	81
Tabla 3.45: Pesos relativos para el balanceo de la muestra.	82
Tabla 3.46: Lógica según la cual se seleccionan las variables para cada modelo.	83
Tabla 3.47: Conjunto de variables incluidas en cada modelo construido.	84
Tabla 3.48: Matriz de confusión.	85
Tabla 3.49: Resultados de validación cruzada para los modelos M1 a M2, respecto al valor de AUC.	86
Tabla 3.50: Coeficientes de la nueva configuración.	87
Tabla 3.51: Selección de parámetros para el escalamiento	88
Tabla 3.52: Escala, factor y score neutro.	88
Tabla 3.53 Interpretación de rangos de score con respecto al promedio.	88
Tabla 3.54: Puntajes para cada variable	89

Tabla 3.55: Puntos de corte	90
Tabla 7.1: Bases de Datos del Sistema SAC, Emprende Microfinanzas.....	103
Tabla 7.2: Tablas principales de la base de datos de Emprende Microfinanzas y sus claves primarias.	104
Tabla 7.3: Clasificación de las tablas de las Bases de Datos: Contabilidad MF.	105
Tabla 7.4: Pasos para anonimizar los datos.....	106
Tabla 7.5: Identificadores únicos de cada fila de la base de datos.	107
Tabla 7.6: Variables asociadas a características del cliente, en orden alfabético.....	107
Tabla 7.7: Variables asociadas con características del negocio del cliente.....	109
Tabla 7.8: Variables asociadas con características del crédito otorgado al cliente.	110
Tabla 7.9: Resultados de selección de variables con Forward Selection.	112
Tabla 7.10: Resultados de selección de variables con Backward Elimination.....	113
Tabla 7.11: WOE de variable CredReprogramación y gráfico.	114
Tabla 7.12: WOE de variable SucZona y prueba de otras configuraciones.	114
Tabla 7.13: WOE de variable CredRefinanciamiento.	114
Tabla 7.14: WOE de variable ClienteNuevo2.....	115
Tabla 7.15: WOE de variable CredAntRefinanciamiento.	115
Tabla 7.16: WOE de variable CredAntReprogramación.....	115
Tabla 7.17: WOE de variable CredAntPrepagado2.....	115
Tabla 7.18: WOE de variable ComunaCasaSuc.....	115
Tabla 7.19: WOE de variable ComunaTallerSuc.	116
Tabla 7.20: WOE de variable ComunaTallerCasaSuc.	116
Tabla 7.21: WOE de variable Agrícola.	116
Tabla 7.22: WOE de variable FOSIS.	116
Tabla 7.23: WOE de variable GesCobranzaAnt.....	116
Tabla 7.24: WOE de variable CORFO.....	117
Tabla 7.25: WOE de variable TotFonos2.....	117
Tabla 7.26: WOE de variable Trabajadores2.	117
Tabla 7.27: WOE de variable CredTransferido.....	117
Tabla 7.28: WOE de variable Sexo y prueba de otras configuraciones.	117
Tabla 7.29: WOE de variable Nacionalidad y prueba de otras configuraciones.....	118
Tabla 7.30 WOE: variable de PersHogar.	118
Tabla 7.31: WOE de variable CodEscolaridad e intentos de otras configuraciones.	119
Tabla 7.32: WOE de variable CodOtrosIng e intento de otras configuraciones.	120

Tabla 7.33: WOE de variable CodEstCiv e intento de otras configuraciones.....	121
Tabla 7.34: WOE de variable Mail.....	121
Tabla 7.35: WOE de variable CasaTaller.....	121
Tabla 7.36: WOE de variable RegionCasa y otros intentos de agrupación.....	122
Tabla 7.37: WOE de variable TrbFeria.....	122
Tabla 7.38: WOE de variable Garantía.....	122
Tabla 7.39: WOE de variable TipoCanal.....	123
Tabla 7.40: WOE de variable ClasCAE y prueba de otras configuraciones.....	123
Tabla 7.41: WOE de variable IA.....	124
Tabla 7.42: WOE de variable Activos2R y gráfico.....	125
Tabla 7.43 WOE: de variable AbonoCta6R.....	125
Tabla 7.44: WOE de variable AbonoRelAtrasoCredAnt2.....	125
Tabla 7.45: WOE de variable Autonomía.....	126
Tabla 7.46: WOE de variable CreditosAnteriores y gráfico.....	126
Tabla 7.47: WOE de variable GesCobranza6R.....	126
Tabla 7.48: WOE de variable CtasConAtrasoPorcCredAnt y gráfico.....	127
Tabla 7.49: WOE de variable CtasFaltaPagarR y gráfico.....	127
Tabla 7.50: WOE de variable CtasPagPorc6R y gráfico.....	128
Tabla 7.51: WOE de variable CtasReal y gráfico.....	128
Tabla 7.52: WOE de variable CuotasPagadasPorc y prueba de otras configuraciones.....	129
Tabla 7.53: WOE de variable CuotasPagadasR y gráfico.....	130
Tabla 7.54: WOE de variable DifMtoAprobAntPorcR.....	130
Tabla 7.55: WOE de variable DifCtasCredAntPorcR, prueba de otras configuraciones y gráfico.....	131
Tabla 7.56: WOE de variable MesesEmp2 y gráfico.....	132
Tabla 7.57: WOE de variable MinMora6R y gráfico.....	132
Tabla 7.58: WOE de variable MoraProm6R y gráfico.....	133
Tabla 7.59: WOE de variable MtoAcumCtasPagadasR y gráfico.....	133
Tabla 7.60: WOE de variable MtoAprobAntR y gráfico.....	134
Tabla 7.61: WOE de variable MtoPagadoPactadoPorcR y gráfico.....	135
Tabla 7.62: WOE de variable PatrimonioR.....	135
Tabla 7.63: WOE de variable RemPromR y gráfico.....	136
Tabla 7.64: WOE de variable SobreMoraProm6 y gráfico.....	136
Tabla 7.65: WOE de variable TasaInterés2.....	137

Tabla 7.66:: WOE de variable ValorCtaPorcR y gráfico.	137
Tabla 7.67: WOE de variable ValorCtaPromR6.	137
Tabla 7.68: WOE de variable ValorCtaPromCredAntR y gráfico.	138
Tabla 7.69: WOE de variable VecesMoraBajo6R.....	138
Tabla 7.70: WOE de variable VecesMoraSubio6R y gráfico.	139
Tabla 7.71: WOE de variable VtasPromR.....	139
Tabla 7.72: WOE de variable Leverage.	140

Índice de figuras

Figura 2.1: Secuencia del desarrollo microempresarial según Muhamad Yunus.....	7
Figura 2.2: Cartera y número de clientes de microcrédito en América Latina, 2009 – 2012.....	7
Figura 2.3: Comparación de tamaño entre entidades microfinancieras reguladas y no reguladas en América Latina, según datos de 2012.....	8
Figura 2.4: Gráfico de la distribución sectorial de la microempresa, año 2016.	10
Figura 2.5: Gráfico de la distribución sectorial de la microempresa, según nivel de formalidad, año 2016.	11
Figura 2.6: Estructura operativa del acuerdo de Basilea II.....	12
Figura 2.7: Supuesto fundamental de un Behavioral Scoring.	17
Figura 2.8: Análisis de variables agrupadas.	19
Figura 2.9: Tendencia Lógica del WOE para la edad.....	21
Figura 2.10: Tendencia ilógica del WOE para la edad.....	22
Figura 2.11: Diagrama de las fases del proceso CRISP-DM	28
Figura 2.12: Componentes de la curva ROC.....	32
Figura 3.1: Organigrama de Empresa Microfinanzas.	36
Figura 3.2: Etapas del proceso de otorgación de créditos en EMF.	38
Figura 3.3: Gráfico de la distribución de clientes por tipo de canal, años 2008 a 2016:.....	39
Figura 3.4: Sub-etapas del proceso de Evaluación de créditos de Empresa Microfinanzas.	40
Figura 3.5: Cifras sobre la gestión de cobranza y casos de repactación en créditos cursados entre los años 2007 y 2016.	44
Figura 3.6: Imagen referencial del diccionario de Códigos creado en Excel.....	48
Figura 3.7: Caracterización de cómo es el resultado que entrega una función HASH.....	49
Figura 3.8: Mapa entidad relación de tablas principales y tablas anexas.	52
Figura 3.9: Gráfico que muestra el porcentaje acumulado del total de créditos de EMF según el número de cuotas pactadas.	56
Figura 3.10: Parámetros de comportamiento de Behavioral Scoring.	57
Figura 3.11: Distribución de las variables numéricas TotFonos y Trabajadores.	64
Figura 3.12: Ilustración de correlación entre variables numéricas.....	70
Figura 3.13: Discretización con árbol de decisión para la variable EdadCierre.....	71
Figura 3.14: Gráfico de tendencia de los WOE para los segmentos de EdadCierre.	74
Figura 3.15: Gráfico de tendencia de los WOE para los segmentos de Monto.....	74
Figura 3.16: Gráfico de tendencia de los WOE para los segmentos de CodEscolaridad.	76

Figura 3.17: Selección de variables nominales, según information value.....	78
Figura 3.18: Selección de variables numéricas, según information value.....	79
Figura 3.19: Ilustración de correlación entre variables transformadas por WOE.	80
Figura 3.20: Resultados de backward elimination y forward selection.....	82
Figura 3.21: Curva ROC para el modelo M3.	86
Figura 3.22: Gráfico del punto de corte del modelo, respecto al porcentaje de créditos rechazados y costo promedio del crédito.	91

Capítulo 1

Introducción

La incertidumbre es la fuente originaria de todo riesgo, palabra cuya etimología se asocia a *la proximidad de un daño o peligro* [1]. Toda decisión tomada en el presente involucrará algún riesgo en el futuro; la clave está en comprender cómo distintos factores influyen en que uno u otro resultado sea posible, lo que permitirá diseñar estrategias más acertadas y que aseguren una menor probabilidad de fracasar.

Cuando la realidad es sencilla y la decisión involucra una reducida cantidad de variables, la experiencia suele ser suficiente para tomar una decisión con alta probabilidad de éxito: por ejemplo, decidir beber agua para calmar la sed. Pero, cuando el problema es multidimensional y relativamente complejo, la intuición no es suficiente para identificar todo lo que interviene en que uno u otro evento ocurra; por ejemplo, decidir comprar acciones de cierta compañía o no.

En particular, en el mundo de las finanzas, para las entidades facilitadoras de préstamos es fundamental poder identificar *a priori*: 1) si el nuevo postulante será un buen o mal pagador, y en ese caso, decidir si otorgar o no el crédito, y 2) cuántos de sus clientes vigentes en el tiempo t dejarán de cumplir con sus obligaciones en un plazo futuro $t + \Delta$, de modo de provisionar una cantidad proporcional de capital para cubrir las pérdidas esperadas y aplicar medidas de cobranza temprana.

La toma de este tipo de decisiones se simplifica bastante con la utilización de técnicas de *Credit Scoring*, las que Anderson [2] define como el uso de modelos estadísticos para la transformación de datos relevantes en medidas numéricas, que permiten tomar decisiones crediticias de forma objetiva, rápida y consistente, respecto de una gran cantidad de individuos. Para responder a las inquietudes planteadas anteriormente, sobre las decisiones que debe tomar una institución financiera, el mismo autor menciona dos tipos de enfoques: *Application Scoring* (modelo de originación de nuevos clientes para decidir si aprobar el otorgamiento de un crédito o no) y *Behavioral Scoring* (modelo de comportamiento que permite gestionar el riesgo de crédito de la cartera de clientes vigentes).

Los resultados de un modelo de *Credit Scoring* suelen representarse en forma de *scorecards*, lo que Siddiqi describe en su libro *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, como un grupo de características estadísticamente determinadas para contribuir a la correcta identificación de buenos y malos clientes [3]. Según la metodología propuesta por el mismo autor, cada característica (por ejemplo, la EDAD), es

discretizada para construir atributos (siguiendo el ejemplo, “entre 23-25 años” sería un atributo de EDAD), a los cuales se les asigna un puntaje basado en análisis estadísticos y el poder predictivo de la característica en cuestión. La suma de todos los puntajes asociados a los atributos de interés representará qué tan buen o mal pagador es o será el individuo. Mientras un *Application Scoring* combina información demográfica del cliente, su historial crediticio e información de tipo *bureau* (en ocasiones), un *Behavioral Scoring* incorpora, además, información del comportamiento de pago del cliente en el tiempo.

La probabilidad de que el cliente no pague según lo acordado se conoce en la literatura como *riesgo de crédito*, y su correcto dimensionamiento evita que las entidades financieras quiebren, a la vez que permite mantener la estabilidad del sistema bancario, tanto a nivel país como a nivel internacional. Por esta razón, existen organismos supervisores y fiscalizadores que regulan y validan las metodologías que utilizan los bancos, a la vez que definen un marco regulatorio eficiente en el que éstos puedan operar adecuadamente, sin incurrir en pérdidas para los depositantes. A nivel internacional, el BIS (*Bank for International Settlements*) es el organismo que se preocupa actualmente de guiar a los bancos centrales en su búsqueda de estabilidad monetaria y financiera [4], temática discutida concretamente por el Comité de Supervisión Bancaria de Basilea [5], regulador mundial que ofrece un foro para la cooperación en estas materias y cuyas propuestas sobre reformas a la regulación bancaria quedan establecidas en los Acuerdos de Basilea.

Por otro lado, las instituciones que no captan ahorro, como las dedicadas a la facilitación de microcréditos (y que no corresponden a filiales bancarias), quedan fuera del marco regulatorio, careciendo de incentivos para utilizar metodologías internas en la estimación del riesgo de crédito, situación de “no regulación” que podría revertirse en los próximos años. Si bien, hasta hace poco no existían promulgaciones de Basilea dirigidas específicamente a este tipo de instituciones, la gran cantidad de entidades dedicadas a proveer microcréditos en el mundo ha llevado al Comité a publicar un *Documento Consultivo* en diciembre de 2015, con procedimientos generales para la aplicación de los *Principios Fundamentales de Basilea* a la regulación de las microfinanzas [6]. Según Rudy Araujo, Secretario General de la Asociación de Supervisores Bancarios de las Américas (ASBA), se espera que la normativa sea “*sensata y adecuada*”, pero al ser un primer intento, “*las entidades de regulación tienen que pasar por una empinada curva de aprendizaje para entender los productos y servicios*”, y “*puede ser difícil encontrar un terreno común*” [7].

En este contexto, y considerando que la cantidad de instituciones de microfinanzas no reguladas en América Latina representan el 57% del total de este tipo de entidades [8], se esperaría que los supervisores bancarios deban, en el corto plazo, proponer iniciativas que ajusten razonablemente las normas existentes, probadas con éxito para préstamos comerciales y de consumo, a la realidad de un mundo totalmente diferente como lo son las microfinanzas productivas, para lo cual es necesario entender cómo es el comportamiento financiero de los microempresarios y cuáles son las variables más relevantes para medir el riesgo de crédito en este segmento.

Dado que en Chile no existe evidencia de que alguna institución no regulada de microfinanzas utilice metodologías propuestas por Basilea para estimar el riesgo de crédito, y por lo tanto no existen precedentes para saber qué normas son aplicables o no a la realidad local de estas instituciones, como parte de este trabajo de título se construirá un modelo de *Behavioral Scoring* siguiendo las especificaciones propuestas por la Superintendencia de Bancos e Instituciones Financieras (SBIF) en el *Compendio de Normas Contables, Anexo N°1, Requisitos Específicos de Desarrollo para las metodologías internas* [9], utilizando los datos de una institución representativa de las microfinanzas productivas en Chile, Emprende Microfinanzas S.A. Como resultado, se espera descubrir si las especificaciones bancarias básicas permiten construir un modelo igualmente factible en instituciones de inclusión financiera en Chile.

Cabe destacar que el presente trabajo se enmarca en un proyecto de mayor índole, cuyos resultados contribuirán al desarrollo de una norma que respalde el adecuado funcionamiento de instituciones facilitadoras de microcréditos, tanto en Chile como en otros países con modelo de negocio similar. Por todo lo mencionado, es claro que el desarrollo de este tema de memoria es deseable, tanto para el bien particular de la empresa facilitadora de los datos, la que podrá contar con un modelo de comportamiento que agregará valor a la gestión de su riesgo de crédito, como contribuir, en términos investigativos, a disipar las interrogantes en torno a cómo debería ser un sistema financiero que incorpore regulación a entidades de microfinanzas y en línea con los avances internacionales en estas materias.

Se espera que el resultado de esta memoria no sólo sea relevante para la institución misma, sino que, además, para la Fundación Microfinanzas BBVA y el resto de entidades que forman parte de la Red para el Desarrollo de las Microfinanzas en Chile, las que podrán encontrar en este trabajo una metodología alternativa para su gestión del riesgo de crédito. Por otro lado, los resultados son de interés particular para el profesor Cristián Bravo, quién está trabajando con otras instituciones de Latinoamérica para comparar los resultados obtenidos y contribuir a la elaboración de una norma adecuada a la realidad de este segmento financiero. Finalmente, las conclusiones sobre la aplicabilidad de los requisitos básicos para el desarrollo de modelos internos (según el Compendio de Normas Contables), a instituciones de microfinanzas podrían ser relevante para el organismo supervisor SBIF, en caso de que se exija emitir una futura norma operativa en estas materias.

1.1 Título de la memoria

Desarrollo de un modelo de comportamiento de pago para estimar el riesgo de crédito en microempresarios.

1.2 Objetivos

1.2.1 Objetivo General

Construir un modelo de comportamiento de pago que permita estimar el riesgo de crédito de la cartera de clientes de Emprende Microfinanzas S.A, según los estándares bancarios modernos.

1.2.2 Objetivos Específicos

- I. Comprender la situación actual del tema a desarrollar utilizando bibliografía pertinente.
- II. Construir y preparar la base de datos que permita desarrollar y evaluar un modelo de comportamiento de pago que estime el riesgo de crédito de la cartera de clientes, para Emprénde Microfinanzas.
- III. Identificar el modelo de comportamiento que discrimine mejor entre un buen y mal pagador.
- IV. Transformar los resultados del modelo de forma tal que su comprensión sea transversal a todos los tomadores de decisiones del negocio.
- V. Analizar los resultados del modelo, específicamente las variables más significativas para explicar el fenómeno de *incumplimiento* en microfinanzas.

1.3 Estructura de la memoria

La estructura para el desarrollo de esta memoria se encuentra alineada con los objetivos específicos anteriormente planteados y se basa en la metodología CRISP-DM de minería de datos.

El capítulo de Marco Teórico provee de contexto referencial sobre el tema abordado, que menciona la importancia del desarrollo de la microempresa en América Latina y en Chile, y cómo la falta de un marco regulatorio ha incentivado el uso de metodologías sub-óptimas por parte de las instituciones de inclusión financiera en el país. El capítulo finaliza con un marco temático que describe las metodologías de *Credit Scoring* y CRISP-DM, utilizadas para el desarrollo de esta memoria.

En el capítulo 3 se describe la implementación de cada etapa de la metodología CRISP-DM aplicada a la base de datos de la institución: 1) Comprensión del Negocio, 2) Comprensión de los Datos, 3) Preparación de los Datos, 4) Modelado, 5) Evaluación y 6) Construcción de *Scorecard*.

El capítulo 4 contempla el análisis de resultados enfocado a las variables que resultan discriminantes a la hora de clasificar entre buenos y malos pagadores de la cartera de clientes vigentes, mencionando también qué variables resultan no significativas, hallazgo igualmente importante. Finalmente, en el capítulo 5 se agregan las conclusiones y comentarios finales del trabajo realizado, además de la propuesta de futuras líneas de investigación.

1.3.1 Alcances

Para el desarrollo de la *scorecard* de comportamiento se definen los siguientes alcances del trabajo:

- 1) El modelo se construirá tomando en consideración las exigencias de desarrollo del Anexo 1 del Compendio de Normas Contables publicado por SBIF.
- 2) La metodología utilizada en el modelo es IRB Fundacional, enfocándose sólo en la construcción del modelo de *scorecard* de no pago.
- 3) Sólo se incluirán créditos que no consideran subcréditos (líneas adicionales).
- 4) No se realizará análisis costo beneficio del modelo aplicado a la empresa.
- 5) La implementación del modelo y sus directrices para llevarla a cabo no se mencionarán en este trabajo.
- 6) Solo se implementará el algoritmo de Regresión Logística, sin comparar con resultados que podrían obtenerse con otros modelos de clasificación.

Capítulo 2

Marco Teórico

2.1 La microempresa y su inclusión financiera

Intuitivamente, se esperaría que un resultado de no pago fuese más difícil de predecir si el cliente pertenece a un segmento heterogéneo y económicamente vulnerable, características propias de la microempresa. Este grupo se compone, en su mayoría, por trabajadores independientes que buscan complementar con una actividad productiva sus ingresos familiares, los que a veces son insuficientes para acceder a los servicios básicos. Heterogéneos en términos de ingresos, tamaño, educación, grado de formalidad y ámbito productivo, la principal motivación que tienen es mejorar su calidad de vida y superar la situación de pobreza en la que viven [10].

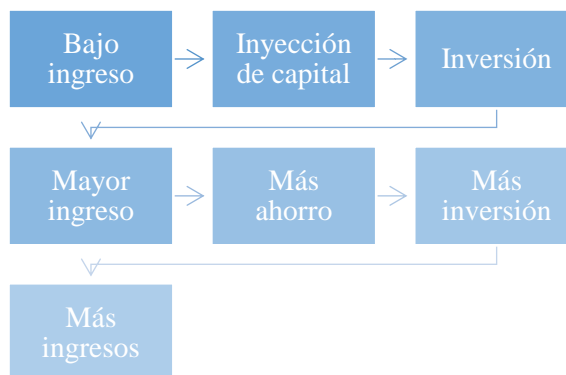
Al no ser un negocio atractivamente rentable para la banca tradicional, la inclusión financiera de la microempresa se mantuvo postergada en el mundo hasta el año 1976, cuando el economista Muhammad Yunus [11] crea en Bangladesh, India, el primer proyecto de responsabilidad social dedicado a otorgar servicios bancarios a los pobres.

El economista vio cómo la crisis financiera que sacudía a su país dejaba a miles de personas sin trabajo. Familias que necesitaban reinventarse y valerse de sus propios medios para generar ingresos no eran apoyadas por la banca tradicional, obligándolos a recurrir a prestamistas que cobraban intereses desmedidos. Ante la problemática, Yunus se propuso comprender la realidad de los pobres y estudiar su comportamiento financiero: prestó dinero propio a 42 personas, con la única condición de reinvertirlo y devolverlo en pequeñas cuotas que se adecuaban a su capacidad de pago, pero a pesar de que todos cumplieron puntualmente, estos antecedentes no fueron suficiente para que los bancos accedieran a entregarles créditos.

Impulsado por los buenos resultados, Yunus crea e institucionaliza el primer banco para los pobres, al que llamó *Grameen Bank*¹. El objetivo del banco es simple: cambiar el antiguo círculo vicioso (bajo ingreso, bajo ahorro, baja inversión) por un nuevo círculo virtuoso:

¹ En el año 2006, el éxito del Grameen Bank, o “Banco de los Pobres”, fue reconocido con el Premio Nobel de la Paz, junto a su fundador. Lemas como: “*las personas no tienen que ir al banco, es el banco el que tiene que ir a las personas*” o “*el derecho al crédito debería ser un derecho humano*”, han inspirado a más de 40 países para imitar la iniciativa.

Figura 2.1: Secuencia del desarrollo microempresarial según Muhamad Yunus.

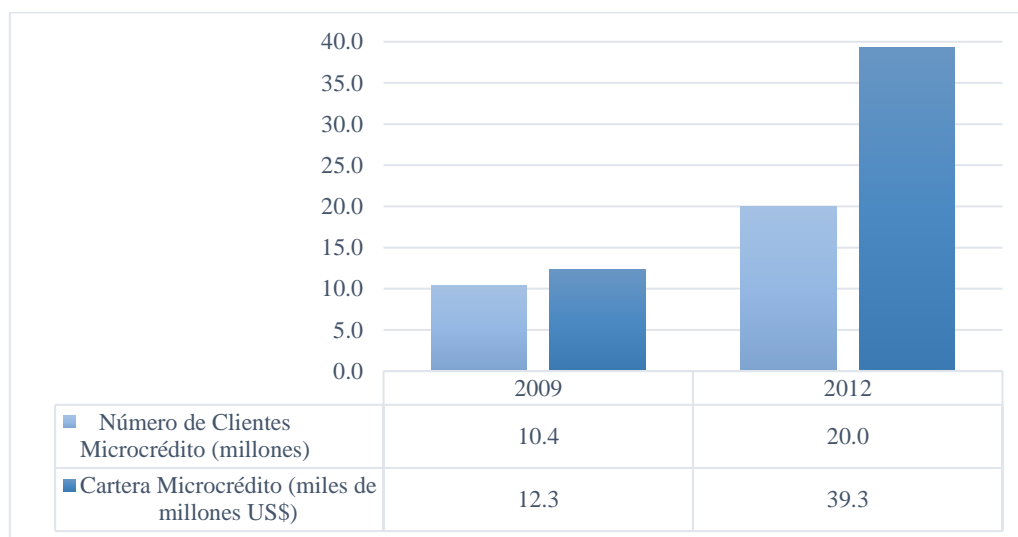


Fuente: Elaboración propia, en base a los objetivos establecidos por *Grameen Bank*.

2.2 Relevancia de la microempresa en América Latina

Desde la creación del *Grameen Bank*, la idea del microcrédito como alternativa para superar la pobreza se expandió rápidamente por el mundo. Según estudios impulsados por el Banco Interamericano de Desarrollo (BID), sólo en América Latina la cantidad de microempresarios atendidos ha crecido al menos un 48% en los últimos años, aumentando de 10.4 millones de clientes en el 2009, a una cobertura de 20.09 millones de clientes en 2012:

Figura 2.2: Cartera y número de clientes de microcrédito en América Latina, 2009 – 2012.



Fuente: Elaboración propia, según datos recolectados por Paola A. Pedroza y Sergio Navajas (FOMIN - 2009) y datos recolectados por Verónica Trujillo y Sergio Navajas (FOMIN – 2012).

Ese mismo año, se registró un total de 1 076 instituciones facilitadoras de microcréditos en el sector, de las cuales un 43% corresponde a instituciones supervisadas y reguladas por una

superintendencia o autoridad financiera equivalente, y el 57% restante a Cooperativas, ONG's y otras formas institucionales [8] que, al no captar ahorros de depositantes, podrían incurrir en quiebra sin desestabilizar el sistema:

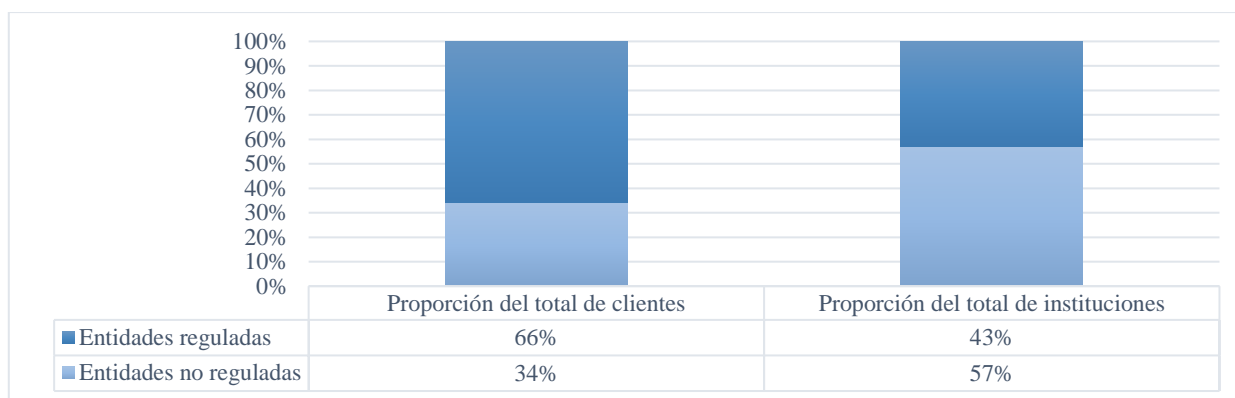
Tabla 2.1: Total de instituciones de microfinanzas en América Latina, por tipo de institución, según datos de 2012.

	Entidades Reguladas	Entidades No Reguladas
Total	461	615
Proporción del total	43%	57%

Fuente: Elaboración propia, según datos recolectados por Verónica Trujillo y Sergio Navajas (FOMIN – 2012).

Un dato importante se revela al observar que, del total de microempresarios atendidos, más de la mitad corresponde a clientes de una institución regulada, cuyos montos promedio de crédito fluctúan alrededor de los US\$2.500, aproximadamente 3 veces el monto promedio de crédito en una institución no regulada, el que se acerca a los US\$800:

Figura 2.3: Comparación de tamaño entre entidades microfinancieras reguladas y no reguladas en América Latina, según datos de 2012.



Fuente: Elaboración propia, según datos recolectados por Verónica Trujillo y Sergio Navajas (FOMIN – 2012).

Este último dato podría implicar que los microempresarios más vulnerables, aquellos que no clasifican como sujeto de crédito ante las instituciones reguladas por sus bajos ingresos, alto riesgo social e imposibilidad de otorgar garantías, son captados el resto de instituciones de inclusión financiera de menor tamaño, las cuales sólo pueden otorgarles bajos montos de financiamiento.

2.3 Caracterización de la microempresa en Chile

Según el BID, en América Latina y el Caribe no existe una única definición sobre lo que es una microempresa, criterio que varía según cada país en función de dos características principales: el número de empleados y el volumen del negocio en ventas. En Chile, según la Ley 20.416 [12], que fija las normas especiales para las empresas de menor tamaño: “son microempresas aquellas empresas cuyos ingresos anuales por ventas y servicios, y otras actividades del giro, no hayan superado las 2.400 unidades de fomento en el último año calendario (...)” Respecto a los empleadores, para efectos laborales, la ley señala que “se entenderá por microempresa aquella que tuviere contratados de 1 a 9 trabajadores (...)”.

Mientras que para estudios descriptivos del segmento se suele utilizar la clasificación por número de trabajadores, por ejemplo, en el caso de la aplicación de la Encuesta de Microemprendimiento (EME), las instituciones que prestan servicios financieros a la microempresa se basan en el criterio de las ventas anuales ya mencionado:

Tabla 2.2: Segmentación de empresas según el nivel de ventas en UF.

Tamaño de la empresa	Ventas anuales (UF)
Microempresa	De 0,1 a 2.400
Pequeña empresa	De 2.401 a 25.000
Mediana empresa	De 25.001 a 100.000
Gran empresa	Superior a 100.001

Fuente: Elaboración propia según datos del Sistema de Impuestos Internos.

En el año 2012, la microempresa representaba el 75,2% del total de empresas formales con ventas mayores a cero en el país, según registros del SII. Si a esto se suma la contribución de la pequeña empresa, la cifra asciende a un 95,5% (contabilizando sólo empresas activas). Dentro del segmento microempresarial, la composición por tramos ha evolucionado de la siguiente manera:

Tabla 2.3: Distribución por tamaño de microempresas formales.

Año	Subtramo N1		Subtramo N2		Subtramo N3	
	Ventas anuales de 0,1 a 200UF		Ventas anuales de 200,1 a 600UF		Ventas anuales de 600,1 a 2400UF	
	Número de empresas	%	Número de empresas	%	Número de empresas	%
2005	264.306	44,0%	163.473	27,2%	172.666	28,8%
2006	266.099	43,7%	165.613	27,2%	177.197	29,1%
2007	263.848	43,2%	165.540	27,1%	180.862	29,6%
2008	262.604	42,7%	167.794	27,3%	184.988	30,1%
2009	259.621	42,2%	169.509	27,5%	186.689	30,3%

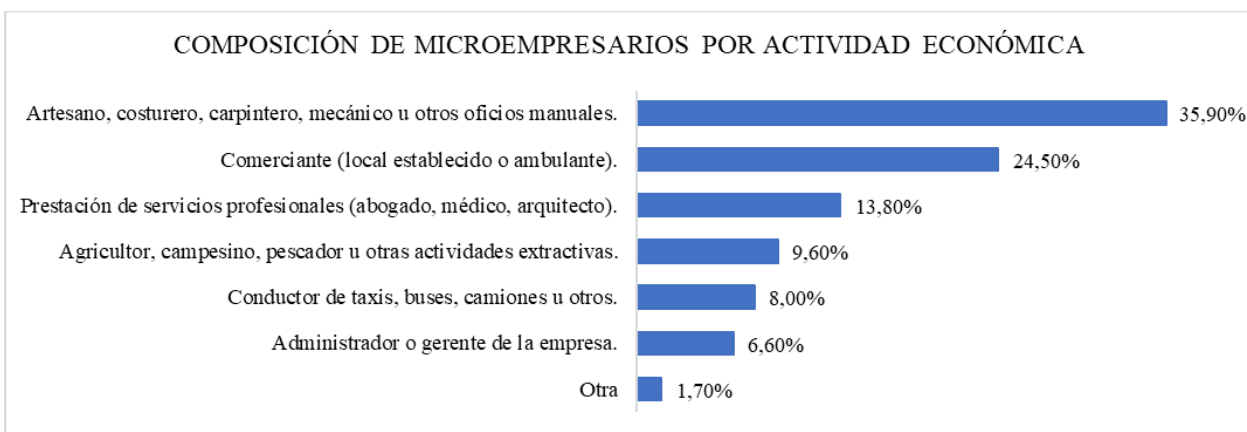
Año	Subtramo N1		Subtramo N2		Subtramo N3	
	Ventas anuales de 0,1 a 200UF		Ventas anuales de 200,1 a 600UF		Ventas anuales de 600,1 a 2400UF	
	Número de empresas	%	Número de empresas	%	Número de empresas	%
2010	255.335	41,3%	169.868	27,4%	193.761	31,3%
2011	252.402	40,1%	173.347	27,5%	203.472	32,3%
2012	250.610	39,1%	176.907	27,6%	213.324	33,3%
2013	249.901	38,4%	179.772	27,6%	221.710	34,0%
2014	259.980	38,6%	185.350	27,5%	227.696	33,8%
2015	261.164	38,2%	187.537	27,4%	234.503	34,3%

Fuente: Elaboración propia a partir de datos publicados por SII, en documento de estadísticas de empresas por tamaño según ventas (2005-2015).

Mientras el segmento N1 es el más vulnerable de la microempresa, el N3 es el que a futuro se espera que se convierta en pequeña empresa. Se puede observar que efectivamente ha ido disminuyendo el primer segmento e incrementándose el tercero, lo que indica que el escalamiento sí se produce. Según datos de la encuesta EME 2016 [13], en Chile existen cerca de 1.8 millones de microempresas, incluyendo a los trabajadores por cuenta propia (TCP), de las cuales un 48,7% tiene iniciación de actividades, mientras que el 51,3% restante opera de manera informal.

Respecto a la distribución por sector de actividad económica de las microempresas formales, el estudio revela que un 35,9% se concentra en el rubro de Artesano(a), Costurero, Carpintero, Mecánico u otros oficios manuales, seguido por Comerciante Establecido o Ambulante (24,5%), Prestación de servicios profesionales (13,8%), por sobre otros, como se observa en el siguiente gráfico:

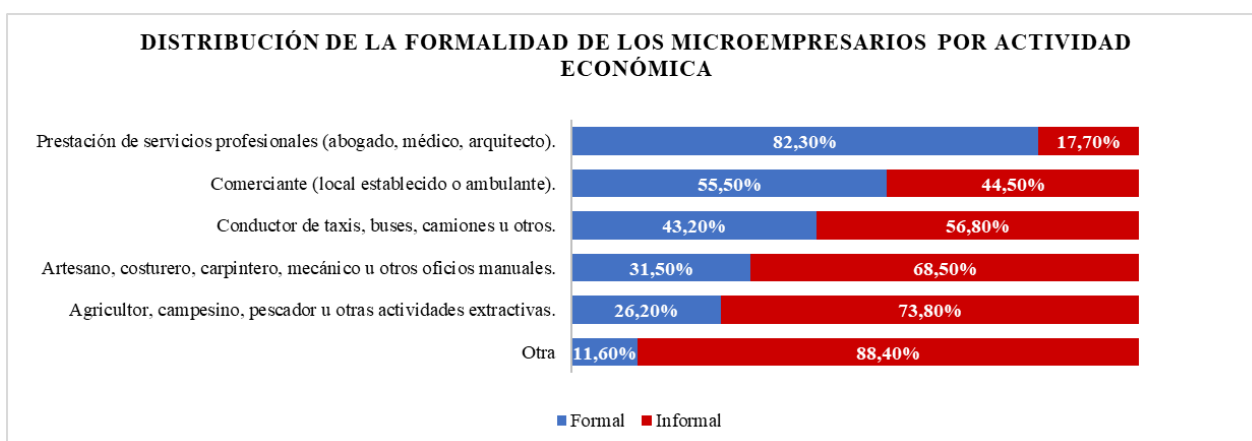
Figura 2.4: Gráfico de la distribución sectorial de la microempresa, año 2016.



Fuente: Elaboración propia a partir de información de la Cuarta EME, 2016.

Respecto a la formalidad, el 48,7% tiene iniciación de actividades ante el SII, mientras el 51,3% restante opera de manera informal, de los cuales el 41,3% considera no esencial su registro. El segmento de la microempresa que presenta un mayor porcentaje de formalidad es el de servicios profesionales (82,3%), seguido por los comerciantes (55,5%), mientras que la informalidad está más presente en actividades como: agricultura, pesca y otras de tipo extractiva, representando un 73,8% del total, como se puede apreciar en el siguiente gráfico:

Figura 2.5: Gráfico de la distribución sectorial de la microempresa, según nivel de formalidad, año 2016.



Fuente: Elaboración propia a partir de información de la Cuarta EME, 2016.

Desde un punto de vista demográfico, la encuesta EME 2016 revela que la edad promedio de los microempresarios es de 50,7 años, que el 61,3% corresponde a hombres y el 38,7% corresponde a mujeres. Otros datos interesantes para comprender al segmento son los siguientes:

- Tienen una experiencia en su negocio de 15,6 años en promedio.
- Un 43% del total ha permanecido 10 o más años en su negocio.
- Un 74,4% tiene una empresa unipersonal, un 13,6% tiene una empresa con dos personas más y el 12% trabaja con 3 o más personas.
- El 46,3% gana mensualmente un ingreso entre \$0 y \$225.000.
- El 27,3% de los microempresarios menciona que el principal aspecto que limita el crecimiento de su empresa es la falta de financiamiento.

2.4 Entorno normativo en Riesgo de Crédito

A nivel mundial, la preocupación por garantizar la solvencia y estabilidad del sistema bancario ante la creciente internacionalización de los mercados financieros, lleva al Comité de

Basilea a publicar el Primer Acuerdo de Capitales [13] en 1988, conocido también como Basilea I. La experiencia posterior puso de manifiesto las debilidades de la norma, que no discriminaba suficientemente bien los distintos niveles de riesgo de crédito ni incentivaba la utilización de técnicas para su reducción.

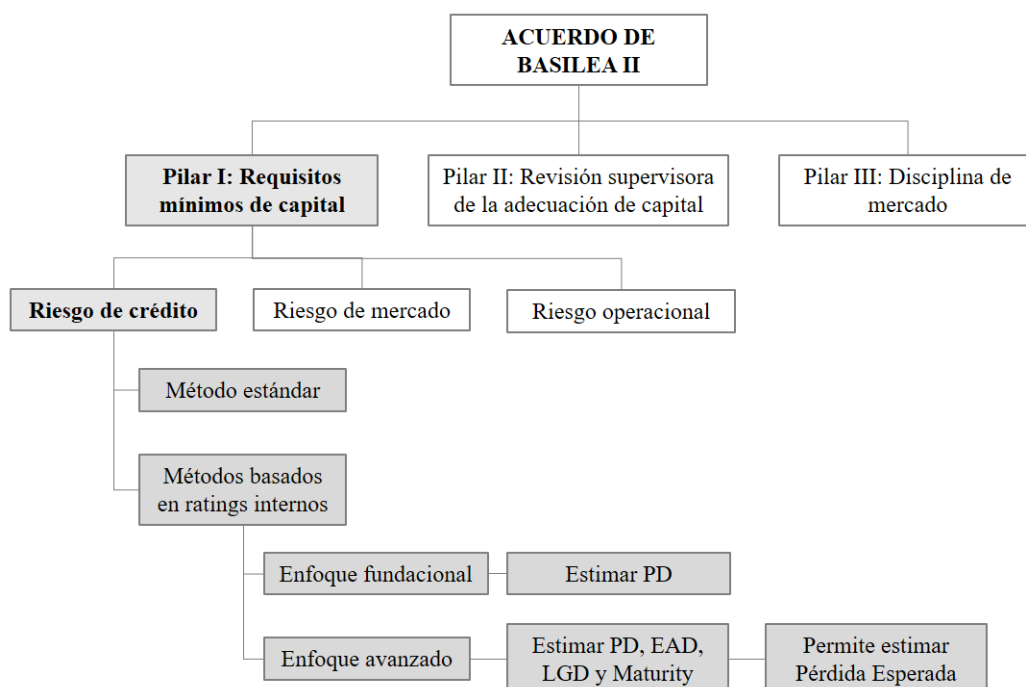
Un segundo acuerdo, conocido como Basilea II y publicado en 2004, introduce el concepto de Pilares y permite la incorporación de ponderaciones externas para el cálculo del riesgo de crédito (dadas por entidades calificadoras) o el desarrollo de métodos internos. Así, la estimación del riesgo de crédito, puede llevarse a cabo mediante tres métodos alternativos [14]:

1. Método estándar: en función de entidades clasificadoras externas y/o ponderaciones establecidas por el supervisor.

Método IRB (*Internal Ratings Based*): utilizando sistemas propios de calificación interna, sujeto a la aprobación explícita del supervisor. Dentro de estos métodos se distinguen:

2. IRB fundacional: Se limita al cálculo de la PD (*Probability of Default o Probabilidad de Incumplimiento*).
3. IRB avanzado: Permite estimar la PD, EAD (*Exposure at Default o Exposición al Incumplimiento*), LGD (*Loss Given Default o Severidad de la Pérdida*) y M (*Maturity o Madurez*).

Figura 2.6: Estructura operativa del acuerdo de Basilea II.



Fuente: Elaboración propia, según mapa conceptual presentado en el documento “Acuerdo de Basilea II: un desafío para la Banca Chilena”, [González y Umañana, 2006].

En diciembre de 2010, un año después de la gran crisis financiera del 2009, el Comité de Basilea publica una nueva norma regulatoria, conocida como Basilea III, la que propone a los organismos supervisores incrementar aún más las exigencias de capital. Esto podría intensificar el desplazamiento de los microempresarios más riesgosos hacia la cartera de instituciones no reguladas, las que deberán idear estrategias para incrementar el retorno de sus colocaciones si quieren continuar siendo rentables.

Para que una colocación genere valor económico, se debe cumplir la siguiente desigualdad:

$$(r_D + s) - \%Costos\ Operativos - PE \geq r_D * \left(\frac{D}{A}\right) + r_E * \left(\frac{E}{A}\right) \quad (1)$$

Donde:

r_A : retorno del activo.

A : activo ponderado por riesgo.

r_D : costo de la deuda (o costo de fondo).

D : deuda.

r_E : costo del patrimonio (dado por el regulador).

E : patrimonio.

PE : pérdidas esperadas.

Como se aprecia en la ecuación anterior, los ingresos de la empresa están dados por la tasa de interés que cobra, menos la proporción de costos operativos y menos las pérdidas esperadas. La mayoría de los términos de la ecuación, exceptuando PE , son valores que no se modifican a corto plazo, mientras que D/A y E/A están fijados por el regulador. Por otro lado, los costos operativos se caracterizan por ser particularmente elevados en este negocio, debido a las grandes asimetrías de información existentes entre el microempresario y la entidad.

De no ser posible disminuir los costos, y si las pérdidas esperadas son muy altas, el resultado podría ser insolvencia o quiebra. Para evitarlo, las entidades no reguladas tienden a modificar la tasa sobre el patrimonio, disminuyendo este valor y con ello la rentabilidad para los accionistas. Aunque la alternativa más sensata pudiese ser mejorar la estimación de las pérdidas esperadas desarrollando modelos de riesgo de crédito más precisos, la falta de incentivos para invertir en modelos internos cuando no se está regulado, sumado a los costos de desarrollo e implementación, imposibilitan seguir esta línea.

2.5 Regulación de las microfinanzas en Chile

En Chile, la experiencia en el ámbito de las microfinanzas se remonta a los años ochenta [6], luego de que el país experimentara una fuerte crisis económica, que tuvo como consecuencia

altos índices de cesantía, la quiebra de grandes empresas y el sistema bancario al borde del colapso. A pesar de lo desfavorable del panorama, miles de chilenos comenzaron a levantar sus propios negocios de subsistencia en poblaciones y barrios, lo que inspiró la necesidad de crear una herramienta financiera diferente, que permitiera inyectarles el capital de trabajo necesario.

En 1992, Basilea I entra en vigor en el país, otorgando facultades a la Superintendencia de Bancos e Instituciones Financieras (SBIF) para una adecuada supervisión preventiva, incorporando la evaluación de gestión y solvencia de las empresas que reciben recursos del público, lo que con los años se convertiría en una desventaja para las instituciones no reguladas, las que carecerían de incentivos para mejorar su gestión del riesgo.

Luego de la publicación de Basilea II, la regulación vigente no ha hecho más que intensificarse; una norma que es compartida por todas las entidades financieras es la fijación de una Tasa Máxima Convencional, que establece un límite al cobro de los intereses por crédito. Un efecto de esta última medida ha sido acotar el retorno sobre la colocación para las instituciones que no cuentan con metodologías sofisticadas para la estimación del Riesgo de Crédito, disminuyendo aún más las utilidades que perciben las instituciones no reguladas.

A pesar de lo anterior, la cantidad de instituciones de microfinanzas en Chile es no menor, alcanzando un total de 34 en el año 2012, según datos del BID. De ellas, 8 corresponden a instituciones no reguladas, con 149.459 de clientes (16% del total), una cartera de 227 millones de dólares (5% del total), y un crédito promedio de US\$ 1.519 por cliente.

En un reporte posterior de FOMIN (Fondo Multilateral de Inversiones, parte del grupo BID, Banco Interamericano de Desarrollo), realizado por *The Economist Intelligence Unit* el 2015, se analiza el entorno para la inclusión financiera de Chile, concluyéndose que en el país “*existe una definición de microcrédito, pero no un marco de gestión de riesgos diferenciado*”, que se “*carece de una normativa específica para los microcréditos y para los microseguros*” y que “*los proveedores más pequeños se topan con obstáculos para ofrecer estos productos debido a la extensa lista de requisitos que deben cumplirse antes de colocarlos en el mercado*” y “*pueden encontrar problemas cuando quieran adaptar sus productos a la realidad del segmento de consumidores de bajos ingresos.*”

Estudios como éste, que dejan entrever que una norma específica para las microfinanzas podría ser una alternativa para nivelar el terreno entre instituciones bancarias e instituciones de inclusión financiera, no son los únicos. Según el documento *Microfinance Consensus Guidelines: Guiding Principles on Regulation and Supervision of Microfinance*, “*la experiencia indica que las microfinanzas alcanzan su mayor potencial de desarrollo cuando operan como entidades reguladas, lo que implica supervisión prudencial*” [15]. Así, en el texto se menciona que el enfoque debe ser lo suficientemente flexible, para no restringir excesivamente su oferta y expansión, pero a su vez rígido, para no permitir que operen instituciones débiles, que a su vez puedan afectar la confianza en esta industria. [16]

Hasta hace poco, no existían promulgaciones de Basilea dirigidas específicamente a instituciones de inclusión financiera. Sólo hace unos meses, en diciembre de 2015, el comité publicó un documento consultivo, con el objetivo de ser una guía sobre cómo aplicar los Principios Fundamentales de la Regulación de Basilea a la supervisión de instituciones relevantes para la inclusión financiera, dentro de las cuales se encuentran las entidades de microfinanzas, sin dejar fuera a aquellas que no reciben ahorros. Si bien no se establece ninguna norma operativa y sólo preceptos para el buen funcionamiento del sistema en general, sí marca un precedente importante que, de una u otra forma, podría significar una próxima nueva regulación enfocada en la realidad de las microfinanzas.

En este contexto, los supervisores bancarios deberán, en el corto plazo, proponer iniciativas que ajusten razonablemente las normas existentes, probadas con éxito para préstamos comerciales y de consumo, a la realidad de un mundo totalmente diferente como lo son las microfinanzas productivas. Para ello, es necesario analizar el mercado y, cómo hizo Muhamad Yunnus en los años 70, comprender cómo es el comportamiento financiero de los microempresarios hoy.

2.5.1 Levantamiento de información de otras instituciones

Como parte de la investigación y con el objetivo de justificar la realización de un modelo de comportamiento de pago para el segmento de las microfinanzas en Chile, se llevó a cabo una serie de entrevistas con agentes de las principales instituciones facilitadoras de microcréditos. Los principales hallazgos se exponen a continuación, e incluyen el testimonio de la cooperativa de créditos que atienden más microcréditos en Chile, Coopeuch, seguida de la institución bancaria con más clientes del segmento microempresarial en Chile, Banco Estado Microempresa:

Pablo Coloma Correa (Gerente General de Empeude Microfinanzas y Presidente del Directorio de Red Microfinanzas) declara que, en general, lo que subsiste en microfinanzas es un modelo basado en la experiencia y capacidad de análisis en terreno de un ejecutivo. Si bien los modelos se utilizan como apoyo, en general lo que manda actualmente es el juicio experto de un ejecutivo y un comité.

Manuel Vivanco Álvarez (Sub-Gerente de Modelos de Evaluación de Coopeuch) aclara que Coopeuch no tiene un modelo de *behavioral scoring*. La razón, dice, es que no tienen una cantidad de variables suficientes y cuentan con escasa data histórica (desde el 2012 en adelante). Actualmente, la institución utiliza una matriz de pérdida para calcular las provisiones del segmento microempresa, siendo su principal desventaja el hecho de que este método no permite proyectar una pérdida esperada futura.

Rodolfo Bahamondes (Sub-Gerente de Modelos Tecnológicos de Riesgos de Banco Estado) aclara que actualmente no cuentan con un modelo de *behavioral scoring* para medir el riesgo de crédito en microempresa y que, si bien, una institución externa construyó uno, aún no se ha implementado. Actualmente utilizan un modelo de regresión logística para el cálculo de provisiones.

Como conclusión, se encuentra que no existen antecedentes de que una institución haya validado ante la SBIF un *behavioral scoring* para el segmento de microempresa en Chile, lo que motiva a entender por qué se ha mantenido esto en el tiempo y si es que la norma bancaria permite llegar a resultados satisfactorios o no.

2.6 Credit Scoring

Corresponde a “un conjunto de modelos de decisiones y sus técnicas subyacentes que ayudan a las instituciones financieras en el proceso de otorgar crédito” [17]. Específicamente, el *Credit Scoring* es utilizado para evaluar el nivel de riesgo asociado a un prospecto, y permite tomar decisiones claras al respecto. Algunos beneficios del *Credit Scoring* son:

- 1) Provee de criterios estándares para predecir el comportamiento de pago del deudor, dado que el modelo provee un análisis objetivo. Se enfoca sólo en información relacionada con el riesgo crediticio y evita la subjetividad de quienes realicen el proceso de evaluación.
- 2) Aumenta la velocidad de análisis de la cartera: permite la automatización del cálculo de provisiones y cuantificación del riesgo asociado a cada crédito vigente.
- 3) Aporta información para gestionar de forma anticipada futuras renegociaciones o acciones que eviten el no pago de algún crédito vigente.
- 4) Fácil de interpretar por una amplia gama de tomadores de decisiones, que no necesariamente tienen conocimientos avanzados de estadística.

Así, un modelo de *Scoring* puede ser diseñado para ser utilizado en la etapa de selección de nuevos clientes (*Application Scoring*) o en el seguimiento a los clientes vigentes (*Behavioral Scoring*). Las diferencias de aplicación entre ambos modelos son las siguientes:

Tabla 2.4: Diferencia entre los modelos *Application Scoring* y *Behavioral Scoring*.

Originación <i>(Application Scoring)</i>	Comportamiento <i>(Behavioral Scoring)</i>
<i>Objetivo: Pronosticar la calidad crediticia de las solicitudes de crédito.</i>	<i>Objetivo: Evaluar contratos vigentes incorporando información de comportamiento del cliente.</i>
<ul style="list-style-type: none"> • Análisis a Priori. 	<ul style="list-style-type: none"> • Análisis a Posteriori.

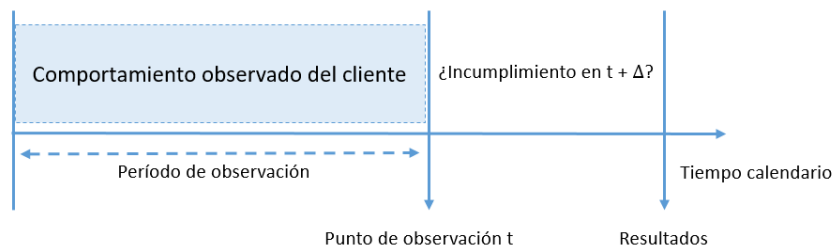
Originación <i>(Application Scoring)</i>	Comportamiento <i>(Behavioral Scoring)</i>
<ul style="list-style-type: none"> Se emplean variables del cliente y tipo de contrato: edad, renta, endeudamiento, tipo de solicitud, canal de captación, etc. 	<ul style="list-style-type: none"> Se emplean variables de la operación y del cliente disponibles internamente: edad, renta, endeudamiento, tipo de solicitud, canal de captación, retraso en pagos, mora, etc.
<ul style="list-style-type: none"> Sus usos son: rechazar o aceptar el servicio, solicitar pago inicial más algo o colaterales, asignar primer crédito pequeño, etc. 	<ul style="list-style-type: none"> Sus usos son: realizar revisión de límite de tarjeta, seguimiento de riesgo, ofrecer mejores tasas o renovar servicio, actualizaciones de productos, etc.

Fuente: Elaboración propia a partir de información publicada en libro *Credit Risk Scorecards* [Naem Siddiqi, 2006].

2.6.1 Behavioral Scoring

Un modelo de *Behavioral Scoring* tiene como supuesto fundamental el hecho de que “*las ocurrencias en el futuro son un reflejo de lo que ocurrió en el pasado*”. Así, para cada instante del tiempo t , se debe determinar una ventana de tiempo futura para definir el incumplimiento y una ventana de tiempo pasada para construir variables de comportamiento que expliquen este evento en t :

Figura 2.7: Supuesto fundamental de un Behavioral Scoring.



Fuente: Elaboración propia a partir de información publicada en libro *Credit Risk Scorecards* [Naem Siddiqi, 2006].

En la ilustración anterior se define:

- **Período de observación:** período de observación histórica que permite resumir el comportamiento observado del crédito para el instante t .
- **Punto de observación:** instante t en que se observa la situación actual del crédito.
- **Resultados:** instante de tiempo futuro en que se observa la evolución del crédito y si es que cayó o no en incumplimiento.

Dependiendo del negocio, el período de observación puede ser 3, 6, 9 o 12 meses, lo mismo para la ventana de observación a futuro. Tomando en cuenta esto, se construyen variables que

permitan representar, para cada instante t , la historia de comportamiento de pago del período de observación. Sea x una variable que registra, por ejemplo, los días de atraso en el período t , algunos ejemplos de variables de comportamiento que se pueden construir son:

- 1) Valor de la variable x en $t - 1$:

$$f(X) = X_{t-1}$$

Ejemplo: Días de atraso en el mes anterior.

- 2) Promedio de la variable x en sobre los últimos n meses:

$$f(X) = \frac{1}{n} \sum_{i=1}^n X_{t-i}$$

Ejemplo: Promedio de los días de atraso en los últimos 6 meses.

- 3) Mínimo de la variable x para los últimos n meses:

$$f(X) = \min_{i=1, \dots, n} (X_{t-i})$$

Ejemplo: Mínima cantidad de días de atraso en los últimos 6 meses.

- 4) Máximo de la variable x para los últimos n meses:

$$f(X) = \max_{i=1, \dots, n} (X_{t-i})$$

Ejemplo: Máxima cantidad de días de atraso en los últimos 6 meses.

- 5) Número de decrecimiento de la variable x durante los últimos n meses:

$$f(X) = \sum_{i=1}^{n-1} [X_{t-i} < X_{t-(i+1)}]$$

Ejemplo: Total de meses en que los días de atraso disminuyeron con respecto al mes anterior, en los últimos 6 meses.

- 6) Número de incrementos de la variable x durante los últimos n meses:

$$f(X) = \sum_{i=1}^{n-1} [X_{t-i} > X_{t-(i+1)}]$$

Ejemplo: Total de meses en que los días de atraso aumentaron con respecto al mes anterior, en los últimos 6 meses.

7) Número de veces que la variable toma un valor positivo en los últimos n meses:

$$f(X) = \sum_{i=1}^n [X_{t-i} > 0]$$

Ejemplo: Total de meses en que existieron días de atraso, en los últimos 6 meses.

2.7 Proceso de elaboración de una scorecard

2.7.1 Medidas estadísticas

La Figura 2.8 muestra un cuadro típico utilizado en el análisis de características agrupadas. El ejemplo muestra la característica “Edad” después de que ha sido agrupada en 5 atributos. En el ejemplo, “Buenos” corresponde a la cantidad de casos que no caen en incumplimiento para cada segmento, “Malos” corresponde a la cantidad de casos que sí caen en incumplimiento para cada segmento y “Total” es la suma de ambos. Luego, “%Total”, “%Buenos”, y “%Malos” corresponden a la distribución porcentual del total y de buenos y de malos, respectivamente, que representa cada atributo. Por ejemplo, para el grupo etario [28-43], se encuentra que concentra un 32% del total de casos observados, 31% del total de buenos y 40% del total de malos, lo que indicaría que, a priori, hay una probabilidad mayor de que un cliente que presente este atributo sea más malo que bueno.

Figura 2.8: Análisis de variables agrupadas.

<i>Edad</i>	<i>Buenos (N)</i>	<i>Malos (S)</i>	<i>Total</i>	<i>%Buenos</i>	<i>%Malos</i>	<i>%Total</i>	<i>WOE</i>
[18 - 27]	2675	785	3460	7%	13%	8%	-0,6393
[28 - 43]	11829	2397	14226	31%	40%	32%	-0,2690
[44 - 54]	11245	1633	12878	29%	27%	29%	0,0642
[55 - 63]	7702	791	8493	20%	13%	19%	0,4106
[64+]	5059	357	5416	13%	6%	12%	0,7858
Total	38510	5963	44473	100%	100%	100%	
Information Value: 0.15							

Fuente: Elaboración propia.

2.7.1.1 WOE (*Weight of Evidence*)

El WOE mide la fuerza de un atributo al distinguir entre buenos y malos clientes. Según Siddiqi [3, p. 81], el WOE también se puede interpretar como una medida de la diferencia entre la

proporción de buenos y malos para un atributo (la razón de probabilidad de que una persona con ese atributo sea bueno o malo). El *WOE* está basado en el cálculo del logaritmo de la razón de probabilidad, también conocido en la literatura como *odds*:

$$Odds = \frac{Distribución\ de\ Buenos}{Distribución\ de\ Malos} \quad (2)$$

Lo anterior mide el riesgo relativo de ser bueno y malo. En el ejemplo de la Figura 2.10, para el atributo [28-43] esto sería $0.31/0.40 = 0.775$, lo que significa que la probabilidad de que una persona con edad entre 28 y 43 años sea buena es 0.775 veces la probabilidad de que una persona con edad entre 28 y 43 años sea mala, lo que se expresa como 0.775:1 *odds* de ser bueno.

Específicamente, el *WOE* para un grupo consistente de n observaciones se obtiene como:

$$WOE = \left[\ln \left(\frac{Distr\ de\ Buenos}{Distr\ de\ Malos} \right) \right] \quad (3)$$

A partir de lo anterior, un *WOE* cercano a 0 indica que la probabilidad de que un individuo que pertenezca a ese segmento sea un buen o mal pagador es muy similar. Por otro lado, un valor negativo de *WOE* indica que el atributo en particular tiende a tener una proporción más alta de malos pagadores, por lo tanto, la probabilidad de que un individuo sea mal pagador es mayor en ese segmento, mientras que un valor positivo de *WOE* indicaría lo contrario.

2.7.1.2 Information Value (*IV*)

El *Information Value* (*IV* o *Valor de la Información*), es una medida del poder predictivo de un atributo, basándose en cuán diferentes son dos distribuciones. La fórmula es la siguiente:

$$IV = \sum_{i=1}^n (Distr\ de\ Buenos_i - Distr\ de\ Malos_i) \cdot WOE_i \quad (4)$$

En consecuencia, calcular el *IV* de una característica como, por ejemplo, la *EDAD*, permite distinguir si efectivamente ésta discrimina significativamente entre buenos y malos pagadores y, en consecuencia, *a priori* agregará información al modelo. Basándose en la metodología anterior, una regla propuesta por Siddiqi (2006) es:

Tabla 2.5: Poder predictivo de una característica según resultados de IV.

Poder Predictivo	Siddiqi (2006)
Ninguno	$x \leq 0,02$
Débil	$0,02 < x \leq 0,1$
Mediano	$0,1 < x \leq 0,3$
Fuerte	$0,3 < x \leq 0,5$
Riesgo de sobre-predicción	$x > 0,5$

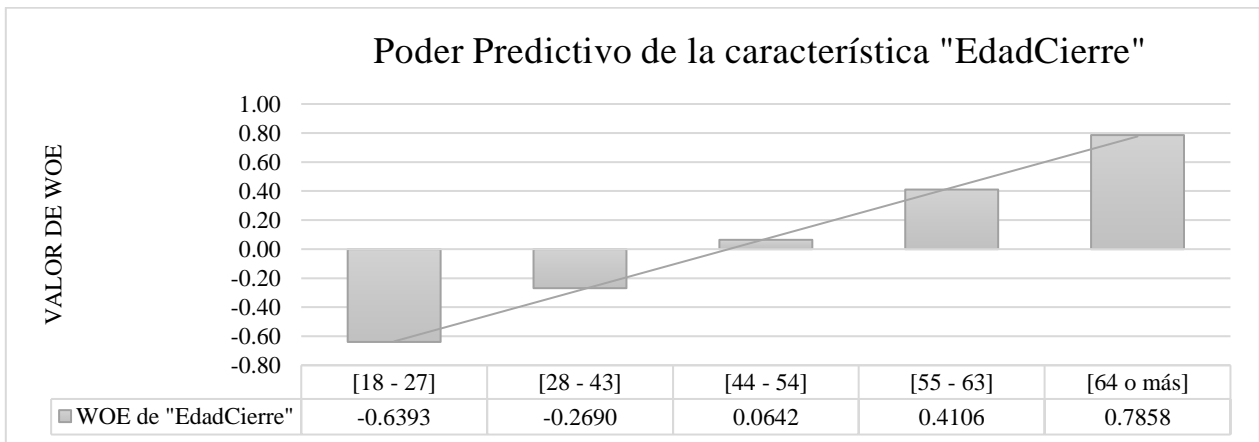
Fuente: Elaboración propia a partir de información publicada en libro *Credit Risk Scorecards* [Naem Siddiqi, 2006].

Siguiendo con el ejemplo, dado que el *IV* de la característica EDAD es 0.15, es posible afirmar que tiene un poder predictivo medianamente fuerte para discriminar entre buenos y malos pagadores. En el caso de que su *IV* fuera mayor a 0,5 se corre riesgo de sobrepredecir, por lo que debe mantenerse fuera del modelo o utilizarse bajo control [3].

2.7.2 Tendencia Lógica

El poder predictivo, medido en términos del *WOE* y *IV*, no es el único factor que se observa para seleccionar una característica. Es muy importante verificar que la distribución del *WOE* sigue un orden lógico y tiene sentido para el negocio. Por ejemplo, en la distribución de la variable EDAD se tiene que:

Figura 2.9: Tendencia Lógica del WOE para la edad.



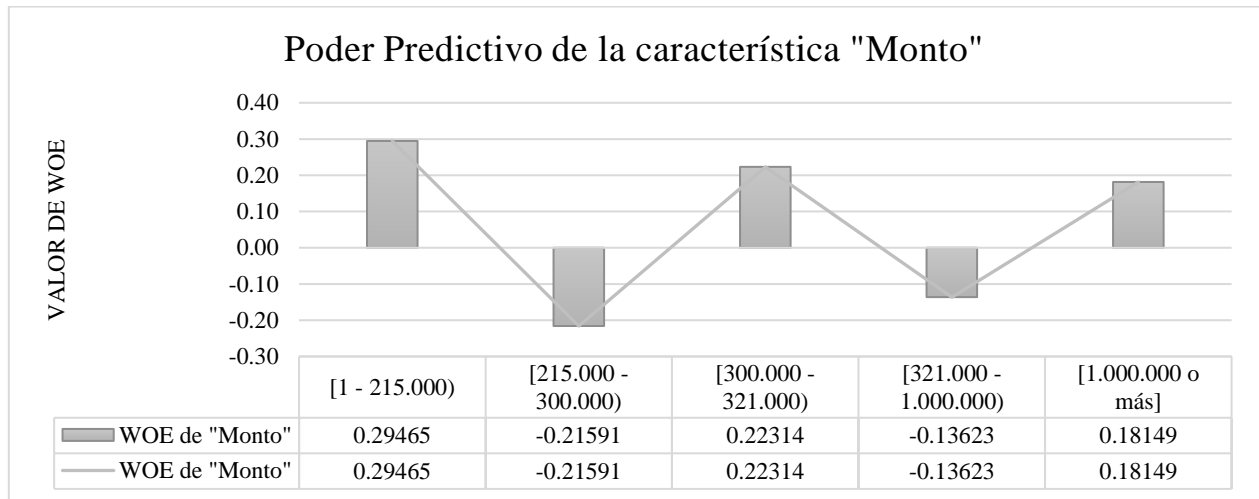
Fuente: Elaboración propia.

Como se observa, todos los grupos en esta característica tienen una relación lineal y lógica entre los atributos de Edad y la proporción de malos, confirmando la experiencia del negocio del microcrédito en que los jóvenes tienden a ser, en general, de mayor riesgo que la población mayor. Establecer relaciones lógicas (que no necesariamente tienen que ser lineales) es el propósito de este análisis, confirmando que lo que se modela es una tendencia y no un caso particular de los datos. Para llegar a la tendencia lógica se utiliza el ensayo y error, buscando equilibrar los grupos y manteniendo un valor de *IV* suficientemente significativo.

Cuando se producen relaciones no lineales válidas, deben utilizarse si es que se pueden explicar con la experiencia o tendencias de la industria. La

Figura 2.10 muestra un ejemplo de una tendencia ilógica. En este conjunto de datos particular, la característica "Monto" representa el monto en pesos chilenos otorgado al cliente. Como se observa, el resultado no muestra ninguna relación lógica entre Monto y el buen o mal comportamiento de pago, con IV igual a 0,03.

Figura 2.10: Tendencia ilógica del WOE para la edad.



Fuente: Elaboración propia.

Otro objetivo de observar la tendencia es agrupar atributos similares para maximizar la diferencia con el grupo siguiente, buscando que la tasa de malos y el *WOE* sean suficientemente distintas. Si bien el valor absoluto del *WOE* es importante, la diferencia entre el *WOE* de los grupos es clave para garantizar la diferenciación. Mientras mayor es la diferencia entre grupos subsiguientes, más alta es la capacidad predictiva de la característica en cuestión.

2.7.3 Regresión Logística

La regresión logística es la técnica más empleada en riesgo de crédito, según Thomas et al. (2002), sobre todo para la creación de *scorecards* en la industria financiera, donde la variable a predecir no es una variable continua, sino categórica.

La regresión logística utiliza un conjunto de características, llamadas también variables explicativas o independientes, para predecir la verosimilitud o probabilidad de obtener un determinado resultado en la variable dependiente (objetivo). Esta técnica transforma la variable dependiente en una función log, estima los coeficientes y maximiza la probabilidad del log de la verosimilitud. El resultado final sigue la siguiente fórmula:

$$Z(odds) = \ln\left(\frac{p(Bueno)}{1 - p(Bueno)}\right) = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_k * x_k + \varepsilon \quad (5)$$

Donde:

p : probabilidad del evento (variable categórica), dadas las variables de entrada.

x_k : variables explicativas no correlacionadas.

β_0 : intercepto de la regresión lineal.

β_k : parámetros.

A partir de este resultado, es posible obtener las probabilidades mediante la ecuación:

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

Los parámetros β_1 a β_k miden la tasa de cambio de la función al aumentar o disminuir en una unidad la variable de entrada (ajustada por el resto de las variables). Los parámetros dependen de la unidad de la entrada, y necesitan ser estandarizados para facilitar el análisis. Las estimaciones estandarizadas son usadas para neutralizar los efectos de las unidades de las variables de entrada.

Una buena forma de tratar con variables agrupadas es usar el *WOE* de cada grupo como una entrada. Esto no solo resuelve los problemas de las diferentes unidades de las entradas, sino que también toma en cuenta la tendencia y escala exactas de la relación de un grupo y el siguiente. Si el agrupamiento está bien hecho, también asegurará que la asignación de puntos para cada grupo en la *scorecard* sea lógica y represente la diferencia de las relaciones entre grupos.

Para automatizar el proceso de selección de variables y encontrar el mejor modelo a partir de todas las combinaciones posibles, se pueden utilizar las siguientes técnicas:

- a. *Selección hacia adelante (Forward Selection)*: comienza sin variables y luego agrega características adicionales, una a una, testeando cada vez si el resultado final del modelo mejora significativamente, utilizando algún criterio de ajuste. Este método es eficiente, pero puede ser débil si hay demasiadas características o alta correlación entre ellas.
- b. *Selección hacia atrás (Backward Elimination)*: al contrario de la técnica anterior, este modelo comienza con todas las características en el modelo y secuencialmente elimina, una a una, las características que son consideradas menos significativas, testeando cada vez si el resultado final del modelo mejora significativamente, utilizando algún criterio de ajuste. Este método permite que las variables de menor significancia tengan una mayor probabilidad de entrar al modelo, mucho más que en las técnicas *Forward Selection* y *Stepwise*.

- c. *Eliminación progresiva (Stepwise)*: es una combinación de las dos técnicas anteriores, por lo que implica agregar y remover características dinámicamente en cada paso hasta encontrar la mejor combinación.

Según Anderson [2], cerca del 90% de las *scorecard* desarrolladas se basan en esta metodología, sobre todo por su flexibilidad. Dentro de las ventajas de utilizar la regresión logística como método de clasificación se encuentran [19]:

- Buen desempeño predictivo.
- Simplicidad en su implementación e interpretación.
- Robustez, al no requerir supuestos muy estrictos.

2.7.4 Producción final del Scorecard

2.7.4.1 Escalamiento

Las *scorecards* pueden ser producidas en muchos formatos y, en algunos casos, deben ser producidas en alguno en particular, ya sea por normativa, facilidad u otros factores, y es necesario aplicar un escalamiento sobre el rango y formato de las puntuaciones encontradas y la tasa de cambio en *odds* por el aumento en la puntuación. Los puntajes de la *scorecard* pueden ser decimales o discretos, considerando que:

- El puntaje es la razón de probabilidad de buenos/malos o la probabilidad de ser malo.
- Hay una determinada escala numérica mínima/máxima (por ejemplo, 0 a 1000, 100 a 800) con un ratio de probabilidad específico en un cierto punto (por ejemplo, razón de probabilidad de 5:1 en 500) y una tasa específica para variar la razón de probabilidad (por ejemplo, se duplican cada 50 puntos).

Es importante destacar que la elección de una escala u otra no afecta el poder predictivo de la *scorecard*. Si bien hay varias escalas utilizadas en la industria, una de las más comunes es la con puntuaciones discretas escaladas logarítmicamente, con la razón de probabilidad duplicándose cada 20 puntos.

2.7.4.2 Cálculo de la escala

En general, la relación entre la razón de probabilidades (*odds*) y los puntajes puede ser presentada como una transformación lineal, de la siguiente forma:

$$Score = Escala + Factor * \ln(odds) \quad (6)$$

Donde la *scorecard* se crea considerando unos *odds* específicos y específicos puntos para duplicar las *odds* (*pdo*), una escala y un factor, que pueden despejarse de la siguiente relación:

$$Score = Escala + Factor * \ln(odds) \quad (7)$$

$$Score + pdo = Escala + Factor * \ln(2 * odds)$$

Resolviendo las ecuaciones, se obtiene:

$$pdo = Factor * \ln(2) \quad (8)$$

Por lo tanto:

$$Factor = pdo / \ln(2) \quad (9)$$

$$Escala = Score - (Factor * \ln(odds)) \quad (10)$$

Donde los *odds* corresponden, en la regresión logística, a la suma de las variables por su coeficiente beta, multiplicado por -1:

$$\ln(odds)_i = -\left(\sum_{i=1}^N \beta_i * x_i + \beta_0\right)$$

Por ejemplo, si una *scorecard* fuera escalada para obtener una probabilidad de 50:1 a 600 puntos, tal que se dupliquen cada 20 puntos, se tendría lo siguiente:

$$Factor = 20 / \ln(2) = 28,8539 \quad (11)$$

$$Escala = 600 - (28,8539 * \ln(50)) = 487,123 \quad (12)$$

Así, cada puntaje correspondiente a cada conjunto de probabilidades puede ser calculado como:

$$Score = 487,123 + 28,8539 * \ln(odds) \quad (13)$$

Dado que esta *scorecard* se desarrolla utilizando el valor de WOE como entrada, la relación anterior puede ser modificada como sigue:

$$\begin{aligned}
score &= \ln(odds) * factor + Escala = \\
& - \left(\sum_{j,i=1}^{k,n} (woe_j * \beta_i) + \beta_0 \right) * factor + Escala = \\
& - \left(\sum_{j,i=1}^{k,n} (woe_j * \beta_i + \frac{\beta_0}{n}) \right) * factor + Escala = \\
& \sum_{j,l=1}^{k,n} \left(- \left(woe_j * \beta_i + \frac{\beta_0}{n} \right) * factor + \frac{Escala}{n} \right)
\end{aligned} \tag{14}$$

Donde:

n = número de características.

k = número de grupos (o atributos) en cada característica.

La fórmula calcula los puntajes que se asignan a cada atributo agrupado, para cada característica en la *scorecard*. Finalmente, la suma de todas las puntuaciones de cada atributo proporcionará la puntuación final. En este punto vale la pena señalar que la tendencia y la diferencia entre los WOE de los atributos agrupados afectará los puntos asignados usando este enfoque.

2.8 Muestreo de Biron

Esta metodología fue planteada por Miguel Biron en 2013 [20], como una forma para asegurar la independencia de las observaciones en aplicaciones de regresión logística para datos de comportamiento crediticio. Este método se basa en la extracción de n subconjuntos de datos de la base original, los cuales tendrán a lo más una observación por cada cliente. Lo anterior permite usar directamente la estimación de máxima verosimilitud de los parámetros de la regresión.

De esta forma, el vector β tendrá n observaciones, denominados $\{\hat{\beta}_i\}_{1 \leq i \leq n}$. Así, es utilizado como estimador de β , el promedio de las observaciones $\hat{\beta}_i$, sobre los n sets de datos.

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n \hat{\beta}_i \tag{15}$$

Debido a que los subconjuntos de datos se generan aleatoriamente siguiendo un mismo algoritmo, y a partir de la misma base de datos, además de que el método de estimación de parámetros que se aplica en cada uno de ellos es idéntico, se puede asumir que los valores $\hat{\beta}_{ki}$

(parámetro estimado en la muestra i , asociado al atributo k), serán variables aleatorias independientes e idénticamente distribuidas para cada atributo k . De esta forma:

$$\hat{\beta}_k = \frac{1}{n} \sum_{i=1}^n \hat{\beta}_{ki} \quad (16)$$

La expresión anterior, según teorema del límite central, tiene una distribución asintóticamente normal, con una media estimada $\hat{\beta}_k$, por lo que es insesgado, y con la siguiente varianza estimada:

$$\text{var}(\hat{\beta}_k) = \frac{1}{n-1} \sum_{i=1}^n (\hat{\beta}_{ki} - \hat{\beta}_k)^2 \quad (17)$$

Según la teoría de muestreo, la construcción de *credit scoring* se caracteriza por tener costos de muestreo iguales y cercanos a cero, y por tener distinta varianza a través de distintos subgrupos de la base de datos. Debido a esto, el problema se reduce a elegir qué observación de cada cliente se utilizará, siempre asegurando que no sea más de una.

2.8.1 Algoritmo

El algoritmo utiliza una base de datos amplia, con la hipótesis de que, para cada cliente, todas las observaciones que tienen más de 12 meses de separación son independientes. Así, la idea principal es seleccionar una vez al cliente por cada 12 meses. El método sigue el siguiente problema de optimización:

$$\max_{S \in P(F)} \text{Card}(S) \quad \text{s. a. } d(i, j) \geq h, \forall i > j \in S \quad (18)$$

Donde F es el conjunto de fechas donde un cliente existe, $P(F)$ es el conjunto de partes de F , y $h \in \mathbb{N}$. La solución entrega un $S \subseteq F$, con todos los elementos dentro de él a una distancia igual o mayor a h . La solución a este problema existe, puesto que el conjunto factible es no vacío. Este problema puede ser expresado como un problema de programación lineal entera, de la siguiente forma:

$$\begin{aligned} & \max_{\{x_i\}_{i \in F}} \sum_{i \in F} x_i \\ & \text{s. a. } M(2 - x_i - x_j) \geq h - d_{ij} \quad \forall i > j \in F \\ & \quad x_{ij} \in \{0,1\} \end{aligned} \quad (19)$$

En el problema anterior, M es un número suficientemente grande. Otra solución factible es tomar un punto i cualquiera, e incorporar todos los puntos $j \in F$ que cumplan la siguiente condición (donde $\%$ es el operador módulo):

$$d(i, j) \% h = 0 \quad (20)$$

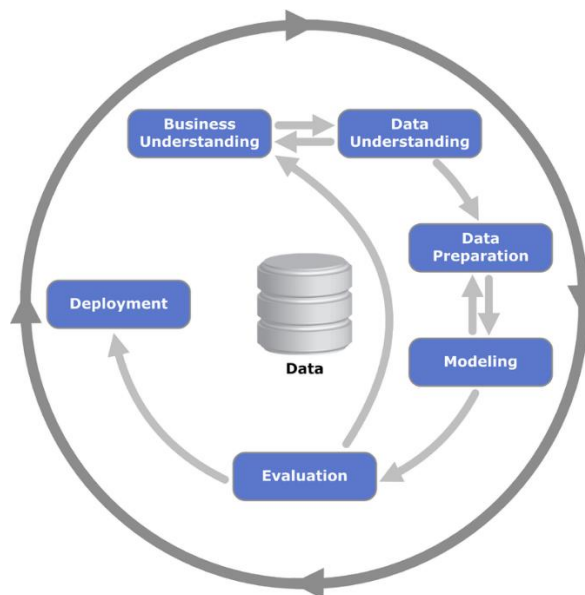
Lo anterior asegura que los elementos que cumplan esta condición estén a una distancia mayor a h entre ellos. Pese a ello, no es óptimo, ya que descarta algunos puntos en casos en que los clientes no aparecen a lo largo de la base de datos.

2.9 Metodología CRISP-DM

La metodología a utilizar sigue las fases del proceso CRISP-DM o *Cross-Industry Standard Process for Data Mining*, el que permitirá orientar el trabajo por medio de etapas, bien definidas, que garanticen un resultado final bien logrado.

El CRISP-DM, como metodología, “*incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas. Como modelo de proceso, CRISP-DM ofrece un resumen del ciclo vital de minería de datos*” [21]. El ciclo se compone de seis fases, con flechas que indican la relación de dependencia más importantes y frecuentes entre ellas:

Figura 2.11: Diagrama de las fases del proceso CRISP-DM



Fuente: *Cross Industry Standard Process for Data Mining*, Wikipedia (2009).

2.9.1 Etapa 1: Comprensión del Negocio

En esta etapa se buscan conocer las expectativas de la organización y las razones principales para el desarrollo del proyecto. En este caso, consiste en identificar el rol que cumplirá la *scorecard* en la organización, identificar los riesgos del proyecto, y las responsabilidades de los implicados en su desarrollo. Esta fase garantiza una *Scorecard* balanceada, orientada a los tomadores de decisiones de la compañía. Se busca, por lo tanto:

- Recopilar información de la situación actual de la institución y del contexto.
- Registrar los objetivos y expectativas de la organización y de los tomadores de decisión.
- Acordar con la institución y el investigador a cargo cuáles serán los criterios que se utilizarán para determinar el rendimiento del proceso desde una perspectiva del negocio.

2.9.2 Etapa 2: Comprensión de los Datos

Esta etapa consiste en estudiar más de cerca los datos disponibles y resulta fundamental para evitar problemas durante la siguiente fase, que suele ser la más extensiva en tiempo. El objetivo es entender qué significa cada variable, identificar las relaciones entre ellas y explorarlas con ayuda de tablas y gráficos. Así, el *output* de esta etapa es encontrar respuesta a las siguientes preguntas:

- ¿Cuáles atributos parecen ser más importantes?
- ¿Qué atributos parecen ser no relevantes y/o podrían excluirse del estudio?
- ¿Existe data suficiente para el desarrollo del modelo? Si no, ¿es posible utilizar un subconjunto de los datos?
- Si se están fusionando bases de datos, ¿qué riesgos implica?
- ¿Cuáles son los problemas de calidad de datos? ¿Qué explicación tienen? ¿Cómo se pueden resolver?

Para responder a lo anterior, la etapa de Comprensión de los Datos se puede desagregar en las siguientes tareas:

1. Recopilación de los Datos: Consiste en registrar cómo se obtuvo acceso a los datos y mencionar las distintas fuentes, en caso de haber más de una.
2. Descripción de los Datos: Consiste en describir la cantidad (dimensión de filas y columnas), formato, tipo de valor y esquema de codificación de los datos.

3. Exploración de datos: Utilizando tablas, gráficos y otras herramientas de visualización.
4. Verificación de calidad de datos: Identificar problemas de datos perdidos, errores de datos, errores de medición, incoherencias de codificación y metadatos erróneos.

2.9.3 Etapa 3: Preparación de los Datos

Esta etapa es uno de los aspectos más importantes y, además, la que toma más tiempo en llevarse a cabo. Se estima que, en la mayoría de los casos, el tiempo invertido corresponde al 70% del total trabajado, sobre todo cuando se utilizan bases de datos de instituciones pequeñas que no tienen una alta sofisticación en registros de información. Las principales tareas son:

- Fusión de registros de datos o tablas.
- Selección del subconjunto de datos a utilizar.
- Agregar registros para simplificar el tratamiento de algunas variables.
- Creación de nuevas variables de interés a partir de las existentes.
- Discretización de variables.
- Eliminación o sustitución de valores nulos o en blanco.
- División de la base en conjunto de prueba y de entrenamiento.

2.9.4 Etapa 4: Modelado

En esta etapa se genera el modelo escogido, configurando los parámetros a utilizar y cómo se esperan describir los resultados obtenidos. Para analizar la capacidad discriminante o predictiva del modelo se entrena el algoritmo utilizando un porcentaje cercano al 70% de los datos (de forma tal que la muestra se encuentre balanceada) y un 30% para evaluar si efectivamente los parámetros obtenidos logran predecir los resultados reales.

2.9.5 Etapa 5: Evaluación

En esta etapa se evalúan y comparan los modelos seleccionados en la fase anterior, de acuerdo a criterios de desempeño escogidos. Para modelos de clasificación, algunos criterios existentes son:

1. **Accuracy:** [22] Se refiere a la habilidad del modelo para predecir correctamente la clasificación de la variable dependiente. Se calcula a partir de la matriz de confusión (ver página 85, Tabla 3.48), utilizando la siguiente fórmula:

$$Accuracy = \frac{\text{Número de predicciones correctas}}{\text{Total de todos los casos predichos}} \quad (21)$$

2. **Precisión:** Proporción de casos positivos clasificados correctamente del total de casos clasificados correctamente, reflejando cuán útil es la predicción del modelo [23]. La fórmula es la siguiente:

$$Precisión = \frac{\text{True Positive}}{\text{True Positive} + \text{True Negative}} \quad (22)$$

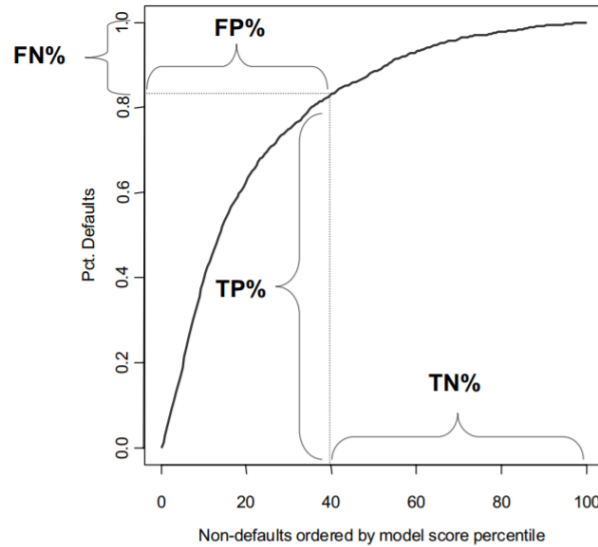
3. **Recall:** Medida de sensibilidad que representa la proporción de predicciones positivas identificadas correctamente, del total de casos predichos como positivos, que refleja cuán completa es la selección del modelo [23]. La fórmula es la siguiente:

$$Recall = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (23)$$

4. **AUC:** Área bajo la curva ROC (*Receiver Operating Characteristics*), en la que se grafica la tasa de datos clasificados como “verdaderos positivos” versus la tasa de “falsos positivos”, que indica una medida de precisión del clasificador que sigue la siguiente lógica: “cuanto más cerca esté la curva ROC de la diagonal (menor AUC), menos preciso es el modelo”. De lo anterior se concluye que un modelo con perfecto ajuste tendrá un AUC igual a 1, y que muchas clasificaciones “verdaderas positivas” se obtienen cuánto más “falsos positivos” se permiten.

La Figura 2.12 muestra las cuatro componentes importantes dentro de la curva ROC, para cada punto de la curva. Considerando que el caso positivo es la presencia de la característica que se busca estudiar, por ejemplo, en *credit scoring*, que el cliente caiga en *default*. Para este caso, %FN y %TP representan la proporción de buenos clientes y malos clientes correctamente clasificados. Por otro lado, %FP y %TN representan la proporción de buenos clientes clasificados como malos y de malos clientes clasificados como buenos. Así, se puede observar que existe un *trade-off* importante al momento de escoger el punto de corte basándose en este método, pues a mayor cantidad de casos TP (en el ejemplo, identificar correctamente a los malos clientes), también es mayor la cantidad de casos FP (en este caso, identificar clientes malos como buenos).

Figura 2.12: Componentes de la curva ROC.



Fuente: *Benchmarking default prediction models: pitfalls and remedies in model validation*. Journal of Risk Model Validation (2007).

2.9.6 Etapa 6: Diseño de *Scorecard*

Una *Scorecard* consiste en un grupo de atributos estadísticamente determinados para clasificar y predecir “los buenos” y los “malos” prospectos o clientes. El *score* más utilizado es el *score logarítmico*, donde el término *odds* representa las *chances* de ocurrencia de un evento (en la regresión logística, corresponden a la suma de las variables por su coeficiente, multiplicado por -1) y *pdo* corresponde a los puntos para duplicar las *chances*:

$$\text{Score} = \text{Escala} + \text{Factor} * \ln(\text{odds}) \quad (24)$$

Toda variable que tenga un *score* por debajo del *score neutro* tiene asociado un riesgo alto. Para poder generar una *scorecard* es necesario trabajar con variables categorizadas en el desarrollo del modelo. Lo más complejo del diseño de una *Scorecard* es la definición del Punto de Corte del modelo, que corresponde al puntaje a partir del cual un cliente pasa de ser un mal pagador a un buen pagador. Una estrategia factible para determinarlo, es determinar el costo para cada punto de corte y encontrar el punto óptimo:

$$\text{Costo}_{\text{corte}} = \text{Costo}_{\text{aceptarMalo}} * (\text{Malos}_{\text{Aceptados}}) + \text{Costo}_{\text{rechazarBueno}} * (\text{Buenos}_{\text{Rechazados}}) \quad (25)$$

Capítulo 3

Metodología

La metodología toma los pasos propuestos por Siddiqi (2006) y los ubica dentro de las etapas de CRISP-DM de minería de datos, donde al cierre de cada una de ellas se obtienen resultados claros y precisos.

3.1 Preliminares: Hipótesis de trabajo

Como se ha mencionado, el objetivo general de este trabajo es construir un *behavioral scoring* según las exigencias básicas de la creación de modelos internos propuestas por SBIF en el Compendio de Nomas Contables, Anexo N°1. Así, se busca que esta herramienta mida el riesgo de crédito de una institución de microfinanzas representativa de Chile, siguiendo las indicaciones del estándar bancario, para poder luego responder a la siguiente inquietud:

¿Las exigencias básicas para la creación de modelos internos en bancos son suficientes para permitir el desarrollo de modelos internos en microfinanzas que midan de forma eficiente el riesgo de crédito en este segmento, para etapas de seguimiento y cobranza?

A partir de lo anterior, y considerando que el ciclo de vida de un microcrédito es muy distinto al de un crédito comercial o de *retail*, sobre todo porque el primero implica una supervisión constante de la institución sobre el cliente, y el pago o no pago de las cuotas está íntimamente ligado con el efecto que el mismo crédito tiene en su negocio, se propone la siguiente hipótesis de trabajo:

Las exigencias básicas para la creación de modelos internos en bancos no son suficientes para el desarrollo de modelos internos en microfinanzas que midan de forma eficiente el riesgo de crédito en el segmento, en etapas de seguimiento y cobranza.

En línea con lo anterior, y con lo mencionado en el Compendio de Normas Contables, Capítulo B-1, Anexo N°1, sobre los *Requisitos Específicos de Desarrollo*, se analizarán las exigencias básicas relacionadas con los siguientes puntos.:

- i. *“La estimación de la pérdida esperada constituye una medida de largo plazo, que considere como mínimo 5 años de información histórica, incluyendo un período recesivo.”*

El modelo a desarrollar considera créditos desarrollados entre el año 2008 y 2016, incluyendo el período recesivo que corresponde a la crisis económica del año 2009.

- ii. *“La cartera en incumplimiento comprende todas las colocaciones y el 100% del monto de los créditos contingentes, de los deudores que al cierre de mes presenten un atraso igual o superior a 90 días en el pago de intereses o capital de algún crédito. También incluirá a los deudores a los que se les otorgue un crédito para dejar vigente una operación que presentaba más de 60 días de atraso en su pago, como asimismo a aquellos deudores que hayan sido objeto de reestructuración forzosa o condonación parcial de una deuda”. Se construye así la variable que indica el incumplimiento crediticio considerando un atraso igual o superior a 90 días en el pago de interés o capital del crédito. También se incluyen casos en que el crédito es reprogramado o refinanciado.*
- iii. *“La definición de incumplimiento considera un horizonte temporal de medición de 12 meses, considerando una frecuencia de observación mensual. Para los créditos con meses de gracia o prórroga contractual en el pago de cuotas, a dicho horizonte se adicionará el número de meses que se utiliza en esa modalidad.”* Siguiendo la idea anterior, el modelo construido define que un crédito cae en *default* en el mes t si en el período de observación equivalente a $t+12$ meses, para alguna cuota superó los 90 días de atraso en el pago.
- iv. *“Las variables evaluadas en las metodologías se basan en las siguientes dos dimensiones: los atributos del deudor o grupo de deudores y las características de sus créditos, cuya elección se sustenta en criterios de predicción y discriminación de riesgo; estabilidad y razonabilidad de las estimaciones. En la primera dimensión, según la cartera evaluada, se consideran al menos las características de riesgo del deudor o grupo de deudores, su carga financiera, leverage, sector de actividad económica, tamaño de la empresa, comportamiento de pago histórico interno y externo, entre otras. Para la segunda dimensión se analizan como mínimo el tipo de operación y sus mitigadores de riesgo (garantías, colaterales y avales/fiadores), la relación saldo del préstamo y valor de la garantía, el grado de prelación de los mitigadores de riesgo, el monto y morosidad de la exposición, el plazo original y residual y características de las reestructuraciones, entre otras.”*

Dado lo anterior, se trabaja en obtener la configuración óptima para un modelo de comportamiento de pago que logre discriminar lo mejor posible entre un buen y un mal pagador en microfinanzas, siguiendo las exigencias de la norma bancaria chilena para la validación de este tipo de modelos. Se espera que, de llegar a buenos resultados, esto motive a las instituciones de microfinanzas a desarrollar modelos similares que permitan, entre otras cosas, tomar decisiones de cobranza a tiempo y decidir de forma racional sobre la renovación o no renovación de créditos a clientes vigentes, de forma rápida y más certera, lo que minimiza el riesgo de continuar con un mal cliente o bien, no actuar a tiempo para evitar que un cliente se convierta efectivamente en un *defaulter*.

3.2 Etapa 1: Comprensión del negocio

3.2.1 Caracterización de la organización

La institución escogida para el desarrollo de este trabajo es Emprende Microfinanzas S.A, institución que nace el año 2009 producto de la asociación entre la Cooperativa de Servicios Liberación (conocida también como Credicoop, fundada en Chile en 1986) y la Fundación Microfinanzas BBVA (entidad española sin fines de lucro, creada por el Grupo BBVA en 2007, con el objetivo de luchar contra la exclusión financiera). A diciembre de 2015, Emprende Microfinanzas presentaba los siguientes datos económicos básicos:

Tabla 3.1: Datos Económicos Básicos de Emprende Microfinanzas S.A.

Clientes vigentes	10.568
Cartera	\$ 7.571.183 (Miles)
Empleados	159
Oficinas	21
Presencia en regiones	V, RM, VI, VII, VIII, IX
Accionariado	La Fundación Microfinanzas BBVA tiene el 100% del capital social

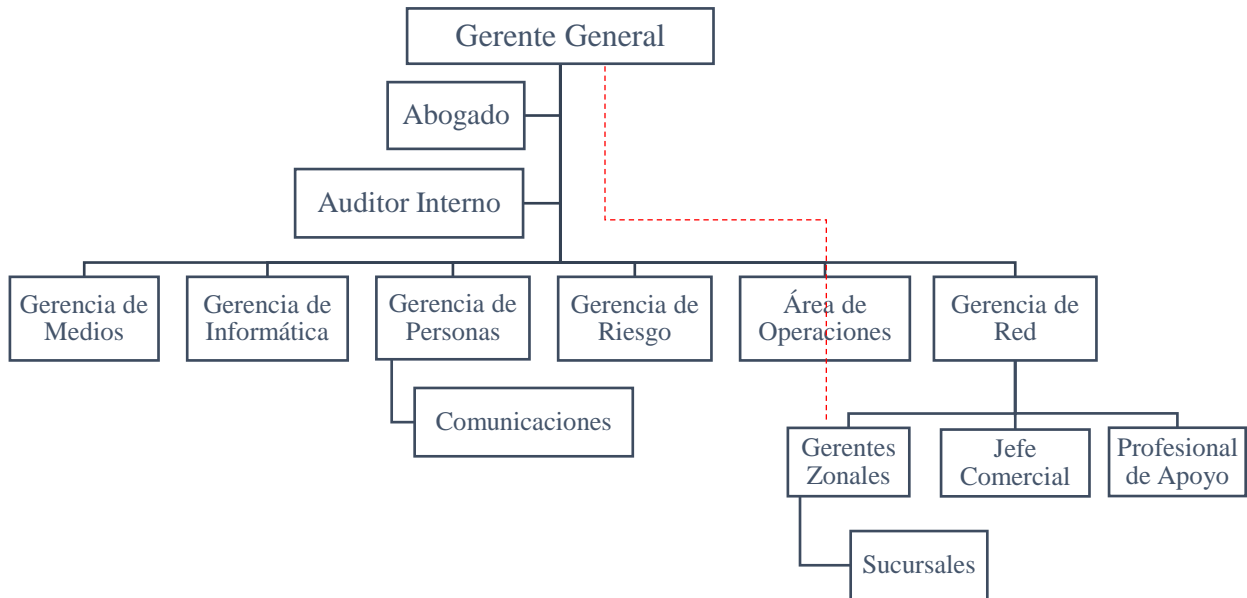
Fuente: Elaboración propia, a partir de la información publicada en el sitio web de Fundación Microfinanzas BBVA.

Se afirma que Emprende Microfinanzas es una institución representativa de las microfinanzas en Chile porque:

- Se especializa en atender clientes que pertenecen exclusivamente al segmento de las microfinanzas, según la clasificación económica del Servicio de Impuestos Internos.
- Tiene sucursales en varias regiones del país, atendiendo microempresarios de distintos sectores económicos y variados perfiles sociales: desde pescadores en zonas rurales hasta comerciantes minoristas en zonas urbanas.
- Es parte de la Red para el Desarrollo de las Microfinanzas en Chile, grupo que busca fortalecer las microfinanzas a nivel público y privado, y que colaboran entre ellas para contribuir conjuntamente al desarrollo social.

La empresa posee una estructura organizacional de índole funcional, donde cada trabajador es supervisado por un superior. Esta estructura es muy típica en empresas que se dedican al mundo financiero, ya que aumenta la especialización de las áreas:

Figura 3.1: Organigrama de Empresa Microfinanzas.



Fuente: Elaboración propia, a partir de la información entregada por la institución.

En relación a los administrativos de gerencia, la mayoría cuenta con estudios universitarios. Los ejecutivos y asistentes comerciales poseen estudios universitarios (20%) o estudios técnicos (80%), y en su mayoría corresponden a mujeres, lo que se explicaría por sus habilidades sociales y la facilidad para mantener una mejor relación con el cliente. Para finalizar, los Jefes de sucursales son aproximadamente 50% hombres y 50% mujeres, en su mayoría con estudios técnicos, y comenzaron trabajando en la institución como ejecutivos.

A lo largo de su historia, la institución ha abierto 29 oficinas, de las cuales 21 se mantienen operativas a la fecha. A partir del año 2014 se agruparon las sucursales vecinas geográficamente en 4 zonas, cada una gestionada por un gerente zonal. A continuación, se presenta la distribución de clientes vigentes en el año 2015, por sucursal, siendo Melipilla la de mayor tamaño:

Tabla 3.2: Clientes vigentes de EMF por sucursal, año 2015.

Sucursal	Porcentaje clientes vigentes 2015	Sucursal	Porcentaje clientes vigentes 2015
Melipilla	10,3%	Recoleta	4,1%
Cañete	7,0%	Los Ángeles	3,8%
Temuco	6,8%	Valparaíso	3,4%
Carahue	6,4%	Coelemu	3,3%
Concepción	6,2%	Santa Bárbara	3,3%
Quinta Normal	5,6%	Santiago Sur	3,3%

Sucursal	Porcentaje clientes vigentes 2015	Sucursal	Porcentaje clientes vigentes 2015
Licantén	5,4%	Chimbarongo	3,2%
Angol	5,3%	Las Cabras	3,1%
Rengo	5,1%	Quinta de Tilcoco	2,7%
La Florida	4,7%	Curacaví	2,3%
Lautaro	4,6%		

Fuente: Elaboración propia a partir de datos entregados por la institución.

Por otro lado, si se observa el tamaño relativo de cada sucursal con respecto a su respectiva zona, se aprecia que las con mayor participación son Melipilla, Licantén, Cañete y Temuco, y que la cantidad de clientes se encuentra bien distribuida a lo largo de las cuatro zonas:

Tabla 3.3: Clientes vigentes de EMF por zona, al año 2015.

Zona	Clientes vigentes por zona 2015	Sucursal	Clientes vigentes por zona 2015 (%)
Metropolitana-Valparaíso	31,3%	Curacaví	7,6%
		La Florida	15,6%
		Melipilla	34,1%
		Quinta Normal	18,5%
		Recoleta	13,5%
		Santiago Sur	10,8%
		Valparaíso	11,4%
Centro Sur	20,3%	Chimbarongo	16,5%
		Las Cabras	15,9%
		Licantén	27,7%
		Quinta de Tilcoco	13,9%
		Rengo	25,9%
Bío Bío	24,4%	Cañete	29,5%
		Coelemu	14,2%
		Concepción	26,1%
		Los Angeles	16,2%
		Santa Bárbara	14,0%
Araucanía	24,0%	Angol	22,7%
		Carahue	27,8%
		Lautaro	20,0%
		Temuco	29,5%

Fuente: Elaboración propia a partir de datos entregados por la institución.

Respecto a los clientes, el mercado objetivo de Empeña Microfinanzas son los trabajadores por cuenta propia (TCP) y empleadores con hasta 9 empleados, que pertenecen a los dos primeros

percentiles económicos y donde se incluyen a pobres e indigentes, cuyas ventas anuales no superan las 2.400 UF. En cifras, los clientes vigentes al 2015 se caracterizan como:

Tabla 3.4: Características principales de la cartera de clientes de Emprende Microfinanzas, año 2015.

	Emprende Microfinanzas 2015
Microempresarios mujer	60,4%
Microempresarios con al menos educación escolar completa	48,7%
Microempresarios con al menos educación superior completa	6,3%
Microempresarias Jefe de Hogar	35,5%
Crédito promedio de un microempresario	\$ 854.000
Comercio minorista	48,2%
Microempresarios informales	89,4%
Población rural	17,6%

Fuente: Memoria anual de Emprende Microfinanzas, 2015.

3.2.2 Proceso de otorgación y seguimiento de créditos

El proceso de otorgación de créditos en microfinanzas es bastante más exhaustivo que en la banca tradicional, con énfasis en reunir la mayor cantidad de información posible sobre la realidad social y económica cliente. Este proceso para Emprende Microfinanzas es como sigue:

Figura 3.2: Etapas del proceso de otorgación de créditos en EMF.



Fuente: Elaboración propia a partir de información entregada por EMF.

A continuación, se describe brevemente en que consiste cada etapa, con énfasis en identificar en que instancia se registran los distintos datos que permiten a la institución tomar la decisión sobre si el prospecto es un sujeto de crédito o no.

3.2.2.1 Captación

Etapa en que la empresa establece contacto con el cliente. Los medios más frecuentes que utiliza Emprende Microfinanzas para este objetivo son:

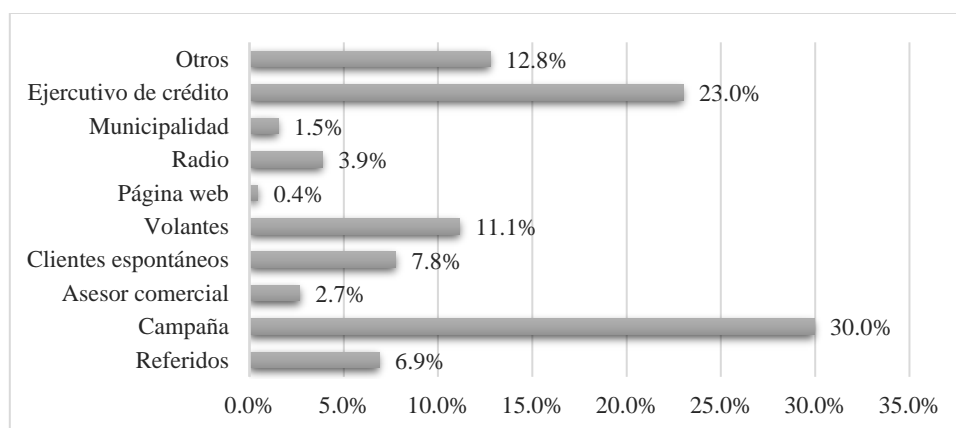
Tabla 3.5: Medios de captación de clientes de Empresa Microfinanzas.

Volanteo	Ejecutivos comerciales reparten volantes en ferias, centros de reuniones, puerta a puerta, etc.
Campaña	Campañas de marketing que buscan atraer nuevos clientes. Ejemplo: <i>Empresa Mujer</i> , foco en la mujer emprendedora.
Referidos	Ejecutivos comerciales solicitan a clientes catalogados como buenos pagadores que recomienden otros futuros clientes.
Otros	Página web, campañas de difusión en radio, campañas con stand físico en Municipalidad o centros de reunión, afiches, etc.

Fuente: Elaboración propia.

De los casos de clientes en que se ha registrado esta variable en la base de datos², se encuentra que la gran mayoría es resultado de una campaña o de la búsqueda de un ejecutivo comercial. Es importante destacar el rol que juegan los clientes *referidos*, que para la institución representan un segmento con alta probabilidad de corresponder a buenos clientes, al corresponder a una recomendación de un conocido cliente con buen comportamiento de pago.

Figura 3.3: Gráfico de la distribución de clientes por tipo de canal, años 2008 a 2016:



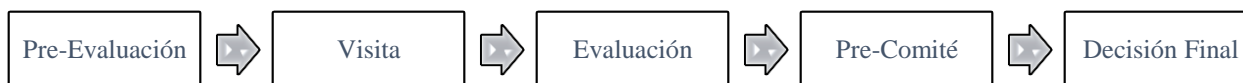
Fuente: Elaboración propia a partir de los datos de la empresa.

3.2.2.2 Evaluación

Segunda etapa del proceso, tiene por objetivo determinar la capacidad de pago del cliente (solventía económica) y su voluntad de pago (solventía moral). La política de evaluación considera a su vez las siguientes sub-etapas:

² El 51.7% tiene este registro en blanco.

Figura 3.4: Sub-etapas del proceso de Evaluación de créditos de Emprénde Microfinanzas.



Fuente: Elaboración propia.

En la pre-evaluación se revisan características básicas del cliente y su negocio, para filtrar aquellos casos que no cumplen los requisitos mínimos. Estando todo en orden el ejecutivo procede a agendar una visita al hogar y al lugar de trabajo, tanto para verificar la existencia física del negocio y su funcionamiento, como para capturar información de la realidad familiar del cliente. A continuación, en la evaluación se revisa y analiza toda la información recolectada para determinar la voluntad de pago y la capacidad de pago del cliente, y presentar la solicitud del crédito al Jefe de Sucursal (pre-comité) y finalmente escalar la solicitud a la oficina central de la institución.

A continuación, se mencionan las principales variables de la base de datos que son registradas durante los procesos de Pre-Evaluación, Visita y Evaluación:

Tabla 3.6: Principales variables que se ingresan a la base de datos en cada etapa de la otorgación de créditos.

Pre-Evaluación:
Datos del cliente: edad, sexo, domicilio, contacto, antigüedad en el rubro.
Datos del negocio: actividad económica, dirección.
Datos del aval: edad, domicilio, registro de DICOM, acreditación de ingresos.
Monto que solicita y destino del financiamiento.
Referencias del cliente con vecinos del sector.
(* Si el cliente es antiguo, se evalúa, además: categoría de pago, situación de mora, antigüedad del cliente.
Visita:
Grupo familiar del cliente: estado civil, integrantes del grupo familiar, nacionalidad, otros ingresos del cliente, estado de propiedad de la vivienda, revisión del pago atrasado de cuentas básicas.
Negocio del cliente: existencia de inversión en mercaderías o maquinaria, nivel de formalidad del negocio, nivel de ventas, patrimonio, activos, pasivos.

Evaluación:

Voluntad de pago: Se determina según la revisión del perfil del cliente, la gestión de su negocio y la situación familiar. Queda a criterio del ejecutivo la respuesta y sólo a veces se registra en la base de datos, como comentario de campo abierto.

Capacidad de pago: Se analiza el balance del negocio y el estado de resultados, el que se construye observando un mes representativo de las ventas del negocio.

Se define la capacidad de pago del cliente (Límite de endeudamiento) como Utilidad Neta*0.5 - Gasto, donde el factor 0.5 es determinado por la política de créditos de la institución.

Fuente: Elaboración propia.

3.2.2.3 Formalización del crédito

En esta etapa se genera toda la documentación que permite cerrar el contrato del crédito, desde un punto de vista legal, emitiendo documentos tales como: pagaré, seguro de desgravamen, contrato de garantías y cuponera. Con la formalización del crédito se da origen en la base de datos al registro del plan de pago del cliente, el que contiene toda la información que permite realizar un seguimiento al comportamiento de pago, cuota a cuota:

Tabla 3.7: Ejemplo de registro del plan de pago de un crédito en la base de datos.

Suc	Cr	Ncta	FVcto	VCta	FPag	TInt	MtoTr	Amort	Int	ST
20	294	1	20-01-2014	31726	21-01-2014	2,9	31726	17812	13914	0
20	294	2	20-02-2014	31726	24-02-2014	2,9	31726	25556	6170	0
20	294	3	20-03-2014	31726	24-03-2014	2,9	31912	26686	5040	0
20	294	4	21-04-2014	31726	22-04-2014	2,9	31726	27865	3861	0
20	294	5	20-05-2014	31726	22-05-2014	2,9	31726	29097	2629	0
20	294	6	20-06-2014	31727	23-06-2014	2,9	31727	30384	1343	0

Fuente: Elaboración propia.

El ejemplo (ficticio) muestra cómo se registra un plan de pago en la base de datos. El crédito n° 294, correspondiente a la sucursal 20, tiene 6 cuotas pactadas de forma mensual entre el 20 de enero de 2014 y el 20 de junio del mismo año, cuya fecha de vencimiento se registra en *FVcto*, y además se observa el valor de la cuota (*VCta*), la fecha en que el cliente pagó la cuota (*FPag*), la tasa de interés (*TInt*), el monto pagado incluyendo intereses por atraso (*MtoTr*), la amortización de la cuota (*Amort*) y los intereses que fueron pagados en la cuota sin incluir el interés penal (*Int*).

3.2.2.4 Cobranza y Estados de Emergencia

Si un crédito presenta problemas de atraso en los pagos de las cuotas, la institución recurre a acciones de recuperación de tipo extra judicial como: envío de cartas de cobranza, llamadas telefónicas al cliente y a su aval, visitas al hogar o al lugar de trabajo. Si bien no existe un protocolo de cobranza que indique cuál medida aplicar sobre qué tipo de cliente y en qué momento, sí se registran en la base de datos los tipos de gestiones de cobranza realizadas, el cliente, la fecha y el resultado de la operación como comentario de campo abierto:

Tabla 3.8: Ejemplo de registro de gestiones de cobranza en la base de datos.

Fecha	Suc	Cred	Hash_Usuario	Observaciones
01-07-2014	12	1670	AA3441069C1B6C0FE	Contacto directo, compromiso para cancelar en oficina el 2014-07-04.
01-07-2014	10	1582	BB3541070C2B6C0CD	Se deja carta en domicilio de cliente.

Fuente: Elaboración propia.

Cuando el cliente definitivamente deja de pagar, al cumplir los 180 días de atraso se da por perdido el crédito y se traspasa a la cartera de créditos castigados. Aun así, antes de llegar a este punto, el cliente puede solicitar reprogramar las cuotas pendientes (vencidas y no vencidas), y generar un nuevo plan de pago por este total. De ser aceptado, se consolida el crédito en la base de datos, cambiando su estado a *reprogramado* y agregando una nueva cuota con fecha al día y que refleja todo el monto pendiente. Luego, se ingresa el nuevo plan de pago en estado de *reprogramación*. En el ejemplo de la Tabla 10, la variable *ST* indica el estado de la cuota del crédito. Como el plan de pago nunca se modificó, el estado de todas las cuotas es cero. Por otro lado, si el mismo cliente hubiese dejado de pagar a partir de la segunda cuota y se reprograma su crédito en el cierre de mes de la cuota 4, este evento se registraría como sigue:

Tabla 3.9: Ejemplo del registro de un crédito reprogramado en el plan de pago.

Suc	Cr	Ncta	FVcto	VCta	FPag	TInt	MtoTr	Amort	Int	ST
20	294	1	20-01-2014	31726	21-01-2014	2,9	31726	17812	13914	0
20	294	2	20-02-2014	31726	NULL	2,9	NULL	25556	6170	5
20	294	3	20-03-2014	31726	NULL	2,9	NULL	26686	5040	5
20	294	4	21-04-2014	31726	NULL	2,9	NULL	27865	3861	5
20	294	7	21-04-2014	158.800	17-04-2014	2,9	158.800	NULL	NULL	0
20	294	5	20-05-2014	31726	NULL	2,9	NULL	29097	2629	5
20	294	6	20-06-2014	31727	NULL	2,9	NULL	30384	1343	5

Fuente: Elaboración propia.

Lo anterior muestra un plan de pago reprogramado, al que se le agrega una cuota final por un total equivalente a las 5 cuotas pendientes más intereses, las cuotas consolidadas tienen fecha de pago NULL y el estado *ST* se actualiza según la siguiente regla:

Tabla 3.10: Código de estado de la cuota en el plan de pago, variable *ST*.

<i>ST</i>	
5	Si la cuota corresponde a una consolidación por refinanciamiento o reprogramación. Se agrega una cuota con <i>ST</i> =0 para marcar el instante del refinanciamiento o reprogramación.
6	Si la cuota corresponde al instante del prepago de un crédito con efectivo o efectivo equivalente. Se marca el resto de cuotas consolidadas con <i>ST</i> =0.
7	Cuota de crédito transferido por cierre de sucursal.
0	Sin modificación, o instante de refinanciamiento o reprogramación.

Fuente: Elaboración propia a partir de información entregada por la institución.

Mientras una reprogramación es asocia a un mal cliente, existe también la opción de solicitar un *refinanciamiento*, alternativa que se ofrece a los *buenos* clientes que cuentan con un crédito vigente, pero que necesitan más capital para cubrir una situación de emergencia que perjudica directamente el funcionamiento regular del negocio (robo de mercadería, incendio de local, situación de catástrofe natural, accidente laboral, entre otros). Ante esta situación, la empresa consolida todas las cuotas pendientes del crédito antiguo, cambia su estado a *refinanciado*, y genera un nuevo crédito por el total del monto adeudado más un monto de capital extra. Siguiendo con el ejemplo anterior, un crédito que es refinanciado se registra en la base de datos como sigue:

Tabla 3.11: Ejemplo del registro de un crédito refinanciado en el plan de pago.

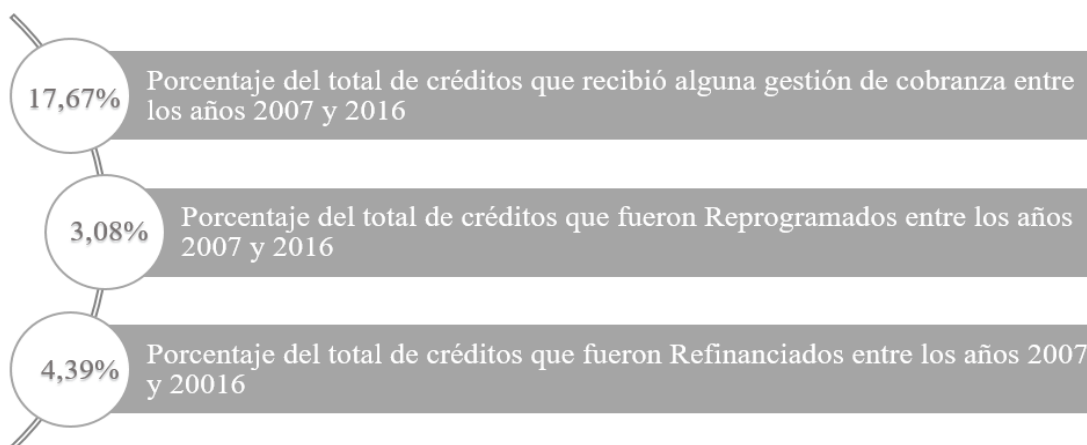
Suc	Cr	NCTa	FVcto	VCta	FPag	TInt	MtoTr	Amort	Int	ST
20	294	1	20-01-2014	31726	21-01-2014	2,9	31726	17812	13914	0
20	294	2	20-02-2014	31726	21-01-2014	2,9	NULL	25556	6170	0
20	294	3	20-03-2014	31726	24-02-2014	2,9	NULL	26686	5040	0
20	294	4	21-04-2014	31726	NULL	2,9	NULL	27865	3861	0
20	294	7	21-04-2014	91.207	17-04-2014	2,9	95.179	NULL	NULL	6
20	294	5	20-05-2014	31726	NULL	2,9	NULL	29097	2629	0
20	294	6	20-06-2014	31727	NULL	2,9	NULL	30384	1343	0

Fuente: Elaboración propia.

El ejemplo anterior muestra el mismo crédito anterior, pero en una situación en que el cliente solicita un refinanciamiento a partir de la cuota cuatro. Así, se consolidan las 3 últimas cuotas pendientes en un único pago que corresponde a la suma de los valores de la cuota cuatro más la amortización de la cuota cinco y seis.

De un total de 75.587 créditos cursados entre el año 2007 y 2016, se obtienen las siguientes cifras con respecto a la gestión de cobranza y repactaciones:

Figura 3.5: Cifras sobre la gestión de cobranza y casos de repactación en créditos cursados entre los años 2007 y 2016.



Fuente: Elaboración propia a partir de los datos entregados por la institución.

Finalmente, cabe destacar que cambiar el estado de un crédito a consolidado y marcar la variable *ST* como refinanciamiento, reprogramación o prepago, NO es un procedimiento que se encuentre automatizado en la institución, por lo que, para identificar efectivamente estos casos, es necesario revisarlos en detalle. Un procedimiento adicional fue la revisión uno a uno de 300 casos, correspondientes a créditos anteriores al año 2014, en que la institución agregaba sobre el plan de pago existente las nuevas cuotas del crédito refinanciado o reprogramado, por lo que coexistían en una misma historia cuotas vigentes y consolidadas.

3.2.3 Variables representativas

Según juicio experto³ y las reglas de negocio respecto al otorgamiento de créditos de la institución, se concluye que las variables que podrían resultar determinantes para explicar el comportamiento de pago en microempresa son:

³ Conversaciones con el gerente general de Emprande Microfinanzas (Pablo Coloma), experiencia del profesor guía con el trabajo en microempresa (Cristián Bravo), levantamiento de información con ejecutivos de Emprande Microfinanzas, gerente de riesgo de Emprande Microfinanzas.

Tabla 3.12: Variables que podrían ser significativas para explicar el comportamiento de pago de un microempresario.

VARIABLES	RELEVANCIA
Edad del cliente	Basándose en la experiencia de los ejecutivos de la institución, los adultos jóvenes entre 18 y 30 años se preocupan menos por estar en DICOM que otros segmentos, razón por la cual se espera que su probabilidad de caer en <i>default</i> sea mayor.
Experiencia bancaria	Según juicio experto, un cliente bancarizado y ya familiarizado con la responsabilidad que implica realizar pagos periódicos, presentaría un menor riesgo de incumplir el pago de sus cuotas que uno que pide un crédito por primera vez.
Integrantes del grupo familiar	Dado que la realidad laboral del microempresario se encuentra íntimamente ligada con la estabilidad de su hogar, un cliente con alta carga familiar está expuesto a más eventualidades que puedan influir en el incumplimiento futuro de sus pagos.
Coincidencia del lugar de trabajo con el hogar	Por la misma razón anterior, el microempresario que trabaja desde el hogar está expuesto a más distracciones y se esperaría que los problemas familiares repercutan mucho más en su rendimiento laboral, influyendo negativamente en su productividad y, por lo tanto, en su capacidad de generar ingresos que le permitan cubrir el pago de sus cuotas.
Nivel de endeudamiento del cliente y del grupo familiar	Se esperaría que un cliente con mayor nivel de deuda externa se vea mucho más expuesto a la necesidad de posponer el pago de algunos compromisos por sobre otros, lo que podría afectar negativamente su capacidad de pago futura.
Gestión de cobranza	Por definición, recibir algunas o muy pocas gestiones de cobranza implica que existen indicios de un mal comportamiento de pago. Si el cliente recibe reiteradas gestiones de cobranza, se esperaría que termine cayendo en <i>default</i> .
Medios de contacto disponibles	Se esperaría que si el cliente entrega una mayor cantidad de datos de contacto (teléfono de casa, celular, correo electrónico), sea más expedito realizar gestiones de fidelización o cobranza y, por lo tanto, evitar que el cliente caiga a futuro en <i>default</i> .
Transferencia de sucursal	Dado que en microfinanzas es muy importante el vínculo que establece el ejecutivo con el cliente, podría suceder que un cliente que es transferido de una sucursal a otra empeore su comportamiento de pago. En la institución, un crédito es transferido por cierre de sucursal o cambio de domicilio del cliente.
Cercanía a la sucursal	Se esperaría que un cliente que vive lejos de la sucursal mantenga una relación más distante con su ejecutivo y, por lo tanto, más débil con la institución. Así, podría ocurrir que el comportamiento de pago de estos clientes sea peor que el resto.
Cantidad de cuotas	Según la experiencia de la institución, la tasa de incumplimiento crediticio de los clientes debería aumentar a medida que aumenta la cantidad de cuotas pactadas.
Medio de captación	Según la experiencia de la institución, un cliente que se ingresa como referido tiene baja probabilidad de caer en incumplimiento, en comparación a los clientes captados por otros tipos de canales.

VARIABLES	RELEVANCIA
Meses de experiencia en el negocio	Según la experiencia de la institución, mientras más experiencia tiene un cliente en el rubro, mejor conoce su negocio y más capacitado está para enfrentar eventuales crisis. Luego, el riesgo de caer en <i>default</i> es menor.
Cliente nuevo	Se esperaría que la probabilidad de que un cliente antiguo y ya fidelizado por la institución caiga en <i>default</i> sea menor que la de un cliente nuevo.
Actividad Económica	Según la experiencia de la institución, y otras instituciones dedicadas al microcrédito, los clientes que dependen de actividades estacionales como la pesca o restaurantes y hoteles son más riesgosos que los que pertenecen a otros segmentos.
Iniciación de actividades	Se esperaría que un cliente formalizado tenga menor probabilidad de incumplir en sus pagos, ya que es menos probable que deje de trabajar de un día para otro.
Nivel educacional	Según juicio experto, se esperaría que un cliente con menor nivel educacional tenga menos oportunidades de obtener crédito en distintas instituciones que alguien mejor calificado y, por lo tanto, tienda a ser más responsable con el pago de un crédito.
Reprogramación o Refinanciamiento	Se esperaría que la probabilidad de caer en <i>default</i> para un crédito reprogramado sea mayor que para un crédito que no ha sido repactado y, a su vez, mayor que en el caso de un crédito refinanciado. Por otro lado, se esperaría que un crédito refinanciado no fuese tanto más riesgoso que un crédito sin repactaciones.

Fuente: Elaboración propia.

Adicionalmente, y a partir de la información existente en la base de datos, se debe considerar la creación de variables de comportamiento relacionadas con los siguientes conceptos:

Tabla 3.13: Variables relacionados con el comportamiento de pago.

VARIABLES	RELEVANCIA
Morosidad	Un supuesto básico de los modelos de comportamiento es que la información pasada sirve para predecir lo que sucederá en el futuro. Así, se esperaría que los clientes que, por ejemplo, tienen una mayor mora promedio de sus últimas cuotas, sean mucho más riesgosos que otros.
Evolución de los pagos	Se esperaría que el porcentaje de cuotas que falta por pagar o el monto acumulado del crédito pagado sean variables relevantes para determinar si, respecto a estas cifras, es más o menos probable que el crédito caiga en <i>default</i> a futuro.

Variables	Relevancia
Crédito anterior	Conocer las características del crédito anterior y compararlas con el crédito actual podría permitir identificar cuándo es más probable que un crédito caiga en <i>default</i> . Si, por ejemplo, esto ocurre por lo general cuando el cliente pide un monto que representa el doble de lo que solicitó anteriormente o en otros casos.

Fuente: Elaboración propia.

3.3 Etapa 2: Comprensión de los datos

3.3.1 Almacenamiento

Toda la información que registra Emprende Microfinanzas, desde los años de Credicoop hasta hoy, está contenida en 5 bases de datos, las que pertenecen al Sistema de Administración de Créditos (SAC). Si bien, existe un nuevo sistema impuesto por la Fundación Microfinanzas BBVA y que contiene una mayor cantidad de variables, el MAS (Macrosistema de Admisión y Seguimiento), su registro histórico comienza recién el año 2014 y se considera insuficiente para este estudio. La descripción de las bases de datos que contiene el sistema SAC se encuentran en Anexo 1, Tabla 7.1.

El acceso a los datos fue directo y por medio de una copia de seguridad completa, facilitada por el Gerente del Área de Informática de la institución. Así, se recibió la información en formato *backup*, sin modificaciones previas, y se manipuló únicamente utilizando medios encriptados. Esta es la única fuente de datos considerada en el estudio.

Para leer y procesar las bases de datos se utilizó, en primera instancia, el administrador de bases de datos SQL Server, versión *Management Studio 2008*, por su simpleza y facilidad para manipular simultáneamente la información entre las distintas tablas. Para etapas posteriores, que requieren mayor capacidad de procesamiento, se exporta la base de datos consolidada a *R Studio*, versión 0.99.903.

3.3.2 Descripción de los datos

Ahora que se tiene conocimiento sobre cómo opera el otorgamiento de créditos en Emprende Microfinanzas y qué tipo de información puede ser relevante para el estudio, se procede a buscar en las bases de datos las variables de interés, para luego consolidar todo en una única tabla, la que finalmente se utilizará para modelar el comportamiento de pago. En reuniones con la institución se identificaron cuáles son las tablas más utilizadas y que almacenan la información necesaria para la gestión diaria de los créditos, consideradas como *tablas principales*, y también sus respectivas llaves primarias. Esta información se encuentra en Anexo 2, Tabla 7.2.

El primer paso sobre las *tablas principales* es entender las variables que contienen y descubrir cómo se relacionan entre ellas, ya que la institución no cuenta con un diccionario de bases de datos ni con un mapa entidad relación. Por otro lado, la mayoría de las variables que están en las *tablas principales* se encuentran codificadas, y el significado de algunas, la mayoría, está en distintas tablas repartidas dentro de las cinco bases de datos existentes, mientras que el resto sólo es conocido por una persona de la institución dedicada a administrar los sistemas SAC y MAS.

Como resultado del levantamiento de información, se crea un diccionario en Excel que describe cada variable de las cinco *tablas principales*, como se muestra en el ejemplo de Anexo 2, Tabla 7.3 (se excluyen las variables de las cuales la empresa desconoce el significado). Adicionalmente, todas las tablas que representan códigos se agrupan en un libro de Excel con vínculos a su respectivo significado, tal y como muestra la siguiente representación:

Figura 3.6: Imagen referencial del diccionario de Códigos creado en Excel

Tabla	Prefijo de la variable	Descripción
Actividad	a	Tipo de actividad relacionada con el microempresario.
Bancos	bco	Nombres de bancos vinculados a EMF.
CAL	cal	Datos del CAL.
Campañas	cam	Características de campañas realizadas por EMF.
Canales	c	Canales de captación de clientes.
Cargos	car	Nombres de cargos adicionales asociados al crédito.
Categorías	cat	Clasificación de tipo de información.
CodActEco	cae	Códigos de actividad económica del SIL.
Comité	com	Comité de evaluación de riesgo.
ComTipCliente	ctc	Tipo de cliente según clasificación de riesgo.
Comunas	c	Detalle de Comunas, densidad, Región, Código de Área, entre otros.
ConFianza	cf	Tipo de fianza asociada al crédito.
Correlativos	corr	Correlativo banco, sucursal.
CtasBancos	cb	Bancos asociados a cada cuenta del Plan Mayor.
CtasEstSit	ces	Nombre de cálculos menores del Estado de Resultados.
Cultivos	cul	Tipo de cultivo agrícola.
EmisoresDicom	emi	Emisores de DICOM.
Escolaridad	es	Nivel de escolaridad del cliente.
EstCivil	ec	Estado civil del cliente.
Etnias	et	Etnia del cliente.

cCod (PK)	cDesc
1	Referidos
2	Campaña uno + uno
3	Canal Alternativo
4	Asesor Comercial
5	Clientes Espontáneos
6	Otros (especificar)
7	Volantes
8	Página WEB
9	Radio
10	Municipalidad
11	Convenio
12	Ejecutivo de Crédito
13	Campaña Gift Card

Fuente: Elaboración propia.

3.3.3 Anonimización

Un requisito impuesto por la institución para acceder a realizar este estudio fue que toda la información que permitiese identificar unívocamente a un cliente estuviese previamente anonimizada, con el objetivo de garantizar la confidencialidad de los datos. Para cumplir con este requerimiento se tomaron las siguientes medidas:

- a) Eliminar datos de dirección, nombre, apellidos, razones sociales, nombres de fantasía y todo dato que pueda identificar unívocamente al titular de la cuenta o aval.
- b) Reemplazar identificadores únicos (como el RUT) por un indicador numérico independiente no invertible (hash), el que funciona de la siguiente forma:

Figura 3.7: Caracterización de cómo es el resultado que entrega una función HASH.



Fuente: Elaboración propia.

A pesar de que en el punto a) se menciona que es necesario eliminar los datos identificadores, hay registros que por el solo hecho de existir podrían significar algo interesante para el estudio. A modo de ejemplo, si bien el número de teléfono en sí es irrelevante, el hecho de que ese cliente haya registrado un número de contacto, no lo es. Así, en la mayoría de los casos, antes de eliminar las variables se toman medidas que permiten rescatar la máxima cantidad de información asociada, definiendo un set de reglas de anonimización, adjuntado en Anexo 3, Tabla 7.4. Las variables creadas como parte de este proceso son:

Tabla 3.14: Variables creadas como resultado de la anonimización de datos relevantes.

Variable No Anónima	Variable Anónima
<i>RUT cliente</i>	<i>PersNatural</i> $\begin{cases} 1 \text{ si el cliRUT es tal que } 999.999 < \text{cliRUT} < 30.000.000 \\ 0 \text{ si no} \end{cases}$
	<i>PersJuridica</i> $\begin{cases} 1 \text{ si el cliRUT es tal que } \text{cliRUT} \geq 30.000.000 \\ 0 \text{ si no} \end{cases}$
	<i>EIRL</i> $\begin{cases} 1 \text{ si el cliRUT es tal que } 49.999.999 < \text{cliRUT} < 60.000.000 \\ 0 \text{ si no} \end{cases}$
<i>Mail</i>	<i>MAIL</i> $\begin{cases} 1 \text{ si la variable cliMail contiene el caracter @} \\ 0 \text{ si no} \end{cases}$
<i>Teléfono</i>	<i>FonoN_i</i> $\begin{cases} 1 \text{ si Fono}_i > 99999 \\ 0 \text{ si no} \end{cases}$

Fuente: Elaboración propia.

Finalmente, se crea la variable *CasaTaller*, que indica si la dirección de trabajo del microempresario coincide con la dirección particular. Como las direcciones, al igual que la gran mayoría del resto de los datos, se ingresan de forma manual, existe una gran cantidad de errores humanos. Al no ser eficiente analizar todos los casos uno a uno, se recurre a dos funciones que permiten identificar con un 90% de efectividad si dos campos de texto contienen la misma información⁴: *Soundex*, función que convierte una cadena alfanumérica en un código de cuatro caracteres que se basa en cómo suena la cadena cuando se pronuncia, dado un idioma y *Difference*,

⁴ Esta medida se obtiene al contrastar los casos correctamente identificados de una muestra aleatoria de 100 clientes a los que se aplicó esta metodología

que entrega una medida, de 0 a 4, de la diferencia entre los resultados de dos cadenas de caracteres analizadas con *Soundex*, donde 4 indica que la diferencia es la mínima posible [24].

Para obtener mayor sensibilidad en los resultados se decide utilizar ambas funciones: en primera instancia, se utiliza *Soundex* para identificar si el contenido de ambas variables es el mismo, guardando el registro en la variable *clidir_sound*:

$$clidir_{sound} \begin{cases} 1 & \text{si } Soundex(DirCasa) = Soundex(DirTaller) \\ 0 & \text{si no} \end{cases}$$

Como se mencionó anteriormente, los campos de dirección particular y comercial son de texto abierto, y la probabilidad de que el ejecutivo ingrese valores distintos, a pesar de corresponder a la misma dirección, es muy alta, sobre todo por el uso de abreviaciones u omisión de letras. Por esta razón, existe la posibilidad de que los resultados de *Soundex* no coincidan para casos en que, al modificar la escritura, cambie mucho el sonido al pronunciar. A modo de ejemplo, si se pide a la función comparar “*Santo domingo 2348, pasaje 10*” con “*sto. domingo 2348, psj. 10*”, la función *Soundex* entrega diferentes resultados para cada una de ellas y, en consecuencia, *clidir_sound* sería 0.

Para discretizar este resultado, se recurre a la función *Difference*. Para el mismo ejemplo anterior, la diferencia fonética entre “*Santo domingo 2348, pasaje 10*” y “*sto. domingo 2348, psj. 10*” según *Difference* es 3, es decir, son muy similares. Siguiendo esta idea, se guarda esta información en una nueva variable llamada *clidifference*:

$$clidifference \begin{cases} 0 & \text{Si } (DirCasa) \text{ es muy distinto a } (DirTaller) \\ 1 \\ 2 \\ 3 \\ 4 & \text{Si } (DirCasa) \text{ es muy similar a } (DirTaller) \end{cases}$$

Finalmente, la combinación de ambas soluciones y que maximiza los casos bien clasificados se resume en la siguiente variable:

$$casataller \begin{cases} 1 & \text{si } [clidir_{sound} \text{ es } 1 \text{ y } clidifference \text{ es } 4] \text{ ó } [si } clidir_{sound} \text{ es } 0 \text{ y } clidifference \text{ es } 4] \\ 0 & \text{si no} \end{cases}$$

3.4 Etapa 3: Preparación de los datos

3.4.1 Construcción de base de datos tipo panel

Para predecir el evento de incumplimiento crediticio es necesario construir una base de datos tipo panel, la que contiene toda la información de las variables de interés para cada instante de observación, en este caso, para la fecha de vencimiento de cada cuota. Un ejemplo ficticio de cómo se debería ver esta base de datos es el siguiente:

Tabla 3.15: Ejemplo de configuración de datos tipo panel.

ID	Año	Mes	Sucursal	Monto	Cuota	Valor Cta	Vencimiento	Pago	Mora	Valor Pagado	Pago Acum	Sexo	Edad	CAE
1025	2014	3	La Florida	250.000	1	27.500	05-03-2014	06-03-2014	1	27.520	27.520	M	34	Madera y Muebles
1025	2014	4	La Florida	250.000	2	27.500	05-04-2014	05-04-2014	0	27.500	55.020	M	34	Madera y Muebles
1025	2014	5	La Florida	250.000	3	27.500	05-05-2014	07-05-2014	2	27.570	82.590	M	34	Madera y Muebles
1025	2014	6	La Florida	250.000	4	27.500	05-06-2014	10-06-2014	5	27.590	110.180	M	34	Madera y Muebles
1025	2014	7	La Florida	250.000	5	27.500	05-07-2014	04-07-2014	0	27.500	137.680	M	34	Madera y Muebles
1025	2014	8	La Florida	250.000	6	27.500	05-08-2014	03-08-2014	0	27.500	165.180	M	35	Madera y Muebles
1025	2014	9	La Florida	250.000	7	27.500	05-09-2014	05-09-2014	0	27.500	192.680	M	35	Madera y Muebles
1025	2014	10	La Florida	250.000	8	27.500	05-10-2014	02-10-2014	0	27.500	220.180	M	35	Madera y Muebles
1025	2014	11	La Florida	250.000	9	27.500	05-11-2014	01-11-2014	0	27.500	247.680	M	35	Madera y Muebles
1025	2014	12	La Florida	250.000	10	27.500	05-12-2014	01-12-2014	0	27.500	275.180	M	35	Madera y Muebles

Fuente: Elaboración propia.

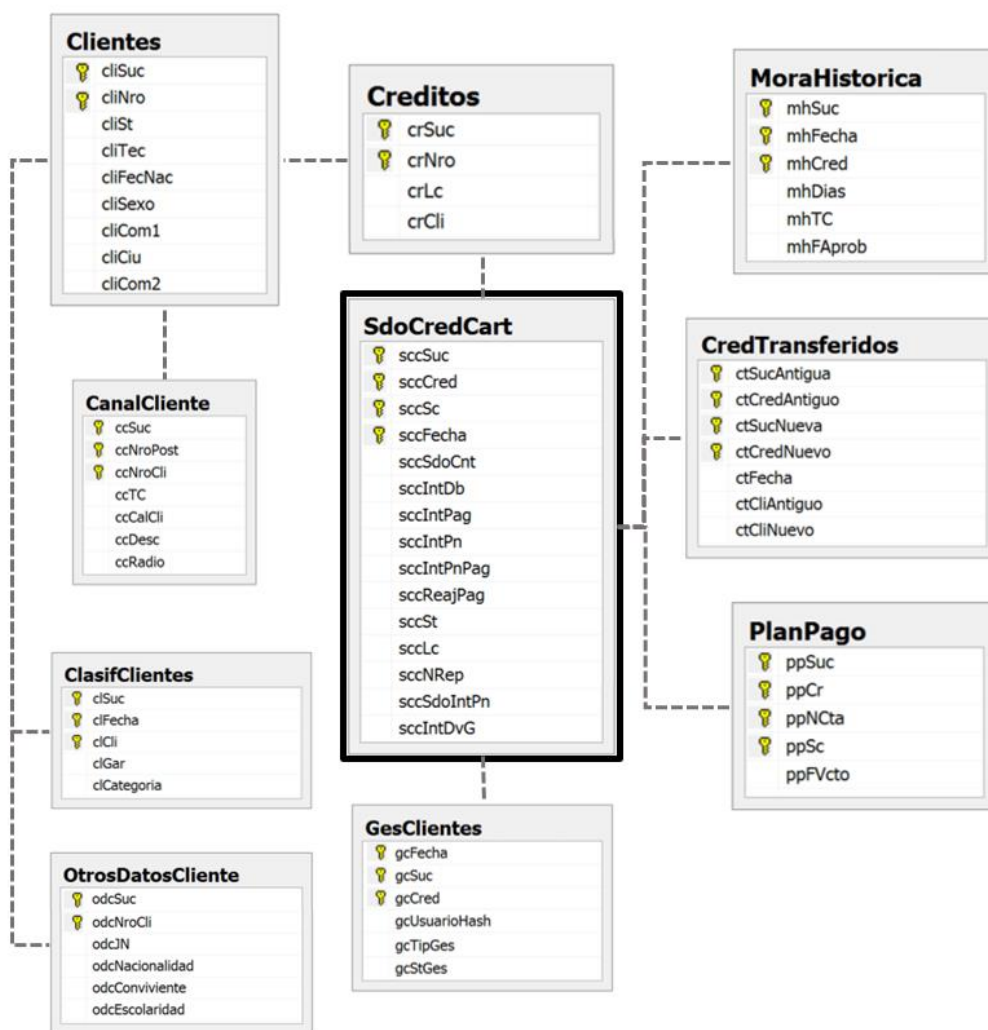
Dado que la institución no cuenta con un mapa entidad relación, se exploran diversas alternativas para unir las *tablas principales* utilizando SQL Server. Considerando que cada crédito se identifica unívocamente por el número de la sucursal y el número de crédito, es decir, el par (Suc, Cred), se juntan las tablas como sigue:

1. A la tabla *SdoCredCart* se agrega la información de las tablas *MoraHistórica*, *CredTransferidos*, *PlanPago*, *GesClientes* y *Créditos*. Como resultado, se obtiene en una misma tabla el estado de todos los créditos al cierre de cada mes, más la información adicional de aquellos que han estado morosos, más los nuevos números de Sucursal, Crédito y Cliente para créditos transferidos, más toda la información del plan de pago, más la información de cobranza y todas las características del crédito pactado.

2. A la tabla resultante se agrega la información de *Clientes*. Dado que la tabla *Créditos* contiene el identificador único de cada cliente, permite unir a *SdoCredCart* con toda la información relacionada con la persona que recibe el crédito.

3. Finalmente, se agrega el resto de información relacionada con el cliente como el canal de captación, la clasificación de riesgo y otros datos adicionales como el estado civil, el nivel de educación, entre otros.

Figura 3.8: Mapa entidad relación de tablas principales y tablas anexas.



Fuente: Elaboración propia.

Como resultado de juntar las tablas, se encuentran varios créditos que aparecen registrados en *SdoCredCart*, pero que carecen de información de *Cliente* y/o de *Plan de Pago*, razón por la se completan con NULL. Al consultar a la institución las razones, ellos informan que es posible que la mayoría de éstos créditos correspondan a antiguos clientes de la cooperativa cuya información no fue traspasada a la base de datos actual, razón por la cual se decide eliminarlos del estudio. La *Tabla Maestra* tiene las siguientes características:

Tabla 3.16: Tabla Maestra N°1.

Filas:	1.216.339
Créditos:	91.221
Cientes:	45.879

Fuente: Elaboración propia.

Luego, dado que la persona encargada de las bases de datos recomendara sólo mantener los créditos posteriores al año 2007, ya que los anteriores podrían contener muchos errores e inconsistencias (se ingresaban a una base de datos antigua), se decide eliminar todos los créditos aprobados antes de marzo de ese año. Por otro lado, dado que no sirven los créditos que finalizan luego de marzo 2016 (fecha en que se recibe la base de datos), pues no se observa el comportamiento completo, se eliminan aquellos tal que la última cuota es posterior a esa fecha. Al mismo tiempo, se decide dejar fuera todos los créditos que no son mensuales. Para ello, se crea una variable que calcula la cantidad de meses entre la primera cuota del plan de pago y la última, y se divide por el total de cuotas pactadas. Si el resultado es mayor a 1, el crédito se deja fuera del estudio. La *Tabla Maestra* queda como sigue:

Tabla 3.17: Tabla Maestra N°2.

Filas:	1.137.621
Créditos:	80.852
Clientes:	36.673

Fuente: Elaboración propia.

3.4.2 Descripción y construcción de variables de tipo *application*

A continuación, se describen las variables de la *Tabla Maestra*, agrupadas según el tipo de información que entregan, detallando si la variable es construida u original de la institución (identificada como EMF) y si es nominal, binaria o numérica.

En una primera tabla, en Anexo 4, Tabla 7.5, se registran los identificadores únicos para cada crédito. Es necesario mencionar que cuando un crédito es transferido de una sucursal a otra, se le asigna un nuevo número de crédito y de cliente, dando origen a un nuevo plan de pago que es la continuación del anterior. Así, para obtener la historia completa de un crédito que fue transferido, es necesario consolidar ambos créditos en uno solo.

Respecto a las variables que indican características demográficas del cliente, se mantienen todas aquellas de las cuales se conoce su codificación y se tiene relativa certeza de que los datos se ingresaron correctamente y son consistentes en el tiempo. De un total de 22 variables, 11 son de construcción propia. Adicional a *CasaTaller* se crean variables binarias que intentan capturar si existe un comportamiento de pago distinto para los clientes que viven o trabajan lejos de la sucursal: *ComunaCasaSuc*, *ComunaTallerSuc*, *ComunaCasaTallerSuc*. A continuación, se muestra la clasificación de estas variables por tipo, mientras que el diccionario completo se encuentra en Anexo 4, Tabla 7.6:

Tabla 3.18: Variables asociadas a características del cliente, clasificadas por tipo.

Variables asociadas al cliente		
Binarias	Nominales	Numéricas
<i>ComunaCasaSuc</i>	<i>CodEscolaridad</i>	<i>EdadCierre</i>
<i>ComunaTallerCasaSuc</i>	<i>CodEstCiv</i>	<i>EdadFechaInc</i>
<i>ComunaTallerSuc</i>	<i>CodOtrosIng</i>	<i>Hijos</i>
<i>CasaTaller</i>	<i>ComunaCasa</i>	<i>MtoOtrosIng</i>
<i>EIRL</i>	<i>Hash_Rut</i>	<i>PersHogar</i>
<i>Mail</i>	<i>Nacionalidad</i>	<i>TotFonos</i>
<i>PersNatural</i>	<i>RegionCasa</i>	--
<i>PersJurídica</i>	<i>Sexo</i>	--

Fuente: Elaboración propia.

Respecto a las características del negocio, de un total de 11 variables, 4 son de construcción propia. Cabe destacar que las variables *Patrimonio* y *Activos* son determinadas por el ejecutivo comercial que evalúa al cliente, y no necesariamente son valores estandarizados en su cálculo. Lo mismo sucede con la variable *VtasProm*, por lo que pueden existir casos en que el valor registrado corresponda a un monto anual y no mensual, sobre todo si el cliente tiene un negocio estacional. A pesar de esto, se construyen: *Autonomía* (nivel de independencia financiera del cliente) y *Leverage* (nivel de endeudamiento). Para esta última se toman dos supuestos: primero, dado que no se conoce la deuda externa del cliente, se considera solo el monto del crédito, lo que tiene sentido para un segmento con capacidad de endeudamiento relativamente baja y, segundo, se consideran solo las ventas promedio y no la utilidad neta, pues no existe información de los costos del negocio. A continuación, se presentan estas variables y su clasificación por tipo, encontrándose en Anexo 4, Tabla 7.7, el detalle completo:

Tabla 3.19: Variables asociadas a características del negocio del cliente, clasificadas por tipo.

Variables asociadas al negocio del cliente		
Binarias	Nominales	Numéricas
<i>Agrícola</i>	<i>TrbFeria</i>	<i>MesesEmp</i>
<i>IA</i>	<i>ClasCAE</i>	<i>VtasProm</i>
--	--	<i>Trabajadores</i>
--	--	<i>RemProm</i>
--	--	<i>Patrimonio</i>
--	--	<i>Activos</i>
--	--	<i>Autonomía</i>
--	--	<i>Leverage</i>

Fuente: Elaboración propia.

Finalmente, respecto a las variables asociadas a las características del crédito, de 27 variables, 9 son construidas a partir de las existentes. Se agrega la variable *CtasReal*, que es una modificación de la variable original que indica el número de cuotas del crédito, pero cuyo valor se

altera cuando el crédito sufre una repactación. Así, esta variable corrige ese efecto no considerando esas nuevas cuotas.

También se agrega, entre otras, una variable que indica si la cuota en observación recibió un abono, lo que ocurre cuando el primer monto registrado es menor al valor total de la cuota pactada. Finalmente, se crean variables sobre el valor de la cuota, como *ValorCtaProm*, que indica el valor de la cuota promedio del crédito, *ValorCtaPorc*, que indica el porcentaje que representa la cuota del monto aprobado (sin interés) y *ValorCtaPorcPactado*, que considera el porcentaje que representa la cuota del monto pactado final, sumando el interés. La motivación es buscar medidas comparativas entre los distintos créditos, además de que se observan casos en que el cliente concentra un gran porcentaje del monto total en la cuota final. A continuación, se clasifican las variables por tipo, pudiendo encontrar información detallada de cada una en Anexo 4, Tabla 7.8:

Tabla 3.20: Variables asociadas con características del crédito otorgado al cliente, clasificadas por tipo.

Variables asociadas al Cliente			
Binarias	Fechas	Nominales	Numéricas
<i>Garantía</i>	<i>FechaCierre</i>	<i>Año</i>	<i>CtasReal</i>
<i>GesCobranza</i>	<i>FechaPago</i>	<i>Mes</i>	<i>Monto</i>
<i>AbonoCuota</i>	<i>FechaVcto</i>	<i>ClasifRiesgo</i>	<i>MoraAlCierre</i>
--	<i>FecAprob</i>	<i>CredTransferido</i>	<i>PrimerMtoPagado</i>
--	--	<i>Lc</i>	<i>TasaInt</i>
--	--	<i>ST</i>	<i>ValorCta</i>
--	--	<i>StCuota</i>	<i>ValorCtaPorc</i>
--	--	<i>SucCerrada</i>	<i>ValorCtaPorcPactado</i>
--	--	<i>SucCom</i>	<i>ValorCtaProm</i>
--	--	<i>SucZona</i>	--
--	--	<i>TipoCanal</i>	--

Fuente: Elaboración propia.

A este punto se tiene una *tabla maestra* con datos que sirven para desarrollar un modelo de originación o *application scoring*, y en un formato en que los datos preliminares permiten construir un posterior modelo de comportamiento: por ejemplo, con la fecha de vencimiento de cada cuota y la fecha de pago es posible determinar los días de atraso y conocer en qué momento el crédito cae en *default*.

3.4.3 Definición de parámetros de comportamiento

Para desarrollar un modelo de comportamiento es necesario analizar, justamente, el comportamiento de los créditos vigentes en un período de tiempo y utilizar esta información para predecir el comportamiento futuro. Para estandarizar este análisis es necesario utilizar créditos que se originan en un período de tiempo específico, en este caso se utilizarán aquellos de los cuales se

observa su historia completa entre marzo de 2007 y marzo de 2016. Siguiendo esta lógica, la ventana muestral para el desarrollo queda definida como:

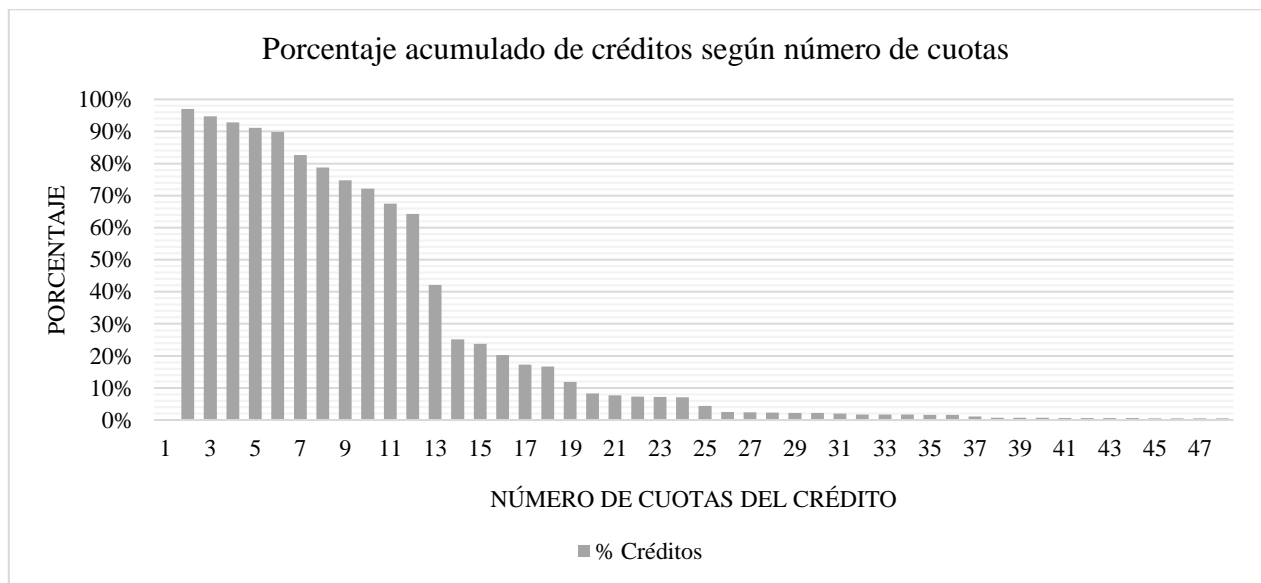
Ventana muestral (Sample window):

Créditos cuya fecha de aprobación es posterior a marzo de 2007 y cuya última cuota es anterior a marzo de 2016.

Estos créditos son observados por un período de observación definido, con el objetivo de poder determinar, en algún punto de observación presente, si en el futuro será un buen o mal pagador. Definir un mismo período de observación para todos los créditos permite comparar el comportamiento observado en distintos instantes de tiempo, así, por ejemplo, una variable que mida la mora promedio en el período de observación, mide lo mismo para un crédito que es observado en su cuota 7 y para uno que se observa en su cuota 23.

Adicionalmente, una ventana de observación debe considerar un mínimo de tiempo necesario para que las variables de comportamiento puedan capturar patrones y tendencias que discriminen entre un buen y mal pagador. En este caso, se considera que entre 3 y 6 meses de historia es un período razonable para este estudio y con el que se conserva un 93% del total de créditos y se elimina el 26.3% del total de filas:

Figura 3.9: Gráfico que muestra el porcentaje acumulado del total de créditos de EMF según el número de cuotas pactadas.



Fuente: Elaboración propia.

Período de observación (Observation period):

Para cada cuota t , se observa el comportamiento entre $t - 1$ y $máx\{t - 6, t - 3\}$. Esto implica que quedan fuera del modelo las cuotas uno, dos y tres de cada crédito y, por lo tanto, los créditos de 3 cuotas o menos.

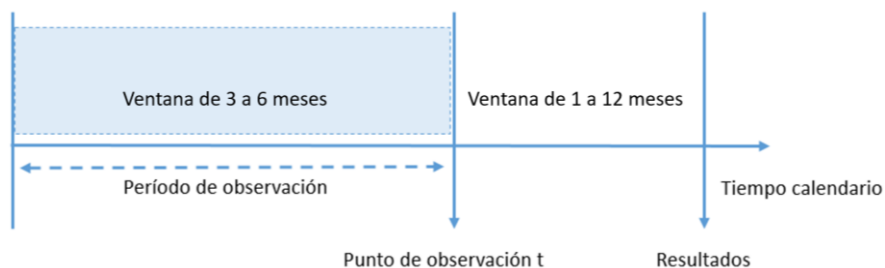
Además de la ventana de tiempo histórica, lo segundo fundamental es la ventana de tiempo de desarrollo, la que define un período de observación hacia adelante en el que se evalúa si el cliente es buen o mal pagador. Según la norma bancaria chilena, el período de desarrollo debe ser de 12 meses, lo que tiene relación con el cálculo de provisiones, y este será el valor a considerar para el desarrollo del modelo.

Ventana de desarrollo (*performance window*):

Para cada cuota t , se observa si el crédito cae en incumplimiento entre t y $\text{máx}\{CtasReal, t + 11\}$

La descripción gráfica de los parámetros anteriormente definidos es la siguiente:

Figura 3.10: Parámetros de comportamiento de Behavioral Scoring.



Fuente: Elaboración propia.

Finalmente, se decide qué se considerará como un mal cliente y, en consecuencia, quiénes serán los buenos clientes. Siguiendo también la norma bancaria, se define un mal cliente como aquel que en la ventana de desarrollo supera, para alguna cuota, los 90 días de atraso en el pago (mora). Este es el *target* final que define los valores de la variable objetivo del modelo, y las variables que permitirán discriminar entre aquellos que serán buenos y malos pagadores son las que finalmente se ingresan en la *scorecard*.

Definición de Incumplimiento (*Default*):

Se define que el crédito cae en incumplimiento en la cuota t si alguna cuota en la ventana de desarrollo iguala o supera los 90 días de mora. Este estado de incumplimiento se considera un estado absorbente y se asume que desde ese momento el crédito no se recupera.

Cabe destacar que quedan excluidos del modelo los clientes que corresponden a empresas y los que funcionan con pagos mediante líneas de crédito, ya que no corresponden al segmento objetivo que se busca representar.

3.4.4 Definición de variable objetivo y mora

En línea con lo anterior, se define el evento de incumplimiento como:

$$Y_t = \begin{cases} 1 & \text{Si entre } t \text{ y } (t + 11), \text{ MoraDefault} \geq 90 \\ 0 & \text{Si no} \end{cases}$$

donde t representa el período o cuota observada, y $MoraDefault$ los días de atraso en el pago de la cuota, construida de la siguiente forma:

$MoraDefault_t$:

$$= \begin{cases} \text{máx}\{0, (FechaPago_t - FechaVcto_t)\} & \text{si } FechaPago_t \text{ no es NULL} \\ \text{DíasEntreVctos}_t & \text{si } FechaPago_t \text{ es NULL y } FechaPago_{t-1} \text{ no es NULL} \\ (MoraDefault_{t-1} + \text{DíasEntreVctos}_t) & \text{si } FechaPago_t \text{ es NULL y } FechaPago_{t-1} \text{ es NULL} \end{cases}$$

donde $DíasEntreVctos_t$ corresponde a la diferencia en días entre la fecha de vencimiento en el período observado t y la fecha de observación en t , que para cada crédito es la fecha de vencimiento de la cuota siguiente. A continuación, se detalla un ejemplo:

Tabla 3.21: Ejemplo de cálculo de $MoraDefault$.

ID	Ncta	FechaVcto	FechaPago	FechaObservación	DíasEntreVctos	MoraDefault
15719	4	05-01-2009	30-01-2009	05-02-2009	31	25
15719	5	05-02-2009	06-04-2009	05-03-2009	28	60
15719	6	05-03-2009	30-04-2009	06-04-2009	32	56
15719	7	06-04-2009	NULL	05-05-2009	29	29
15719	8	05-05-2009	NULL	05-06-2009	31	60
15719	9	05-06-2009	NULL	06-07-2009	31	91

Fuente: Elaboración propia.

En la tabla se puede apreciar que el valor de $MoraDefault$ para la cuota $t=9$ corresponde al valor de $MoraDefault$ para la cuota anterior (dado que tampoco la pagó) más el valor de $DíasEntreVctos$ del período en observación t . Ahora, se calcula la variable $Default90$ para una variación del ejemplo anterior, obteniendo los siguientes resultados:

Tabla 3.22: Ejemplo de cálculo de Default90.

ID	Ncta	FechaVcto	FechaPago	FechaObservación	DíasEntreVctos	MoraDefault	Default90
15719	4	05-01-2009	30-01-2009	05-02-2009	31	25	0
15719	5	05-02-2009	06-04-2009	05-03-2009	28	60	1
15719	6	05-03-2009	30-04-2009	06-04-2009	32	56	1
15719	7	06-04-2009	29-05-2009	05-05-2009	29	53	1
15719	8	05-05-2009	30-06-2009	05-06-2009	31	56	1
15719	9	05-06-2009	31-07-2009	06-07-2009	31	56	1
15719	10	06-07-2009	02-10-2009	05-08-2009	30	88	1
15719	11	05-08-2009	06-08-2009	07-09-2009	33	1	1
15719	12	07-09-2009	07-08-2009	05-10-2009	28	0	1
15719	13	05-10-2009	08-08-2009	05-11-2009	31	0	1
15719	14	05-11-2009	NULL	07-12-2009	32	32	1
15719	15	07-12-2009	NULL	05-01-2010	29	61	1
15719	16	05-01-2010	NULL	06-02-2010	32	93	1
15719	17	06-02-2010	NULL	05-03-2010	27	120	1
15719	18	05-03-2010	NULL	05-04-2010	31	151	1

Fuente: Elaboración propia.

Se puede apreciar que para la primera cuota el crédito no ha caído en incumplimiento, pues entre t y $t + 11$ (cuota 4 a cuota 15) *MoraDefault* es menor a 90 días. Al contrario, para la cuota 5 el valor de *Default90* es 1 ya que la cuota 16 tiene una *MoraDefault* de 93 días. Cabe destacar que para este modelo no se considera la posibilidad de salir del estado de incumplimiento (lo que se conoce como recuperación o “cura”).

Es necesario mencionar también que no todas las cuotas de cada crédito sirven para entrenar el modelo. En el ejemplo anterior, a partir de la cuota 17 ya no agregan información relevante pues desde ese momento el crédito se da por perdido. Así, para entrenar el modelo, del ejemplo sólo es necesario considerar hasta la cuota 16 y el resto se elimina de la base de datos. De todas formas, se recomienda realizar la eliminación de estos datos después del análisis de valores perdidos, fuera de rango y *outliers*, ya que esto garantiza que queden fuera del estudio todos los casos que representan patrones muy distintos a lo que se busca estudiar.

Por otro lado, para el cálculo de las variables de comportamiento, la mora se calcula de forma distinta, cuidando de no incluir información futura y solo la que corresponde al período de observación. En este caso, para la cuota= t , se calcula la mora de todas las cuotas anteriores, dentro del período de observación, como la diferencia en días entre la fecha de vencimiento respectiva y la fecha de observación de la cuota t , la que queda fija. Aplicando esto al ejemplo, si se desea calcular, por ejemplo, la mora promedio del período de observación para la cuota $t=4$, es necesario construir la siguiente matriz:

Tabla 3.23: Ejemplo de cálculo de mora para construcción de variables de comportamiento.

Ncta	Ncta2	FechaVcto	FechaPago	FechaObservación	DiasVctoObs	Mora1	Cond	Mora
4	1	02-11-2008	02-11-2008	05-02-2009	95	0	0	0
4	2	03-12-2008	03-12-2008	05-02-2009	64	0	0	0
4	3	04-01-2009	30-01-2009	05-02-2009	32	26	0	26

Fuente: Elaboración propia.

En la tabla se ve que la fecha de observación con respecto a la cual se calculan los días de mora de cada período anterior es la de la cuota 4. Así, $DiasVctoObs_t$ es la diferencia en días entre la fecha de vencimiento y la fecha de observación. Luego, $Mora1_t$ es la diferencia en días entre la fecha de vencimiento y la fecha de pago. La variable condicional $Cond_t$ indica si $FechaPago_t > FechaObservación_t$. Finalmente:

$$Mora_t = \begin{cases} Mora1_t & \text{si } Cond_t = 1 \\ DiasVctoObs & \text{si } Cond_t = 0 \end{cases}$$

Para el mismo ejemplo, si se quiere calcular la mora de comportamiento para la cuota 10, el resultado sería el siguiente:

Tabla 3.24: Ejemplo de cálculo de mora para construcción de variables de comportamiento (2).

Ncta	Ncta2	FechaVcto	FechaPago	FechaObservación	DiasVctoObs	Mora1	Cond	Mora
10	4	05-02-2009	06-04-2009	05-08-2009	181	60	0	60
10	5	05-03-2009	30-04-2009	05-08-2009	153	56	0	56
10	6	06-04-2009	29-05-2009	05-08-2009	121	53	0	53
10	7	05-05-2009	30-06-2009	05-08-2009	92	56	0	56
10	8	05-06-2009	31-07-2009	05-08-2009	61	56	0	56
10	9	06-07-2009	02-10-2009	05-08-2009	30	88	1	30

Fuente: Elaboración propia.

En este ejemplo se puede ver cómo la fecha de pago de la cuota 9 es posterior a la fecha de observación y, por lo tanto, la contribución de días de atraso debe ser sólo 30 días y no 88, de lo contrario se estaría incluyendo información futura.

Finalmente, una vez construida esta matriz de comportamiento para todos los créditos, es posible calcular variables sobre la ventana de observación como por ejemplo la mora promedio, la mora máxima, la mora mínima, entre otras. Para el último ejemplo, la mora promedio en la ventana de observación para la cuota 9 es 52 días, la mora máxima es 60 y la mínima es 30. Cabe destacar que la variable de comportamiento no necesariamente es la mora, como se verá en la siguiente sección.

3.4.5 Variables de comportamiento

A partir de la matriz de comportamiento y basándose en las construcciones más típicas de variables para un *behavioral scoring*, se construyen variables de comportamiento sobre el período de observación, respecto al período anterior y, si existe, sobre el crédito anterior.

Respecto de la morosidad:

Tabla 3.25: Variables de comportamiento relacionadas con la morosidad.

VARIABLE	DESCRIPCIÓN
<i>MoraProm6</i>	Mora promedio del período de observación.
<i>MoraProm6Up</i>	Aproximación hacia arriba de <i>MoraProm6</i> .
<i>MoraProm6Ant</i>	Mora promedio del período de observación del mes anterior.
<i>MoraProm6UpAnt</i>	Aproximación hacia arriba de la mora promedio del período de observación del mes anterior.
<i>MinMora6</i>	Mínimo de mora del período de observación.
<i>MaxMora6</i>	Máximo de mora del período de observación.
<i>ValorCtaProm6</i>	Promedio del valor de las cuotas del período de observación.
<i>VecesMoraBajo6</i>	Veces que disminuyó la mora con respecto al mes anterior durante el período de observación.
<i>VecesMoraSubio6</i>	Veces que aumentó la mora con respecto al mes anterior durante el período de observación.
<i>BajoMoraProm6</i>	Veces que la mora del período de observación estuvo bajo el promedio del mismo período.
<i>SobreMoraProm6</i>	Veces que la mora del período de observación estuvo sobre el promedio del mismo período.

Fuente: Elaboración propia.

Respecto de la cobranza:

Tabla 3.26: Variables de comportamiento relacionadas con la cobranza.

VARIABLE	DESCRIPCIÓN
<i>GesCobranza6</i>	Veces que el cliente recibe gestión de cobranza en el período de observación.
<i>GesCobranzaAnt</i>	Si en el mes anterior el cliente recibió gestión de cobranza.

Fuente: Elaboración propia.

Respecto de los pagos realizados en el período de observación o acumulados:

Tabla 3.27: Variables de comportamiento relacionadas con los pagos del crédito.

VARIABLE	DESCRIPCIÓN
<i>AbonoCta6</i>	Veces que el cliente abonó al pago de la cuota en el período de observación.
<i>CtasPagPorc6</i>	Porcentaje de cuotas pagadas del total de cuotas en el período de observación.
<i>CuotasPagadas</i>	Cuotas pagadas del total de cuotas al mes en observación.
<i>CuotasPagadasPorc</i>	$(CuotasPagadas) * 100 / CtasReal$
<i>CtasFaltaPagar</i>	Cuotas que faltan por pagar, considerando las no pagadas y vencidas, al mes en observación.
<i>CtasFaltantesPorc</i>	$(CuotasFaltaPagar) * 100 / CtasReal$
<i>MtoAcumCtasPagadas</i>	Monto pagado correspondiente a las cuotas pagadas del total de cuotas al mes en observación.
<i>MtoPagadoPactadoPorc</i>	$(MtoAcumCtasPagadas) * 100 / MtoPactado$

Fuente: Elaboración propia.

Respecto del crédito anterior:

Tabla 3.28: Variables de comportamiento relacionadas con el crédito anterior.

VARIABLE	DESCRIPCIÓN
<i>CtasCredAnt</i>	Cuotas del crédito anterior. 0 si el cliente es nuevo.
<i>CreditosAnteriores</i>	Número de créditos anteriores que tiene el cliente. 0 si el cliente es nuevo.
<i>AbonoRelAtraso</i>	$(CtasAbonadas) * 100 / CtasConAtraso$ Relación entre cuotas abonadas y cuotas pagadas con atraso en el crédito anterior. 0 si el cliente no pagó cuotas con atraso o es nuevo.
<i>CtasConAtrasoCredAnt</i>	Cuotas que no fueron pagadas antes de la fecha de vencimiento en el crédito anterior. 0 si el cliente es nuevo.
<i>CtasConAtrasoPorcCredAnt</i>	$(CtasConAtraso) * 100 / CtasReal$ Porcentaje de cuotas que no fueron pagadas antes de la fecha de vencimiento en el crédito anterior con respecto al total de cuotas. 0 si el cliente es nuevo.
<i>DifCtasCredAntPorc</i>	$(CtasReal - CtasCredAnt) * 100 / CtasCredAnt$ Porcentaje de aumento en cuotas pactadas con respecto al crédito anterior.
<i>MtoPactadoCredAnt</i>	Monto pactado del crédito anterior. 0 si el cliente es nuevo.
<i>DifMtoPactadoAntPorc</i>	$(MontoPactado - MtoPactadoCredAnt) * 100 / MtoPactadoCredAnt$ Porcentaje de aumento del monto adeudado con respecto al crédito anterior.
<i>MtoAprobAnt</i>	Monto aprobado del crédito anterior. 0 si el cliente es nuevo.
<i>DifMtoAprobAntPorc</i>	$(Monto - MtoAprobAnt) * 100 / MtoAprobAnt$ Porcentaje de aumento del crédito aprobado con respecto al crédito anterior.
<i>ValorCtaPromCredAnt</i>	Valor promedio de la cuota del crédito anterior. 0 si el cliente es nuevo.

Fuente: Elaboración propia.

3.4.6 Limpieza de los datos

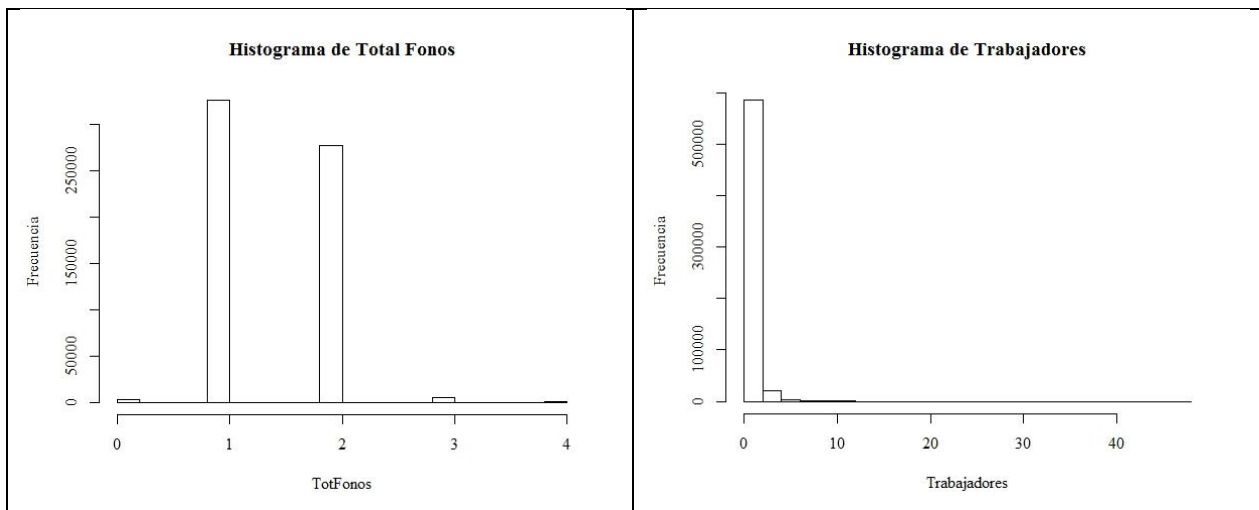
Una vez consolidada la *tabla maestra* con todas las variables tentativas, se procede a identificar y corregir problemas de valores perdidos, errores de registro, errores fuera de rango y *outliers*. También, se identifican cuáles variables quedan fuera por estar concentradas en un mismo valor (su varianza tiende a cero), y así garantizar que sólo ingresen al modelo aquellas con suficiente potencial discriminante para distinguir entre un buen o mal pagador.

3.4.6.1 Análisis de variables numéricas

Respecto de las variables numéricas, los indicadores más claros son calcular la varianza y observar la distribución mediante un histograma, su división en cuartiles y gráfico de *boxplot*. En total, se analizan 52 variables numéricas, de las cuáles: “*Hijos*” y “*PersHogar*” quedan fuera del estudio por incluir un alto porcentaje de valores perdidos, equivalente al 40% del total de los datos. “*MontoOtrosIng*” se analiza junto a la variable nominal “*CodOtrosIng*”, la que es descartada por la imposibilidad de re-agrupar los valores nulos dentro de alguna categoría sin modificar el sentido del patrón de comportamiento, quedando ambas variables fuera del modelo.

“*Total Fonos*” y “*Trabajadores*” se re-definen como variables binarias, pues en ambos casos se observa una alta concentración en el caso 0 o 1. Por otro lado, para el caso específico de *Trabajadores*, se busca corregir la posibilidad de que el ejecutivo incluya en el valor final de la variable al cliente como un trabajador más. La distribución de ambas variables es la siguiente:

Figura 3.11: Distribución de las variables numéricas TotFonos y Trabajadores.



Fuente: Elaboración propia.

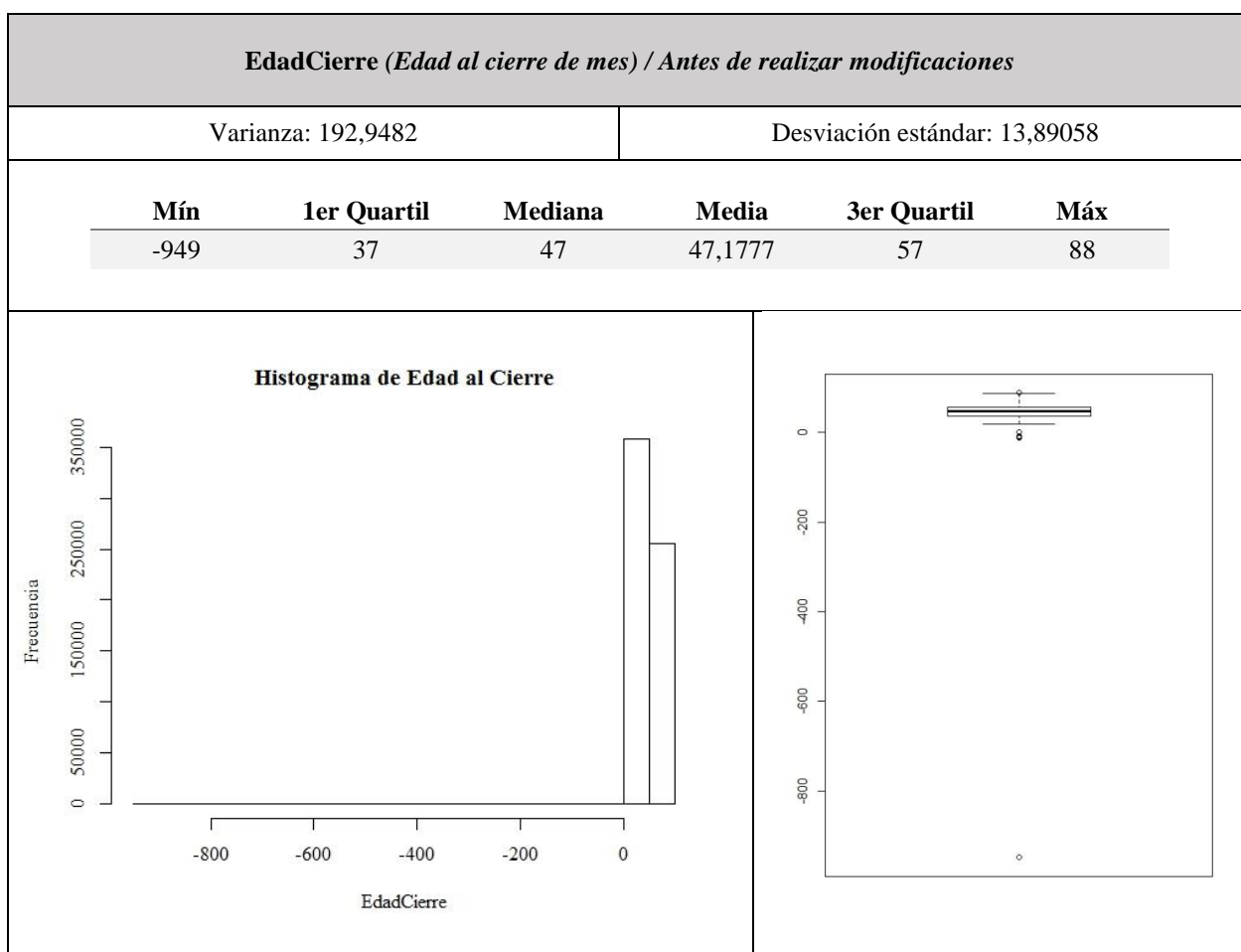
Las nuevas variables definidas son:

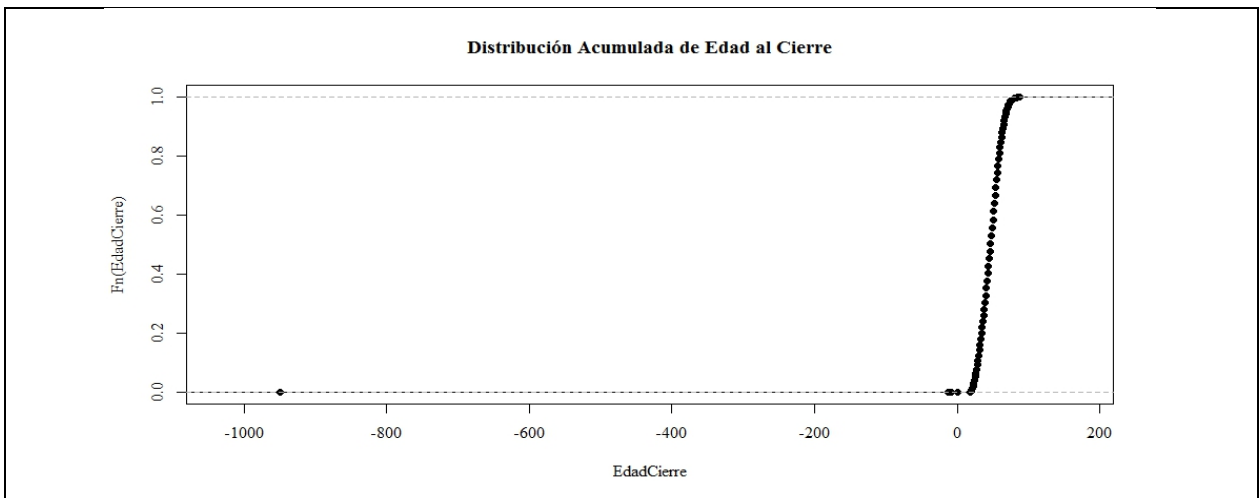
$TotalFonos_bin \begin{cases} 1 \\ 0 \end{cases} \begin{cases} \text{si el cliente declara 1 o ningún teléfono válido.} \\ \text{si el cliente declara 2 o más teléfonos válidos.} \end{cases}$

$Trabajadores_bin \begin{cases} 1 \\ 0 \end{cases} \begin{cases} \text{si el cliente declara 1 o ningún trabajador.} \\ \text{si el cliente declara 2 o más trabajadores.} \end{cases}$

A modo de ejemplo, se mostrará a continuación el caso de análisis de la variable “*Edad al Cierre de Mes*”, (el resto de las variables se analiza de la misma forma). Se identifica que el 0,01% del total de filas registra edad inferior a 18 años, valores fuera de rango que distorsionan la distribución de la variable. El primer paso consiste en reemplazar estos valores como valor perdido, para luego recalculer el valor de la mediana y reemplazar estos datos faltantes por este valor, que es más robusto que el promedio cuando la distribución no es normal, aunque para este caso particular ambos valores coinciden. Como resultado, se observa una distribución continua y similar a una variable normal, con un *outlier* estadístico en los 88 años. En este caso no es necesario eliminar este cliente, pues su comportamiento de crédito es similar al grupo etario más cercano (60 a 70 años) y, además, en la discretización de la variable es muy probable que este dato se agrupe en un mismo segmento como “65 años o más”.

Tabla 3.29: Descripción de variable EdadCierre, antes de realizar modificaciones.

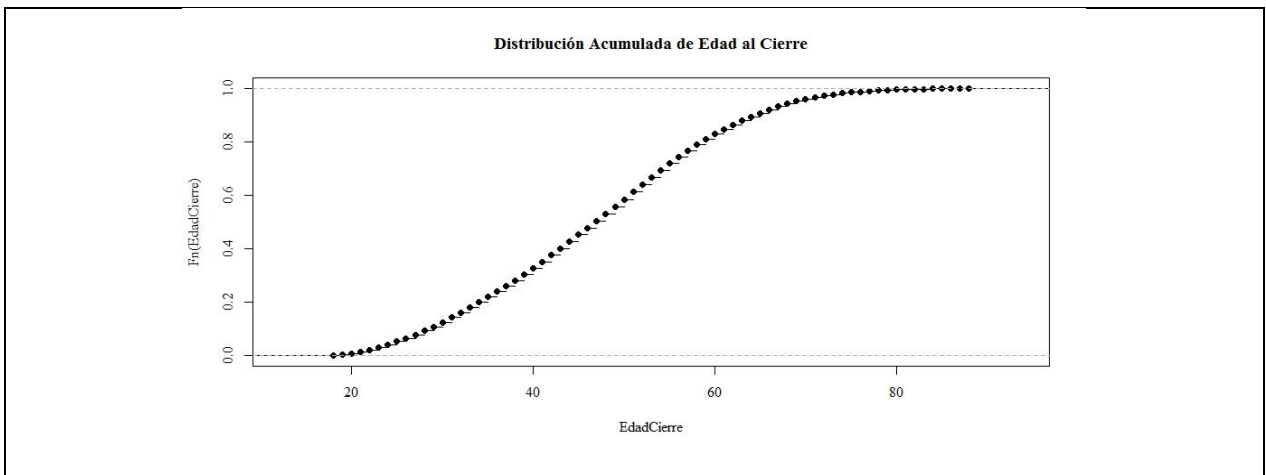




Fuente: Elaboración propia.

Tabla 3.30: Descripción de variable EdadCierre, después de las modificaciones.

EdadCierre (Edad al cierre de mes) / Después de las modificaciones					
Varianza: 178,1381			Desviación estándar: 13,34684		
Mín	1er Quartil	Mediana	Media	3er Quartil	Máx
18	37	47	47,19685	57	88
<p style="text-align: center;">Histograma de Edad al Cierre</p>					



Fuente: Elaboración propia.

Es importante destacar que existen otros casos en que sí se recurre a la eliminación de *outliers* como, por ejemplo, en el análisis de la variable “*Meses de Emprendimiento*”.

Tabla 3.31: Distribución en cuartiles de la variable MesesEmp.

Mín	1er Cuartil	Mediana	Media	3er Cuartil	Máx
-2260	57	109	143,6021	193	1401

Fuente: Elaboración propia.

En este caso, el máximo registrado alcanza 1.401 meses, lo que equivale a 116 años. Considerando que la mediana representa 9 años de experiencia en el rubro, se decide eliminar los créditos en que se registra un valor superior a los 75 años (un 0,02% del total). Por otro lado, para la misma variable, se observan valores fuera de rango en el extremo inferior, donde un 0.3% del total registra menos de 6 meses de experiencia en el rubro (lo que por políticas de la empresa no es válido). Todos estos valores se reemplazan como valores perdidos y luego por la mediana recalculada. Cabe mencionar que, dado que se consideró que tener más de 75 años de experiencia en el rubro es un *caso extremo*, es necesario mencionar a la institución que este tipo de cliente se debe evaluar de forma manual y no utilizando la *scorecard*.

Como resultado de esta etapa, la base de datos para el desarrollo de *Behavioral* disminuye en un 1,3% y queda como se muestra a continuación:

Tabla 3.32: Caracterización numérica de la tabla maestra para el desarrollo del modelo.

<i>Filas</i>	605.260
<i>Créditos únicos</i>	68.489
<i>Clientes</i>	32.607

Fuente: Elaboración propia.

3.4.6.2 Análisis de variables nominales y binarias

En esta etapa se analizan los segmentos definidos por cada variable, dejando fuera todos aquellos casos en el prácticamente el total de los casos se concentra en un único valor. Así, de un total de 34 variables categóricas (nominales y binarias), se eliminan: *CredRefinanciamiento*, *CredAntRefinanciamiento*, *CredAntReprogramación*, *CredAntPrepagado*, *CredTransferido*, *GesCobranzaAnt*, *Agrícola*, *Nacionalidad*, *Mail*.

Un análisis adicional que se realiza sobre las variables categóricas es observar, para cada segmento definido, el porcentaje de buenos y malos clientes que contiene, con el objetivo de reagrupar segmentos similares en riesgo. Tomando esto en consideración, antes de analizar a fondo las variables nominales y binarias, es necesario separar la base de datos en entrenamiento y testeo, y realizar el análisis tomando en consideración únicamente la primera muestra, de lo contrario, el modelo podría alcanzar un alto poder predictivo por razones equivocadas, ya que se estaría entrenando sobre categorías definidas con conocimiento *ex post*.

Por razones similares, es necesario aplicar antes las técnicas de muestreo que se requieran, ya que las proporciones de buenos y malos clientes para cada segmento pueden variar. Un muestreo que permite garantizar la independencia de las observaciones, supuesto teórico que permite estimar los parámetros de una regresión logística utilizando el método de máxima verosimilitud, es el muestreo de Biron [20], el que selecciona, para observaciones del mismo cliente con más de 12 meses de distancia temporal, una única observación aleatoria por cada uno. Aplicando el muestreo a la base de datos (por definición, sobre el segmento de buenos y malos clientes por separado) se obtiene la siguiente muestra, que representa el 10,5% del total:

Tabla 3.33: Muestreo de la base de datos. Representa el 10,5% del total.

	Muestreo sobre datos con <i>Default90=0</i>	Muestreo sobre datos con <i>Default90=1</i>
Filas	54.987	8.547
Créditos	54.987	8.547
Clientes	24.722	8.291

Fuente: Elaboración propia.

Por otro lado, la proporción de buenos y malos clientes de la muestra total es:

Tabla 3.34: Proporción de buenos y malos clientes de la muestra.

Filas totales	63.534
Porcentaje total de malos clientes	13,5%
Porcentaje total de buenos clientes	86,5%

Fuente: Elaboración propia.

Así, se observa una base de datos que si bien, no está idealmente balanceada, mantiene una proporción de casos de interés suficiente para entrenar un modelo que pueda identificar correctamente a un mal cliente. Teniendo esto en consideración, se descarta utilizar técnicas de sobre muestreo para el entrenamiento.

Finalmente, dado que el paso siguiente es discretizar también las variables numéricas, el análisis de segmentos se realizará para todas las variables de forma posterior, utilizando el cálculo de *Weight of Evidence*.

3.4.7 Transformación de variables

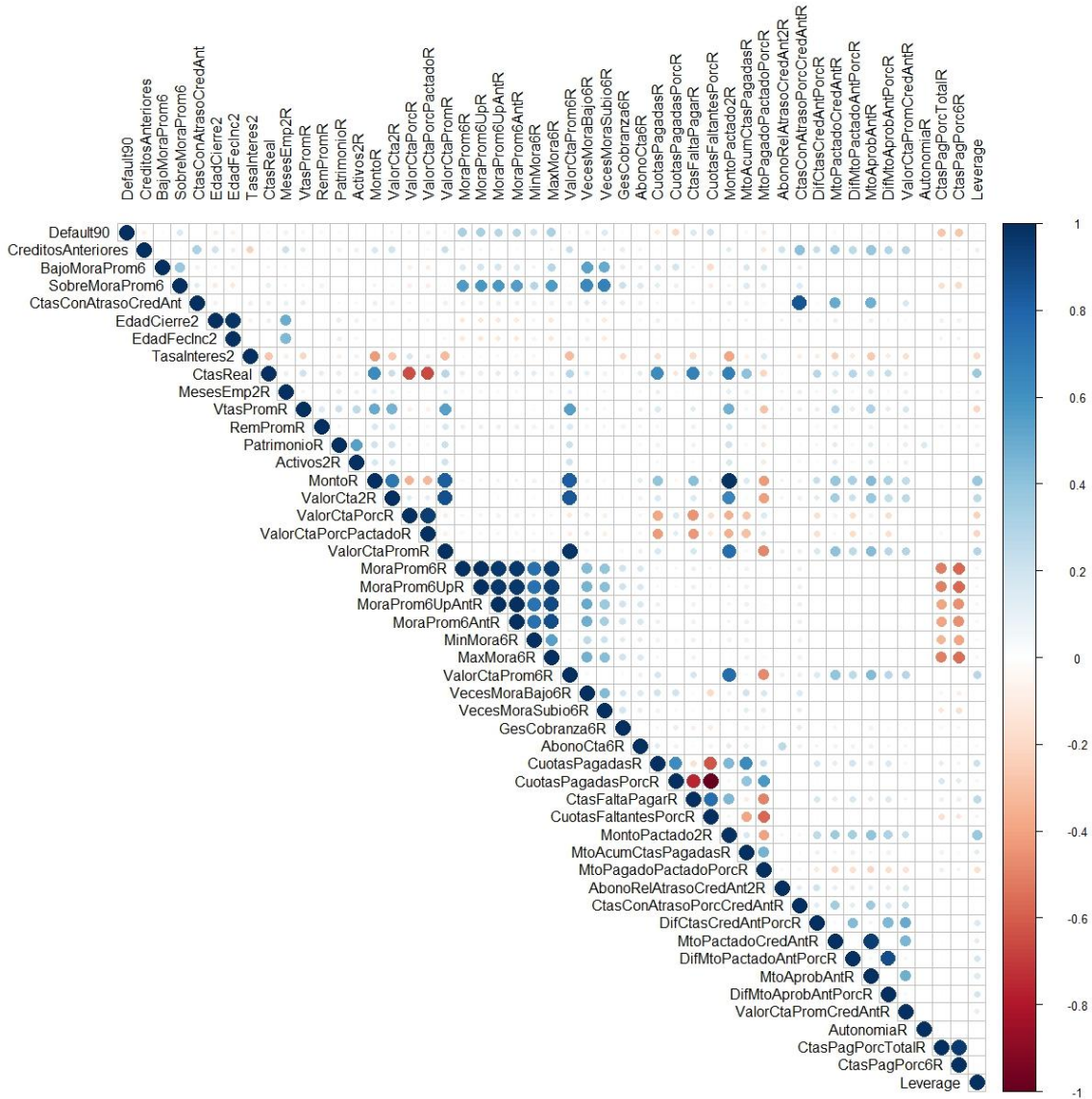
3.4.7.1 Discretización de variables numéricas

Teniendo en cuenta que el resultado final se expresará en forma de *scorecard*, es necesario discretizar las variables numéricas en segmentos que sean muy similares entre ellos y distintos al resto, con respecto a la variable objetivo. Para este procedimiento se utilizará el algoritmo de árboles de decisión, el que permite generar cortes en la variable continua de forma rápida y simple, los que servirán para definir rangos y transformar la variable en nominal.

Antes de intentar obtener los cortes de las variables, se recomienda redondear algunos valores según la cantidad de cifras significativas que se consideren relevantes, con el objetivo de evitar que el algoritmo entregue cortes poco accionables como “*segmento 3: monto de crédito entre 254.899 y 514.325*”, siendo preferible un resultado que considere 3 cifras significativas como “*segmento 3: monto de crédito entre 255.000 y 514.000*”. Esta transformación es aplicable a variables como “*Patrimonio*”, “*Monto*”, “*ValorCta*”, entre otras, mientras que hay casos que como, por ejemplo, “*Edad al Cierre*” o “*CtasReal*”, que no necesitan esta transformación. También, se recomienda identificar las variables con correlación excesivamente alta (sobre 0.9 o bajo -0.9), y eliminar aquella menos correlacionada con la variable objetivo. El resto de los casos se vuelve a evaluar más adelante, pues al transformar las variables con la discretización suele desaparecer, en algunos casos, el efecto de correlación.

El siguiente gráfico ilustra las correlaciones entre las variables numéricas consideradas. El radio del círculo indica la magnitud de la correlación y el color el sentido. Se pueden ver focos altos en las variables que miden la morosidad, el porcentaje de cuotas pagadas o el monto. Como resultado de esta selección se eliminan 15 variables.

Figura 3.12: Ilustración de correlación entre variables numéricas.



Fuente: Elaboración propia.

A continuación, se debe aplicar el algoritmo de árboles de decisión a cada variable numérica por separado, es decir, por cada variable que se desee discretizar se construye un modelo de árboles de decisión tal que la variable explicativa sea la variable numérica y la variable independiente sea la variable objetivo, en este caso *Default90*. Como el algoritmo utilizará los valores de la variable numérica para predecir la variable objetivo, es necesario utilizar únicamente la muestra de entrenamiento, porción de la base de datos que permite ajustar los parámetros del modelo. La segmentación se realiza siguiendo una proporción de 70% entrenamiento y 30% testeo, mediante un muestreo aleatorio simple:

Tabla 3.35: Muestra de entrenamiento y Muestra de testeo

<i>Muestra de entrenamiento</i>	Filas	44.473		<i>Muestra de testeo</i>	Filas	19.061	
	Buenos	38.510	86,6%		Buenos	16.477	86,4%
	Malos	5.963	13,4%		Malos	2.584	13,6%

Fuente: Elaboración propia.

Luego de segmentar, se decide incorporar a la base de testeo los datos de comportamiento que el muestreo de Biron deja fuera, quedando esta muestra con la siguiente proporción:

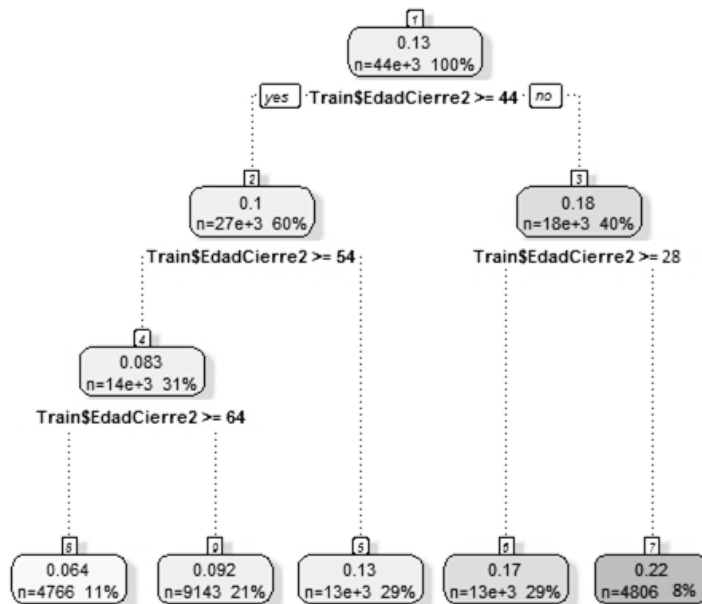
Tabla 3.36: Muestra luego de incorporar los datos que el Muestreo de Biron dejó fuera.

<i>Muestra de testeo</i>	Filas	462.405	
	Buenos	428.589	93,7 %
	Malos	33.816	7,3%

Fuente: Elaboración propia.

A continuación, se procede a discretizar las variables utilizando el árbol de decisión. A modo de ejemplo, se muestra el resultado que entrega el algoritmo para la variable “*EdadCierre*”, el que encuentra 5 grupos significativos (cada uno con un 5% o más de datos).

Figura 3.13: Discretización con árbol de decisión para la variable *EdadCierre*.



Fuente: Elaboración propia.

3.4.7.2 Cálculo de *Weight of Evidence* e *Information Value*:

A continuación, se analiza el poder individual de cada característica como predictor de comportamiento. El poder predictivo de una variable se mide utilizando dos criterios: primero, observando el poder predictivo de cada atributo de la variable, luego, observando como varía el poder predictivo a lo largo de los distintos atributos, luego, el poder predictivo de la característica observando el IV, y utilizando consideraciones del negocio.

Para analizar el poder predictivo de la variable tras la discretización se debe observar que se cumplan las siguientes reglas básicas:

- Mínimo de 5% de datos en cada segmento.
- Similar proporción de buenos y malos. No hay grupos con 0% buenos o 0% malos.
- El valor de WOE es significativamente diferente entre los distintos grupos, lo que garantiza que cada grupo se haya configurado maximizando la diferencia entre buenos y malos.
- Finalmente, se debe experimentar con diferentes agrupaciones, eliminando aquellas que no tienen sentido lógico.

Tabla 3.37: Cálculo de WOE para los atributos de la variable *EdadCierre*.

<i>EdadCierre</i>	<i>Buenos</i>	<i>Malos</i>	<i>Total</i>	<i>%Buenos</i>	<i>%Malos</i>	<i>%Total</i>	<i>%Buenos-%Malos</i>	<i>WOE</i>
[18 - 27]	2675	785	3460	7%	13%	8%	-6%	-0,6393
[28 - 43]	11829	2397	14226	31%	40%	32%	-9%	-0,2690
[44 - 54]	11245	1633	12878	29%	27%	29%	2%	0,0641
[55 - 63]	7702	791	8493	20%	13%	19%	7%	0,4105
[64 o más]	5059	357	5416	13%	6%	12%	7%	0,7858
Total	38510	5963	44473	100%	100%	100%		

Fuente: Elaboración propia.

La primera columna muestra los cortes definidos por la segmentación. La columna *Buenos* indica la cantidad de buenos clientes del segmento (tal que *Default90=0*), mientras que la columna *Malos* indica la cantidad de malos clientes (tal que *Default90=1*). Las siguientes tres columnas representan los respectivos porcentajes, donde *%Buenos* indica la proporción de buenos clientes del segmento (para [18 – 27] es 2.675), con respecto a los buenos totales (38.510), y el mismo cálculo se realiza para obtener *%Malos*. Finalmente, el tamaño de cada segmento se observa en la columna *%Total*. Como se puede observar, ningún segmento tiene una proporción de buenos y malos excesivamente polarizada y cada uno representa al menos el 5% del total, lo que indica que, en primera instancia, la agrupación es correcta.

Siguiendo con el análisis, la columna de *%Buenos-%Malos* permite observar si el segmento está más cargado hacia los malos clientes o hacia los buenos, según sea mayor o menor a 0 el resultado. Si el signo es negativo, entonces hay una mayor concentración de malos pagadores. En el ejemplo, se puede observar que los dos primeros grupos tienen una tendencia a ser peores pagadores que las personas de 44 años o más.

Finalmente, se analiza el valor de WOE para cada segmento, el que mide la diferencia entre la proporción de buenos y malos clientes para cada atributo. Mientras mayor es el valor del WOE del segmento, mayor es la probabilidad de que alguien que pertenezca a ese grupo sea un buen pagador. Contrariamente, mientras menor sea el valor del WOE del grupo, las probabilidades de que alguien que pertenezca a ese segmento sea un mal pagador son menores, con respecto al resto de los grupos. El valor de WOE se calcula como sigue:

$$WOE_i = \ln \left(\frac{\%Buenos_i}{\%Malos_i} \right) \quad (26)$$

En este caso, como se construye dividiendo la distribución de buenos clientes por la distribución de malos clientes, el WOE para cada segmento representa una medida de la “*probabilidad de ser efectivamente un buen pagador en el segmento i*”. Siguiendo con el ejemplo, para el primer grupo de la variable edad, el WOE es:

$$WOE_{[18-27]} = \ln \left(\frac{7\%}{13\%} \right) = \ln(0,53) = -0.6393 \quad (27)$$

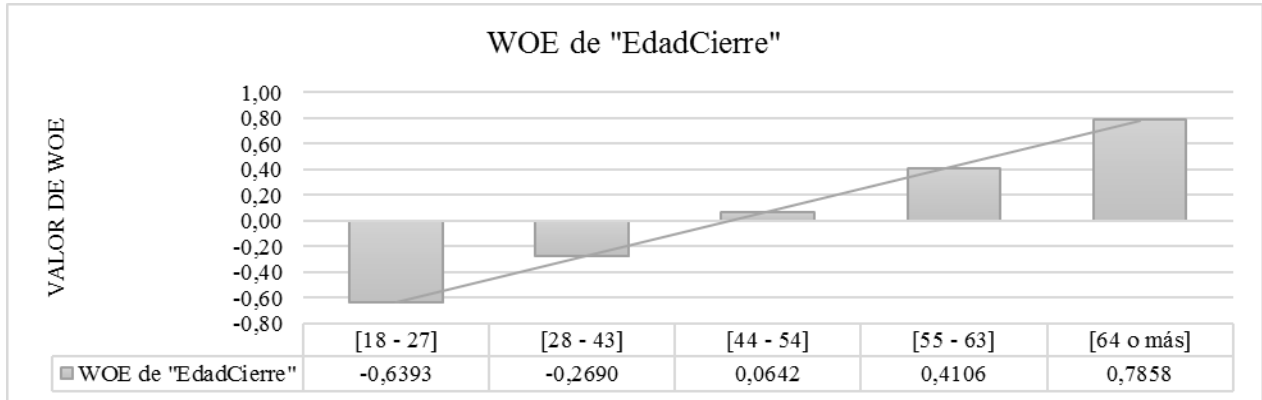
Así, un cliente de 27 años o menos, escogido de forma aleatoria, tiene 0.53 veces más oportunidad de ser un buen pagador que alguien que no presenta esa característica (en este caso, que alguien mayor). En términos estadísticos, las *odds* de que un cliente de 27 años sea un buen cliente es 0.53:1, o bien 53:100. Otra forma de interpretarlo es que, de cada 153 clientes de 27 años, escogidos aleatoriamente, 53 sean buenos pagadores, lo que representa una probabilidad de 35%. Realizando el mismo análisis sobre el segmento de 55 a 63 años, se ve claramente que este grupo es relativamente mejor:

$$WOE_{[55-63]} = \ln \left(\frac{20\%}{13\%} \right) = \ln(1,54) = 0.4105 \quad (28)$$

En este ejemplo, un cliente con edad entre los 55 y 63 años tiene 1,54:1 veces más oportunidad de ser un buen pagador que alguien fuera del segmento, es decir, de cada 254 clientes escogidos del segmento, 154 sean buenos pagadores, lo que representa una probabilidad del 60%, mucho mejor que el caso analizando anteriormente.

A continuación, una vez que se encuentra una discretización que hace sentido, es necesario verificar que la interpretación sea lógica y tenga sentido, lo que se observa mejor al graficar los valores de WOE y observar la tendencia:

Figura 3.14: Gráfico de tendencia de los WOE para los segmentos de EdadCierre.

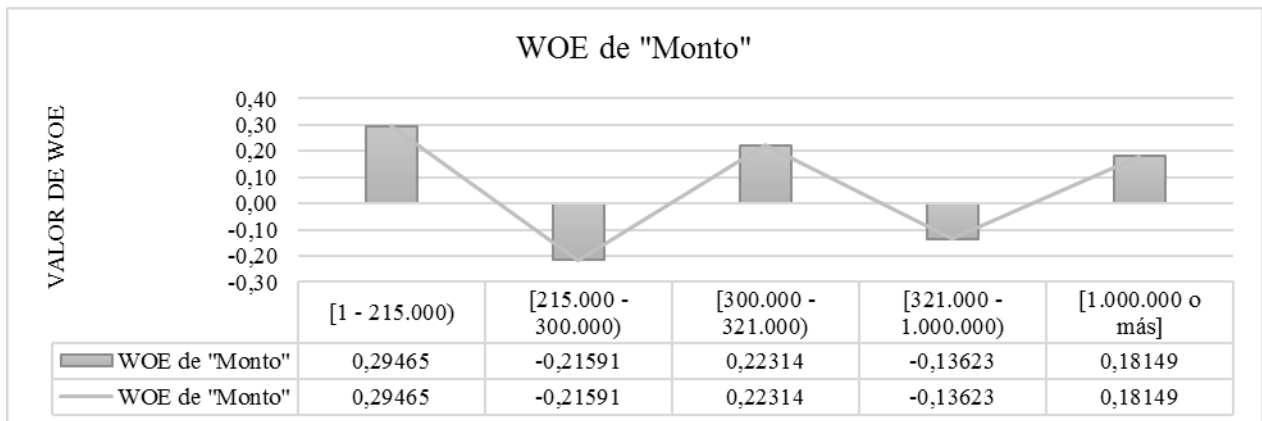


Fuente: Elaboración propia.

Para el mismo ejemplo, la interpretación del gráfico muestra que, a mayor edad, menor es el riesgo de que el cliente sea un mal pagador. De acuerdo al levantamiento de información realizado en etapas previas, para la institución este resultado es algo esperado ya que, por lo general, el segmento más joven tiene menos experiencia en el rubro y son más adversos al riesgo de caer en DICOM. Por otro lado, las personas de mayor edad, de las cuales se puede esperar que tengan su vida más o menos “resuelta”, son pagadores más responsables o maduros.

Cabe mencionar que hay casos en que la tendencia que sigue el WOE de la variable a lo largo de los segmentos no tiene comportamiento lógico, como, por ejemplo, lo que sucede al analizar la variable “Monto”:

Figura 3.15: Gráfico de tendencia de los WOE para los segmentos de Monto.



Fuente: Elaboración propia.

Casos como el anterior se pueden intentar corregir al reagrupar los segmentos. Si el comportamiento de la variable no mejora, se elimina del estudio. Por otro lado, también hay casos mucho más complicados y que requieren de un análisis más profundo, como el caso de las variables

nominales, las que tienen predefinidos los segmentos, y cuya discretización original no sirve tal como está para que la variable sea predictiva. Un buen ejemplo de esto es el caso de la variable nominal “*CodEscolaridad*”:

Tabla 3.38: Cálculo de WOE para los atributos de la variable *CodEscolaridad*.

<i>CodEscolaridad</i>	<i>Buenos</i>	<i>Malos</i>	<i>Total</i>	<i>%Buenos</i>	<i>%Malos</i>	<i>%Total</i>	<i>%Buenos-%Malos</i>	<i>WOE</i>
<i>NA</i>	1484	778	2261	4%	13%	5%	-9%	-1,2196
<i>0 (sin estudios)</i>	777	138	915	2%	2%	2%	0%	-0,1372
<i>1 (básica inc)</i>	11416	1373	12789	30%	23%	29%	7%	0,2527
<i>2 (básica com)</i>	4243	472	4715	11%	8%	11%	3%	0,3307
<i>3 (media inc)</i>	5788	780	6568	15%	13%	15%	2%	0,1389
<i>4 (media com)</i>	11914	1870	13784	31%	31%	31%	0%	-0,0136
<i>5 (tec. Prof. Inc)</i>	647	137	784	2%	2%	2%	-1%	-0,3130
<i>6 (tec. Prof. Com)</i>	1251	222	1473	3%	4%	3%	0%	-0,1363
<i>7 (Superior com)</i>	562	92	654	1%	2%	1%	0%	-0,0556
<i>8 (Superior inc)</i>	428	102	530	1%	2%	1%	-1%	-0,4312
Total	38510	5964	44473	100%	100%	100%		

Fuente: Elaboración propia.

Esta variable manifiesta varios problemas: primero, contiene un segmento de *NA* (valores perdidos), el que idealmente se debe juntar con otro segmento. En segundo lugar, se observan segmentos con menos de un 5%, los que no son significativos por si solos. Lo bueno de esta variable es que, desde un punto de vista lógico y operativo, es relativamente sencillo decidir qué grupos se pueden agrupar (esto es decisión del analista y de los expertos del negocio, y puede existir más de una configuración válida). Así, lo primero es garantizar que no exista ningún grupo con menos de un 5% de los datos, ya sea juntándolos entre ellos o agrupándolos con categorías mayores:

Tabla 3.39: Cálculo de WOE para los atributos de la variable *CodEscolaridad*, agrupación de segmentos.

<i>CodEscolaridad</i>	<i>Buenos</i>	<i>Malos</i>	<i>Total</i>	<i>%Buenos</i>	<i>%Malos</i>	<i>%Total</i>	<i>%Buenos-%Malos</i>	<i>WOE</i>
<i>NA</i>	1484	778	2261	4%	13%	5%	-9%	-1,2196
<i>0 - 1</i>	12193	1511	13704	32%	25%	31%	6%	0,2227
<i>2 (básica com)</i>	4243	472	4715	11%	8%	11%	3%	0,3307
<i>3 (media inc)</i>	5788	780	6568	15%	13%	15%	2%	0,1389
<i>4 (media com)</i>	11914	1870	13784	31%	31%	31%	0%	-0,0136
<i>5 - 6 - 7 - 8</i>	2888	553	3441	7%	9%	8%	-2%	-0,2124
Total	38510	5964	44473	100%	100%	100%		

Fuente: Elaboración propia.

Lo anterior se logra tras decidir juntar a los clientes sin estudios con aquellos que tienen educación básica incompleta, ya que no se espera una diferencia significativa entre tener un nivel mínimo de educación y nada. Luego, se decide juntar a los clientes con educación superior o

técnica, completa o incompleta, en un mismo grupo. Ahora que todos los grupos tienen al menos un 5% del total de datos, se observa que el segmento [0 – 1], es decir, aquellos que muy bajo nivel de educación, tienen un WOE que no es muy diferente del segmento 2 [básica completa], por lo que es posible juntarlos en un mismo grupo:

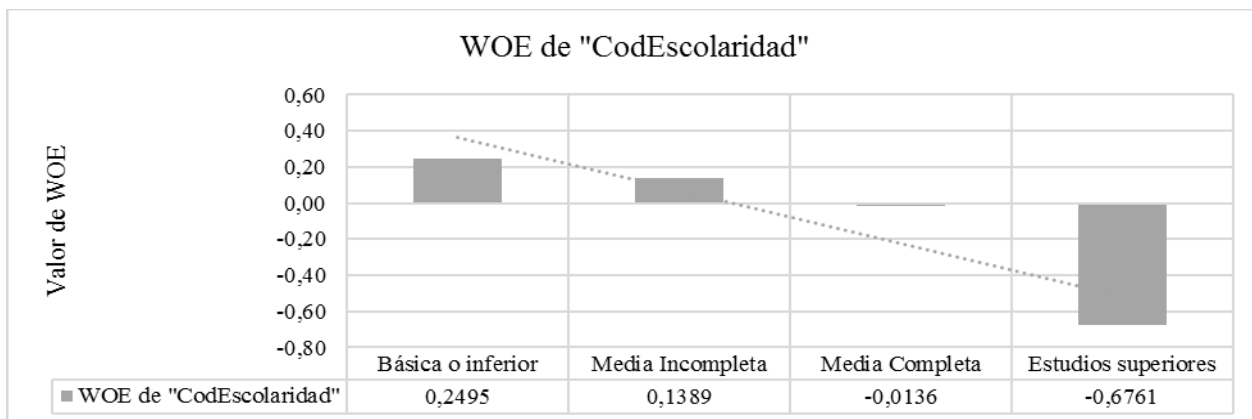
Tabla 3.40: Cálculo de WOE para los atributos de la variable CodEscolaridad, agrupación de segmentos (2).

<i>CodEscolaridad</i>	<i>Buenos</i>	<i>Malos</i>	<i>Total</i>	<i>%Buenos</i>	<i>%Malos</i>	<i>%Total</i>	<i>%Buenos-%Malos</i>	<i>WOE</i>
NA	1484	778	2261	4%	13%	5%	-9%	-1,2196
0 - 1 - 2	16436	1983	18419	43%	33%	41%	9%	0,2495
3 (<i>media inc</i>)	5788	780	6568	15%	13%	15%	2%	0,1389
4 (<i>media com</i>)	11914	1870	13784	31%	31%	31%	0%	-0,0136
5 - 6 - 7 - 8	2888	553	3441	7%	9%	8%	-2%	-0,2124
Total	38510	5964	44473	100%	100%	100%		

Fuente: Elaboración propia.

Finalmente, queda intentar agrupar el segmento de NA. Aún sin juntarlo con otro grupo, es clara la tendencia: a mayor nivel de educación, los clientes de la institución son peores pagadores. Este patrón es claro y hace sentido, ya que se esperaría que los clientes mejor calificados laboralmente dentro de la cartera sean todos aquellos que el resto de instituciones financieras no atiende, por considerarlos demasiado riesgosos o malos pagadores. Luego de intentar incorporar a los NA con el resto de grupos, se encuentra que unirlos con el grupo que tiene “algún grado de educación superior” es la mejor opción para mantener el patrón identificado:

Figura 3.16: Gráfico de tendencia de los WOE para los segmentos de CodEscolaridad.



Fuente: Elaboración propia.

El análisis de *Weight of Evidence* se realizó para un total de 35 variables numéricas y 18 variables categóricas (nominales y binarias). Las tablas y gráficos del cálculo de WOE se encuentran en Anexo 5 (variables nominales y binarias) y Anexo 6 (variables numéricas).

3.4.8 Selección de variables

3.4.8.1 Selección por nivel de la información.

A continuación, se procede a calcular el valor de información (*information value* o simplemente IV) para cada característica. Para este modelo, el representa el poder predictivo total de cada característica i para discriminar entre un buen y mal pagador, y se construye utilizando la siguiente fórmula:

$$\sum_{i=1}^n (\%Buenos_i - \%Malos_i * \ln\left(\frac{\%Buenos_i}{\%Malos_i}\right)) = \sum_{i=1}^n (\%Buenos_i - \%Malos_i * WOE_i) \quad (29)$$

Volviendo al ejemplo de la variable “*EdadCierre*”, el valor de IV de esta variable es la suma ponderada de la columna %Buenos-%Malos de cada segmento, por el respectivo valor de su WOE. Realizando el cálculo, se obtiene un valor de IV=0,15. Para saber si la variable resulta predictiva se utiliza la siguiente regla, basada en lo propuesto por *Siddiqi (2006)*:

Tabla 3.41: Regla sobre poder predictivo de IV en una variable.

Poder Predictivo	Valor de IV
Muy bajo	< 0.1
Bajo	0.1 <= IV < 0.3
Mediano	0.3 <= IV < 0.5
Fuerte	0.5 <= IV < 1
Riesgo de sobre-predicción	> 1

Fuente: Elaboración propia.

Dado que el valor de IV depende directamente de la proporción de buenos y malos dentro de un segmento, es posible mejorarlo cambiando la discretización. Esto es muy útil cuando se tiene una variable que tiene lógica y cumple con todos los requisitos necesarios según WOE, pero que su IV muestra un sobre-ajuste o muy alto poder predictivo en la muestra de entrenamiento que no generaliza a casos reales. Esto se observa en el siguiente ejemplo:

Tabla 3.42: Ejemplo de modificación de segmentación de variable para mejorar el IV.

<i>CtasPagadasPorc</i>	<i>Buenos</i>	<i>Malos</i>	<i>Total</i>	<i>%Buenos</i>	<i>%Malos</i>	<i>%Total</i>	<i>%Buenos-%Malos</i>	<i>WOE</i>
[0% - 33%)	4911	2805	7716	13%	47%	17%	-34%	-1,3053
[33% - 50%)	7168	1777	8945	19%	30%	20%	-11%	-0,4706
[50% - 60%)	7507	858	8365	19%	14%	19%	5%	0,3036
[60% - 80%)	10658	479	11137	28%	8%	25%	20%	1,2370
[80% o más)	8266	44	8310	21%	1%	19%	21%	3,3704
Total	38510	5963	44473	100%	100%	100%		
IV	1,46							

Fuente: Elaboración propia.

La variable anterior indica el porcentaje de cuotas que el cliente lleva pagadas, del total de cuotas pactadas, al momento observado. Si no fuera por el valor de IV que indica que puede existir riesgo de sobre ajuste, la variable estaría perfectamente segmentada. En vez de eliminarla a priori, se puede intentar proponer una segmentación más robusta:

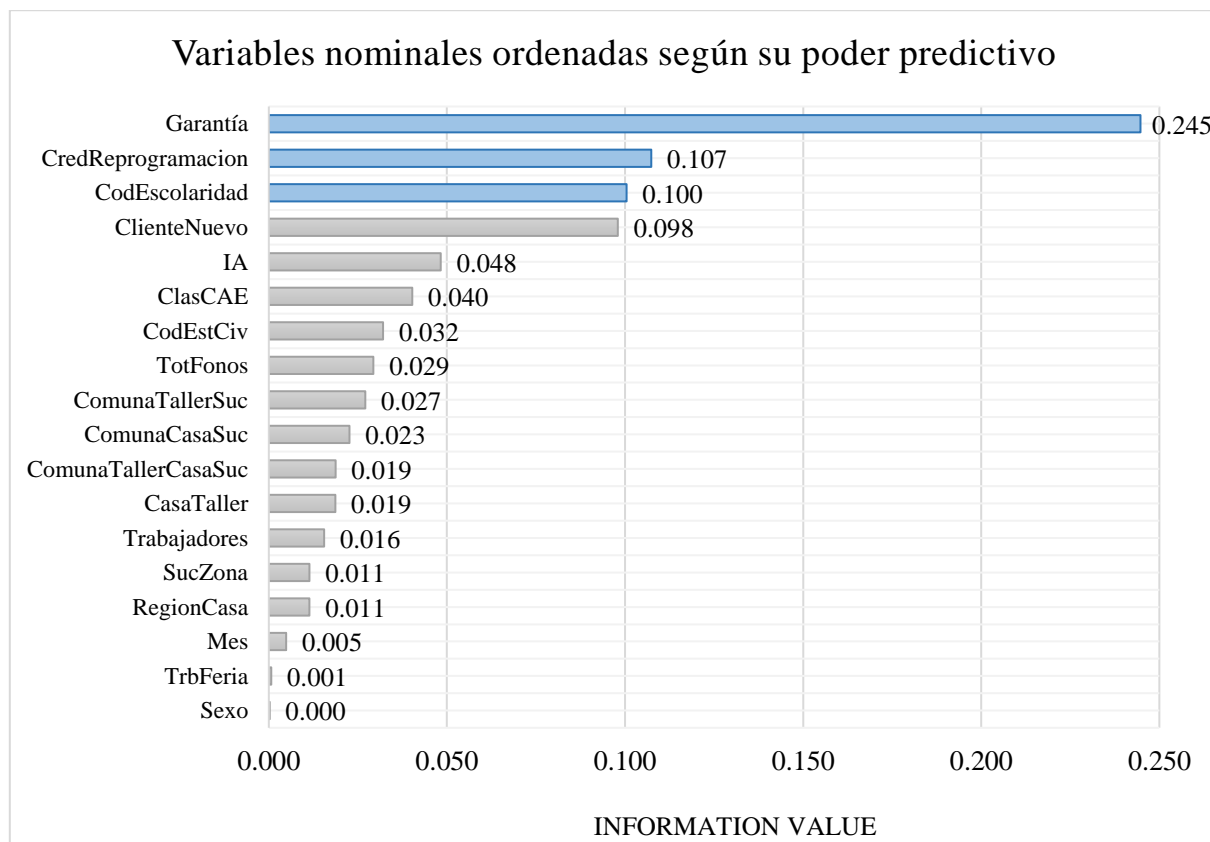
Tabla 3.43: Ejemplo de modificación de segmentación de variable para mejorar el IV (2).

<i>CtasPagadasPorc</i>	<i>Buenos</i>	<i>Malos</i>	<i>Total</i>	<i>%Buenos</i>	<i>%Malos</i>	<i>%Total</i>	<i>%Buenos-%Malos</i>	<i>WOE</i>
[0% - 33%)	4911	2805	7716	13%	47%	17%	-34%	-1,3053
[33% - 50%)	7168	1777	8945	19%	30%	20%	-11%	-0,4706
[50% o más]	26431	1381	27812	69%	23%	63%	45%	1,0864
Total	38510	5963	44473	100%	100%	100%		
IV	0,99							

Fuente: Elaboración propia.

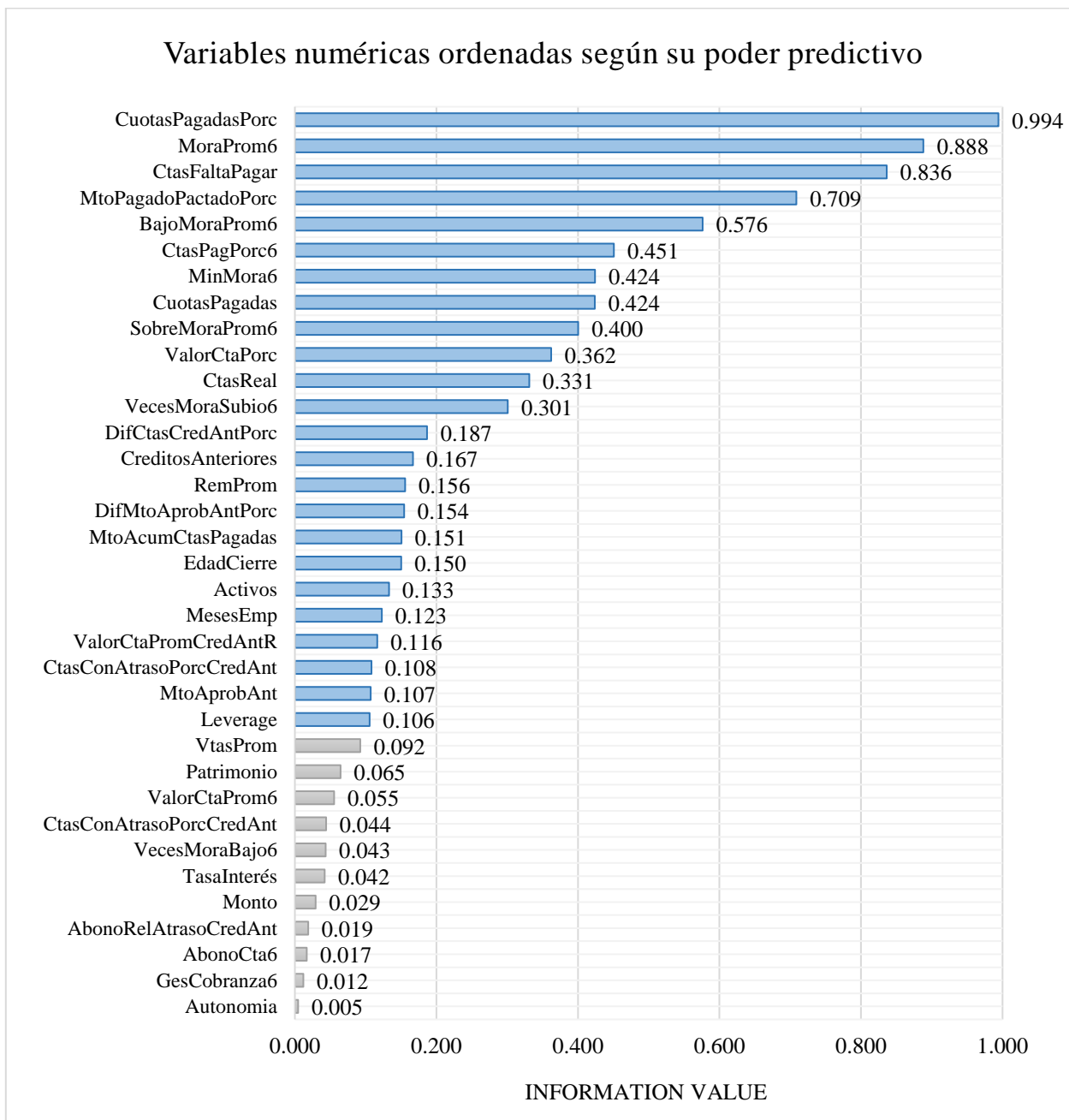
Teniendo todo lo anterior en consideración, y realizando los ajustes pertinentes, se ordenan todas las variables según su IV, para seleccionar aquellas con valor sobre 0,1:

Figura 3.17: Selección de variables nominales, según information value.



Fuente: Elaboración propia.

Figura 3.18: Selección de variables numéricas, según information value.



Fuente: Elaboración propia.

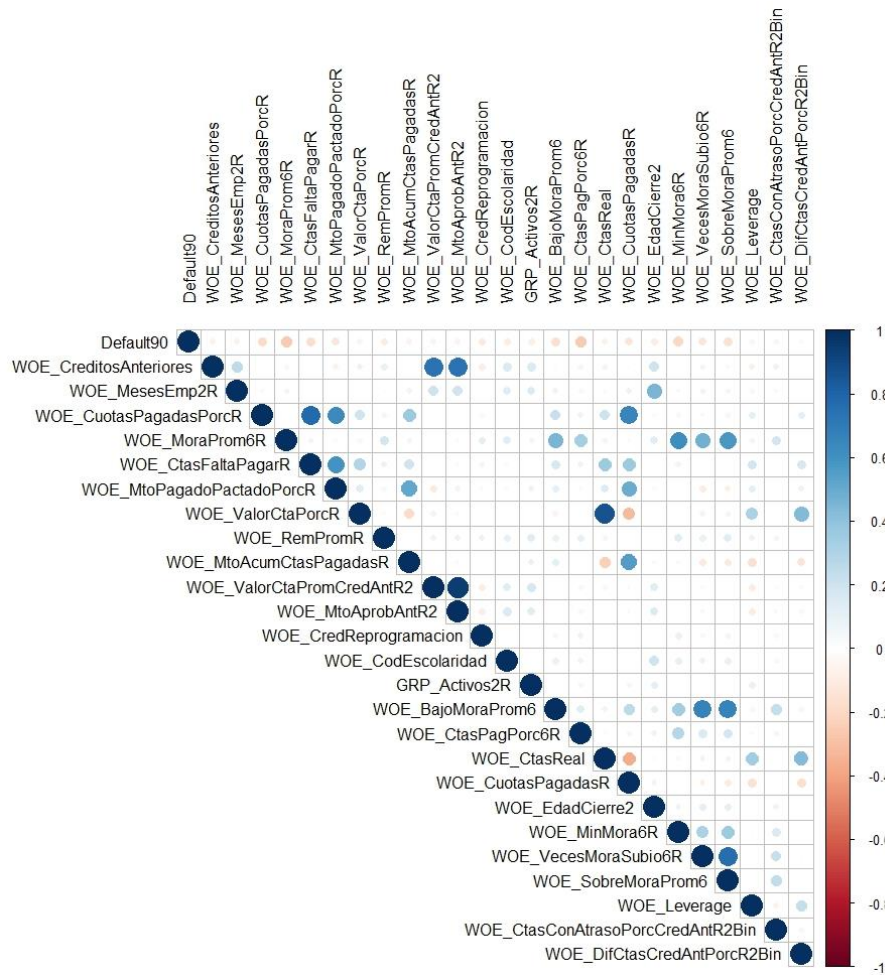
Una decisión adicional en este punto fue eliminar la variable *Garantía* pues, a pesar de que manifiesta un poder predictivo relativamente considerable, se encontró que podía estar sesgada por reglas internas de la institución, dado que el segmento menos riesgoso resultó ser aquel que no dejaba bienes en garantía. Lo anterior se explica porque la variable presenta una causalidad inversa con el comportamiento de ser mal pagador, ya que la institución sólo solicita garantías a los clientes que considera que son, efectivamente, potenciales incumplidores.

Como resultado de todo este se eliminan 27 variables, quedando 26 variables que tienen algún grado de poder predictivo para discriminar entre buenos y malos pagadores. Las variables que pasan el filtro por IV se mantienen en la base de datos reemplazando cada valor por su WOE respectivo, según el segmento al que pertenezca, y las variables que dependen de si el cliente es nuevo o antiguo, como *CtasConAtrasoPorcCredAnt* y *DifCtasCredAntPorc* se conservan como la multiplicación de la variable por la respectiva binaria.

3.4.8.2 Selección por correlación.

A continuación, se procede a realizar un segundo filtro por correlación, para verificar que la transformación aplicada por WOE no haya incrementado generando efectos inesperados en algunas variables y revisar si, para aquellas variables que en la etapa de análisis se encontró correlación, esto se mantuvo o desapareció. A continuación, se presenta el gráfico visual de las correlaciones:

Figura 3.19: Ilustración de correlación entre variables transformadas por WOE.



Fuente: Elaboración propia.

NOTA: la variable con R significa que está redondeada, y la variable con sufijo “2” es porque ha sido modificada de alguna forma por los procesos anteriores.

Se observa que existen dos pares de variables muy correlacionadas: $WOE_ValorCtaPorc$ con $WOE_CtasReal$ (0,863) y $WOE_ValorCtaPromCredAnt$ con $WOE_MtoAprobAnt$ (0,945). Definitivamente, es necesario eliminar una del último par, ya que la correlación es excesivamente alta. Al observar la correlación de estas variables con $Default90$ se encuentra que:

Tabla 3.44: Correlación de variables muy correlacionadas con respecto a la variable objetivo.

Variable	Correlación con Variable Objetivo
$WOE_ValorCtaPorc$	-0,0689
$WOE_CtasReal$	-0,0735
$WOE_ValorCtaPromCredAnt$	-0,0681
$WOE_MtoAprobAnt$	-0,0684

Fuente: Elaboración propia.

Mirando estos resultados, a priori sería correcto eliminar $WOE_ValorCtaPorc$ y $WOE_ValorCtaPromCredAnt$, aunque en el primer caso la correlación es alta pero no en exceso y, por otro lado, en el segundo caso la diferencia que motiva a eliminar una variable por sobre otra es prácticamente mínima. Dado esto, se decide consultar primero los resultados de *backward elimination* y *forward selection*.

3.4.8.3 Selección por *backward* y *forward*.

Una buena estrategia para analizar la significancia global de las variables de forma rápida es la utilización de algoritmos que comparan los resultados de distintos modelos y proveen de configuraciones que maximizan criterios de ajuste como R^2 Ajustado, AIC, entre otros.

Por su parte, *forward selection* parte de un primer modelo con N variables o ninguna, y prueba la capacidad predictiva mejora significativamente al comenzar a agregar variables progresivamente. Por el contrario, *backward elimination* busca la mejor configuración, pero iterando en la dirección contraria, es decir, tomando como primer candidato un modelo con todas las variables, y luego eliminando de una en una. Ambos algoritmos se detienen cuando ya no se puede mejorar el ajuste de forma significativa. Dado que no necesariamente se obtiene el mismo sub conjunto de variables con ambos métodos, se recomienda aplicar los dos y probar distintas configuraciones según las combinaciones de resultados.

La regresión logística sobre la cual se aplican los algoritmos de selección es la siguiente:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^N \beta_i x_i)}}$$

Donde $p(x)$ corresponde a la probabilidad de ocurrencia del evento estudiado que, en este caso, es el evento de *default*, en función de los valores de las variables dependientes $x = (x_1, \dots, x_N)$, que para este modelo corresponde a 26 variables discretas transformadas según el WOE, y un intercepto β_0 que corresponde al coeficiente de los pesos relativos de cada clase ($Default90=0$, $Default90=1$). Esta última variable se agrega en caso de tener muestras desbalanceadas, con el objetivo de dar el mismo peso relativo a los de errores de la clase con menos datos por sobre la clase mayoritaria. Por lo general los pesos recomendados son: peso de 1 a la clase con menos casos (en este caso a $Default90=1$) y peso de $Total\ de\ Buenos/Total\ de\ Malos$ a la clase con más casos. Para este modelo los valores son:

Tabla 3.45: Pesos relativos para el balanceo de la muestra.

<i>Default90</i>	<i>Peso relativo de la clase</i>
0	1
1	$(5.963/38.510) = 0,1548$

Fuente: Elaboración propia.

En Anexo 5, Tabla 7.9 y Tabla 7.10, se presentan los resultados que entregan los algoritmos de *forward* y *backward*, respectivamente. La tabla a continuación muestra la significancia estadística de cada variable de acuerdo cada uno, para un total de 5 iteraciones:

Figura 3.20: Resultados de backward elimination y forward selection.

<i>Variable</i>	<i>Significancia según backward</i>	<i>Significancia según forward</i>
<i>ActivosR</i>	***	***
<i>BajoMoraProm6</i>	No significativo	No significativo
<i>CtasPagPorc6R</i>	***	***
<i>CtasReal</i>	***	***
<i>CuotasPagadasR</i>	***	***
<i>EdadCierre</i>	***	***
<i>MinMora6R</i>	***	***
<i>VecesMoraSubio6R</i>	***	***
<i>CreditosAnteriores</i>	***	***
<i>MesesEmpR</i>	***	***
<i>CuotasPagadasPorcR</i>	No significativo	No significativo
<i>MoraProm6R</i>	***	***
<i>CtasFaltaPagarR</i>	***	***
<i>MtoPagadoPactadoPorcR</i>	***	***
<i>SobreMoraProm6</i>	**	***
<i>ValorCtaPorcR</i>	***	***
<i>RemPromR</i>	***	***
<i>MtoAcumCtasPagadasR</i>	***	***
<i>CredReprogramacion</i>	***	***
<i>CodEscolaridad</i>	***	***

<i>Variable</i>	<i>Significancia según backward</i>	<i>Significancia según forward</i>
<i>ValorCtaPromCredAntR</i>	**	***
<i>MtoAprobAntR</i>	No significativo	No significativo
<i>Leverage</i>	***	***
<i>CtasConAtrasoPorcCredAntRBin</i>	.	.
<i>DifCtasCredAntPorcRBin</i>	*	*
<i>Intercepto</i>	No significativo	

Niveles de significancia:

‘***’ 0.1%	‘**’ 1%	‘*’ 5%	‘.’ 10%
------------	---------	--------	---------

Fuente: Elaboración propia.

Donde una significancia del 5% representa que existe un riesgo del 5% de concluir que efectivamente la variable permite discriminar entre buenos y malos pagadores cuando no existe una diferencia real entre estos segmentos respecto a esta característica.

Una forma más técnica de entender el nivel de significancia es que, para el mismo caso de 5%, es el máximo nivel aceptable de riesgo para cometer un error de tipo I (en este caso, clasificar como malo a un buen cliente). El *trade off* presente en esta decisión es que mientras más pequeño sea el nivel de significancia escogido, menor es la probabilidad de cometer error tipo I, pero mayor es la probabilidad de cometer error tipo II (en este caso, clasificar como buen cliente a un mal pagador). Para decidir cuál conjunto de características permite discriminar mejor entre buenos y malos pagadores se prueban distintos modelos.

3.5 Etapa 4: Modelado

A continuación, se presentan los resultados para siete modelos, configurados a partir de los resultados anteriormente obtenidos respecto a la significancia de las variables. Los estimadores β se ajustan de acuerdo al método de máxima verosimilitud, el que permite encontrar los parámetros que maximizan la probabilidad de, efectivamente, observar lo que se está observando. Como se utilizó el muestreo de independencia de *Biron* [20], se cumple el supuesto para utilizar la función de máxima verosimilitud.

A partir de los resultados de selección que entregaron los algoritmos *backward* y *forward*, se proponen las siguientes configuraciones, dando origen a los modelos siguientes:

Tabla 3.46: Lógica según la cual se seleccionan las variables para cada modelo.

Modelo	Variables incluidas
M1:	Todas las variables.

Modelo	VARIABLES INCLUIDAS
M2:	VARIABLES QUE PARA AMBOS ALGORITMOS TIENEN UN NIVEL DE SIGNIFICANCIA DEL 10% O MÁS.
M3:	VARIABLES QUE PARA AMBOS ALGORITMOS TIENEN UN NIVEL DE SIGNIFICANCIA DEL 5% O MÁS.
M4:	VARIABLES QUE PARA AMBOS ALGORITMOS TIENEN UN NIVEL DE SIGNIFICANCIA DEL 1% O MÁS.
M5:	VARIABLES QUE PARA AMBOS ALGORITMOS TIENEN UN NIVEL DE SIGNIFICANCIA DEL 0.1% O MÁS.
M6:	VARIABLES QUE PARA AMBOS ALGORITMOS TIENEN UN NIVEL DE SIGNIFICANCIA DEL 0.1% O MÁS, Y <i>SobreMoraProm6</i> .
M7:	VARIABLES QUE PARA AMBOS ALGORITMOS TIENEN UN NIVEL DE SIGNIFICANCIA DEL 0.1% O MÁS, Y <i>ValosCtaPromCredAnt</i> .

Fuente: Elaboración propia.

Visto de otra forma, la configuración de cada modelo es la siguiente:

Tabla 3.47: Conjunto de variables incluidas en cada modelo construido.

VARIABLES	M1	M2	M3	M4	M5	M6	M7
<i>ActivosR</i>	*	*	*	*	*	*	*
<i>CtasPagPorc6R</i>	*	*	*	*	*	*	*
<i>CtasReal</i>	*	*	*	*	*	*	*
<i>CuotasPagadasR</i>	*	*	*	*	*	*	*
<i>EdadCierre</i>	*	*	*	*	*	*	*
<i>MinMora6R</i>	*	*	*	*	*	*	*
<i>VecesMoraSubio6R</i>	*	*	*	*	*	*	*
<i>CreditosAnteriores</i>	*	*	*	*	*	*	*
<i>MesesEmpR</i>	*	*	*	*	*	*	*
<i>MoraProm6R</i>	*	*	*	*	*	*	*
<i>CtasFaltaPagarR</i>	*	*	*	*	*	*	*
<i>MtoPagadoPactadoPorcR</i>	*	*	*	*	*	*	*
<i>ValorCtaPorcR</i>	*	*	*	*	*	*	*
<i>RemPromR</i>	*	*	*	*	*	*	*
<i>CredReprogramacion</i>	*	*	*	*	*	*	*
<i>CodEscolaridad</i>	*	*	*	*	*	*	*
<i>Leverage</i>	*	*	*	*	*	*	*
<i>MtoAcumCtasPagadasR</i>	*	*	*	*	*	*	*
<i>SobreMoraProm6</i>	*	*	*	*	---	*	---
<i>ValorCtaPromCredAntR</i>	*	*	*	*	---	---	*
<i>DifCtasCredAntPorcRBin</i>	*	*	*	---	---	---	---
<i>CtasConAtrasoPorcCredAntRBin</i>	*	*	---	---	---	---	---
<i>BajoMoraProm6</i>	*	---	---	---	---	---	---
<i>CuotasPagadasPorcR</i>	*	---	---	---	---	---	---
<i>MtoAprobAntR</i>	*	---	---	---	---	---	---
Total	25	22	21	18	19	19	20

Fuente: Elaboración propia.

Adicionalmente, observando el signo de los coeficientes en los modelos se encuentra que la variable *MtoAcumCtasPag* es significativa, pero tiene signo contrario a lo esperado, en todos los casos. Esta variable, que indica el monto acumulado de cuotas pagadas en porcentaje con respecto al total adeudado, tiene asociado un β positivo, lo que se interpretaría de la siguiente forma: “a medida que mayor es el porcentaje acumulado del monto de cuotas pagadas, la magnitud del efecto en el incumplimiento es mayor”. Esto se produce porque la variable se construye considerando una ventana de observación que no necesariamente es la misma para todos los clientes y, en consecuencia, existe un tercer efecto asociado a la cantidad de períodos pasados, siendo lógico esperar que se observe una mayor cantidad de clientes que caen en incumplimiento para períodos más largos de tiempo pasado. Para evitar esto, se decide eliminar la variable.

3.6 Etapa 5: Evaluación

Para analizar la capacidad predictiva de un modelo existen diversas alternativas, siendo una muy utilizada el cálculo del área bajo la curva ROC (*Receiver Operating Characteristics*), conocido como AUC, y que entrega una medida de la precisión del modelo para clasificar entre las clases estudiadas. Por otro lado, la curva ROC se construye como la relación entre la sensibilidad del modelo (casos positivos correctamente determinados) y uno menos la especificidad (casos negativos correctamente determinados), para cada punto de corte posible, según los resultados de la matriz de confusión o *tabla de contingencia*, la que permite contrastar los resultados predichos con la realidad de cada clase:

Tabla 3.48: Matriz de confusión

		Valor Predicho	
		<i>No Default</i>	<i>Default</i>
Valor Real	<i>No Default</i>	Verdaderos positivos	Falsos negativos
	<i>Default</i>	Falsos positivos	Verdaderos negativos

Fuente: Elaboración propia.

Al calcular los valores de AUC para cada modelo se encuentra que todos son muy similares. Por esta razón, y para poder discriminar entre ellos, se decide realizar validación cruzada sobre la muestra de entrenamiento en 10 subconjuntos disjuntos, estimando el modelo sobre el 90% del total de los datos cada vez y testeando sobre el 10% restante. Así, se obtiene un estimador del desempeño basado en la desviación estándar, que permitirá encontrar aquel que sea más estable.

Los resultados del AUC sobre la base de testeo y los resultados sobre la base de entrenamiento en la validación cruzada se reportan en la siguiente tabla:

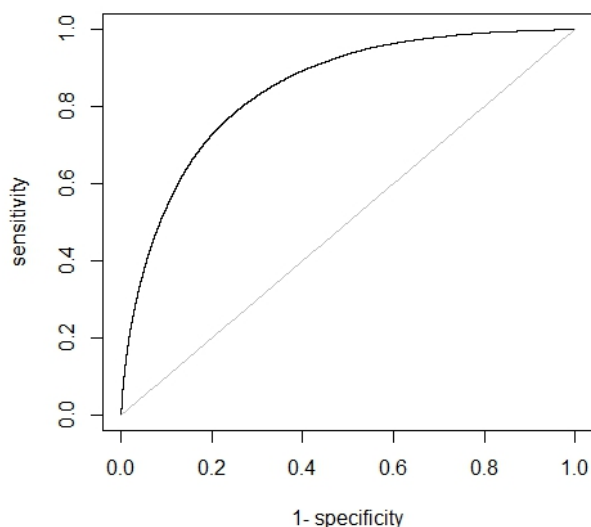
Tabla 3.49: Resultados de validación cruzada para los modelos M1 a M2, respecto al valor de AUC.

	AUC	Promedio (AUC)	Desviación Estándar (AUC)	Varianza (AUC)
M1	0.8446	0,88335	0,01121	0,000049
M2	0.8443	0,88338	0,01118	0,000049
M3	0.8442	0,88335	0,01120	0,000049
M4	0.8440	0,88322	0,01138	0,000049
M5	0.8435	0,88269	0,01152	0,000048
M6	0.8439	0,88297	0,01134	0,000049
M7	0.8436	0,88296	0,01157	0,000049

Fuente: Elaboración propia.

De lo anterior se desprende que todos los modelos siguen siendo muy similares entre sí y no hay diferencias significativas en rendimiento entre uno y otro, a pesar de que M2 destaca por tener la menor desviación estándar. Aun así, para efectos de análisis se escogerá el modelo M3, que garantiza un error del 5%, lo más utilizado en la literatura. Es importante mencionar que la empresa en cuestión podría escoger otro modelo si consideran que se ajusta mejor en términos operativos. A continuación, se presenta el gráfico de la curva ROC para el modelo M3:

Figura 3.21: Curva ROC para el modelo M3.



Fuente: Elaboración propia.

En el gráfico de la curva ROC el modelo perfecto es aquel que tiene un 100% de sensibilidad y un 100% de especificidad, es decir, aquel que clasifica correctamente todos los casos positivos (y, por lo tanto, no hay falsos negativos), y los casos negativos (por lo que no existirían falsos positivos). Por el contrario, la diagonal representa un modelo que clasifica aleatoriamente. En este caso, un AUC de 84% representa que hay un 34% de probabilidad de clasificar correctamente a un cliente, por sobre discriminación aleatoria.

Finalmente, los coeficientes que entrega este modelo son:

Tabla 3.50: Coeficientes de la nueva configuración.

Variable	β	σ_{β_i}	P – Valor	Pr(> z)
WOE_Activos2R	-0,73586	0,06981	<0,0000000000000002	***
WOE_CtasPagPorc6R	-0,3009	0,04198	0,000000000001	***
WOE_CtasReal	-0,36684	0,09997	0,000243	***
WOE_CuotasPagadasR	-1,03939	0,05654	<0,0000000000000002	***
WOE_EdadCierre2	-0,29817	0,0727	0,000041048822	***
WOE_MinMora6R	-0,19484	0,04506	0,000015314358	***
WOE_VecesMoraSubio6R	-0,32758	0,06402	0,000000311086	***
WOE_CreditosAnteriores	-0,39635	0,09863	0,000058587495	***
WOE_MesesEmp2R	-0,2727	0,08026	0,00068	***
WOE_MoraProm6R	-0,64429	0,03721	<0,0000000000000002	***
WOE_CtasFaltaPagarR	-0,23757	0,03962	0,000000002027	***
WOE_MtoPagadoPactadoPorcR	-0,41136	0,04406	<0,0000000000000002	***
WOE_SobreMoraProm6	-0,21341	0,05968	0,000349	***
WOE_ValorCtaPorcR	-0,56471	0,08871	0,000000000194	***
WOE_RemPromR	-0,39662	0,06342	0,000000000400	***
WOE_CredReprogramacion	-0,94021	0,08562	<0,0000000000000002	***
WOE_CodEscolaridad	-0,39763	0,08083	0,000000869951	***
WOE_ValorCtaPromCredAntR2	-0,43046	0,11398	0,000159	***
WOE_Leverage	-0,36342	0,08386	0,000014662388	***
WOE_DifCtasCredAntPorcR2Bin	-0,24219	0,10255	0,018187	*

Fuente: Elaboración propia.

3.7 Etapa 6: Construcción de Scorecard

Antes de analizar los resultados se procederá a transformar las probabilidades en puntajes, lo que permite simplificar mucho más la comprensión de los patrones identificados. Para este caso se utilizará la siguiente transformación logarítmica:

$$Score = Escala + Factor * \ln(odds) \quad (30)$$

3.7.1 Definición de parámetros.

La elección de los parámetros para el escalamiento a puntaje es arbitraria y toma en consideración la tendencia en la industria, que es utilizar PDO de 20 y chances de 50 (que puede cambiar según requerimientos de la empresa):

Tabla 3.51: Selección de parámetros para el escalamiento

Cantidad variables:	20
PDO	20
Score Base	650
Chances Base	50

Fuente: Elaboración propia.

Lo anterior significa que cada 20 puntos (PDO) se duplicarán las chances de que un cliente sea efectivamente un mal pagador, mientras que el Score Base es solo un valor para definir un puntaje a partir del cual se comienza a escalar el *score*. En este caso, un cliente con 650 puntos es un microempresario tal que, si se toman 50 clientes, donde todos tienen 650 puntos, se esperaría que uno de ellos caiga en incumplimiento. Respecto a las chances, al moverse 20 puntos hacia arriba estas se duplican, resultando que de 100 clientes con 670 puntos se esperaría que uno de ellos caiga en incumplimiento. Por otro lado, al moverse 20 puntos hacia abajo, las chances se dividen, resultando que, de 25 clientes con 630 puntos, uno de ellos sea mal pagador. Como resultado, se obtienen los siguientes valores para la escala, el factor y el score neutro. Todas las variables que tengan valores bajo el *score neutro* tienen asociado un riesgo alto:

Tabla 3.52: Escala, factor y score neutro

Escala	537,1
Factor	28,9
Score Neutro	26,9

Fuente: Elaboración propia.

Dado que la definición de WOE es, justamente, $\ln(odds)$, es relativamente simple obtener los puntajes para cada variable, los que se observan en la Tabla 3.53. Para hacer más intuitivo el análisis de resultados, se ha calculado cuántos puntos por encima y por debajo del score neutro aumentan las chances en un 10%, encontrándose un rango de ± 4 puntos redondeados. El aumento de las chances en un x% se obtiene como:

$$\text{Aumento en } x\% \text{ de las chances} = \frac{\ln(1+x)}{\ln(2)} * \text{Score Base} = \frac{\ln(1,1)}{\ln(2)} * 26,9 = 3,69 \sim 4$$

Tabla 3.53 Interpretación de rangos de *score* con respecto al promedio.

Rango	Interpretación
[0 – 22]	Mayor que 1,1 veces más riesgoso que el promedio.
[23 – 30]	Dentro de 1,1 veces el riesgo promedio.
[30 o más]	Mayor que 1,1 veces menos riesgoso que el promedio.

Fuente: Elaboración propia.

Tabla 3.54: Puntajes para cada variable

	Rangos	Score		Rangos	Score
Activos	[0 - 390.000)	18	MoraProm6R	[0 - 3)	47
	[390.000 - 810.000)	24		[3 - 5)	29
	[810.000 - 1.650.000)	30		[5 o más]	8
	[1.650.000 - 6.420.000)	34	CtasFaltaPagarR	[1 - 6]	34
	[6.420.000 o más]	39		[7 o más]	21
CtasPagPorc6R	[0% - 100%)	11	MtoPagadoPactadoPorcR	[0% - 32%)	20
	[100%]	29		[32% - 41%)	24
CtasReal	[1 - 8]	42		[41% - 54%)	32
	[9 - 11]	28		[54% o más]	47
	12	25	SobreMoraProm6	0	34
	[13 o más]	23		1	26
CuotasPagadasR	[0 - 3]	3		2	25
	4	20		[3 o más]	23
	[5 - 6]	30	ValorCtaPorR	[0% - 10%)	21
	[7 - 8]	41		[10% - 14%)	24
	[9 o más)	61		[14% - 17%)	36
[17% - o más]	58				
EdadCierre2	[18 - 27]	21	RemPromR	[0 - 30.000)	31
	[28 - 43]	25		[30.000 - 300.000)	21
	[44 - 54]	27		[300.000 o más]	24
	[55 - 63]	30	CredReprogramacion	0	29
	[64 o más]	34		1	0
MinMora6R	0	29	CodEscolaridad	0 - 1 - 2	30
	1	24		3 (media inc.)	28
	[2 - 3]	22		4 (media comp.)	27
	[4 o más]	20		5 - 6 - 7 - 8 + NA	19
VecesMoraSubio6R	0	37		Leverage	(0 - 0,3]
	1	26	(0.3 - 0.5]		27
	2	24	(0.5 - 0.8]		25
	[3 - 5]	23	(0,8 o más)		24
CreditosAnteriores	0	23	ValorCtaPromCredAntR		0
	1	27		(0 - 36.000)	29
	2	30		[36.000 - 57.000)	30
	3	34		[57.000 - 69.000)	32
	[4 o más]	40		[69.000 o más]	34
MesesEmp2R	[6 - 49]	23	DifCtasCredAntPorcR	[menor a -40%)	36
	[50 - 89]	26		[-40% - 6.5%)	27
	[90 - 159]	27		[6.5% - 96%)	24
	[160 - 299]	30		[96% o más]	26
	[300 o más]	32			

Fuente: Elaboración propia.

Con el puntaje calculado para cada cliente es posible definir, desde un punto de vista del negocio, cuál sería el punto de corte adecuado para la institución, el que depende de los objetivos del área de riesgo o comercial. Por un lado, un punto de corte demasiado alto asegura que la cantidad de casos no deseados sea muy baja, pero también reduce el alcance del total de casos, lo que podría no resultar rentable. Para este caso en particular, la decisión que podría depender del punto de corte es la renovación o el rechazo de un nuevo crédito a un cliente vigente, así, el objetivo es disminuir las pérdidas por renovar créditos a clientes que son malos pagadores y maximizar las utilidades por entregar créditos a los buenos pagadores.

Para tener una noción de cuánto le cuesta a la empresa dejar de dar un crédito a un buen cliente, se toman todos los créditos de clientes antiguos (que ya tenían un crédito vigente) y se calcula la utilidad promedio por cada colocación que nunca cayó en *default*, encontrándose un valor de \$243.604. Por otro lado, para calcular cuánto le cuesta a la empresa renovar un crédito a un mal pagador, se toman nuevamente todos los créditos de clientes antiguos y se calcula, para todos los que caen en *default*, el monto promedio que nunca pagaron, lo que da un total de \$678.207. Se construye con esta información la siguiente tabla:

Tabla 3.55: Puntos de corte

Punto de Corte	Buenos aceptados	Malos aceptados	%Buenos aceptados	%Malos rechazados	Costo promedio por crédito
500	428589	33816	100,0%	0,00%	\$49.598
525	428586	33798	100,0%	0,05%	\$49.575
550	428425	33342	100,0%	1,40%	\$49.063
575	425636	30198	99,3%	10,70%	\$46.608
600	411206	22703	95,9%	32,86%	\$45.365
605	405670	20770	94,7%	38,58%	\$46.059
610	398837	18834	93,1%	44,30%	\$47.493
615	390657	16722	91,1%	50,55%	\$49.399
620	380516	14637	88,8%	56,72%	\$52.446
625	368080	12625	85,9%	62,67%	\$56.883
650	274882	4527	64,1%	86,61%	\$98.353
675	164660	1137	38,4%	96,64%	\$154.665
700	67781	196	15,8%	99,42%	\$207.034
725	9903	15	2,3%	99,96%	\$239.002
750	520	0	0,1%	100,00%	\$243.309

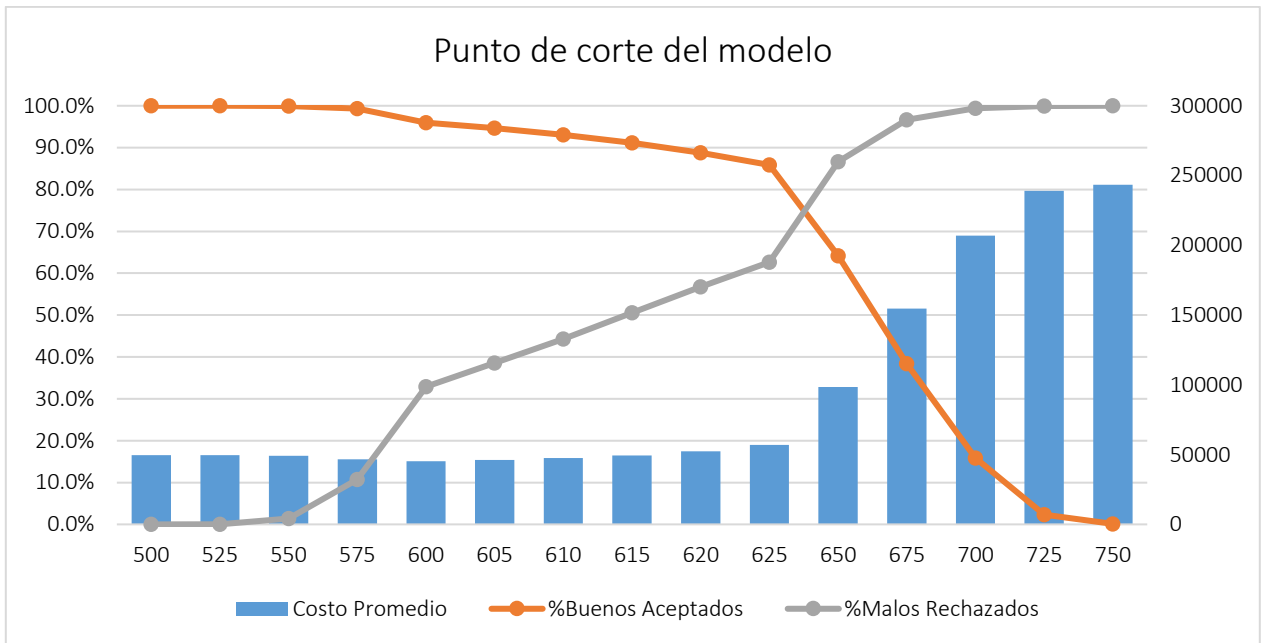
Fuente: Elaboración propia.

Donde el costo promedio se calcula como:

$$\text{Costo Promedio} = \$ 678.207 * \text{MalosAceptados} + \$243.604 * \text{BuenosRechazados}. \quad (31)$$

Finalmente, graficando para cada punto de corte la tasa de buenos clientes que se aceptan y de malos clientes que se rechazan, se obtiene el siguiente resultado:

Figura 3.22: Gráfico del punto de corte del modelo, respecto al porcentaje de créditos rechazados y costo promedio del crédito.



Fuente: Elaboración propia.

Del gráfico se infiere que a partir de los 625 puntos el costo promedio por renovación de créditos comienza a aumentar de forma más drástica. A ese nivel, el costo promedio por crédito es de \$56.883 por renovación, filtrando al 62,7% del total de malos pagadores. Nuevamente, el punto de corte dependerá de la estrategia de la empresa. Un caso extremo sería aquel en que la institución desee filtrar al 100% de los malos pagadores, lo que, en ningún caso sería sustentable en el tiempo, pues su cartera se reduciría a 0.

Cabe mencionar que este gráfico no considera los créditos que se recuperan, factor importante dentro del cálculo de la exposición al incumplimiento. Aun así, es una aproximación útil para la empresa en caso de no tener la capacidad de estimar de forma más compleja el punto de corte, y que siempre puede modificar agregando más información.

Capítulo 4

Análisis de resultados

Un hallazgo general importante, a partir de los resultados de la *scorecard*, es que no hay una única variable que dirija el enfoque del riesgo de crédito hacia una u otra característica por sobre el resto, sino que existe un efecto multivariado en el que todas las variables se relacionan entre ellas y la discriminación proviene del conjunto completo.

Aun así, es posible identificar segmentos accionables de alto riesgo, como por ejemplo aquellos créditos con menos de 4 cuotas pagadas, y aquellos en que la mora promedio de los últimos 6 meses es de 5 días o más, o lo que es lo mismo, créditos con 30 días o más de mora en los últimos 6 meses. Teniendo esto en consideración, se esperaría que la empresa tenga principal cuidado con estos grupos, identificándolos correctamente y realizando un seguimiento adecuado de esta cartera, pudiendo realizar gestiones de cobranza temprana para evitar que caigan efectivamente en incumplimiento.

Por otro lado, a partir de los resultados del punto de corte, la empresa puede definir políticas de pre-aprobación o mantención de cartera para los clientes con puntajes altos, como por ejemplo sobre los 675 puntos, donde la tasa de rechazo de malos clientes es muy alta, sobre el 96%.

4.1 Variables que no discriminan

El hecho de que una variable resulte no discriminante también es un resultado revelador. Por ejemplo, en este caso se encontró que la característica sexo no influye en que el cliente presente un mejor o peor comportamiento de pago. Lo mismo sucede con el caso de tener o no iniciación de actividades, ya que el segmento que atiende la institución concentra un porcentaje muy bajo de microempresarios formales.

Por otro lado, un resultado importante es que la variable *CasaTaller*, que indica si el cliente trabaja desde su casa o no, no resultó predictivamente relevante para discriminar entre un buen o mal pagador, contrario a lo que los expertos del negocio esperaban, siendo irrelevante desde el punto de vista del riesgo de crédito.

4.2 Variables que resultan discriminantes

A continuación, se analizarán los resultados de la *scorecard* construida en la Tabla 3.54, interpretando la lógica de los segmentos encontrados desde el punto de vista del riesgo.

Activos. La variable activos indica el monto en pesos chilenos que el ejecutivo de la institución registra en la base de datos, y que es determinado por él. En un principio se tuvieron dudas sobre el poder predictivo de esta variable por la incapacidad de verificar que los montos registrados fueran una verdadera aproximación de la realidad, pero finalmente resultó significativa. Observando el resultado de la *scorecard*, se encuentra que los clientes con menos activos son más riesgosos que el resto de los grupos, lo que se explicaría por la menor capacidad de transformar de forma relativamente fácil sus bienes en liquidez y, en caso de ser necesario, recurrir a esto para pagar su deuda.

CtasPagPorc6R. Esta variable, que indica el porcentaje de cuotas pagadas del total de cuotas vencidas en los últimos 6 meses, termina transformándose en una variable binaria donde aquel cliente que está al día (es decir, $CtasPagPorc6R=100\%$) es menos riesgoso que alguien que está atrasado en alguna cuota, resultado esperable y razonable.

CtasReal. Esta variable indica el total de cuotas pactadas en el crédito vigente en observación. A partir de los resultados de la *scorecard* se encuentra que los créditos que se pactan en 8 cuotas o menos son mucho menos riesgosos que el promedio, mientras que aquellos que se pactan en 13 cuotas o más de 12 cuotas están en el límite de superarlo. Este resultado sigue la lógica esperada de que, a mayor cantidad de cuotas pactadas, mayor es la probabilidad de que el cliente deje en algún momento de pagar.

CuotasPagadasR. Indica el total de cuotas pagadas a la fecha de observación. Esta variable muestra que a medida que más cuotas ha pagado el cliente, menor es la probabilidad de incumplimiento futura. Considerando que el promedio de cuotas pactadas de los créditos de la institución es 12, este resultado muestra que es poco probable que los clientes que ya han pagado más de la mitad de sus cuotas dejen de hacerlo. Así, se podría inferir que el esfuerzo adicional por pagar una cuota más es cada vez menor, considerando que el cliente tiene acceso restringido al financiamiento bancario y el beneficio de mantener un buen comportamiento de pago es alto. Esto puede ser revelador para la empresa, en el sentido de que puede considerar ofrecer renovaciones de crédito a segmentos que hayan pagado 7 cuotas o más. Cabe destacar que el hecho de que $cuotaspagadas=4$ sea un grupo en sí mismo es por construcción del modelo de comportamiento, que considera los créditos con 3 cuotas o más de historia.

EdadCierre2. Esta variable, que indica la edad del cliente al cierre de mes, cumple con lo esperado por los expertos del negocio, pues muestra que existe un segmento más joven, que es peor pagador que el resto, y un segmento sobre los 64 años que es tiene un comportamiento de pago mejor. Esto se explicaría, por un lado, porque los jóvenes microempresarios tienen menos experiencia en su negocio y les importa menos caer en DICOM por primera vez. Por otro lado, los

adultos mayores se preocupan más de mantener un buen comportamiento dado que, dada su avanzada edad, es más difícil que consigan financiamiento en otras instituciones mayores.

MinMora6R. Esta variable indica la mínima mora de los últimos 6 meses. Los resultados del modelo muestran que mientras mayor sea el mínimo de días de mora, mayor es la probabilidad de que el cliente caiga en incumplimiento, mientras el caso de $MinMora6R=0$ es el mejor escenario posible, aquel en que el cliente no se ha atrasado en sus pagos en los últimos 6 meses.

VecesMoraSubio6R. Esta variable indica las veces en los últimos 6 meses que la mora se incrementó de forma consecutiva, lo que refleja una clara tendencia hacia un futuro incumplimiento. Así, a medida que más veces seguidas se haya registrado un aumento en los días de mora, mayor es el riesgo de que el cliente deje de pagar el crédito en el futuro, siendo el mejor de los casos aquel con $VecesMoraSubio6R = 0$ (el cliente se ha mantenido al día en los últimos 6 meses).

CreditosAnteriores. Representa el total de créditos anteriores que el cliente ha obtenido en la institución, y muestra que mientras más créditos haya finalizado, menor es el riesgo de que en algún momento el cliente deje de pagar un crédito. Esto podría corresponder al resultado de un efecto de fidelización de la institución sobre el cliente, el que efectivamente funciona, o bien a su acumulación de experiencia crediticia, esto es, el cumplimiento de plazos de pago y ahorro para el pago de las cuotas, capacidad que no todos los clientes nuevos tienen pues muchos de ellos son no bancarizados, y por lo mismo son el segmento más riesgoso dentro de esta clasificación.

MesesEmp2R. Indica los meses de experiencia del cliente en su negocio. La *scorecard* muestra que mientras más años de experiencia tiene el cliente en su microempresa, menor es la probabilidad de que deje de pagar el crédito, lo que es esperable dado que aumenta su capacidad de reaccionar asertivamente ante momentos de crisis o bien, ha alcanzado cierta estabilidad económica. Concretamente, el resultado muestra que clientes con más de 13 años de experiencia tienen un riesgo bajo el promedio. El resultado de esta variable sigue la tendencia esperada por la institución.

MoraProm6R. Corresponde a la mora promedio de los últimos 6 meses, o bien, si se multiplica por 6, la mora promedio. A partir de los resultados de la *scorecard* se observa que mientras mayor sea la mora promedio, mayor es el riesgo de caer en incumplimiento, resultado que sigue una lógica esperada. Dentro de los segmentos, el grupo de clientes que tiene 30 días de mora o más es el más riesgoso, y en general, es altamente riesgoso, muy por debajo del promedio. Así, dado esto, la empresa debería gestionar con cuidado esta cartera de créditos, aplicando cobranza temprana en cuanto el cliente pase los 30 días de mora (o incluso evitar que llegue a este punto).

CtasFaltaPagarR. Cuotas totales que faltan por pagar. Esta variable muestra que si el cliente ha pagado más de la mitad del crédito (considerando que el total de cuotas promedio de la institución es 12), entonces la probabilidad de caer en incumplimiento es mucho menor que si no.

Nuevamente, esta es una variable que permite a la institución segmentar su cartera e identificar potenciales clientes a los cuales podría ofrecer renovaciones, una vez pasado este umbral.

MtoPagadoPactadoPorcR. Porcentaje pagado del total del monto pactado. Esta variable muestra que si el cliente ha pagado menos del 32% del total del monto del crédito pactado, es más riesgoso que el promedio y se debería mantener en observación. Por otro lado, si el cliente ya ha pagado el 41% o más del monto pactado, entonces el riesgo de que caiga en incumplimiento disminuye por debajo del promedio, y la empresa podría considerar ofrecer nuevos productos o renovaciones, una vez alcanzado este punto.

SobreMoraProm6. Veces que la mora superó la mora promedio en los últimos 6 meses. Se observa que a medida que la mora del cliente superó más veces la mora promedio en los últimos 6 meses, mayor es el riesgo de incumplimiento futuro, siguiendo la lógica esperada, siendo el mejor de los casos aquel en que *SobreMoraProm6* = 0.

ValorCtaPorR. Porcentaje que representa la cuota del monto pactado. Al observar el resultado de esta variable se podría interpretar que mientras mayor es el valor de la cuota pactada, menor es el riesgo de caer en incumplimiento futuro, pero existen dos efectos adicionales a considerar y que no permiten inferir directamente este resultado: primero, que hay varios casos de clientes que prepagan el crédito efectuando una última cuota equivalente al saldo completo de la deuda, y segundo, que el valor de la cuota depende de la cantidad de cuotas pactadas.

RemPromR. Remuneración promedio que declara el microempresario. Esta variable muestra un comportamiento particularmente no lineal con respecto al riesgo de crédito, en que existe un primer segmento en que el microempresario declara realizar alguna remuneración por menos de \$30.000 al mes, y que es de menor riesgo promedio que el resto de los grupos, un segundo segmento con una remuneración cercana al sueldo mínimo, y que resulta ser el segmento más riesgoso, seguido por un grupo de remuneración mayor al anterior. La interpretación lógica que tendría este resultado es que hay microempresarios que por necesidades laborales deben contratar a un trabajador medianamente calificado, aumentando sus obligaciones mensuales y disminuyendo la probabilidad de que ante un mes de bajas utilidades pueda priorizar por el pago del crédito. Por otro lado, el primer caso, que es el menos riesgoso, podría corresponder a un cliente que paga a un familiar o amigo por servicios esporádicos, de los que podría prescindir en caso de tener bajas utilidades durante un mes.

CredReprogramacion. Si el crédito actual corresponde a la reprogramación de un crédito anterior. Esta variable es categórica al mostrar que un crédito reprogramado es, por naturaleza, altamente riesgoso, y la institución debería constantemente monitorear esta cartera, pues existe una probabilidad muy alta de que caigan en incumplimiento.

CodEscolaridad. Nivel educacional del cliente al solicitar el crédito. Esta variable muestra que, respecto al nivel de escolaridad, los clientes más riesgosos de la institución son los

profesionales con educación superior o técnica, completa o incompleta. Esto se debe, probablemente, a la institución califica como sujetos de crédito a individuos que pertenecen a un segmento de la microempresa más calificado, pero que justamente son aquellos que la banca deja de atender por mantener mal comportamiento de pago y, ante esto, optan por solicitar créditos en instituciones menores. Ante esta situación, se recomendaría a la empresa revisar sus políticas de admisión e identificar por qué no están aplicando mejores filtros de selección en este segmento.

Leverage. Nivel de endeudamiento del cliente, se calcula como el monto del crédito sobre las ventas mensuales. Siguiendo la lógica esperada, los clientes que tienen un *leverage* menor son menos riesgosos que el resto, pues su nivel de endeudamiento es menor y tienen la capacidad para cumplir con sus pagos.

ValorCtaPromCredAntR. Valor promedio de la cuota del crédito anterior. Esta variable muestra que aquellos clientes cuyo valor de la cuota promedio del crédito anterior es mayor, tienen un menor riesgo de caer en incumplimiento en el crédito actual. Esto podría explicarse por el hecho de que un cliente que pudo pagar cuotas altas, tiene alta capacidad de ahorro y solvencia para sostener una deuda.

DifCtasCredAntPorcR. Diferencia porcentual del aumento en cuotas pactadas con respecto al crédito anterior. Esta variable muestra que, si el crédito actual del cliente aumenta en casi un 100% el total de cuotas pactadas, con respecto a su crédito anterior, entonces es más riesgoso que un cliente que pacta un nuevo crédito, pero disminuyendo el total de cuotas. Este resultado no se puede analizar aislado del monto del crédito pues puede existir el caso de que el cliente incremente el total de cuotas dado que también pidió un crédito mayor.

Es clave destacar que los patrones encontrados en los datos provienen de un comportamiento observado de los clientes y que, si bien la institución podría estar tentada a querer modificar este comportamiento para disminuir el riesgo, de hacerlo estaría modificando la ocurrencia del patrón y, en consecuencia, alterando la validez del modelo. Ante esta situación, se recomienda a la institución no influir directamente en el comportamiento del cliente (por ejemplo, no decidir dejar de otorgar créditos de 12 cuotas o más porque el riesgo de este grupo es mayor), sino que modificar su propio comportamiento por medio de estrategias de mitigación, como cobranza temprana. Ahora bien, en caso de que la institución insista en querer modificar directamente el comportamiento del cliente, puede hacerlo, pero deberá volver a evaluar y calibrar el modelo a corto plazo.

A modo de conclusión, y basándose en los resultados anteriores, se recomienda a la institución observar principalmente los siguientes segmentos accionables de alto riesgo: créditos reprogramados, créditos con menos de 4 cuotas pagadas a la fecha y créditos que a la fecha presentan 30 días de mora o más. Por otro lado, se recomienda identificar qué segmentos podrían representar una oportunidad para renovar créditos: créditos con más de la mitad del total de cuotas pagadas o bien que lleven acumulado un pago que represente el 41% del total pactado o más.

Capítulo 5

Conclusiones y recomendaciones

El segmento de microempresa, tanto en Chile como en América Latina, se ha incrementado notablemente en los últimos años, llegando a convertirse incluso en la principal fuente laboral de la mayoría de los países, contribuyendo a la disminución de la pobreza y el desempleo. Desde este punto de vista, es deseable impulsar el crecimiento y progreso de los microempresarios, garantizando un entorno seguro para su desarrollo, lo que involucra entre otras cosas, el correcto funcionamiento de instituciones de microfinanzas, sean bancarias o no.

Por un lado, las instituciones bancarias en Chile son bastante estables ante eventos de crisis económica, pues han ido avanzando relativamente a la par con el resto de instituciones internacionales que han implementado modelos de riesgo de crédito según estándares de Basilea II y, ahora, en miras de Basilea III. Si bien estas instituciones también entregan créditos a la microempresa, el cliente objetivo no es el microempresario más vulnerable social y económicamente, individuo que termina siendo atendido por otras empresas de microfinanzas de tipo no bancarias.

Instituciones como Emprende Microfinanzas son las que se encargan de inyectar capital al microempresario que los bancos rechazan por considerar demasiado riesgoso y, por lo tanto, aportan a que exista efectivamente la inclusión financiera en el país y se trabaje por la erradicación de la pobreza. Por esto, es fundamental que estas instituciones continúen creciendo y puedan atender a la mayor cantidad de clientes posibles, lo que implica, justamente, ser capaces de medir el riesgo de crédito de forma adecuada y precisa.

Si bien la mayoría de las instituciones no bancarias cuenta con modelos de originación, en Chile no se encontró ninguna que utilizara un modelo de *behavioral scoring* para la gestión de su riesgo de crédito y, por lo tanto, decisiones como renovar o no el crédito a un nuevo cliente sólo se basan en apreciaciones de juicio experto. Desde el punto de vista de una institución, desarrollar e implementar un modelo de comportamiento para estimar el riesgo de crédito es altamente costoso, y desde el punto de vista de la regulación, es algo que aún no es requisito para las instituciones no bancarias, situación que podría cambiar a futuro, dada la publicación del documento consultivo promulgado en diciembre de 2015 por el BCBS [6].

Como resultado del estudio se construyó un modelo de comportamiento con 20 variables significativas, todas adecuadamente interpretables con respecto al riesgo de crédito (es decir, todas siguen una tendencia lógica dentro de lo que se espera en la industria del microcrédito) y fáciles de

capturar (es decir, la institución no debe realizar esfuerzos desmedidos para construir estas variables). Así, se encontró que variables como la edad al cierre de mes, créditos anteriores, meses de emprendimiento al cierre de mes, mora promedio al cierre de mes, porcentaje de cuotas pagadas al cierre de mes, nivel de escolaridad del cliente, entre otras, son significativas para discriminar entre un buen y mal pagador: *los clientes más jóvenes son más riesgosos que los adultos mayores; un cliente nuevo es más riesgoso que uno antiguo y a medida que concreta o finaliza más créditos con la institución, menor es el riesgo de incumplimiento esperado; clientes con más experiencia en el rubro son menos riesgosos, por la capacidad de sobrellevar crisis del negocio; clientes con mayor mora promedio al cierre de mes son más riesgosos; clientes que han pagado más de la mitad del crédito son menos riesgosos que quienes aún no lo hacen; la institución capta los profesionales riesgosos que la banca decide no atender, encontrándose que a mayor nivel educacional de los clientes de la empresa, mayor es el riesgo de que no paguen.* Es clave mencionar que los patrones encontrados provienen de los datos observados, dado el comportamiento actual de los clientes. Si la empresa influye directamente buscando modificar el comportamiento, se observarán nuevos patrones y deberá calibrar nuevamente un modelo.

En términos de ajuste predictivo, el modelo escogido tiene un rendimiento de un 84% según el resultado de la curva ROC, lo que indica que, si la empresa clasifica a sus clientes utilizando los puntajes asociados a los segmentos encontrados por variable, tiene un 34% de probabilidad de clasificar correctamente a un cliente por sobre discriminación aleatoria. Así, se llega a un resultado que es tan bueno como los que desarrollan las instituciones que otorgan créditos de tipo *retail*.

Respecto a la simplicidad de la aplicación, se logró escalar los resultados del modelo a un formato *scorecard*, que es más fácil de interpretar por los tomadores de decisiones y que, además, permite transparentar ante el cliente situaciones en que, por ejemplo, la empresa decide no renovar el contrato de crédito, existiendo ahora una forma de visibilizar las decisiones de rechazo.

También, a partir de la *scorecard*, se identificaron segmentos accionables con respecto al riesgo de crédito, a los que la empresa debería sí o sí hacer un seguimiento y mantener en observación, cuidando de que no caigan en incumplimiento: cartera de créditos reprogramados, créditos con 30 días de mora o más y créditos con menos de 4 cuotas pagadas. Para mitigar el riesgo, la empresa puede definir metodologías de cobranza temprana basándose en la segmentación encontrada: por ejemplo, si la mora promedio en la fecha de observación es de 30 días o más, visitar, si está entre 18 y 30 días, enviar una carta, si está entre 0 y 18 días, realizar una llamada para recordar la fecha de pago. Esto no sólo permite definir políticas de cobranza racionales, sino que además permite invertir estos recursos eficientemente.

En términos monetarios, se encontró un resultado que permite a la empresa minimizar los costos promedio por crédito renovado, algo que sin un modelo de este tipo no se puede decidir *a priori*, sino que solo cuando las pérdidas ya se manifiestan. Así, a partir del punto de corte, la empresa puede decidir renovar un crédito sólo a aquellos clientes tal que su *score* de comportamiento sea mayor o igual a 625 puntos. A este nivel, la institución deja fuera al 62,7% del total de malos pagadores a un costo relativamente bajo, de \$56.883 por operación renovada, que aumenta a \$98.353 (un 73%) en caso de cortar en 650 puntos. Otro resultado relevante a partir

del punto de corte es que la empresa puede definir políticas de pre-aprobación para clientes con *score* de comportamiento alto, por ejemplo, sobre los 675 puntos. Es importante recordar que el punto de corte dependerá de la estrategia de negocio de la empresa, la que debe tener presente el *trade off* entre aumentar el punto de corte y disminuir la tasa de malos aceptados, pero también disminuyendo la tasa total de clientes atendidos.

Finalmente, los resultados encontrados demuestran que las exigencias básicas para la creación de modelos internos en microfinanzas, que midan de forma eficiente el riesgo de crédito en etapas de seguimiento y cobranza, sí son suficientes para construir un modelo de comportamiento de pago con alta capacidad discriminante en microempresa, utilizando los parámetros exigidos por SBIF a instituciones bancarias, esto es, definición de incumplimiento a 90 días de mora, con una ventana de observación de 12 meses. Tomando los resultados encontrados, la empresa podría estimar el parámetro PD de forma interna, basándose en una metodología IRB fundacional, que le permitirá tener una primera aproximación de lo que, a futuro, se esperaría sea exigible a instituciones de inclusión financiera, dado el documento consultivo publicado por la BCBS que se refiere al tema.

Por otro lado, los resultados encontrados pueden ser generalizables a otras entidades de inclusión financiera con modelo de negocio similar, basado en créditos individuales y sin captación de ahorros. Esto, porque se escogió una institución representativa de las microfinanzas en Chile, que atiende al segmento exclusivo de microempresa y tiene sucursales en varias regiones del país (lo que le permite atender distintos sectores económicos y variados perfiles sociales). Así, otras instituciones similares podrían tomar esta metodología propuesta, la que es ordenada y clara en sus etapas, para ir desde una base de datos tipo panel a una *scorecard* de comportamiento, partiendo de la base de que en este estudio ya hay evidencia de que ciertas variables podrían resultar discriminante para medir el riesgo de crédito en microempresarios.

Queda propuesto como línea de trabajo futura el cálculo de la probabilidad de incumplimiento, PD, trabajo que debería ser directo a partir de los resultados del *score*. Otros trabajos podrían ir más allá y desarrollar un *behavioral scoring* pero considerando el incumplimiento a 30 días, 60 días o 120 días, y comparar lo obtenido con los resultados de este trabajo, testeando si es que existe una ventana temporal más adecuada.

Capítulo 6

Bibliografía

- [1] REAL ACADEMIA ESPAÑOLA, 2016. [En línea]. <<http://dle.rae.es/?id=WT8tAMI>>. [Consulta: 30 Mayo 2016].
- [2] ANDERSON, Raymond. The Credit Scoring Toolkit, 1ª ed. New York, Oxford University Press Inc., 2007, 731p.
- [3] SIDDIQI, Naem. Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring, New Jersey, SAS Institute Inc., 2006, 210p.
- [4] BIS. About the BIS. [En línea]. <<https://www.bis.org/about/index.htm?m=1%7C1>>. [Consulta: 27 abril 2017].
- [5] BIS. About the Basel Committee, 2017. [En línea]. <<https://www.bis.org/bcbs/about.htm>>. [Consulta: 27 abril 2017].
- [6] BCBS. Guidance on the application of the Core principles for effective banking supervision to the regulation and supervision of institutions relevant to financial inclusion. Basilea, Suiza, Bank for International Settlements, 2015, 55p.
- [7] THE ECONOMIST INTELLIGENCE UNIT, Microscopio global sobre el entorno de negocios para las microfinanzas 2013, Londres, Reino Unido, The Economist Group, 2013, 89p.
- [8] TRUJILLO, Verónica. Microfinanzas en América Latina y El Caribe: El sector en Cifras. New York, Washintong D.C., FOMIN, 2013. 20p.
- [9] SBIF. Compendio de Normas Contables. Sitio web SBIF, 2014. [En línea]. <<http://www.sbif.cl>>. [Consulta: 20 11 2016].
- [10] EMPRENDE MICROFINANZAS. La Microempresa. Sitio web Emprende Microfinanzas, 2015. [En línea]. <<http://emprendemf.cl/la-microempresa.html>>. [Consulta: 30 05 2016].
- [11] YUNUS, Muhammad. Creer, Crear y Crecer. Muhammad Yunus en Chile 2003, Santiago, Chile, Desafío S.A., 2004, 98p.
- [12] MINISTERIO DE ECONOMÍA, FOMENTO Y RECONSTRUCCIÓN. Ley 20.416: *Fija normas especiales para las empresas de menor tamaño*. Biblioteca del Congreso Nacional,

2014. [En línea]. < <https://www.leychile.cl/Navegar?idNorma=1010668>>. [Consulta: 30 05 2016].
- [13] ARELLANO, P. y SCHUSTER, T., Informe de resultados: El microemprendedor en Chile. Cuarta Encuesta de Microemprendimiento 2015, Santiago, Chile, Ministerio de Economía, Fomento y Turismo, 2016, 25p..
- [14] HELD, Günther. Suficiencia de Capital de los Bancos: Estándares del Comité de Basilea. Sitio web de SBIF, 2012 [En línea]. <https://www.sbif.cl/sbifweb3/internet/archivos/publicacion_10076.pdf>. [Consulta: 30 05 2016].
- [15] MICROFINANCE CONSENSUS guidelines. 2002. Por Robert Peck Christen "et al". New York, Washington D.C., 120p.
- [16] LARRAÍN, Christian. ¿Existe un modelo de microfinanzas en América Latina?, 2009, Santiago, Chile, CEPAL, 43p.
- [17] PRINCIPIOS Y PRÁCTICAS PARA LA REGULACIÓN y supervisión de las microfinanzas. 2003. Por Jansson "et al". Banco Interamericano de Desarrollo. New York, Washington D.C., 114p.
- [18] BRAVO, Cristián. Introducción al Riesgo de Crédito en Carteras Masivas, 2014, Santiago, Chile, Departamento de Ingeniería Civil Industrial.
- [19] BRAVO, C., WEBER, R. y MALDONADO, S. Experiencias prácticas: riesgo crediticio de microempresarios mediante Credit Scoring. *Revista Ingeniería de Sistemas* 24, 69-88, 2010.
- [20] BIRON, Miguel. Desarrollo y evaluación de metodologías para la aplicación de regresiones logísticas en modelos de comportamiento bajo supuesto de independencia, 2012, Santiago, Chile, 86p.
- [21] IBM. Manual CRISP-DM de IBM SPSS Modeler. IBM Corporation, 2012. [En línea]. <<ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>>. [Consulta: 30 05 2016].
- [22] GERARD, Nicolás. Data Mining, 2008. Sitio web de Gerardnico. [En línea]. http://gerardnico.com/wiki/data_mining/accuracy. [Consulta: 9 12 2015].
- [23] BERZAL, Fernando. Clasificación y predicción. Departamento de Ciencias de la Computación, Universidad de Granada [En línea]. <<http://elvex.ugr.es/decsai/intelligent/slides/dm/D4%20Classification.pdf>> [Consulta: 9 12 2015].
- [24] MICROSOFT SQL SERVER. Soundex (Transact - SQL), Microsoft, 2016. [En línea]. <<https://msdn.microsoft.com/es-cl/library/ms187384.aspx>>. [Consulta: 12 10 2016].

- [25] MARSHALL, Enrique. Superintendencia de Bancos e Instituciones Financieras, SBIF, 2014. [En línea].<<http://www.sbif.cl/>>. [Consulta: 30 05 2016].
- [26] RED DE MICROFINANZAS, Sitio web de Red de Microfinanzas. [En línea]. <http://www.redmicrofinanzas.cl/web/?page_id=5>. [Consulta: 30 05 2016].
- [27] BARRIGA. 2010. [En línea].<<http://historiabarriga.blogspot.cl/2010/12/neoliberalismo-en-chile-crisis-de-1982.html>>. [Consulta: 30 05 2016].
- [28] COOPERATIVA LIBERACIÓN, 2016. [En línea]. <<http://coopliberacion.cl/historia/>>. [Consulta: 30 05 2015].
- [29] MODELING CREDIT RISK FOR SMWS: Evidence from the US Market. 2007. Por Sabato "et al", 37p.
- [30] MORALES, L. y Yáñez, Á. Microfinanzas en Chile. Resultados de la Encuesta de Colocaciones en Segmentos Microempresariales, Serie Técnica de Estudios N°011, SBIF, 2007, Santiago, Chile, 28p.
- [31] ASBA. Guía de Principios para una Efectiva Regulación y Supervisión de las Operaciones de Microfinanzas. Banco Interamericano de Desarrollo, 2010. [En línea]. <publications.iadb.org>. [Consulta: 29 Junio 2016].
- [32] THOMAS, Lyn. Consumer Credit Models: Pricing, Profit and Portfolios., New York, Oxford University Press Inc., 2009, 400p.

Capítulo 7

Anexos

7.1 Anexo 1: Almacenamiento

Tabla 7.1: Bases de Datos del Sistema SAC, Emprande Microfinanzas

BASE DE DATOS	TABLAS	INFORMACIÓN
1. COBRANZAMF	37	<i>La mayoría de estas tablas ya no se utiliza y corresponde a gestión antigua, que hoy ya no se realiza.</i>
2. CONTABILIDADMF	27	
3. GESTIÓN_HISTÓRICO	17	<i>La mayoría de estas tablas se utilizan y contienen toda la información del ciclo de vida del crédito.</i>
4. GESTIONMF	103	
5. OTRASTABLASMf	103	

Fuente: Elaboración propia a partir de información entregada por Emprande Microfinanzas.

7.2 Anexo 2: Descripción de los datos

Tabla 7.2: Tablas principales de la base de datos de Empresa Microfinanzas y sus claves primarias.

Base de Datos	Tabla	Llave Primaria	Filas	Columnas
GESTIÓN HISTÓRICA	<i>SdoCredCart</i> Estado de cada crédito al cierre de mes.	<i>sscSuc: ID Sucursal</i>	6.874.346	15
		<i>sccCred: ID Crédito</i>		
		<i>sccSc: Subcrédito</i>		
		<i>sccFecha: Fecha al cierre de mes</i>		
	<i>MoraHistórica</i> Información de cada crédito que ha estado en mora al cierre de mes.	<i>mhSuc: ID Sucursal</i>	818.890	16
		<i>mhFecha: Fecha al cierre de mes</i>		
<i>mhCred: ID Crédito</i>				
GESTIÓNMF	<i>Créditos</i> Características adicionales del crédito.	<i>crSuc: ID Sucursal</i>	105.572	34
		<i>crNro: ID Crédito</i>		
	<i>Clientes</i> Características adicionales del crédito.	<i>cliSuc: ID Sucursal</i>	78.850	55
		<i>cliNro: ID Cliente</i>		
	<i>PlanPago</i> Plan de pago de todos los créditos de la institución.	<i>ppSuc: ID Sucursal</i>	1309848	26
		<i>ppCr: ID Crédito</i>		
<i>ppFVcto: Fecha de vencimiento de la cuota</i>				
OTRASTABLASMf	<i>Créditos Transferidos</i> Identificador de datos del crédito anterior, en caso de cierre de sucursal.	<i>ctSucAntigua: ID Sucursal Anterior</i>	4947	7
		<i>ctCredAntiguo: ID Crédito Anterior</i>		
		<i>ctCliAntiguo: ID Cliente Anterior</i>		
		<i>ppFVcto: Fecha de vencimiento de la cuota</i>		

Fuente: Elaboración propia.

Tabla 7.3: Clasificación de las tablas de las Bases de Datos: Contabilidad MF.

BASE DE DATOS: COBRANZA MF						ANÁLISIS DE TABLAS		
NOMBRE TABLA	COLUMNAS	FILAS	ANONIMO	ESTADO	DETALLE	Primer Proceso de Selección	Proceso de Anonimizado	Segundo Proceso de Selección/Anonimizado
Autorizaciones	3	4348	NO	VIGENTE	Indica a qué puede acceder cada usuario en el sistema (atribuciones).	Contiene Usuarios definidos por RUT. Se agregan a tabla Usr. Se junta con la tabla Autorizaciones de GestionMF. Se reemplaza esta tabla por la nueva, que contiene ambas.	Se reemplaza el usuario por su hash respectivo. Antes: autUsr Ahora: autUsrHash Se elimina Autorizaciones y se reemplaza por Autorizaciones_Hash No se pudo restaurar las PK.	
Camara	21	3559053	NO	VIGENTE	Información que se envía a la cámara de comercio, cada 2 semanas, de gente morosa.	Contiene Ruts, se agregan a tabla Ruts. No tiene DV ni punto ni guión.	Se reemplazan los nombres, direcciones y otros por variables binarias o hash. Se restauran las PK.	
GesClientes1	21	312609	NO	VIGENTE	Gestiones que se realizan en Sistema Gestión de Crédito (para un sistema que está en internet).	Contiene Usuarios definidos por RUT. Se agregan a tabla Usr.	Se reemplaza el usuario por su hash respectivo. Antes: gcUsuario Ahora: gcUsuarioHash Se elimina GesClientes1 y se reemplaza por GesClientes1_Hash Se restauran las PK.	Contiene observaciones. Se anonimiza utilizando variables binarias, por la gran cantidad de filas que posee la tabla y la imposibilidad de revisar 1 a 1. Falta eliminar definitivamente la glosa.
StGestion	3	2	SÍ	VIGENTE	Registra si fue revisada la gestión levantada.	Se traspasa a BBDD Códigos		
TipoGestion	3	42	SÍ	VIGENTE	Códigos, para tabla Gesclientes.	Se traspasa a BBDD Códigos		
CE_Estados	3	---	---	VACÍA	Sistema Antiguo de cobranza del cliente.	Se elimina		
CE_Juzgados	3	---	---	VACÍA	Sistema Antiguo de cobranza del cliente.	Se elimina		
CE_Procuradores	3	---	---	VACÍA	Sistema Antiguo de cobranza del cliente.	Se elimina		
CobExt	6	---	---	VACÍA	Vacía. Sin Información.	Se elimina		
DatPagNC	11	---	---	VACÍA	Vacía. Sin Información.	Se elimina		
Protesto	3	---	---	VACÍA	Vacía. Sin Información.	Se elimina		
Rubros	3	---	---	VACÍA	Vacía. Sin Información.	Se elimina		
SoCarVen	23	---	---	VACÍA	Vacía. Sin Información.	Se elimina		
usuarios	9	---	---	VACÍA	Vacía. Sin Información.	Se elimina		
Caja	12	1939	SÍ	NO VIGENTE	No actualizada. Hasta el 2009.	Se elimina		
CE_Demanda	11	474	NO	NO VIGENTE	Sistema Antiguo de cobranza del cliente.	Se elimina		
CE_Gestiones	7	918	NO	NO VIGENTE	Sistema Antiguo de cobranza del cliente. Solo año 2010 y 2011.	Se elimina		
Cheques	16	907	NO	NO VIGENTE	Cheques, se maneja en BASE DE DATOS GESTIONMF. La información está en GestionMF.Cheques	Se elimina		
Cobrador	3	18	NO	NO VIGENTE	No actualizada.	Se elimina		
Comunas	3	42	SÍ	NO VIGENTE	No actualizada. En GestionMF está completa.	Se elimina		
CorrCamara	6	28	NO	NO VIGENTE	No actualizada.	Se elimina		
Correlativos	5	93	SÍ	NO VIGENTE	No actualizada.	Se elimina		
Creditos	6	11159	SÍ	NO VIGENTE	Créditos, se maneja en BASE DE DATOS GESTIONMF.	Se elimina		
Diligencias	4	150	SÍ	NO VIGENTE	No actualizada.	Se elimina		
Funciones	3	53	---	NO VIGENTE	No actualizada.	Se elimina		
GesClientes	21	3454411	NO	NO VIGENTE	No actualizada. Año 2012, 2013 y 2014.	Se elimina		
GesClientes_BACKUP	21	159918	NO	NO VIGENTE	No actualizada.	Se elimina		
GesClientesIQA	21	3454411	NO	NO VIGENTE	No actualizada.	Se elimina		
GesClientesAux	2	3454411	SÍ	NO VIGENTE	No actualizada.	Se elimina		
GesClientesHist	20	426700	NO	NO VIGENTE	No actualizada.	Se elimina		
Gestion	13	18733	NO	NO VIGENTE	No actualizada. Fechas 2007 a 2011.	Se elimina		
Listados	15	12	SÍ	NO VIGENTE	No actualizada.	Se elimina		
NoClientes	18	4	NO	NO VIGENTE	Solo tiene 4 filas.	Se elimina		
PreGestion	9	127956	NO	NO VIGENTE	No todos los clientes tienen registrada esta información de pre-gestión.	Se elimina		
Sancion	26	122	NO	NO VIGENTE	No actualizada. Solo datos de 2009 y 2010.	Se elimina		
TpDiligencias	3	6	SÍ	NO VIGENTE	No actualizada.	Se elimina		
Tribunal	3	3	SÍ	NO VIGENTE	No actualizada.	Se elimina		

Fuente: Elaboración propia.

7.3 Anexo 3: Anonimización

Tabla 7.4: Pasos para anonimizar los datos.

Pasos para anonimizar el RUT
Objetivo: Crear una tabla con todos los RUTS, sin DV ni puntos ni guión, de todas las bases de datos. La tabla servirá para categorizar los RUTS antes de eliminarlos y hashearlos.
Paso 1: Verificar que el RUT no incluye puntos, guión ni DV, si no, limpiar. Seleccionar los diferentes y guardar en tabla RUTS.
Paso 2: Cuando se tengan todos los RUT en la tabla RUTS, crear las variables: PersNatural, PersJurídica, EIRL.
Paso 3: Hashear los RUT utilizando la función hash + salt. Eliminar los RUT que no tienen sentido (0, 00, 0000, etc) para no reemplazarlos.
Paso 4: Reemplazar en cada tabla con RUT su respectivo hash y eliminar RUT.

Pasos para anonimizar la dirección
Objetivo: Verificar si la dirección del cliente es igual que su dirección de trabajo, o similar a la de su aval, etc.
Paso 1: Crear una variable que indique si se registra una dirección.
Paso 2: Si hay más direcciones de aval o referido, o particular y comercial del cliente, comparar utilizando la función Soundex de SQL.
Paso 3: Eliminar las direcciones

Pasos para anonimizar el Mail
Objetivo: Identificar si el cliente entrega un correo electrónico válido.
Paso 1: Verificar si el correo electrónico es válido (si contiene un @). Crear una variable que indique 1 si entrega Email, 0 si no.
Paso 2: Eliminar el correo electrónico.

Pasos para anonimizar el Teléfono
Objetivo: Identificar si el cliente entrega un teléfono válido, y cuántos números de contacto entrega.
Paso 1: Verificar si el número de teléfono es válido. Crear una variable que indique 1 si entrega número válido, 0 si no.
Paso 2: Si entrega más de un número en la misma tabla, sumar la cantidad de números de contactos en una variable.

Fuente: Elaboración propia.

7.4 Anexo 4: Descripción y construcción de variables de tipo *application*.

Tabla 7.5: Identificadores únicos de cada fila de la base de datos.

Nombre	Tipo de recurso	Descripción
<i>ID</i>	Construida	Identificador único para cada crédito distinto (para cada tripleta Sucursal, Número de crédito, Número de cliente).
<i>SucID</i>	EMF	Sucursal antigua (para créditos transferidos) o sucursal actual (para créditos no transferidos).
<i>CredID</i>	EMF	Número de crédito antiguo (para créditos transferidos) o número de crédito actual (para créditos no transferidos)
<i>ClIID</i>	EMF	Número de cliente antiguo (para créditos transferidos) o número de cliente actual (para créditos no transferidos)
<i>NCta</i>	EMF	Número de la cuota en observación.

Fuente: Elaboración propia.

Tabla 7.6: Variables asociadas a características del cliente, en orden alfabético.

Nombre	Tipo de variable	Tipo de recurso	Descripción
<i>CodEscolaridad</i>	Nominal	EMF	Código de nivel de escolaridad del cliente.
<i>CodEstCiv</i>	Nominal	EMF	Código de estado civil del cliente.
<i>CodOtrosIng</i>	Nominal	EMF	Código de otros ingresos que recibe el cliente.
<i>ComunaCasa</i>	Nominal	EMF	Código de comuna en que vive el cliente.
<i>ComunaCasaSuc</i>	Binaria	Construida	$\left\{ \begin{array}{l} 1 \text{ si el cliente vive en la comuna donde se ubica} \\ \text{la sucursal.} \\ 0 \text{ si no.} \end{array} \right.$
<i>ComunaTallerCasaSuc</i>	Binaria	Construida	$\left\{ \begin{array}{l} 1 \text{ si el cliente vive o trabaja en la comuna donde se} \\ \text{se ubica la sucursal.} \\ 0 \text{ si no.} \end{array} \right.$
<i>ComunaTallerSuc</i>	Binaria	Construida	$\left\{ \begin{array}{l} 1 \text{ si el cliente trabaja en la comuna donde se ubica} \\ \text{la sucursal.} \\ 0 \text{ si no} \end{array} \right.$

<i>Nombre</i>	<i>Tipo de variable</i>	<i>Tipo de recurso</i>	<i>Descripción</i>
<i>CasaTaller</i>	Binaria	Construida	Variable que indica si el cliente trabaja desde el hogar en su emprendimiento (ver pág. 54).
<i>EdadCierre</i>	Numérica	Construida	Edad del cliente al cierre de mes. Se calcula como la diferencia en años entre la fecha al cierre de mes y la fecha de nacimiento.
<i>EdadFechaInc</i>	Numérica	Construida	Edad del cliente al momento de su incorporación. Se calcula como la diferencia en años entre la fecha de incorporación y la fecha de nacimiento.
<i>EIRL</i>	Binaria	Construida	Variable que indica si el cliente corresponde a una EIRL, resultado de la anonimización de RUT (ver tabla 19, pág.46).
<i>Hash_Rut</i>	Nominal	Construida	Transformación no invertible del RUT del cliente (ver tabla 19, pág. 46).
<i>Hijos</i>	Numérica	EMF	Número de hijos que declara el cliente.
<i>Mail</i>	Binaria	Construida	Variable resultado de la anonimización de MAIL (ver tabla 19, pág. 46). Indica si el cliente registra un correo electrónico válido.
<i>MtoOtrosIng</i>	Numérica	EMF	Monto de otros ingresos que declara recibir el cliente, en pesos chilenos.
<i>Nacionalidad</i>	Nominal	EMF	Nacionalidad del cliente.
<i>PersHogar</i>	Numérica	EMF	Número de personas que conviven con el cliente, incluyendo al cliente.
<i>PersNatural</i>	Binaria	Construida	Variable que indica si el cliente es una persona natural, resultado de la anonimización de RUT (ver tabla 19, pág.46).
<i>PersJurídica</i>	Binaria	Construida	Variable que indica si el cliente es una persona jurídica, resultado de la anonimización de RUT (ver tabla 19, pág.46).
<i>RegionCasa</i>	Nominal	EMF	Código de la región en que vive el cliente.
<i>Sexo</i>	Nominal	EMF	Sexo del cliente.

<i>Nombre</i>	<i>Tipo de variable</i>	<i>Tipo de recurso</i>	<i>Descripción</i>
<i>TotFonos</i>	Numérica	Construida	Números total de teléfonos válidos entregados por el cliente.

Fuente: Elaboración propia.

Tabla 7.7: Variables asociadas con características del negocio del cliente.

<i>Nombre</i>	<i>Tipo de variable</i>	<i>Tipo de recurso</i>	<i>Descripción</i>
<i>MesesEmp</i>	Numérica	Construida	Meses de experiencia del cliente en su negocio. Se calcula como la diferencia entre la fecha en que inicia su emprendimiento y la fecha de cierre de mes.
<i>TrbFeria</i>	Nominal	Construida	$\begin{cases} 1 & \text{si el cliente declara que trabaja en la feria.} \\ 0 & \text{si no.} \end{cases}$
<i>Agrícola</i>	Binaria	Construida	$\begin{cases} 1 & \text{si el cliente declara declara que trabaja en rubro agrícola.} \\ 0 & \text{si no.} \end{cases}$
<i>VtasProm</i>	Numérica	EMF	Monto total de ventas promedio que declara el cliente, en pesos chilenos.
<i>Trabajadores</i>	Numérica	EMF	Número de trabajadores que operan junto al cliente.
<i>RemProm</i>	Numérica	EMF	Monto promedio de remuneración a trabajadores que operan junto al cliente, en pesos chilenos.
<i>Patrimonio</i>	Numérica	EMF	Patrimonio del negocio del cliente, en pesos chilenos.
<i>Activos</i>	Numérica	EMF	Activos del negocio del cliente, en pesos chilenos.
<i>ClasCAE</i>	Nominal	EMF	Código de clasificación de tipo de actividad económica del negocio, según agrupación mayor utilizada por la institución.
<i>IA</i>	Binaria	EMF	$\begin{cases} 1 & \text{si el cliente tiene iniciación de actividades.} \\ 0 & \text{si no.} \end{cases}$
<i>Autonomía</i>	Numérica	Construida	$\begin{cases} \frac{\text{Patrimonio}}{\text{Activos}}, & \text{si el cliente declara activos} > 0. \\ 0 & \text{si no.} \end{cases}$ Representa el nivel de autonomía del cliente.

<i>Nombre</i>	<i>Tipo de variable</i>	<i>Tipo de recurso</i>	<i>Descripción</i>
<i>Leverage</i>	Numérica	Construida	$\frac{\text{Monto del crédito}}{\text{Ventas Promedio}}$ Representa una medida de cuan endeudado está el cliente.

Fuente: Elaboración propia.

Tabla 7.8: Variables asociadas con características del crédito otorgado al cliente.

<i>Nombre</i>	<i>Tipo de variable</i>	<i>Tipo de recurso</i>	<i>Descripción</i>
<i>Año</i>	Nominal	EMF	Año del vencimiento de la cuota en observación, coincide con el año del cierre de mes.
<i>Mes</i>	Nominal	EMF	Mes del vencimiento de la cuota en observación, coincide con el mes del cierre de mes.
<i>FechaCierre</i>	Date	EMF	Fecha de cierre de mes: año-mes-día.
<i>FechaPago</i>	Date	EMF	Fecha en que se registra el pago correspondiente a la cuota del mes en observación: año-mes-día.
<i>FechaVcto</i>	Date	EMF	Fecha de vencimiento de la cuota del mes en observación: año-mes-día.
<i>FecAprob</i>	Date	EMF	Fecha de aprobación del crédito: año-mes-día.
<i>ClasifRiesgo</i>	Nominal	EMF	Clasificación de pago del cliente según área de riesgo de la empresa.
<i>CredTransferido</i>	Nominal	Construida	$\left\{ \begin{array}{l} 1 \text{ si el crédito fue transferido de una sucursal a otra} \\ \text{en el período de observación o antes.} \\ 0 \text{ si no.} \end{array} \right.$
<i>CtasReal</i>	Numérica	Construida	Cantidad de cuotas pactadas originalmente en el plan de pago. No incluye cuotas que son ingresadas para identificar una repactación.
<i>Garantía</i>	Binaria	Construida	$\left\{ \begin{array}{l} 1 \text{ si el crédito incluye garantías por no pago, dadas} \\ \text{por el cliente.} \\ 0 \text{ si no.} \end{array} \right.$
<i>GesCobranza</i>	Binaria	Construida	$\left\{ \begin{array}{l} 1 \text{ si el cliente recibió alguna gestión de cobranza,} \\ \text{para la cuota en observación.} \\ 0 \text{ si no.} \end{array} \right.$
<i>Lc</i>	Nominal	EMF	Código de tipo de crédito que se otorga al cliente.

<i>Nombre</i>	<i>Tipo de variable</i>	<i>Tipo de recurso</i>	<i>Descripción</i>
<i>Monto</i>	Numérica	EMF	Monto total del crédito aprobado, en pesos chilenos.
<i>MoraAlCierre</i>	Numérica	EMF	Días de mora al cierre de mes en observación.
<i>PrimerMtoPagado</i>	Numérica	EMF	Primer pago registrado para la cuota del mes en observación.
<i>AbonoCuota</i>	Binaria	Construida	$\begin{cases} 1 & \text{si } \text{PrimerMtoPagado} < \text{ValorCta, para la cuota} \\ & \text{en observación.} \\ 0 & \text{si no.} \end{cases}$
<i>ST</i>	Nominal	EMF	Estado del crédito en observación. Ver tabla 13, pág. 37.
<i>StCuota</i>	Nominal	EMF	Estado actual de la cuota del mes en observación.
<i>SucCerrada</i>	Nominal	Construida	Si en el mes en observación la sucursal se encuentra cerrada.
<i>SucCom</i>	Nominal	EMF	Comuna a la que pertenece la sucursal asociada a la cuota del crédito en observación.
<i>SucZona</i>	Nominal	EMF	Zona a la que pertenece la sucursal, según agrupación de la empresa.
<i>TasaInt</i>	Numérica	EMF	Tasa de interés porcentual aplicada al crédito.
<i>TipoCanal</i>	Nominal	EMF	Código de tipo de canal por el que fue captado el cliente.
<i>ValorCta</i>	Numérica	EMF	Monto de la cuota que corresponde ser cancelada en el mes en observación.
<i>ValorCtaPorc</i>	Numérica	Construida	Porcentaje que representa el valor cuota del mes en observación, del monto aprobado.
<i>ValorCtaPorcPactado</i>	Numérica	Construida	Porcentaje que representa el valor cuota del mes en observación, del monto pactado.
<i>ValorCtaProm</i>	Numérica	Construida	Valor cuota promedio del crédito en observación.

Fuente: Elaboración propia.

7.5 Anexo 5: Selección de variables

Tabla 7.9: Resultados de selección de variables con Forward Selection.

```
Call:
glm(formula = Default90 ~ WOE_Activos2R + WOE_BajoMoraProm6 +
    WOE_CtasPagPorc6R + WOE_CtasReal + WOE_CuotasPagadasR + WOE_EdadCierre2 +
    WOE_MinMora6R + WOE_VecesMoraSubio6R + WOE_CreditosAnteriores +
    WOE_MesesEmp2R + WOE_CuotasPagadasPorcR + WOE_MoraProm6R +
    WOE_CtasFaltaPagarR + WOE_MtoPagadoPactadoPorcR + WOE_SobreMoraProm6 +
    WOE_ValorCtaPorcR + WOE_RemPromR + WOE_MtoAcumCtasPagadasR +
    WOE_CredReprogramacion + WOE_CodEscolaridad + WOE_ValorCtaPromCredAntr2 +
    WOE_MtoAprobAntr2 + WOE_Leverage + WOE_CtasConAtrasoPorcCredAntr2Bin +
    WOE_DifCtasCredAntPorcR2Bin, family = "binomial", data = ModeloTrain2,
    weights = ModeloTrain2$Balanceo)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.27383  -0.38785  -0.20819  -0.09933   2.88682

Coefficients:
              Estimate Std. Error z value      Pr(>|z|)
(Intercept)    0.01583    0.02506    0.632    0.527701
WOE_Activos2R -0.76397    0.07092  -10.773 < 0.0000000000000002 ***
WOE_BajoMoraProm6  0.01816    0.04972    0.365    0.714941
WOE_CtasPagPorc6R -0.30192    0.04236   -7.128    0.00000000000102 ***
WOE_CtasReal    -0.34851    0.10040   -3.471    0.000518 ***
WOE_CuotasPagadasR -1.15651    0.07834  -14.763 < 0.0000000000000002 ***
WOE_EdadCierre2  -0.29529    0.07291   -4.050    0.00005123224493 ***
WOE_MinMora6R   -0.18925    0.04551   -4.158    0.00003206571420 ***
WOE_VecesMoraSubio6R -0.32266    0.06849   -4.711    0.00000246156713 ***
WOE_CreditosAnteriores -0.41084    0.10138   -4.052    0.00005069646367 ***
WOE_MesesEmp2R  -0.28301    0.08048   -3.516    0.000437 ***
WOE_CuotasPagadasPorcR  0.03998    0.05821    0.687    0.492263
WOE_MoraProm6R  -0.64741    0.03727  -17.370 < 0.0000000000000002 ***
WOE_CtasFaltaPagarR -0.24324    0.05024   -4.841    0.00000129104258 ***
WOE_MtoPagadoPactadoPorcR -0.49091    0.04787  -10.256 < 0.0000000000000002 ***
WOE_SobreMoraProm6 -0.21056    0.06484   -3.248    0.001164 **
WOE_ValorCtaPorcR -0.57395    0.09100   -6.307    0.00000000028382 ***
WOE_RemPromR    -0.41337    0.06366   -6.493    0.00000000008415 ***
WOE_MtoAcumCtasPagadasR  0.37392    0.08475    4.412    0.00001023083079 ***
WOE_CredReprogramacion -0.94466    0.08598  -10.987 < 0.0000000000000002 ***
WOE_CodEscolaridad -0.39254    0.08107   -4.842    0.00000128411399 ***
WOE_ValorCtaPromCredAntr2 -0.77130    0.24972   -3.089    0.002011 **
WOE_MtoAprobAntr2  0.33441    0.25733    1.300    0.193768
WOE_Leverage    -0.33505    0.08463   -3.959    0.00007528818926 ***
WOE_CtasConAtrasoPorcCredAntr2Bin -0.22225    0.11559   -1.923    0.054509 .
WOE_DifCtasCredAntPorcR2Bin -0.23249    0.10331   -2.250    0.024432 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16533  on 44472  degrees of freedom
Residual deviance: 10186  on 44447  degrees of freedom
AIC: 4998

Number of Fisher Scoring iterations: 5
```

Fuente: Elaboración propia.

Tabla 7.10: Resultados de selección de variables con Backward Elimination.

```

Call:
glm(formula = Default90 ~ WOE_Activos2R + WOE_CtasPagPorc6R +
     WOE_CtasReal + WOE_CuotasPagadasR + WOE_EdadCierre2 + WOE_MinMora6R +
     WOE_VecesMoraSubio6R + WOE_CreditosAnteriores + WOE_MesesEmp2R +
     WOE_MoraProm6R + WOE_CtasFaltaPagarR + WOE_MtoPagadoPactadoPorcR +
     WOE_SobreMoraProm6 + WOE_ValorCtaPorcR + WOE_RemPromR + WOE_MtoAcumCtasPagadasR +
     WOE_CredReprogramacion + WOE_CodEscolaridad + WOE_ValorCtaPromCredAntR2 +
     WOE_Leverage + WOE_CtasConAtrasoPorcCredAntR2Bin + WOE_DifCtasCredAntPorcR2Bin,
     family = "binomial", data = ModeloTrain2, weights = ModeloTrain2$Balanceo)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.27446  -0.38762  -0.20823  -0.09998   2.90310

Coefficients:
                Estimate Std. Error z value      Pr(>|z|)
(Intercept)      0.01534    0.02502   0.613    0.539765
WOE_Activos2R   -0.77310    0.07062 -10.947 < 0.0000000000000002 ***
WOE_CtasPagPorc6R -0.30144    0.04201  -7.175  0.0000000000000723 ***
WOE_CtasReal    -0.35264    0.10021  -3.519  0.000433 ***
WOE_CuotasPagadasR -1.11766    0.06008 -18.602 < 0.0000000000000002 ***
WOE_EdadCierre2 -0.29147    0.07283  -4.002  0.000062756014251 ***
WOE_MinMora6R  -0.18714    0.04520  -4.140  0.000034726484612 ***
WOE_VecesMoraSubio6R -0.31582    0.06411  -4.926  0.000000837993356 ***
WOE_CreditosAnteriores -0.38372    0.09918  -3.869  0.000109 ***
WOE_MesesEmp2R -0.28670    0.08044  -3.564  0.000365 ***
WOE_MoraProm6R -0.64595    0.03726 -17.337 < 0.0000000000000002 ***
WOE_CtasFaltaPagarR -0.22313    0.03987  -5.597  0.000000021813239 ***
WOE_MtoPagadoPactadoPorcR -0.47841    0.04694 -10.192 < 0.0000000000000002 ***
WOE_SobreMoraProm6 -0.20159    0.05986  -3.368  0.000758 ***
WOE_ValorCtaPorcR -0.56525    0.08882  -6.364  0.000000000196947 ***
WOE_RemPromR   -0.40992    0.06358  -6.448  0.000000000113557 ***
WOE_MtoAcumCtasPagadasR 0.36523    0.08429   4.333  0.000014722763952 ***
WOE_CredReprogramacion -0.93909    0.08583 -10.942 < 0.0000000000000002 ***
WOE_CodEscolaridad -0.38581    0.08095  -4.766  0.000001876821710 ***
WOE_ValorCtaPromCredAntR2 -0.48547    0.11534  -4.209  0.000025645687369 ***
WOE_Leverage   -0.33522    0.08460  -3.963  0.000074145462167 ***
WOE_CtasConAtrasoPorcCredAntR2Bin -0.22238    0.11553  -1.925  0.054234 .
WOE_DifCtasCredAntPorcR2Bin -0.22851    0.10324  -2.213  0.026868 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 16533 on 44472 degrees of freedom
Residual deviance: 10189 on 44450 degrees of freedom
AIC: 4992.8

Number of Fisher Scoring iterations: 5

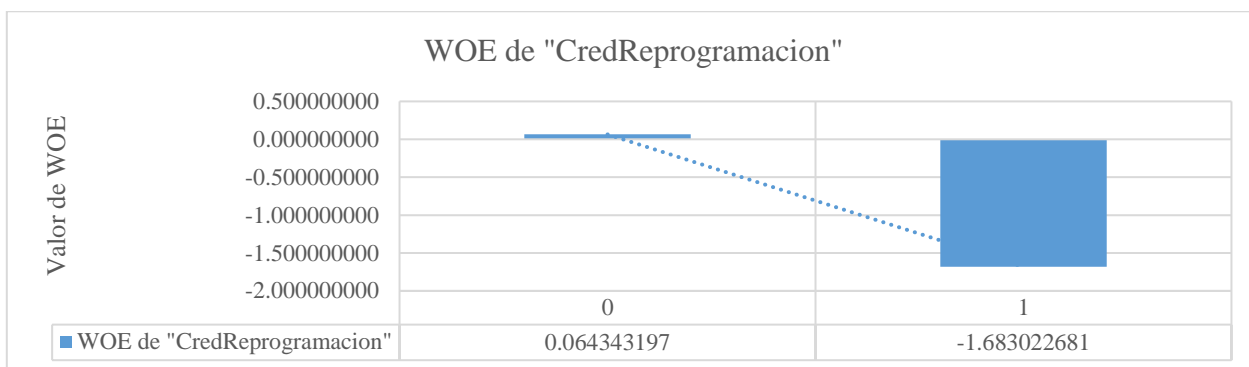
```

Fuente: Elaboración propia.

7.6 Anexo 6: WOE de variables nominales y binarias.

Tabla 7.11: WOE de variable CredReprogramación y gráfico.

CredReprogramacion	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CredReprograma
0	37970	5513	43483	99%	92%	98%	6%	0,064343197
1	540	450	990	1%	8%	2%	-6%	-1,683022681
Total	38510	5963	44473	100%	100%	100%		
IV	0,107363468							



Fuente: Elaboración propia.

Tabla 7.12: WOE de variable SucZona y prueba de otras configuraciones.

SucZona	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "SucZona"
150	15634	2664	18298	41%	45%	41%	-4%	-0,095724961
151	6666	1004	7670	17%	17%	17%	0%	0,027683721
152	8269	1141	9410	21%	19%	21%	2%	0,115264274
153	7915	1091	9006	21%	18%	20%	2%	0,116320749
NA	26	63	89	0%	1%	0%	-1%	-2,750382426
Total	38510	5963	44473	100%	100%	100%		
IV	0,036555839							

Se reemplazan los NA por 150

SucZona	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "SucZona"
150	15660	2727	18387	41%	46%	41%	-5%	-0,117436651
151	6666	1004	7670	17%	17%	17%	0%	0,027683721
152 + 153	16184	2232	18416	42%	37%	41%	5%	0,115780818
Total	38510	5963	44473	100%	100%	100%		
IV	0,011401345							

Se prueban distintas configuraciones. Ninguna consigue un IV sobre 0.

Fuente: Elaboración propia.

Tabla 7.13: WOE de variable CredRefinanciamiento.

CredRefinanciamiento	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CredRefinanciamiento"
0	37377	5904	43281	97%	99%	97%	-2%	-0,019918783
1	1133	59	1192	3%	1%	3%	2%	1,089742579
Total	38510	5963	44473	100%	100%	100%		
IV	0,021667893							

Se encuentra que la variable no aporta información como tal.

Fuente: Elaboración propia.

Tabla 7.14: WOE de variable ClienteNuevo2.

ClienteNuevo2	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ClienteNuevo2"
0	21992	2476	24468	57%	42%	55%	16%	0,318690158
1	16518	3487	20005	43%	58%	45%	-16%	-0,30993531
Total	38510	5963	44473	100%	100%	100%		
IV	0,097968271							

Se encuentra que la variable no aporta información como tal.

Fuente: Elaboración propia.

Tabla 7.15: WOE de variable CredAntRefinanciamiento.

CredAntRefinanciamiento	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CredAntRefinanciamiento"
0	38010	5881	43891	99%	99%	99%	0%	0,000778230
1	500	82	582	1%	1%	1%	0%	-0,057455387
Total	38510	5963	44473	100%	100%	100%		
IV	0,00004							

Se encuentra que la variable no aporta información como tal. Son muy pocos casos en que el anterior es un refinanciamiento.

Fuente: Elaboración propia.

Tabla 7.16: WOE de variable CredAntReprogramación.

CredAntReprogramacion	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CredAntReprogramacion"
0	38331	5930	44261	100%	99%	100%	0%	0,000890518
1	179	33	212	0%	1%	0%	0%	-0,174465993
Total	38510	5963	44473	100%	100%	100%		
IV	0,00016							

Se encuentra que la variable no aporta información como tal. Son muy pocos casos en que el anterior es una reprogramación.

Fuente: Elaboración propia.

Tabla 7.17: WOE de variable CredAntPrepagado2.

CredAntPrepagado2	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CredAntPrepagado2"
0	27034	4786	31820	70%	80%	72%	-10%	-0,133943894
1	11476	1177	12653	30%	20%	28%	10%	0,411944832
Total	38510	5963	44473	100%	100%	100%		
IV	0,05493							

Se encuentra que la variable no aporta información como tal. Son muy pocos casos en que el anterior es un prepagado.

Fuente: Elaboración propia.

Tabla 7.18: WOE de variable ComunaCasaSuc.

ComunaCasaSuc	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ComunaCasaSuc"
0	16935	3070	20005	44%	51%	45%	-8%	-0,157639313
1	21575	2893	24468	56%	49%	55%	8%	0,143896973
Total	38510	5963	44473	100%	100%	100%		
IV	0,02264							

Se encuentra que la variable no aporta información como tal.

Fuente: Elaboración propia.

Tabla 7.19: WOE de variable ComunaTallerSuc.

ComunaTallerSuc	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ComunaTallerSuc"
0	25998	3554	29552	68%	60%	66%	8%	0,124601645
1	12512	2409	14921	32%	40%	34%	-8%	-0,217867778
Total	38510	5963	44473	100%	100%	100%		
IV	0,02709							

Se encuentra que la variable no aporta información como tal.

Fuente: Elaboración propia.

Tabla 7.20: WOE de variable ComunaTallerCasaSuc.

ComunaTallerCasaSuc	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ComunaTallerCasaSuc"
0	16055	2892	18947	42%	48%	43%	-7%	-0,151272214
1	22455	3071	25526	58%	52%	57%	7%	0,124165827
Total	38510	5963	44473	100%	100%	100%		
IV	0,01875							

Se encuentra que la variable no aporta información como tal.

Fuente: Elaboración propia.

Tabla 7.21: WOE de variable Agrícola.

Agrícola	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Agrícola"
0	36012	5782	41794	94%	97%	94%	-3%	-0,036241665
1	2498	181	2679	6%	3%	6%	3%	0,759404422
Total	38510	5963	44473	100%	100%	100%		
IV	0,02746							

Se encuentra que la variable no aporta información como tal.

Fuente: Elaboración propia.

Tabla 7.22: WOE de variable FOSIS.

Fosis	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Fosis"
0	18893	3249	22142	49%	54%	50%	-5%	-0,104900011
1	19617	2714	22331	51%	46%	50%	5%	0,112628738
Total	38510	5963	44473	100%	100%	100%		
IV	0,01180							

Se encuentra que la variable no aporta información como tal.

Fuente: Elaboración propia.

Tabla 7.23: WOE de variable GesCobranzaAnt.

GesCobranzaAnt	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "GesCobranzaAnt"
0	36334	5437	41771	94%	91%	94%	3%	0,034182266
1	2176	526	2702	6%	9%	6%	-3%	-0,445401842
Total	38510	5963	44473	100%	100%	100%		
IV	0,01521							

Se encuentra que la variable no aporta información como tal.

Fuente: Elaboración propia.

Tabla 7.24: WOE de variable CORFO.

CORFO	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Corfo"
0	37702	5773	43475	98%	97%	98%	1%	0,011177030
1	808	190	998	2%	3%	2%	-1%	-0,417806251
Total	38510	5963	44473	100%	100%	100%		
IV	0,00467							

Se encuentra que la variable no aporta información como tal.

Fuente: Elaboración propia.

Tabla 7.25: WOE de variable TotFonos2.

TotFonos2	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "TotFonos2"
0	21819	2869	24688	57%	48%	56%	9%	0,163473375
1	16691	3094	19785	43%	52%	44%	-9%	-0,179939338
Total	38510	5963	44473	100%	100%	100%		
IV	0,02934							

Se encuentra que la variable no aporta información como tal.

Fuente: Elaboración propia.

Tabla 7.26: WOE de variable Trabajadores2.

Trabajadores2	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Trabajadores"
0	31292	5120	36412	81%	86%	82%	-5%	-0,055136203
1	7218	843	8061	19%	14%	18%	5%	0,28202199
Total	38510	5963	44473	100%	100%	100%		
IV	0,01553							

Se encuentra que la variable no aporta información como tal.

Fuente: Elaboración propia.

Tabla 7.27: WOE de variable CredTransferido.

CredTransferido	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CredTransferido"
0	38004	5847	43851	99%	98%	99%	1%	0,006418470
1	506	116	622	1%	2%	1%	-1%	-0,392397760
Total	38510	5963	44473	100%	100%	100%		
IV	0,00252							

Se encuentra que la variable no aporta información.

Fuente: Elaboración propia.

Tabla 7.28: WOE de variable Sexo y prueba de otras configuraciones.

Sexo	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Sexo"
0	1	0	1	0%	0%	0%	0%	#iDIV/0!
F	23241	3645	26886	60%	61%	60%	-1%	-0,012782643
M	15268	2318	17586	40%	39%	40%	1%	0,019710152
Total	38510	5963	44473	100%	100%	100%		
IV	#iDIV/0!							

El cliente que tiene sexo 0 se reemplaza por F.

Sexo	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Sexo"
F	23242	3645	26887	60%	61%	60%	-1%	-0,012739616
M	15268	2318	17586	40%	39%	40%	1%	0,019710152
Total	38510	5963	44473	100%	100%	100%		
IV	0,00025							

Se encuentra que la variable no aporta información.

Fuente: Elaboración propia.

Tabla 7.29: WOE de variable Nacionalidad y prueba de otras configuraciones.

Nacionalidad	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Nacionalidad"
0	0	1	1	0%	0%	0%	0%	#iNUM!
Vacío	20	1	21	0%	0%	0%	0%	1,130388036
ZZZZZZZZZZZZZZZZ	4086	1343	5429	11%	23%	12%	-12%	-0,752683659
argentino	4	3	7	0%	0%	0%	0%	-1,577662165
brasileño	2	0	2	0%	0%	0%	0%	#iDIV/0!
chileno	34293	4606	38899	89%	77%	87%	12%	0,142237213
colombiano	18	1	19	0%	0%	0%	0%	1,025027520
dominicano	9	0	9	0%	0%	0%	0%	#iDIV/0!
ecuatoriano	0	0	0	0%	0%	0%	0%	#iDIV/0!
mexicano	4	0	4	0%	0%	0%	0%	#iDIV/0!
paraguayo	5	1	6	0%	0%	0%	0%	-0,255906325
peruano	67	7	74	0%	0%	0%	0%	0,393438233
uruguayo	0	0	0	0%	0%	0%	0%	#iDIV/0!
venezolano	2	1	3	0%	0%	0%	0%	-1,172197057
Total	38510	5964	44474	100%	100%	100%		
IV	#iNUM!							

Vacío y 0 se reemplazan por chileno. Se prueba juntando el segmento Z con chileno. Todo el resto se junta en extranjeros.

Nacionalidad	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Nacionalidad"
extranjero	111	13	124	0%	0%	0%	0%	0,279236606
chileno	38399	5951	44350	100%	100%	100%	0%	-0,000872093
Total	38510	5964	44474	100%	100%	100%		
IV	0,00020							

La variable queda muy concentrada en chilenos. Se descarta.

Fuente: Elaboración propia.

Tabla 7.30 WOE: variable de PersHogar.

PersHogar	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "PersHogar"
NA	12424	3274	15697	32%	55%	35%	-23%	-0,531726631
0	12549	1232	13781	33%	21%	31%	12%	0,455657878
1	1163	106	1269	3%	2%	3%	1%	0,529974821
2	3972	325	4297	10%	5%	10%	5%	0,637855605
3	4440	470	4910	12%	8%	11%	4%	0,380332723
4	2837	381	3218	7%	6%	7%	1%	0,142358822
5	886	136	1022	2%	2%	2%	0%	0,008717827
6	178	30	208	0%	1%	0%	0%	-0,084758069
7	42	9	51	0%	0%	0%	0%	-0,324899197
8	7	1	8	0%	0%	0%	0%	0,080565911
9	3	0	3	0%	0%	0%	0%	#iDIV/0!
10	8	0	8	0%	0%	0%	0%	#iDIV/0!
13	1	0	1	0%	0%	0%	0%	#iDIV/0!
Total	38510	5964	44473	100%	100%	100%		
IV	#iDIV/0!							

Se observa que los NA son demasiados para agruparlos con otra variable. Se descarta.

Fuente: Elaboración propia.

Tabla 7.31: WOE de variable CodEscarlaridad e intentos de otras configuraciones.

CodEscarlaridad	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodEscarlaridad"
NA	1484	778	2261	4%	13%	5%	-9%	-1,2196
0 (sin estudios)	777	138	915	2%	2%	2%	0%	-0,1372
1 (básica inc)	11416	1373	12789	30%	23%	29%	7%	0,2527
2 (básica com)	4243	472	4715	11%	8%	11%	3%	0,3307
3 (media inc)	5788	780	6568	15%	13%	15%	2%	0,1389
4 (media com)	11914	1870	13784	31%	31%	31%	0%	-0,0136
5 (tec. Prof. Inc)	647	137	784	2%	2%	2%	-1%	-0,3130
6 (tec. Prof. Com)	1251	222	1473	3%	4%	3%	0%	-0,1363
7 (Superior com)	562	92	654	1%	2%	1%	0%	-0,0556
8 (Superior inc)	428	102	530	1%	2%	1%	-1%	-0,4312
Total	38510	5964	44473	100%	100%	100%		
IV	0,14749							

CodEscarlaridad	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodEscarlaridad"
NA	1484	778	2261	4%	13%	5%	-9%	-1,2196
0 - 1 - 2	16436	1983	18419	43%	33%	41%	9%	0,2495
3 (media inc)	5788	780	6568	15%	13%	15%	2%	0,1389
4 (media com)	11914	1870	13784	31%	31%	31%	0%	-0,0136
5 - 6 - 7 - 8	2888	553	3441	7%	9%	8%	-2%	-0,2124
Total	38510	5964	44473	100%	100%	100%		
IV	0,14217							

Se prueba juntar los NA con algún otro segmento. Se decide juntar con los profesionales completos o incompletos, ya que de lo contrario cambia el patrón.

CodEscarlaridad	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodEscarlaridad"
0 - 1	12193	1511	13704	32%	25%	31%	6%	0,222746096
2 (básica com)	4243	472	4715	11%	8%	11%	3%	0,330702622
3 (media inc)	5788	780	6568	15%	13%	15%	2%	0,138903930
4 (media com)	11914	1870	13784	31%	31%	31%	0%	-0,013568489
5 - 6 - 7 - 8 + NA	4372	1331	5702	11%	22%	13%	-11%	-0,676054207
Total	38510	5964	44473	100%	100%	100%		
IV	0,10126							

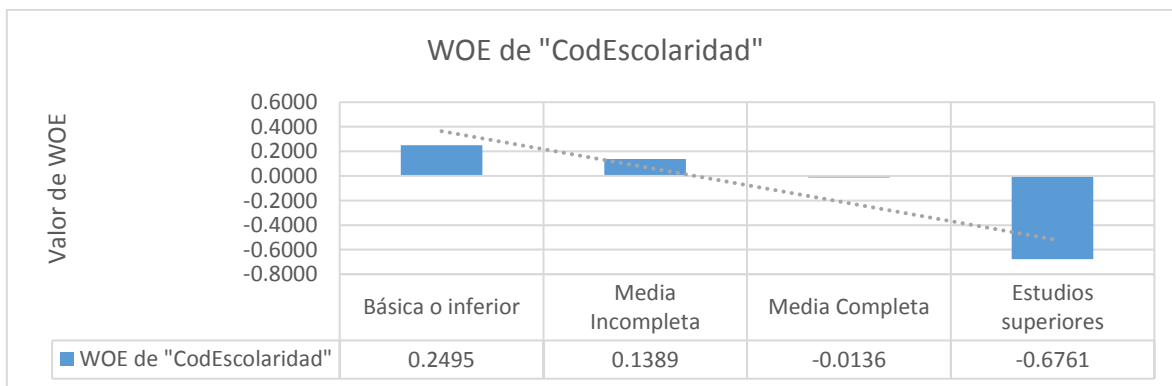
Se prueba a juntar 2 y 3 para ver si mejora el IV. No mejora:

CodEscarlaridad	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodEscarlaridad"
0 - 1	12193	1511	13704	32%	25%	31%	6%	0,222746096
2 - 3	10031	1252	11283	26%	21%	25%	5%	0,215593788
4 (media com)	11914	1870	13784	31%	31%	31%	0%	-0,013568489
5 - 6 - 7 - 8 + NA	4372	1331	5702	11%	22%	13%	-11%	-0,676054207
Total	38510	5964	44473	100%	100%	100%		
IV	0,09918							

Se prueba a juntar 2 y 3 para ver si mejora el IV, pero empeora y la variable deja de aportar información.

CodEscarlaridad	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodEscarlaridad"
Básica o inferior	16436	1983	18419	43%	33%	41%	9%	0,2495
Media Incompleta	5788	780	6568	15%	13%	15%	2%	0,1389
Media Completa	11914	1870	13784	31%	31%	31%	0%	-0,0136
Estudios superiores	4372	1331	5702	11%	22%	13%	-11%	-0,6761
Total	38510	5964	44473	100%	100%	100%		
IV	0,10043							

Se prueba a juntar 0 ,1 y 2 para ver si mejora el IV. Mejora y la variable cobra más sentido, ya que indica que a medida que los clientes de emprende tienen mayor nivel educacional son peores pagadores.



Fuente: Elaboración propia.

Tabla 7.32: WOE de variable CodOtrosIng e intento de otras configuraciones.

CodOtrosing	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodOtrosing"
NA	1484	778	2261	4%	13%	5%	-9%	-1,219574338
0 (ninguno)	20232	2999	23231	53%	50%	52%	2%	0,043642372
1 (empleado)	2879	359	3238	7%	6%	7%	1%	0,216531664
2 (pensión o jubilación)	4640	428	5068	12%	7%	11%	5%	0,518002212
3 (subsídios)	1823	306	2129	5%	5%	5%	0%	-0,080690565
4 (arriendos)	520	72	592	1%	1%	1%	0%	0,111818455
5 (pensión alimenticia)	705	136	841	2%	2%	2%	0%	-0,219801321
6 (otros)	3764	482	4246	10%	8%	10%	2%	0,189949149
7	2412	397	2809	6%	7%	6%	0%	-0,061068961
8	36	4	40	0%	0%	0%	0%	0,331880340
9	15	1	16	0%	0%	0%	0%	0,842705963
10	0	1	1	0%	0%	0%	0%	#¡NUM!
14	0	1	1	0%	0%	0%	0%	#¡NUM!
Total	38510	5964	44473	100%	100%	100%		
IV	#¡NUM!							

CodOtrosing	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodOtrosing"
NA	1484	778	2261	4%	13%	5%	-9,2%	-1,219574338
0 (ninguno)	20232	2999	23231	53%	50%	52%	2,2%	0,043642372
1 (empleado)	2879	359	3238	7%	6%	7%	1,5%	0,216531664
2 (pensión o jubilación)	4640	428	5068	12%	7%	11%	4,9%	0,518002212
3 (subsídios)	1823	306	2129	5%	5%	5%	-0,4%	-0,080690565
4 (arriendos)	520	72	592	1%	1%	1%	0,1%	0,111818455
5 (pensión alimenticia)	705	136	841	2%	2%	2%	-0,5%	-0,219801321
6 (otros)	6227	886	7113	16%	15%	16%	1,3%	0,084588767
Total	38510	5964	44473	100%	100%	100%		
IV	0,14407							

Se agrupan las que tienen menos de un 5%:

CodOtrosing	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodOtrosing"
NA	1484	778	2261	4%	13%	5%	-9%	-1,219574338
0 (ninguno)	20232	2999	23231	53%	50%	52%	2%	0,043642372
1 (empleado)	2879	359	3238	7%	6%	7%	1%	0,216531664
2 (pensión o jubilación)	4640	428	5068	12%	7%	11%	5%	0,518002212
3 - 4 - 5	3048	514	3562	8%	9%	8%	-1%	-0,085326586
6 (otros)	6227	886	7113	16%	15%	16%	1%	0,084588767
Total	38510	5964	44473	100%	100%	100%		
IV	0,14320							

CodOtrosing	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodOtrosing"
NA	1484	778	2261	4%	13%	5%	-9%	-1,219574338
0 (ninguno)	20232	2999	23231	53%	50%	52%	2%	0,043642372
2 (pensión o jubilación)	4640	428	5068	12%	7%	11%	5%	0,518002212
1 - 3 - 4 - 5	5927	873	6800	15%	15%	15%	1%	0,049993668
6 (otros)	6227	886	7113	16%	15%	16%	1%	0,084588767
Total	38510	5964	44473	100%	100%	100%		
IV	0,13982							

CodOtrosing	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodOtrosing"
NA	1484	778	2261	4%	13%	5%	-9%	-1,219574338
0 (ninguno)	20232	2999	23231	53%	50%	52%	2%	0,043642372
1 (empleado)	2879	359	3238	7%	6%	7%	1%	0,216531664
2 (pensión o jubilación)	4640	428	5068	12%	7%	11%	5%	0,518002212
3 - 4 - 5 - otros	9275	1400	10675	24%	23%	24%	1%	0,025506134
Total	38510	5964	44473	100%	100%	100%		
IV	0,14164							

Se juntan los NA con otras categorías pero el IV queda bajo cero. Se descarta la variable.

Fuente: Elaboración propia.

Tabla 7.33: WOE de variable CodEstCiv e intento de otras configuraciones.

CodEstCiv	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodEstCiv"
NA	715	516	1230	2%	9%	3%	-7%	-1,539168460
0 (soltero)	15962	2617	18579	41%	44%	42%	-2%	-0,057161964
1 (casado)	16840	2147	18987	44%	36%	43%	8%	0,194341252
2 (viudo)	1876	223	2099	5%	4%	5%	1%	0,264381120
3 (separado)	2516	386	2902	7%	6%	7%	0%	0,009244011
4 (divorciado)	594	75	669	2%	1%	2%	0%	0,204046968
5 (conviviente)	7	0	7	0%	0%	0%	0%	#iDIV/0!
Total	38510	5964	44473	100%	100%	100%		
IV	#iDIV/0!							

CodEstCiv	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodEstCiv"
NA	715	516	1230	2%	9%	3%	-7%	-1,539168460
0 - 3 - 4	19072	3078	22150	50%	52%	50%	-2%	-0,041402982
1 - 2 - 5	18723	2370	21093	49%	40%	47%	9%	0,201518522
Total	38510	5964	44473	100%	100%	100%		
IV	0,12336							

CodEstCiv	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CodEstCiv"
0 - 3 - 4 + NA	19787	3594	23381	51%	60%	53%	-9%	-0,159584870
1 - 2 - 5	18723	2370	21093	49%	40%	47%	9%	0,201518522
Total	38510	5964	44474	100%	100%	100%		
IV	0,03207							

Se juntan los NA con otras categorías pero el IV queda bajo cero. Se descarta la variable.

Fuente: Elaboración propia.

Tabla 7.34: WOE de variable Mail.

Mail	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Mail"
0	37471	5734	43205	97%	96%	97%	1%	0,011809695
1	1039	229	1268	3%	4%	3%	-1%	-0,353052250
Total	38510	5963	44473	100%	100%	100%		
IV	0,00417							

La variable no aporta información.

Fuente: Elaboración propia.

Tabla 7.35: WOE de variable CasaTaller.

CasaTaller	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CasaTaller"
0	14552	2654	17206	38%	45%	39%	-7%	-0,163683732
1	23958	3309	27267	62%	55%	61%	7%	0,114312031
Total	38510	5963	44473	100%	100%	100%		
IV	0,01868							

Se descarta la variable porque no agrega información.

Fuente: Elaboración propia.

Tabla 7.36: WOE de variable RegionCasa y otros intentos de agrupación.

RegionCasa	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "RegionCasa"
4 (Coquimbo)	2	0	2	0%	0%	0%	0%	#iDIV/0!
5 (Valparaíso)	1629	350	1979	4%	6%	4%	-2%	-0,327555784
6 (O'Higgins)	4264	614	4878	11%	10%	11%	1%	0,072623800
7 (Maule)	2428	416	2844	6%	7%	6%	-1%	-0,101206346
8 (Bio-Bio)	8285	1146	9431	22%	19%	21%	2%	0,112824795
9 (La Araucanía)	7898	1088	8986	21%	18%	20%	2%	0,116924177
12 (Magallanes)	2	0	2	0%	0%	0%	0%	#iDIV/0!
13 (RM)	14000	2349	16349	36%	39%	37%	-3%	-0,080276614
14 (Los Ríos)	2	0	2	0%	0%	0%	0%	#iDIV/0!
Total	38510	5963	44473	100%	100%	100%		
IV	#iDIV/0!							
Se agrupan las que tienen menos de 5%:								
RegionCasa	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "RegionCasa"
4 - 5 - 12 - 14	1635	350	1985	4%	6%	4%	-2%	-0,323879309
6 (O'Higgins)	4264	614	4878	11%	10%	11%	1%	0,072623800
7 (Maule)	2428	416	2844	6%	7%	6%	-1%	-0,101206346
8 (Bio-Bio)	8285	1146	9431	22%	19%	21%	2%	0,112824795
9 (La Araucanía)	7898	1088	8986	21%	18%	20%	2%	0,116924177
13 (RM)	14000	2349	16349	36%	39%	37%	-3%	-0,080276614
Total	38510	5963	44473	100%	100%	100%		
IV	0,01418							
RegionCasa	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "RegionCasa"
4 - 5 - 12 - 14 - 7	4063	766	4829	11%	13%	11%	-2%	-0,196849511
6 (O'Higgins)	4264	614	4878	11%	10%	11%	1%	0,072623800
8 - 9	16183	2234	18417	42%	37%	41%	5%	0,114823370
13 (RM)	14000	2349	16349	36%	39%	37%	-3%	-0,080276614
Total	38510	5963	44473	100%	100%	100%		
IV	0,01276							
RegionCasa	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "RegionCasa"
4 - 5 - 12 - 14 - 7 - RM	18063	3115	21178	47%	52%	48%	-5%	-0,107707743
8 - 9 + 6	20447	2848	23295	53%	48%	52%	5%	0,105874941
Total	38510	5963	44473	100%	100%	100%		
IV	0,01139							
Se prueban distintas configuraciones y ninguna mejora el IV. Se descarta.								

Fuente: Elaboración propia.

Tabla 7.37: WOE de variable TrbFeria.

TrbFeria	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "TrbFeria"
0	36389	5598	41987	94%	94%	94%	1%	0,006512902
1	2121	365	2486	6%	6%	6%	-1%	-0,105598637
Total	38510	5963	44473	100%	100%	100%		
IV	0,00069							
Se descarta la variable por no agregar la información.								

Fuente: Elaboración propia.

Tabla 7.38: WOE de variable Garantía.

Garantía	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Garantia"
0	29215	3129	32344	76%	52%	73%	23%	0,368624574
1	9295	2834	12129	24%	48%	27%	-23%	-0,677556758
Total	38510	5963	44473	100%	100%	100%		
IV	0,24470							
Se ve que si el cliente deja algo en garantía tiene mal comportamiento. Esto puede ser porque el cliente ya se identificaba como malo y se le pida dejar algo en garantía. No se sabe si usar.								

Fuente: Elaboración propia.

Tabla 7.39: WOE de variable TipoCanal.

TipoCanal	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "TipoCanal"
NA	19073	2847	21920	50%	48%	49%	2%	0,036663676
1 (referidos)	1259	189	1448	3%	3%	3%	0%	0,030981781
2 (campana 1+ 1)	5587	938	6525	15%	16%	15%	-1%	-0,080896437
3 (canal alternativo)	1638	269	1907	4%	5%	4%	0%	-0,058824353
4 (asesor comercial)	554	76	630	1%	1%	1%	0%	0,121087109
5 (clientes espontáneos)	1461	278	1739	4%	5%	4%	-1%	-0,206088940
6 (otros)	815	135	950	2%	2%	2%	0%	-0,067430903
7 (volantes)	2127	397	2524	6%	7%	6%	-1%	-0,186812703
8 (página web)	74	19	93	0%	0%	0%	0%	-0,505718124
9 (radio)	751	97	848	2%	2%	2%	0%	0,181350436
10 (municipalidad)	273	56	329	1%	1%	1%	0%	-0,281224133
11 (convenio)	43	6	49	0%	0%	0%	0%	0,104096409
12 (ejecutivo de crédito)	4485	568	5053	12%	10%	11%	2%	0,201028118
13 (campana gift card)	370	88	458	1%	1%	1%	-1%	-0,429178047
Total	38510	5963	44473	100%	100%	100%		
IV	0,01439							

Se descarta la variable porque el 50% corresponde a NA.

Fuente: Elaboración propia.

Tabla 7.40: WOE de variable ClasCAE y prueba de otras configuraciones..

ClasCAE	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ClasCAE"
NA	559	154	713	1%	3%	2%	-1%	-0,576147367
1 (agropecuario)	5338	468	5806	14%	8%	13%	6%	0,568793796
2 (alimentos)	2068	304	2372	5%	5%	5%	0%	0,051965297
3 (comercio en general)	19438	3188	22626	50%	53%	51%	-3%	-0,057508086
4 (confecciones y textil)	2492	415	2907	6%	7%	7%	0%	-0,072781878
5 (construcción)	886	149	1035	2%	2%	2%	0%	-0,082573593
6 (cuero y calzado)	164	22	186	0%	0%	0%	0%	0,143479737
7 (impresión y editoriales)	124	19	143	0%	0%	0%	0%	0,010498349
8 (madera y muebles)	675	94	769	2%	2%	2%	0%	0,106073671
9 (maquinaria y equipo)	209	33	242	1%	1%	1%	0%	-0,019517547
10 (otras manufacturas)	176	26	202	0%	0%	0%	0%	0,047043219
11 (plásticos, químicos y cerámicas)	399	75	474	1%	1%	1%	0%	-0,193870934
12 (productos metálicos)	269	43	312	1%	1%	1%	0%	-0,031832974
13 (restaurantes y hoteles)	1161	187	1348	3%	3%	3%	0%	-0,039415873
14 (servicios para las empresas)	846	109	955	2%	2%	2%	0%	0,183827240
15 (servicios para las personas)	3368	524	3892	9%	9%	9%	0%	-0,004761547
16 (otros)	338	153	491	1%	3%	1%	-2%	-1,072736264
Total	38510	5963	44473	100%	100%	100%		
IV	0,06265							

Se agrupan las que tienen menos de 5%:

ClasCAE	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ClasCAE"
1 (agropecuario)	5338	468	5806	14%	8%	13%	6%	0,568793796
2 (alimentos)	2068	304	2372	5%	5%	5%	0%	0,051965297
3 - 16 + NA	20335	3495	23830	53%	59%	54%	-6%	-0,104334074
4 - 6 - 8 - 10	3507	557	4064	9%	9%	9%	0%	-0,025393227
5 - 9 - 11 - 12	1763	300	2063	5%	5%	5%	0%	-0,094354530
7 - 13 - 14	2131	315	2446	6%	5%	5%	0%	0,046429756
15 (servicios para las personas)	3368	524	3892	9%	9%	9%	0%	-0,004761547
Total	38510	5963	44473	100%	100%	100%		
IV	0,04101							

ClasCAE	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ClasCAE"
1 (agropecuario)	5338	468	5806	14%	8%	13%	6%	0,568793796
3 - 16 + NA	20335	3495	23830	53%	59%	54%	-6%	-0,104334074
2 - 4 - 6 - 8 - 10	5575	861	6436	14%	14%	14%	0%	0,002608854
5 - 9 - 11 - 12	1763	300	2063	5%	5%	5%	0%	-0,094354530
7 - 13 - 14 - 15	5499	839	6338	14%	14%	14%	0%	0,014766592
Total	38510	5963	44473	100%	100%	100%		
IV	0,04072							

ClasCAE	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ClasCAE"
1 (agropecuario)	5338	468	5806	14%	8%	13%	6%	0,568793796
3 - 16 + NA	20335	3495	23830	53%	59%	54%	-6%	-0,104334074
Otro	12837	2000	14837	33%	34%	33%	0%	-0,006159792
Total	38510	5963	44473	100%	100%	100%		
IV	0,04027							

No mejora la variable. Se descarta.

Fuente: Elaboración propia.

Tabla 7.41: WOE de variable IA.

IA	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "IA"
False	32825	4572	37397	85%	77%	84%	9%	0,105895437
True	5685	1391	7076	15%	23%	16%	-9%	-0,457536023
Total	38510	5963	44473	100%	100%	100%		
IV	0,04826							

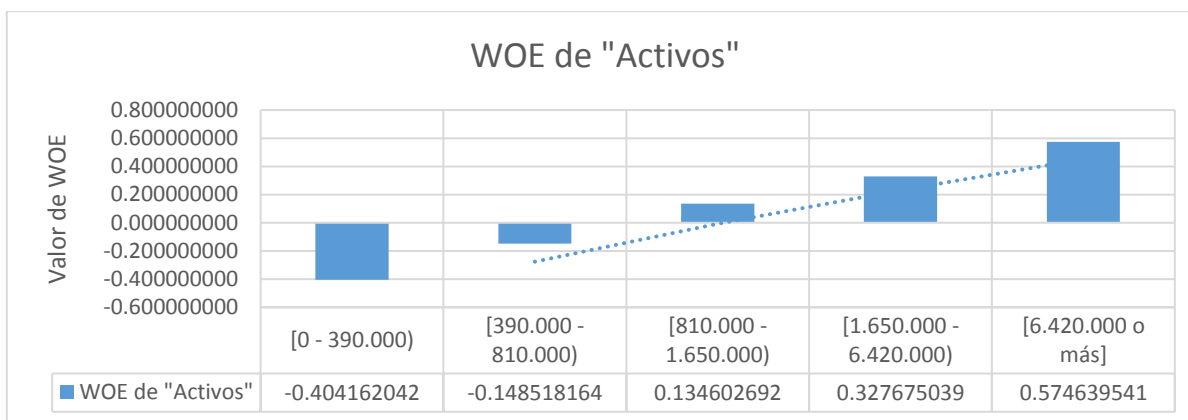
Se descarta la variable por no agregar la información.

Fuente: Elaboración propia.

7.7 Anexo 7: WOE de variables numéricas.

Tabla 7.42: WOE de variable Activos2R y gráfico.

Activos2R	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Activos"
[0 - 390.000)	11545	2678	14223	30%	45%	32%	-15%	-0,404162042
[390.000 - 810.000)	4498	808	5306	12%	14%	12%	-2%	-0,148518164
[810.000 - 1.650.000)	5475	741	6216	14%	12%	14%	2%	0,134602692
[1.650.000 - 6.420.000)	10441	1165	11606	27%	20%	26%	8%	0,327675039
[6.420.000 o más]	6551	571	7122	17%	10%	16%	7%	0,574639541
Total	38510	5963	44473	100%	100%	100%		
IV	0,133082529							



Fuente: Elaboración propia.

Tabla 7.43 WOE: de variable AbonoCta6R.

AbonoCta6R	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "AbonoCta6"
0	26545	3936	30481	69%	66%	69%	3%	0,043332189
1	5902	1124	7026	15%	19%	16%	-4%	-0,206946713
2	2424	463	2887	6%	8%	6%	-1%	-0,209896945
[3 - 6]	3639	440	4079	9%	7%	9%	2%	0,247345233
Total	38510	5963	44473	100%	100%	100%		
IV	0,016766115							

Se prueban distintas configuraciones. En todas se rechaza la variable por tener IV menor a 0.1

Fuente: Elaboración propia.

Tabla 7.44: WOE de variable AbonoRelAtrasoCredAnt2.

AbonoRelAtrasoCredAnt2	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "AbonoRelAtrasoCredAnt"
0%	30207	4757	34964	78%	80%	79%	-1%	-0,016887770
[1% - 3%]	2589	500	3089	7%	8%	7%	-2%	-0,220925356
[3% - 7%]	2394	383	2777	6%	6%	6%	0%	-0,032658341
[7% o más]	3320	323	3643	9%	5%	8%	3%	0,464723501
Total	38510	5963	44473	100%	100%	100%		
IV	0,018856635							

Se prueban distintas configuraciones. En todas se rechaza la variable por tener IV menor a 0.1

Fuente: Elaboración propia.

Tabla 7.45: WOE de variable Autonomía.

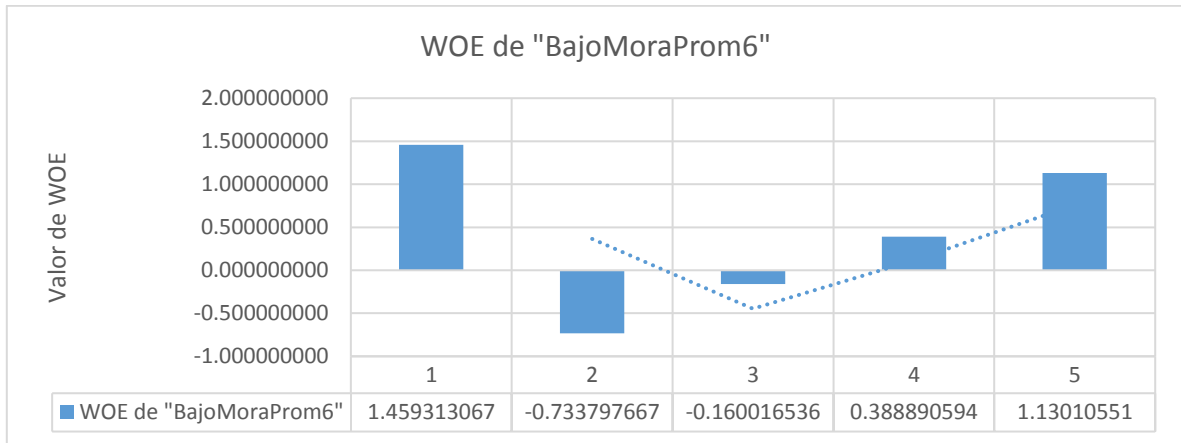
AutonomíaR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Autonomía"
[menos de 5%]	37360	5710	43070	97%	96%	97%	1%	0,013037352
[5% o más]	1150	253	1403	3%	4%	3%	-1%	-0,351216505
Total	38510	5963	44473	100%	100%	100%		
IV	0,00457719							

Se prueban distintas configuraciones. En todas se rechaza la variable por tener IV menor a 0.

Tabla 7.46: WOE de variable CreditosAnteriores y gráfico.

CreditosAnteriores	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CreditosAnteriores"
0	16518	3487	20005	43%	58%	45%	-16%	-0,309935310
1	9123	1365	10488	24%	23%	24%	1%	0,034300031
2	5333	624	5957	14%	10%	13%	3%	0,280174605
3	3102	263	3365	8%	4%	8%	4%	0,602304074
[4 o más]	4434	224	4658	12%	4%	10%	8%	1,120067101
Total	38510	5963	44473	100%	100%	100%		
IV	0,166895629							

A medida que tiene más créditos (más antigüedad), mejor es el cliente en su comportamiento de pago. No lineal pero tiene sentido. En 0 es muy bueno porque nunca se atrasó. Acá se están modelando shock de corto plazo. Si alguien tiene un shock temporal 1 mes (una urgencia), se puede atrasar 1 mes pero el cliente siempre estuvo bien. Cuando el shock temporal dura mucho se transforma en permanente (el comportamiento intermedio).



Fuente: Elaboración propia.

Tabla 7.47: WOE de variable GesCobranza6R.

GesCobranza6R	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "GesCobranza6"
0	34095	5257	39352	89%	88%	88%	0%	0,004245986
1	2025	215	2240	5%	4%	5%	2%	0,377342714
[2 - 6]	2390	491	2881	6%	8%	6%	-2%	-0,282739721
Total	38510	5963	44473	100%	100%	100%		
IV	0,011986438							

Se ve que hay una correlación entre realizar varias gestiones de cobranza y que el cliente sea mal pagador. Es mejor no utilizar, además que su IV es menor a 0.

Fuente: Elaboración propia.

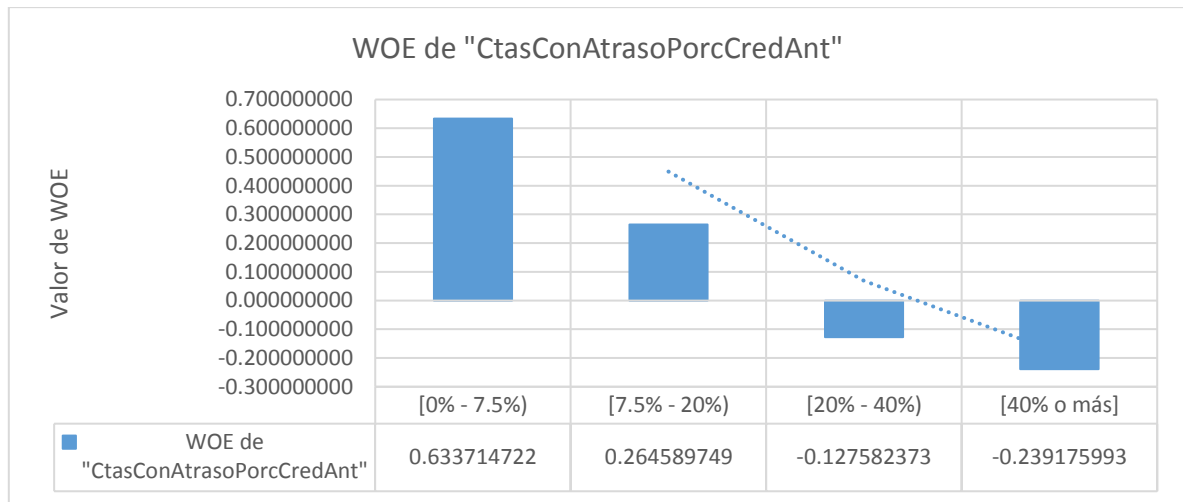
Tabla 7.48: WOE de variable CtasConAtrasoPorCredAnt y gráfico.

CtasConAtrasoPorCredAnt	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CtasConAtrasoPorCredAnt"
[0 - 4%]	21137	3766	24903	55%	63%	56%	-8%	-0,140332609
[4% - 16%]	2242	170	2412	6%	3%	5%	3%	0,713980929
[16% - 25%]	2057	217	2274	5%	4%	5%	2%	0,383762298
[25% - 83%]	10790	1527	12317	28%	26%	28%	2%	0,089970515
[83% o más]	2284	283	2567	6%	5%	6%	1%	0,222892435
Total	38510	5963	44473	100%	100%	100%		
IV	0,044159522							

Esta variable está distorsionada por el hecho de que el cliente nuevo tiene 0 en esta variable y, por lo general, son peores pagadores que los antiguos. Se genera una nueva configuración.

CtasConAtrasoPorCredAnt	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CtasConAtrasoPorCredAnt"
[0% - 7.5%]	5005	299	5304	23%	12%	22%	11%	0,633714722
[7.5% - 20%]	3356	290	3646	15%	12%	15%	4%	0,264589749
[20% - 40%]	4128	528	4656	19%	21%	19%	-3%	-0,127582373
[40% o más]	9503	1359	10862	43%	55%	44%	-12%	-0,239175993
Total	21992	2476	24468	100%	45%	100%		
IV	0,108266699							

Realizando cortes con árbol de decisión sobre los datos de la variable para clientes antiguos.

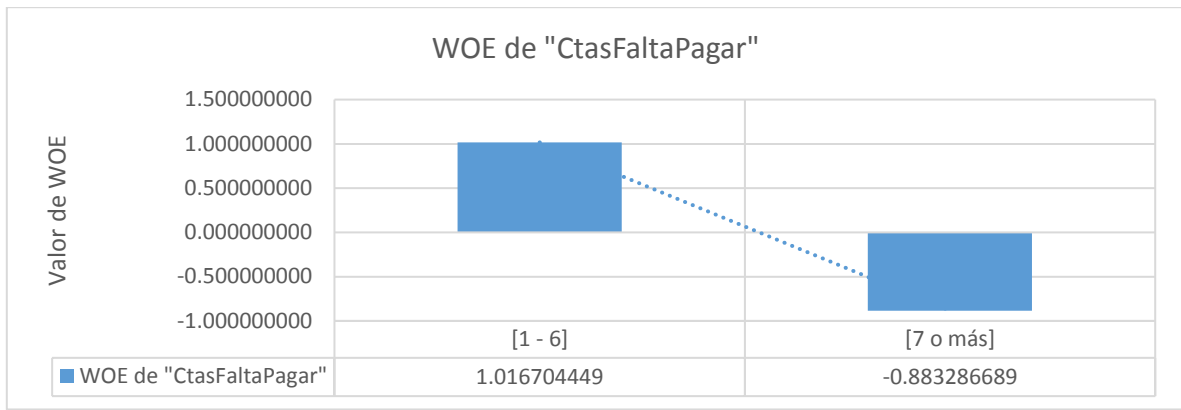


Fuente: Elaboración propia.

Tabla 7.49: WOE de variable CtasFaltaPagarR y gráfico.

CtasFaltaPagarR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CtasFaltaPagar"
[1 - 2]	11389	60	11449	30%	1%	26%	29%	3,380714456
[3 - 4]	8800	492	9292	23%	8%	21%	15%	1,018684046
[5 - 6]	6373	936	7309	17%	16%	16%	1%	0,052865881
[7 - 8]	5405	1446	6851	14%	24%	15%	-10%	-0,54682091
[9 o más]	6543	3029	9572	17%	51%	22%	-34%	-1,095180994
Total	38510	5963	44473	100%	100%	100%		
IV	1,541074337							

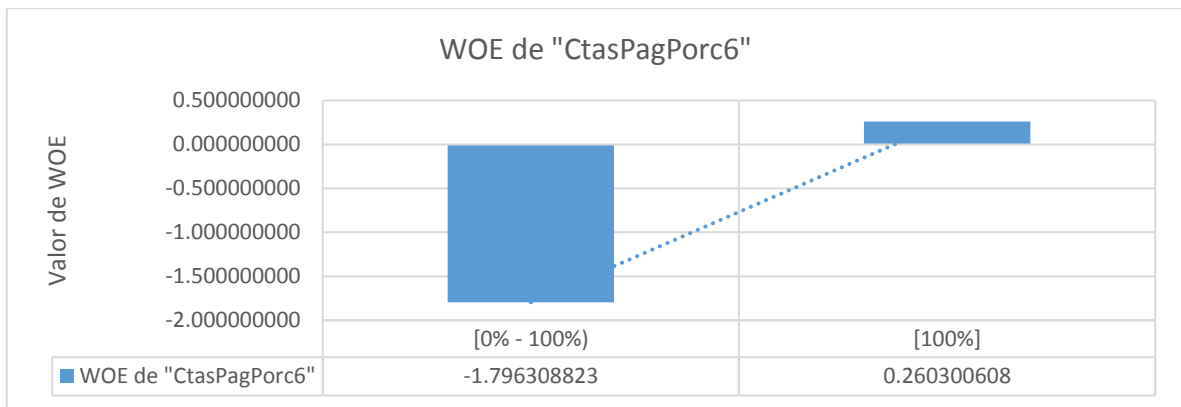
CtasFaltaPagarR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CtasFaltaPagar"
[1 - 6]	26562	1488	28050	69%	25%	63%	44%	1,016704449
[7 o más]	11948	4475	16423	31%	75%	37%	-44%	-0,883286689
Total	38510	5963	44473	100%	100%	100%		
IV	0,836383891							



Fuente: Elaboración propia.

Tabla 7.50: WOE de variable CtasPagPorc6R y gráfico.

CtasPagPorc6R	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CtasPagPorc6"
[0% - 100%]	1679	1567	3246	4%	26%	7%	-22%	-1,796308823
[100%]	36831	4396	41227	96%	74%	93%	22%	0,260300608
Total	38510	5963	44473	100%	100%	100%		
IV	0,45078436							

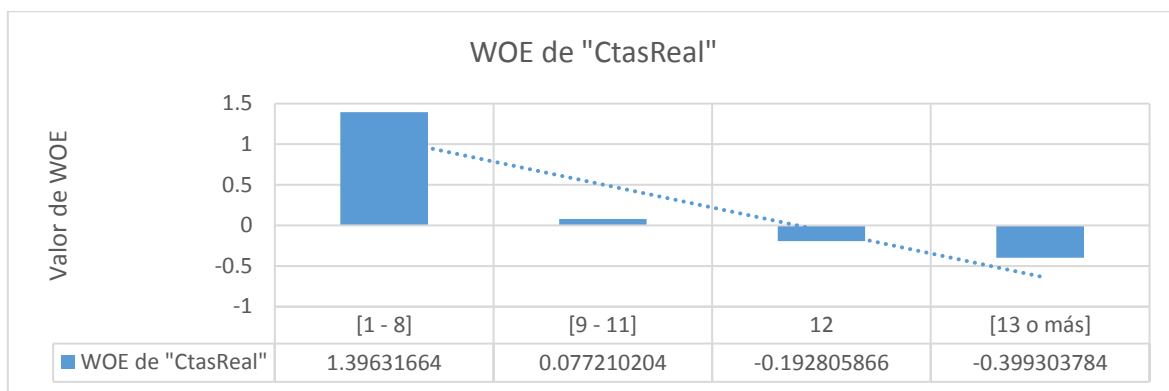


Fuente: Elaboración propia.

Tabla 7.51: WOE de variable CtasReal y gráfico.

CtasReal	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CtasReal"
[1 - 8]	9837	377	10214	26%	6%	23%	19%	1,39631664
[9 - 11]	4165	597	4762	11%	10%	11%	1%	0,077210204
12	15519	2914	18433	40%	49%	41%	-9%	-0,192805866
[13 o más]	8989	2075	11064	23%	35%	25%	-11%	-0,399303784
Total	38510	5963	44473	100%	100%	100%		
IV	0,331282458							

Se prueban distintas configuraciones, pero todas empeoraban el IV.



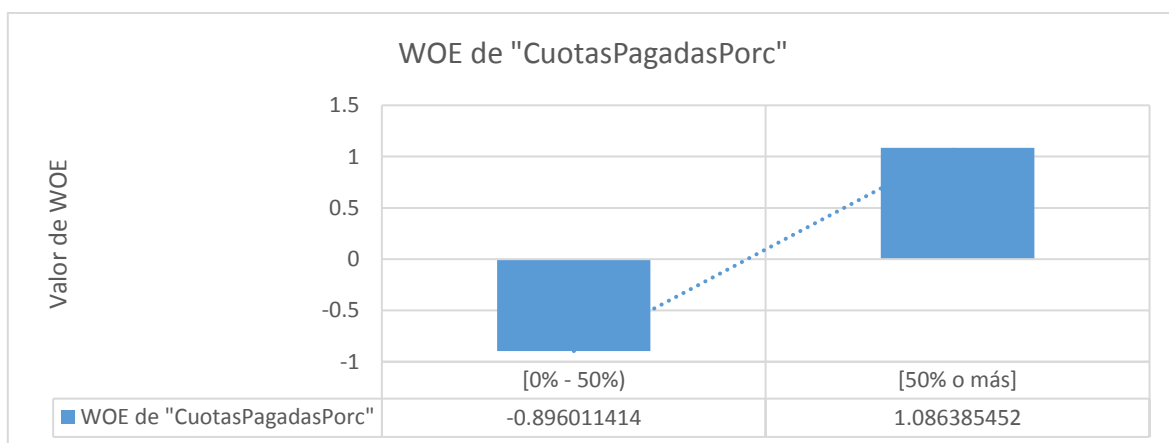
Fuente: Elaboración propia.

Tabla 7.52: WOE de variable CuotasPagadasPorc y prueba de otras configuraciones.

CuotasPagadasPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CuotasPagadasPorc"
[0% - 33%]	4911	2805	7716	13%	47%	17%	-34%	-1,305270190
[33% - 50%]	7168	1777	8945	19%	30%	20%	-11%	-0,470644111
[50% - 60%]	7507	858	8365	19%	14%	19%	5%	0,30364286
[60% - 80%]	10658	479	11137	28%	8%	25%	20%	1,237021228
[80% o más]	8266	44	8310	21%	1%	19%	21%	3,370372123
Total	38510	5963	44473	100%	100%	100%		
IV	1,457251966							

CuotasPagadasPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CuotasPagadasPorc"
[0% - 50%]	12079	4582	16661	31%	77%	37%	-45%	-0,896011414
[50% o más]	26431	1381	27812	69%	23%	63%	45%	1,086385452
Total	38510	5963	44473	100%	100%	100%		
IV	0,901487789							

CuotasPagadasPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CuotasPagadasPorc"
[0% - 33%]	4911	2805	7716	13%	47%	17%	-34%	-1,305270190
[33% - 50%]	7168	1777	8945	19%	30%	20%	-11%	-0,470644111
[50% o más]	26431	1381	27812	69%	23%	63%	45%	1,086385452
Total	38510	5963	44473	100%	100%	100%		
IV	0,994226373							

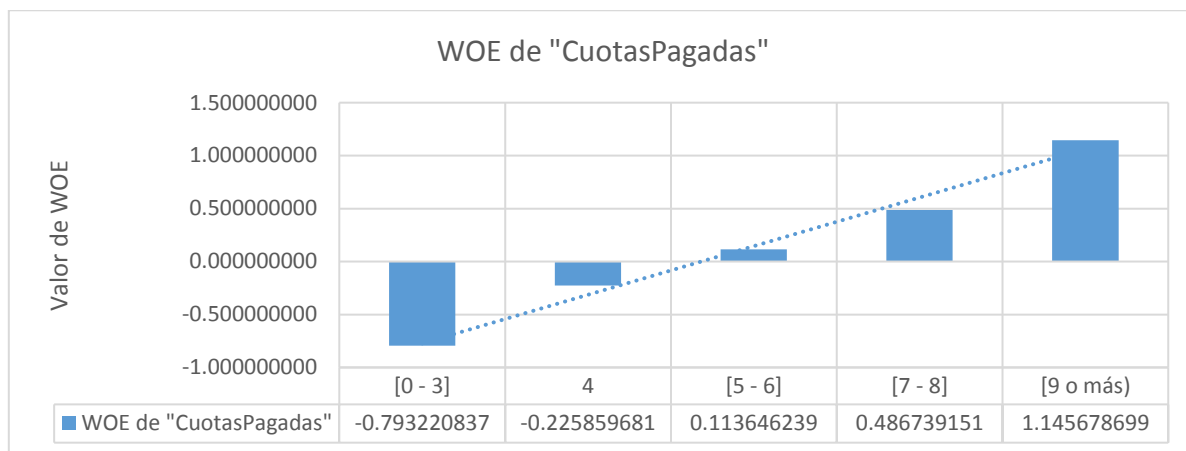


Fuente: Elaboración propia.

Tabla 7.53: WOE de variable CuotasPagadasR y gráfico.

CuotasPagadasR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "CuotasPagadas"
[0 - 3]	7190	2461	9651	19%	41%	22%	-23%	-0,793220837
4	5946	1154	7100	15%	19%	16%	-4%	-0,225859681
[5 - 6]	9066	1253	10319	24%	21%	23%	3%	0,113646239
[7 - 8]	6357	605	6962	17%	10%	16%	6%	0,486739151
[9 o más]	9951	490	10441	26%	8%	23%	18%	1,145678699
Total	38510	5963	44473	100%	100%	100%		
IV	0,423847856							

Se prueban distintas configuraciones, pero todas empeoraban el IV.



Fuente: Elaboración propia.

Tabla 7.54: WOE de variable DifMtoAprobAntPorcR.

DifMtoAprobAntPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "DifCtasCredAntPorc"
[menor a -47%]	3401	135	3536	9%	2%	8%	7%	1,361205769
[-47% - 2%]	21442	4282	25724	56%	72%	58%	-16%	-0,254412812
[2% - 47%]	4892	530	5422	13%	9%	12%	4%	0,357135253
[47% - 100%]	5115	653	5768	13%	11%	13%	2%	0,193011311
[100% o más]	3660	363	4023	10%	6%	9%	3%	0,445471354
Total	38510	5963	44473	100%	100%	100%		
IV	0,163779487							

Esta distorcionado por los clientes nuevos en que el crédito anterior es 0.

DifMtoAprobAntPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "DifCtasCredAntPorc"
[menor a -47%]	3399	135	3534	15%	5%	14%	10%	1,041927375
[-47% - -4%]	2834	566	3400	13%	23%	14%	-10%	-0,573184054
[-4% - 55%]	8122	886	9008	37%	36%	37%	1%	0,031580362
[55% - 99%]	3868	510	4378	18%	21%	18%	-3%	-0,157952265
[99% o más]	3769	379	4148	17%	15%	17%	2%	0,112994392
Total	21992	2476	24468	100%	100%	100%		
IV	0,168575456							

DifMtoAprobAntPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "DifCtasCredAntPorc"
[menor a -44%]	3564	159	3723	16%	6%	15%	10%	0,925700190
[-44% - 0%]	2974	585	3559	14%	24%	15%	-10%	-0,557983116
[0% - 55%]	7817	843	8660	36%	34%	35%	1%	0,043054774
[55% - 99%]	3868	510	4378	18%	21%	18%	-3%	-0,157952265
[99% o más]	3769	379	4148	17%	15%	17%	2%	0,112994392
Total	21992	2476	24468	100%	100%	100%		
IV	0,154417389							

Se descarta la variable.

Fuente: Elaboración propia.

Tabla 7.55: WOE de variable DifCtasCredAntPorcR, prueba de otras configuraciones y gráfico.

DifCtasCredAntPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "DifCtasCredAntPorc"
[menor a -33%]	3382	110	3492	9%	2%	8%	7%	1,560397926
[-33% - -8%]	2404	255	2659	6%	4%	6%	2%	0,378281513
[-8% - 13%]	24200	4332	28532	63%	73%	64%	-10%	-0,145020934
[13% - 100%]	5880	924	6804	15%	15%	15%	0%	-0,014744268
[100% o más]	2644	342	2986	7%	6%	7%	1%	0,179893226
Total	38510	5963	44473	100%	100%	100%		
IV	0,131978452							

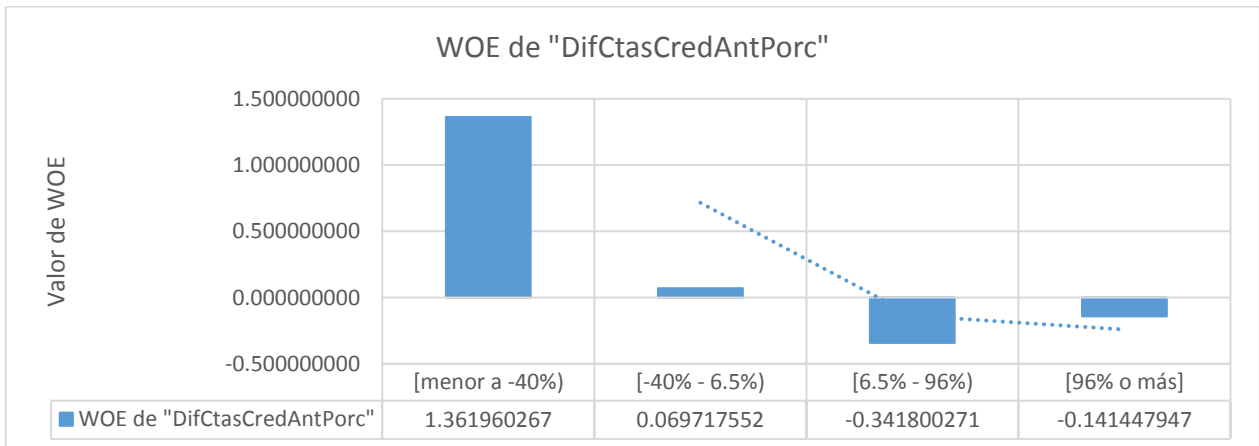
Esta distorsionado por los clientes nuevos en que el crédito anterior es 0.

DifCtasCredAntPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "DifCtasCredAntPorc"
[menor a -40%]	3086	89	3175	14%	4%	13%	10%	1,361960267
[-40% - -6.5%]	2900	302	3202	13%	12%	13%	1%	0,078004603
[-6.5% - 6.5%]	7071	745	7816	32%	30%	32%	2%	0,066338577
[6.5% - 96%]	6298	998	7296	29%	40%	30%	-12%	-0,341800271
[96% o más]	2637	342	2979	12%	14%	12%	-2%	-0,141447947
Total	21992	2476	24468	100%	100%	100%		
IV	0,186763121							

DifCtasCredAntPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "DifCtasCredAntPorc"
[menor a -40%]	3086	89	3175	14%	4%	13%	10%	1,361960267
[-40% - 6.5%]	9971	1047	11018	45%	42%	45%	3%	0,069717552
[6.5% o más]	8935	1340	10275	41%	54%	42%	-13%	-0,286727861
Total	21992	2476	24468	100%	100%	100%		
IV	0,182971131							

DifCtasCredAntPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "DifCtasCredAntPorc"
[menor a -40%]	3086	89	3175	14%	4%	13%	10%	1,361960267
[-40% - 6.5%]	9971	1047	11018	45%	42%	45%	3%	0,069717552
[6.5% - 96%]	6298	998	7296	29%	40%	30%	-12%	-0,341800271
[96% o más]	2637	342	2979	12%	14%	12%	-2%	-0,141447947
Total	21992	2476	24468	100%	100%	100%		
IV	0,186750852							

Dice que alguien que reduce la cantidad de cuotas a casi la mitad tiene un buen comportamiento (se da cuenta de su capacidad de pago). Alguien que reduce un poco, se mantiene o aumenta un poco es bueno, pero no tanto. Los que aumentan mucho, son los peores. Los que aumentan en casi un 100% deben ser clientes que la empresa evalúa muy bien y son malos, pero no tanto como los anteriores.

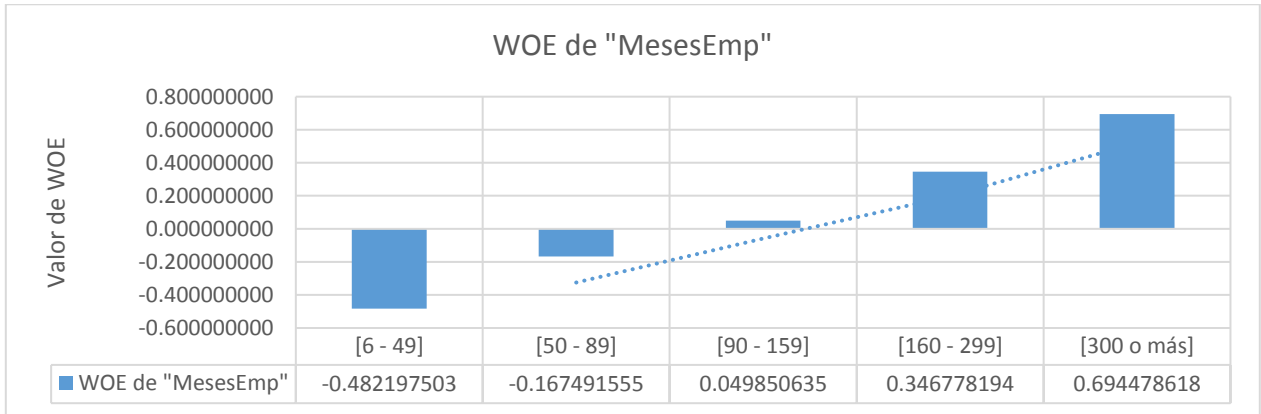


Fuente: Elaboración propia.

Tabla 7.56: WOE de variable MesesEmp2 y gráfico.

MesesEmp2R	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "MesesEmp"
[6 - 49]	6344	1591	7935	16%	27%	18%	-10%	-0,482197503
[50 - 89]	8816	1614	10430	23%	27%	23%	-4%	-0,167491555
[90 - 159]	9137	1346	10483	24%	23%	24%	1%	0,049850635
[160 - 299]	9738	1066	10804	25%	18%	24%	7%	0,346778194
[300 o más]	4475	346	4821	12%	6%	11%	6%	0,694478618
Total	38510	5963	44473	100%	100%	100%		
IV	0,122887624							

Se ve que a mayor experiencia en el rubro, mejor pagador es el cliente.

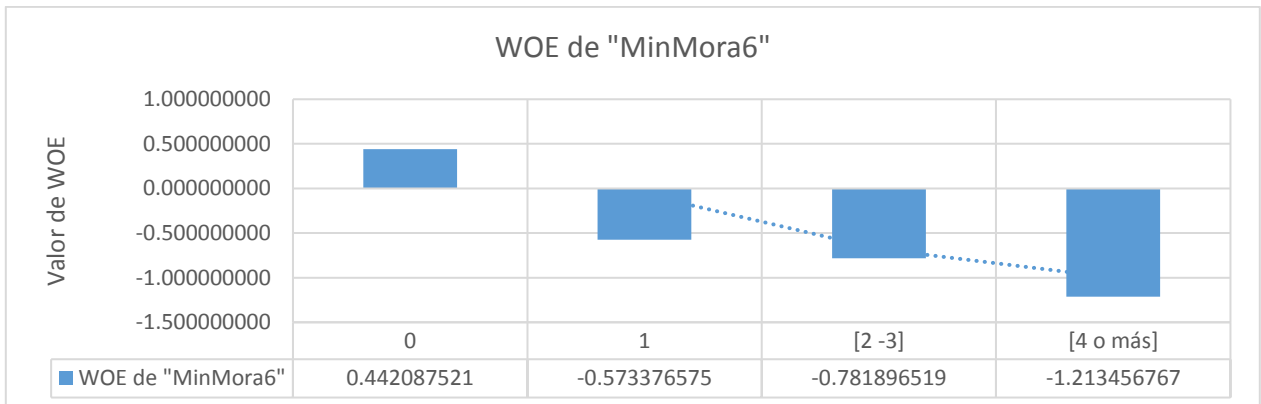


Fuente: Elaboración propia.

Tabla 7.57: WOE de variable MinMora6R y gráfico.

MinMora6R	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "MinMora6"
0	31231	3108	34339	81%	52%	77%	29%	0,442087521
1	2497	686	3183	6%	12%	7%	-5%	-0,573376575
[2 -3]	1767	598	2365	5%	10%	5%	-5%	-0,781896519
[4 o más]	3015	1571	4586	8%	26%	10%	-19%	-1,213456767
Total	38510	5963	44473	100%	100%	100%		
IV	0,424116208							

Mientras mayor sea el mínimo de mora promedio de los últimos 6 meses, peor pagador es.



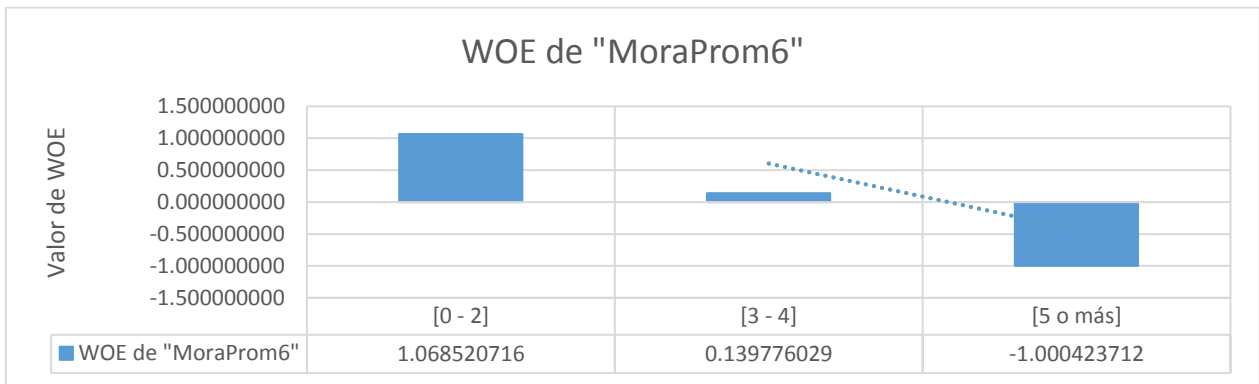
Fuente: Elaboración propia.

Tabla 7.58: WOE de variable MoraProm6R y gráfico.

MoraProm6R	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "MoraProm6"
[0 - 1]	21486	1012	22498	56%	17%	51%	39%	1,190128752
[2 - 4]	7273	845	8118	19%	14%	18%	5%	0,287243275
[5 - 9]	4745	1168	5913	12%	20%	13%	-7%	-0,46354569
[10 - 14]	2246	983	3229	6%	16%	7%	-11%	-1,039047223
[15 o más]	2760	1955	4715	7%	33%	11%	-26%	-1,520503751
Total	38510	5963	44473	100%	100%	100%		
IV	1,009474863							

MoraProm6R	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "MoraProm6"
0	16355	666	17021	42%	11%	38%	31%	1,335655032
[1 - 2]	8386	650	9036	22%	11%	20%	11%	0,692002327
[3 - 4]	4018	541	4559	10%	9%	10%	1%	0,139776029
[5 - 10]	5366	1379	6745	14%	23%	15%	-9%	-0,506620084
[11 o más]	4385	2727	7112	11%	46%	16%	-34%	-1,390356716
Total	38510	5963	44473	100%	100%	100%		
IV	1,019321143							

MoraProm6R	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "MoraProm6"
[0 - 2]	24741	1316	26057	64%	22%	59%	42%	1,068520716
[3 - 4]	4018	541	4559	10%	9%	10%	1%	0,139776029
[5 o más]	9751	4106	13857	25%	69%	31%	-44%	-1,000423712
Total	38510	5963	44473	100%	100%	100%		
IV	0,888121167							



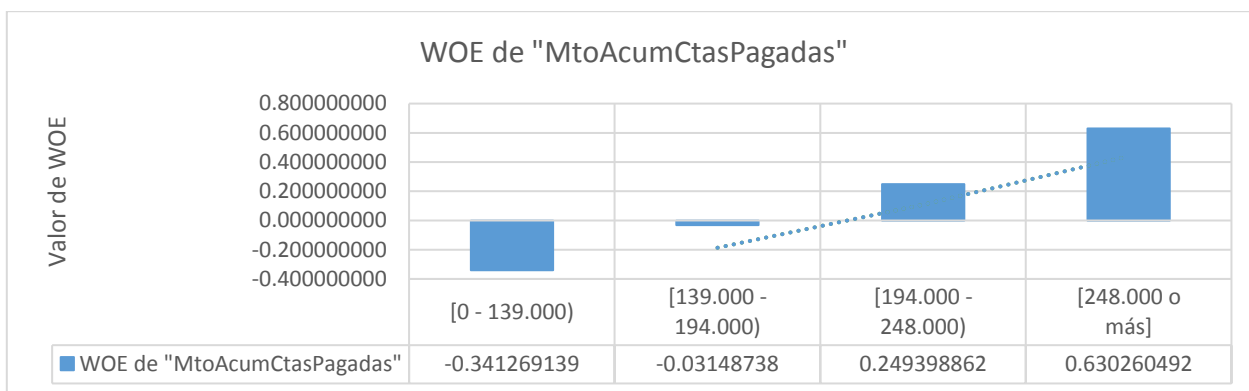
Fuente: Elaboración propia.

Tabla 7.59: WOE de variable MtoAcumCtasPagadasR y gráfico.

MtoAcumCtasPagadasR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "MtoAcumCtasPagadas"
[0 - 51.000)	4524	783	5307	12%	13%	12%	-1%	-0,111325096
[51.000 - 139.000)	10805	2556	13361	28%	43%	30%	-15%	-0,423778784
[139.000 - 194.000)	6865	1097	7962	18%	18%	18%	-1%	-0,03148738
[194.000 - 248.000)	4757	574	5331	12%	10%	12%	3%	0,249398862
[248.000 o más]	11559	953	12512	30%	16%	28%	14%	0,630260492
Total	38510	5963	44473	100%	100%	100%		
IV	0,159716126							

Si se deja como está igual tiene sentido, ya que el primer segmento debe corresponder al primer pago. Por lo general los clientes no fallan a la primera cuota. Aún así, puede ser mejor asumir que se van a caer.

MtoAcumCtasPagadasR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "MtoAcumCtasPagadas"
[0 - 139.000)	15329	3339	18668	40%	56%	42%	-16%	-0,341269139
[139.000 - 194.000)	6865	1097	7962	18%	18%	18%	-1%	-0,03148738
[194.000 - 248.000)	4757	574	5331	12%	10%	12%	3%	0,249398862
[248.000 o más]	11559	953	12512	30%	16%	28%	14%	0,630260492
Total	38510	5963	44473	100%	100%	100%		
IV	0,150680176							



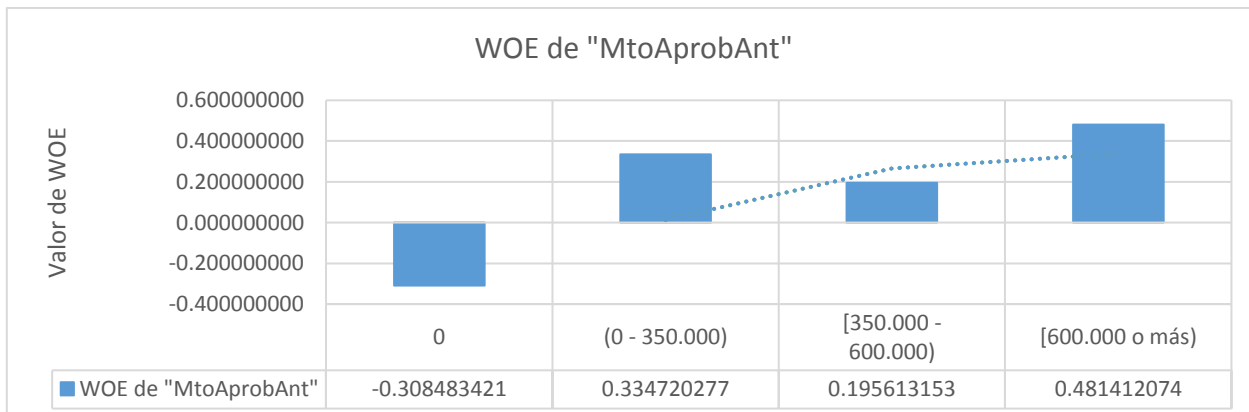
Fuente: Elaboración propia.

Tabla 7.60: WOE de variable MtoAprobAntR y gráfico.

MtoAprobAntR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "MtoAprobAnt"
[0 - 110.000)	17713	3630	21343	46%	61%	48%	-15%	-0,280278053
[110.000 - 350.000)	9444	1072	10516	25%	18%	24%	7%	0,310509319
[350.000 - 600.000)	5521	703	6224	14%	12%	14%	3%	0,195613153
[600.000 - 950.000)	2559	272	2831	7%	5%	6%	2%	0,376225532
[950.000 o más]	3273	286	3559	8%	5%	8%	4%	0,572126226
Total	38510	5963	44473	100%	100%	100%		
IV	0,096036437							

MtoAprobAntR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "MtoAprobAnt"
0	16931	3569	20500	44%	60%	46%	-16%	-0,308483421
(0 - 350.000)	10226	1133	11359	27%	19%	26%	8%	0,334720277
[350.000 - 600.000)	5521	703	6224	14%	12%	14%	3%	0,195613153
[600.000 o más]	5832	558	6390	15%	9%	14%	6%	0,481412074
Total	38510	5963	44473	100%	100%	100%		
IV	0,107132083							

El cero se deja aparte porque corresponde a los créditos que son nuevos y que por lo general esos clientes son más riesgosos. Si el cliente recibió un monto alto, es mejor que uno que recibió un monto medianamente alto, y si es muy bajo, también es mejor.



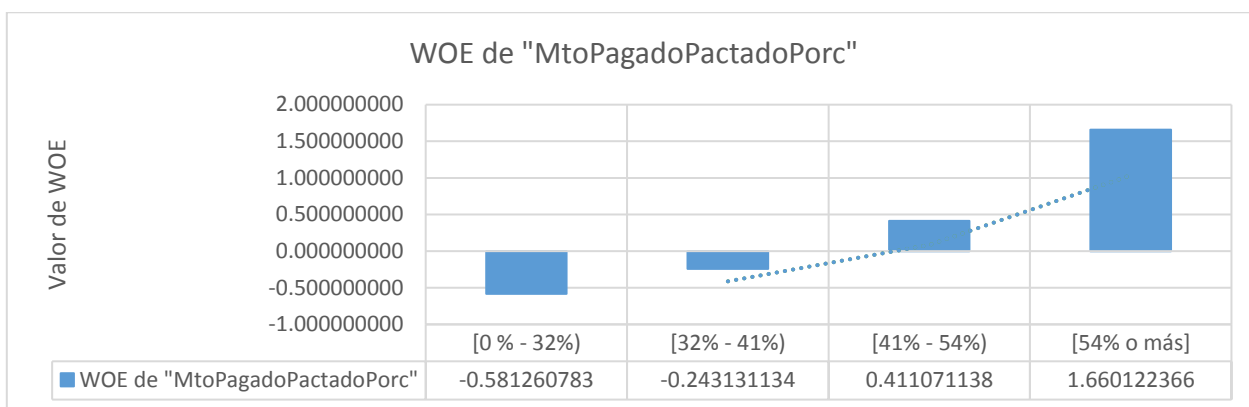
Fuente: Elaboración propia.

Tabla 7.61: WOE de variable MtoPagadoPactadoPorcR y gráfico.

MtoPagadoPactadoPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "MtoPagadoPactadoPorc"
[0 % - 7%]	4730	474	5204	12%	8%	12%	4%	0,435128922
[7% - 32%]	8188	3441	11629	21%	58%	26%	-36%	-0,998436697
[32% - 41%]	4254	840	5094	11%	14%	11%	-3%	-0,243131134
[41% - 54%]	7920	813	8733	21%	14%	20%	7%	0,411071138
[54% o más]	13418	395	13813	35%	7%	31%	28%	1,660122366
Total	38510	5963	44473	100%	100%	100%		
IV	0,887077863							

MtoPagadoPactadoPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "MtoPagadoPactadoPorc"
[0 % - 32%]	17172	4755	21927	45%	80%	49%	-35%	-0,581260783
[32% - 41%]	4254	840	5094	11%	14%	11%	-3%	-0,243131134
[41% - 54%]	7920	813	8733	21%	14%	20%	7%	0,411071138
[54% o más]	13418	395	13813	35%	7%	31%	28%	1,660122366
Total	42764	6803	49567	111%	114%	111%		
IV	0,70867021							

Cuando lleva un porcentaje bajo puede ser porque pactó en muchas cuotas y estamos observando la cuota 4 a lo más, o porque pactó montos bajos y casi el total del crédito en una cuota final. Se decide juntar con el segundo segmento asumiendo que es más lógico que sea más riesgoso ese cliente que uno que lleva casi la mitad pagada.



Fuente: Elaboración propia.

Tabla 7.62: WOE de variable PatrimonioR.

PatrimonioR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Patrimonio"
[menor a 380.000]	14703	2981	17684	38%	50%	40%	-12%	-0,269551498
[380.000 - 600.000]	2243	366	2609	6%	6%	6%	0%	-0,052408037
[600.000 - 1.280.000]	4798	661	5459	12%	11%	12%	1%	0,116856366
[1.280.000 - 13.780.000]	14326	1724	16050	37%	29%	36%	8%	0,252084658
[13.780.000 o más]	2440	231	2671	6%	4%	6%	2%	0,49199137
Total	38510	5963	44473	100%	100%	100%		
IV	0,066618167							

PatrimonioR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "Patrimonio"
[menor a 250.000]	13394	2727	16121	35%	46%	36%	-11%	-0,273739496
(250.000 - 580.000]	3489	612	4101	9%	10%	9%	-1%	-0,124706079
(580.000 - 1.220.000]	4651	654	5305	12%	11%	12%	1%	0,09638594
(1.220.000 - 2.980.000]	5886	664	6550	15%	11%	15%	4%	0,316705542
(2.980.000 o más]	11090	1306	12396	29%	22%	28%	7%	0,273730533
Total	38510	5963	44473	100%	100%	100%		
IV	0,064565207							

Se prueban distintas configuraciones pero no se logra mejorar el IV.

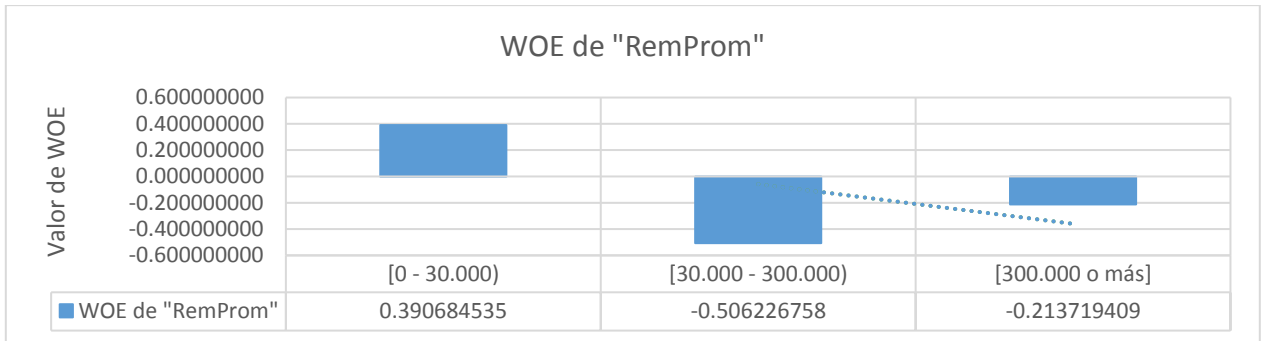
Fuente: Elaboración propia.

Tabla 7.63: WOE de variable RemPromR y gráfico.

RemPromR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "RemProm"
[0 - 30.000]	22536	2361	24897	59%	40%	56%	19%	0,390684535
[30.000 - 300.000]	8276	2126	10402	21%	36%	23%	-14%	-0,506226758
[300.000 - 400.000]	2914	596	3510	8%	10%	8%	-2%	-0,278302918
[400.000 - 600.000]	2422	406	2828	6%	7%	6%	-1%	-0,079348473
[600.000 o más]	2362	474	2836	6%	8%	6%	-2%	-0,259287563
Total	38510	5963	44473	100%	100%	100%		
IV	0,157512114							

RemPromR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "RemProm"
[0 - 30.000]	22536	2361	24897	59%	40%	56%	19%	0,390684535
[30.000 - 300.000]	8276	2126	10402	21%	36%	23%	-14%	-0,506226758
[300.000 o más]	7698	1476	9174	20%	25%	21%	-5%	-0,213719409
Total	38510	5963	44473	100%	100%	100%		
IV	0,155814542							

Si da muy baja remuneración, tiene mejor comportamiento (quizás puede reducir fácilmente los costos de personal o bien trabaja con familiares), por otro lado, si da una remuneración sobre el sueldo mínimo es más riesgoso. Si el sueldo es muy alto, es más seguro.

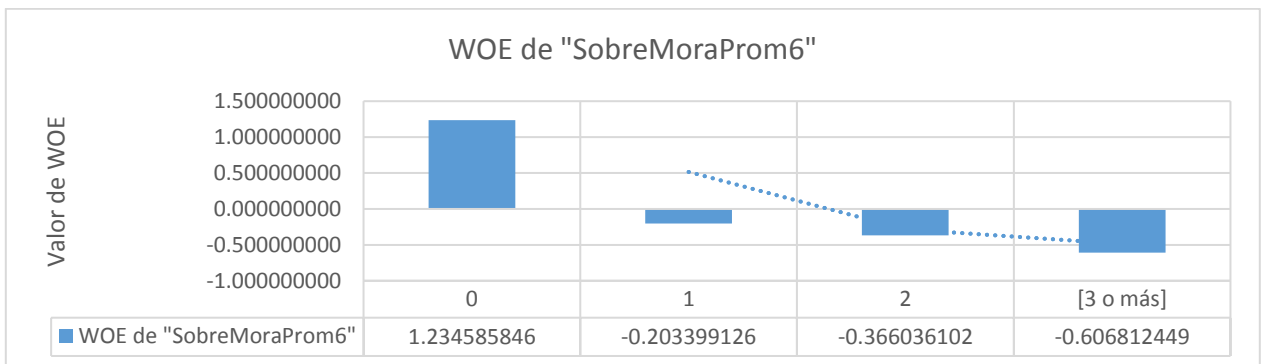


Fuente: Elaboración propia.

Tabla 7.64: WOE de variable SobreMoraProm6 y gráfico.

SobreMoraProm6	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "SobreMoraProm6"
0	13562	611	14173	35%	10%	32%	25%	1,234585846
1	12668	2404	15072	33%	40%	34%	-7%	-0,203399126
2	8890	1985	10875	23%	33%	24%	-10%	-0,366036102
[3 o más]	3390	963	4353	9%	16%	10%	-7%	-0,606812449
Total	38510	5963	44473	100%	100%	100%		
IV	0,405301735							

SobreMoraProm6	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "SobreMoraProm6"
0	13562	611	14173	35%	10%	32%	25%	1,234585846
1	12668	2404	15072	33%	40%	34%	-7%	-0,203399126
[2 o más]	12280	2948	15228	32%	49%	34%	-18%	-0,438499289
Total	38510	5963	44473	100%	100%	100%		
IV	0,40033023							



Fuente: Elaboración propia.

Tabla 7.65: WOE de variable TasaInterés2.

TasaInterés2	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "TasaInteres"
[0% - 3%)	4197	891	5088	11%	15%	11%	-4%	-0,315563402
[3% - 3,38%)	9524	1150	10674	25%	19%	24%	5%	0,248708749
[3,38% - 3,6%)	12156	2130	14286	32%	36%	32%	-4%	-0,123643342
[3,6% - 3,7%)	2164	206	2370	6%	3%	5%	2%	0,486493233
[3,7% o más]	10469	1586	12055	27%	27%	27%	1%	0,021859148
Total	38510	5963	44473	100%	100%	100%		
IV	0,042100373							

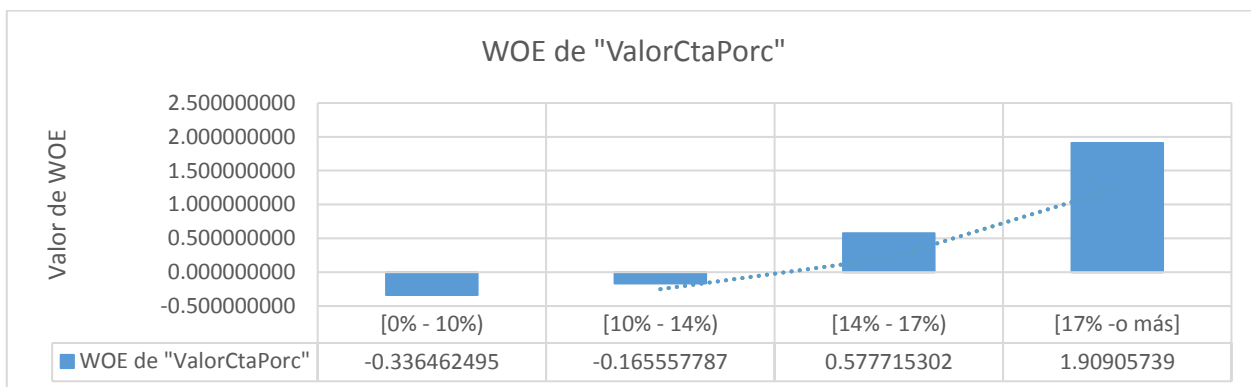
No se encuentra otra configuración que incremente el IV. Finalmente, el monto de la tasa (bastante estable) no es determinante.

Fuente: Elaboración propia.

Tabla 7.66:: WOE de variable ValorCtaPorcR y gráfico.

ValorCtaPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ValorCtaPorc"
[0% - 10%)	9286	2013	11299	24%	34%	25%	-10%	-0,336462495
[10% - 14%)	19401	3545	22946	50%	59%	52%	-9%	-0,165557787
[14% - 17%)	2808	244	3052	7%	4%	7%	3%	0,577715302
[17% - 20%)	4753	142	4895	12%	2%	11%	10%	1,645359981
[20% o más]	2262	19	2281	6%	0%	5%	6%	2,91422144
Total	38510	5963	44473	100%	100%	100%		
IV	0,391736838							

ValorCtaPorcR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ValorCtaPorc"
[0% - 10%)	9286	2013	11299	24%	34%	25%	-10%	-0,336462495
[10% - 14%)	19401	3545	22946	50%	59%	52%	-9%	-0,165557787
[14% - 17%)	2808	244	3052	7%	4%	7%	3%	0,577715302
[17% -o más]	7015	161	7176	18%	3%	16%	16%	1,90905739
Total	38510	5963	44473	100%	100%	100%		
IV	0,362164919							



Fuente: Elaboración propia.

Tabla 7.67: WOE de variable ValorCtaPromR6.

ValorCtaPromR6	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ValorCtaPorc"
(0 - 39.000)	16432	3019	19451	43%	51%	44%	-8%	-0,171039235
[39.000 - 58.000)	9286	1490	10776	24%	25%	24%	-1%	-0,035612468
[58.000 - 85.000)	6257	832	7089	16%	14%	16%	2%	0,152279437
[85.000 - 115.000)	3185	369	3554	8%	6%	8%	2%	0,290066686
[115.000 o más]	3350	253	3603	9%	4%	8%	4%	0,717981898
Total	38510	5963	44473	100%	100%	100%		
IV	0,05545511							

No se encuentra otra configuración que incremente el IV.

Fuente: Elaboración propia.

Tabla 7.68: WOE de variable ValorCtaPromCredAntR y gráfico.

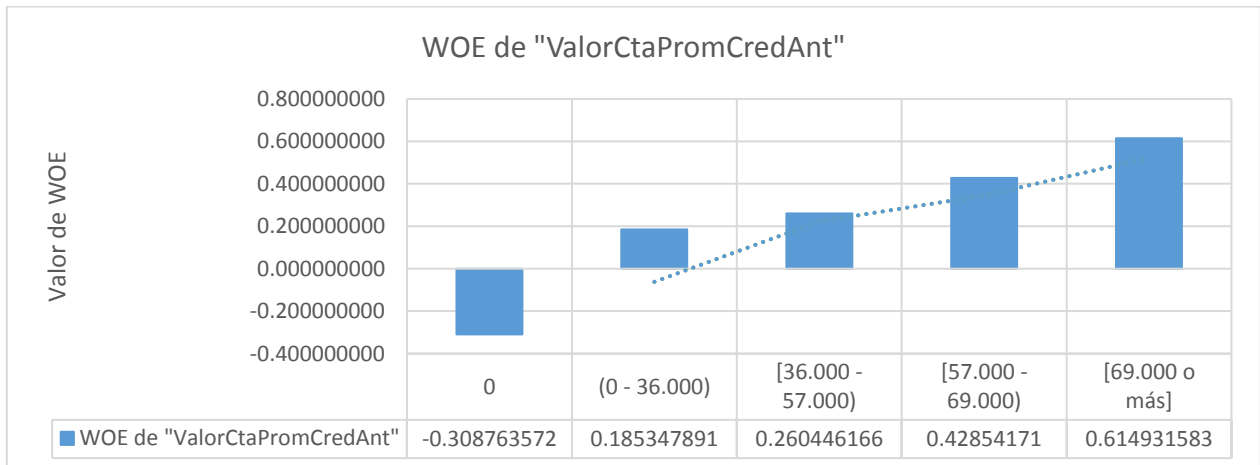
ValorCtaPromCredAntR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ValorCtaPromCredAnt"
[0 - 16.000)	17574	3639	21213	46%	61%	48%	-15%	-0,290632619
[16.000 - 36.000)	7037	919	7956	18%	15%	18%	3%	0,170306862
[36.000 - 57.000)	5564	664	6228	14%	11%	14%	3%	0,260446166
[57.000 - 69.000)	2518	254	2772	7%	4%	6%	2%	0,42854171
[69.000 o más]	5817	487	6304	15%	8%	14%	7%	0,614931583
Total	38510	5963	44473	100%	100%	100%		
IV	0,110665107							

Esta variable está distorciónada por los clientes nuevos en que el valor de la cuota del crédito anterior es 0. Se separa la variable.

ValorCtaPromCredAntR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ValorCtaPromCredAnt"
0	16931	3570	20501	44%	60%	46%	-16%	-0,308763572
(0 - 36.000)	7680	988	8668	20%	17%	19%	3%	0,185347891
[36.000 - 57.000)	5564	664	6228	14%	11%	14%	3%	0,260446166
[57.000 - 69.000)	2518	254	2772	7%	4%	6%	2%	0,42854171
[69.000 o más]	5817	487	6304	15%	8%	14%	7%	0,614931583
Total	38510	5963	44473	100%	100%	100%		
IV	0,116418735							

ValorCtaPromCredAntR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "ValorCtaPromCredAnt"
0	16931	3570	20501	44%	60%	46%	-16%	-0,308763572
(0 - 57.000)	13244	1652	14896	34%	28%	33%	7%	0,216213707
[57.000 o más]	8335	741	9076	22%	12%	20%	9%	0,554873932
Total	38510	5963	44473	100%	100%	100%		
IV	0,114706976							

Si en el crédito anterior pactó en cuotas altas y sobrevivió, es mejor que alguien que pidió cuotas bajas.



Fuente: Elaboración propia.

Tabla 7.69: WOE de variable VecesMoraBajo6R.

VecesMoraBajo6R	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "VecesMoraBajo6"
0	13259	1550	14809	34%	26%	33%	8%	0,281077399
1	13014	2515	15529	34%	42%	35%	-8%	-0,22159134
2	8688	1294	9982	23%	22%	22%	1%	0,038860329
[3 - 5]	3549	604	4153	9%	10%	9%	-1%	-0,094497283
Total	38510	5963	44473	100%	100%	100%		
IV	0,043485888							

La variable no tiene lógica.

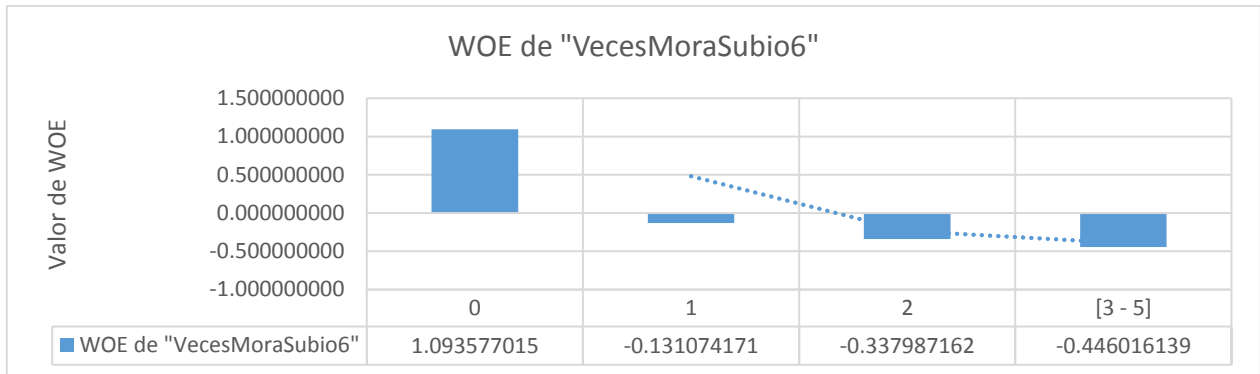
Fuente: Elaboración propia.

Tabla 7.70: WOE de variable VecesMoraSubio6R y gráfico.

VecesMoraSubio6R	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "VecesMoraSubio6"
0	12241	635	12876	32%	11%	29%	21%	1,093577015
1	12083	2133	14216	31%	36%	32%	-4%	-0,131074171
2	9539	2071	11610	25%	35%	26%	-10%	-0,337987162
[3 - 5]	4647	1124	5771	12%	19%	13%	-7%	-0,446016139
Total	38510	5963	44473	100%	100%	100%		
IV	0,300832278							

VecesMoraSubio6R	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "VecesMoraSubio6"
0	12241	635	12876	32%	11%	29%	21%	1,093577015
1	12083	2133	14216	31%	36%	32%	-4%	-0,131074171
[2 - 5]	14186	3195	17381	37%	54%	39%	-17%	-0,374675763
Total	38510	5963	44473	100%	100%	100%		
IV	0,299647986							

Esta variable indica si la morosidad va subiendo y a medida que lo hace la calidad del cliente empeora.



Fuente: Elaboración propia.

Tabla 7.71: WOE de variable VtasPromR.

VtasPromR	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "VtasPromR"
[1 - 360.000]	4787	1215	6002	12%	20%	13%	-8%	-0,494184404
[360.000 - 630.000]	8354	1590	9944	22%	27%	22%	-5%	-0,206337788
[630.000 - 1.000.000]	9339	1350	10689	24%	23%	24%	2%	0,06875035
[1.000.000 - 1.570.000]	7128	920	8048	19%	15%	18%	3%	0,182068061
[1.570.000 o más]	8902	888	9790	23%	15%	22%	8%	0,439715269
Total	38510	5963	44473	100%	100%	100%		
IV	0,092401801							

La variable no agrega información.

Fuente: Elaboración propia.

Tabla 7.72: WOE de variable Leverage.

Leverage	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "VtasPromR"
(0 - 0,3]	10393	874	11267	27%	15%	25%	12%	0,610463168
(0,3 - 0,4]	5382	786	6168	14%	13%	14%	1%	0,058514301
(0,4 - 0,53]	5901	941	6842	15%	16%	15%	0%	-0,02941027
(0,53 - 0,6]	3699	804	4503	10%	13%	10%	-4%	-0,339125715
(0,6 - 0,8]	5300	952	6252	14%	16%	14%	-2%	-0,148447173
[0,8 o más)	7835	1606	9441	20%	27%	21%	-7%	-0,280489977
Total	38510	5963	44473	100%	100%	100%		
IV	0,110771171							

Leverage	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "VtasPromR"
(0 - 0,3]	10393	874	11267	27%	15%	25%	12%	0,610463168
(0,3 - 0,4]	5382	786	6168	14%	13%	14%	1%	0,058514301
(0,4 - 0,5]	4737	772	5509	12%	13%	12%	-1%	-0,051169485
(0,5 - 0,6]	4863	973	5836	13%	16%	13%	-4%	-0,256317509
(0,6 - 0,8]	5300	952	6252	14%	16%	14%	-2%	-0,148447173
[0,8 o más)	7835	1606	9441	20%	27%	21%	-7%	-0,280489977
Total	38510	5963	44473	100%	100%	100%		
IV	0,107272914							

Leverage	Goods (N)	Bads (S)	Total General	%Good	%Bad	%Total	%Good-%Bad	WOE de "VtasPromR"
(0 - 0,3]	10393	874	11267	27%	15%	25%	12%	0,610463168
(0,3 - 0,5]	10119	1558	11677	26%	26%	26%	0%	0,00566766
(0,5 - 0,8]	10163	1925	12088	26%	32%	27%	-6%	-0,201516531
(0,8 o más)	7835	1606	9441	20%	27%	21%	-7%	-0,280489977
Total	38510	5963	44473	100%	100%	100%		
IV	0,105633098							

Fuente: Elaboración propia.