ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

CONSTRUCCIÓN DE UN MODELO DE INFERENCIA DIFUSA PARA LA EVALUACIÓN CREDITICIA DE POTENCIALES CLIENTES DE UNA CARTERA DE CONSUMO

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO MATEMÁTICO

PROYECTO INTEGRADOR

ALEX SANTIAGO LOACHAMÍN SUNTAXI

alex.1603@live.com

DIRECTOR: HOLGER EFRAÍN BENALCÁZAR PALADINES, MSC.

holger.benalcazar@epn.edu.ec

QUITO, NOVIEMBRE 2016

DECLARACIÓN

Yo Alex Santiago Loachamín Suntaxi, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la siguiente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

_____

Alex Santiago Loachamín Suntaxi

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Alex Santiago Loachamín Suntaxi, bajo mi supervisión.

_____

MSc. Holger Efraín Benalcázar Paladines

DIRECTOR DEL TRABAJO

AGRADECIMIENTO

Es muy extensa la lista de las personas que han aportado en el desarrollo de mis labores personales, académicas y profesionales; no puedo mencionar a todas, pero quiero agradecer en este trabajo a algunas de ellas.

Fiel a mi fe cristiana, agradezco a Dios, mi motivación para continuar siempre adelante.

A la Escuela Politécnica Nacional, a la Facultad de Ciencias, sus autoridades y maestros, en especial al MSc. Holger Benalcázar, a quien debo el haber confirmado mi decisión de estudiar la carrera de Ingeniería Matemática y por supuesto el haber concluido este trabajo exitosamente.

A la institución que me ha brindado la oportunidad de poner en práctica los conocimientos adquiridos en la carrera y plasmarlos dentro de este proyecto, merece especial mención Cecilia Vallejo, quién ha compartido conmigo su conocimiento y experiencia, y además ha confiado en mis habilidades profesionales.

A mis padres, primos, tíos, amigos y en general a todas las personas que siempre me han contagiado de su alegría y han mostrado su apoyo incondicional y desinteresado.

Mi reconocimiento a todas y cada una de las personas mencionadas y también a las que no he alcanzado a citar.

DEDICATORIA

A Luz María y Margarita, mis abuelas materna y paterna, como homenaje post mortem. A ellas debo el haber comprendido el verdadero significado de los buenos valores y principios, por sobre todo el haber despertado un gran sentimiento de amor hacia toda mi familia.

A mis padres, Blanca y Carlos, quienes con su apoyo sincero, permanente y por supuesto desinteresado han contribuido a que este objetivo profesional concluya exitosamente.

CONTENIDO

CAPÍTULO 1 LOS PROCESOS DE EVALUACIÓN CREDITICIA DE POTENCIALES CLIENTES. 1

1.1 LA EVALUACIÓN CREDITICIA DENTRO DEL MARCO REGULATORIO ECUATORIANO. 2

1.2 EL CICLO DE CRÉDITO. 4

1.2.1 PREVENTA. 5

1.2.2 ANÁLISIS. 6

1.2.3 SEGUIMIENTO. 8

1.2.4 RECUPERACIÓN. 9

1.3 GESTIÓN INTEGRAL Y CONTROL DE RIESGOS. 10

1.3.1 DEFINICIÓN DE RIESGO. 10

1.3.2 CLASIFICACIÓN DEL RIESGO. 11

1.3.3 ADMINISTRACIÓN DEL RIESGO DE CRÉDITO. 13

1.4 PROCESOS DE SEGMENTACIÓN DE POTENCIALES CLIENTES. 15

1.4.1 PROCESAMIENTO DE INFORMACIÓN POR LOTES. 15

1.4.2 CAPTACIÓN DE CLIENTES NUEVOS. 17

1.5 PLANTEAMIENTO DEL PROBLEMA. 19

1.6 JUSTIFICACIÓN DEL TRABAJO. 23

1.6.1 MODELOS SCORING DE ORIGINACIÓN. 23

1.6.2 MODELO DE INFERENCIA DIFUSA. 25

1.7 HIPÓTESIS DEL TRABAJO. 28

1.8 OBJETIVOS DEL ESTUDIO. 29

CAPÍTULO 2 ESTIMACIÓN DEL PERFIL DE RIESGO.....30

2.1 DEFINICIÓN DEL UNIVERSO DE INDIVIDUOS Y VARIABLES. 31

2.2 ELECCIÓN DE LA MUESTRA DE ESTUDIO. 33

2.2.1 MADUREZ DE LA CARTERA..... 33

2.2.2 REPRESENTATIVIDAD DE LA MUESTRA. 34

2.2.3 MUESTRAS DE APRENDIZAJE Y TESTING. 34

2.3 ELECCIÓN DE LA VARIABLE DEPENDIENTE DEL MODELO. 35

2.3.1 INDICADOR DE PÉRDIDA POR OPERACIÓN..... 35

2.3.2 DEFINICIÓN DE LA VARIABLE DEPENDIENTE..... 37

2.3.3 ESTABILIDAD DE LA DEFINICIÓN. 42

2.4 ANÁLISIS DESCRIPTIVO DE LAS VARIABLES. 45

2.4.1 IDENTIFICACIÓN DE VALORES NULOS..... 46

2.4.2 ANÁLISIS DE LA CONCENTRACIÓN DE INFORMACIÓN. 47

2.5 SELECCIÓN DE VARIABLES REGRESORAS DEL MODELO. 50

2.5.1 CATEGORIZACIÓN DE VARIABLES..... 50

2.5.2 CRITERIOS PARA LA SELECCIÓN DE VARIABLES REGRESORAS. 54

2.5.3 SELECCIÓN DE VARIABLES REGRESORAS DICOTÓMICAS..... 59

2.5.4 SELECCIÓN DE VARIABLES REGRESORAS RECODIFICADAS..... 60

2.5.5 SELECCIÓN DE VARIABLES REGRESORAS LOGARÍTMICAS. 62

2.5.6 ESTUDIO DEL PROBLEMA DE MULTICOLINEALIDAD EN LAS VARIABLES PREDICTORAS. 62

2.6 ESTIMACIÓN DEL MODELO DE REGRESIÓN LOGÍSTICA. 64

2.6.1 BOOTSTRAP PARA ESTIMAR INTERVALOS DE CONFIANZA DE LOS COEFICIENTES DEL MODELO. 66

2.6.2 DESCRIPCIÓN DE LAS VARIABLES DEL MODELO..... 70

2.7 MEDIDAS DE BONDAD DE AJUSTE DEL MODELO..... 76
2.7.1 COEFICIENTES R2..... 76
2.7.2 BONDAD DE AJUSTE DE HOSMER-LEMESHOW..... 77
2.7.3 ESTADÍSTICO DE WALD. 78
2.7.4 PRUEBA DE OMNIBUS. 79
2.8 CORRECCIÓN DE LA CONSTANTE DEL MODELO. 80
2.8.1 EFECTOS DE CORREGIR LA CONSTANTE DEL MODELO..... 81
2.9 VALORACIÓN DE LA CALIDAD DE DISCRIMINACIÓN. 85
2.9.1 TEST DE KOLMOGOROV-SMIRNOV (KS). 85
2.9.2 COEFICIENTE DE GINI. 85
2.10 INTERVALOS DE CONFIANZA PARA LOS INDICADORES KS Y GINI.. 88
CAPÍTULO 3 CONSTRUCCIÓN DEL MODELO DE INFERENCIA DIFUSA.....90
3.1 ETAPAS DE UN MODELO DE INFERENCIA DIFUSA..... 91
3.1.1 DEFINICIÓN DE VARIABLES. 91
3.1.2 FUSIFICACIÓN DE VARIABLES..... 94
3.1.3 MECANISMOS DE INFERENCIA..... 95
3.1.4 AGREGACIÓN. 96
3.1.5 CONCRECIÓN. 101
3.2 MOTIVACIÓN PARA LA CONSTRUCCIÓN DE UN MODELO DE INFERENCIA DIFUSA. 103
3.3 CONSTRUCCIÓN DEL MODELO DE INFERENCIA DIFUSA..... 104
3.3.1 DEFINICIÓN DEL UNIVERSO DE INDIVIDUOS..... 104
3.3.2 SELECCIÓN DE LAS VARIABLES DEL MODELO. 107
3.3.3 DEFINICIÓN DE CONJUNTOS DIFUSOS..... 109
3.3.4 REGLAS TIPO MANDANI. 121

3.3.5 AGREGACIÓN. 123

3.3.6 VALORES CONCRETOS PARA EL CUPO DE CRÉDITO. 126

3.4 EVALUACIÓN DE RESULTADOS. 126

CAPÍTULO 4 CONCLUSIONES Y RECOMENDACIONES. 132

4.1 CONCLUSIONES 132

4.2 RECOMENDACIONES. 136

LISTA DE FIGURAS

Figura 1.1 Ciclo de crédito según la visión de Equifax 4
Figura 1.2 Administración integral del riesgo..... 11
Figura 1.3 Clasificación del riesgo, según la SBS. 12
Figura 1.4 Etapas de preventa y análisis del ciclo de crédito de ABC..... 23
Figura 2.1 Frecuencia de clientes nuevos por mes de estudio..... 31
Figura 2.2 Frecuencia y pérdida de la variable peor calificación de clientes nuevos 38
Figura 2.3 Promedio de los días de mora de los créditos otorgados a los clientes nuevos..... 39
Figura 2.4 Proporción de malos en la muestra de aprendizaje 44
Figura 2.5 Punto de corte de la variable saldo por vencer histórico 52
Figura 2.6 Reagrupación de las categorías de la variable tipo de vivienda..... 53
Figura 2.7 Combinación de las variables antigüedad de residencia y tipo de vivienda 53
Figura 2.8 Histograma del coeficiente de la variable antigüedad de residencia (WOE) 67
Figura 2.9 Poder predictivo de modelos scoring. Fuente: Equifax..... 76
Figura 2.10 Desplazamiento de la curva logística 80
Figura 2.11 Curva de aciertos en la predicción de clientes buenos y malos luego de corregir la constante del modelo 84
Figura 2.12 Curva de aciertos en la predicción de clientes buenos y malos antes de corregir la constante del modelo 84
Figura 2.13 Indicador KS del modelo 86
Figura 2.14 Coeficiente de Gini 87
Figura 3.1 Etapas de un Sistema de Inferencia Difusa. Fuente: [1] Medina & Paniagua 91
Figura 3.2 Función de pertenencia en la Lógica Clásica..... 92
Figura 3.3 Función de pertenencia en la Lógica Difusa..... 93
Figura 3.4 Funciones de pertenencia de perfil de riesgo y cupo de crédito..... 101

Figura 3.5 Concreción de conjuntos difusos 102
Figura 3.6 Serie temporal de endeudamiento de un potencial cliente de ABC 104
Figura 3.7 Variables seleccionadas para los modelos I y II 109
Figura 3.8 Histograma y función de pertenencia del perfil de riesgo – Modelo I 111
Figura 3.9 Histograma y función de pertenencia de carga financiera máxima – Modelo II 113
Figura 3.10 Histograma y función de pertenencia de endeudamiento – Modelo I .. 113
Figura 3.11 Histograma y función de pertenencia de variación de la deuda – Modelo I 115
Figura 3.12 Histograma y función de pertenencia de perfil de riesgo – Modelo II .. 116
Figura 3.13 Histograma y función de pertenencia de máximo endeudamiento – Modelo II 117
Figura 3.14 Histograma y función de pertenencia de saldo adeudado – Modelo II 118
Figura 3.15 Histograma y función de pertenencia de variación de endeudamiento – Modelo II..... 119
Figura 3.16 Conjuntos difusos del cupo de crédito – Modelo I y II 120
Figura 3.17 T-norma del mínimo sobre las reglas tipo Mandani 125
Figura 3.18 Agregación de la T-norma 126
Figura 3.19 Evaluación de resultados – Modelo I 128
Figura 3.20 Comparación de los cupos de crédito reales y estimados – Modelo I 129
Figura 3.21 Evaluación de resultados – Modelo II 131
Figura 3.22 Comparación de los cupos de crédito reales y estimados – Modelo II 131

LISTA DE TABLAS

Tabla 2.1 Frecuencia de clientes nuevos de ABC, por mes de estudio 32
Tabla 2.2 Muestras de aprendizaje y testing..... 35
Tabla 2.3 Calificación de créditos de consumo según la SBS 36
Tabla 2.4 Constitución de provisiones de créditos de consumo 36
Tabla 2.5 Peor calificación de clientes nuevos en la muestra de aprendizaje 37
Tabla 2.6 Promedio de los días de mora de los créditos otorgados a los clientes nuevos 38
Tabla 2.7 Pérdida de ABC respecto a la peor calificación y promedio de los días de mora 40
Tabla 2.8 Frecuencia de la variable dependiente sobre la muestra de aprendizaje . 41
Tabla 2.9 Proporción de malos por mes en la muestra de aprendizaje 43
Tabla 2.10 Resultados del test de Dickey Fuller 45
Tabla 2.11 Frecuencia y frecuencia relativa de datos válidos y nulos 46
Tabla 2.12 Resumen del análisis de los valores nulos 47
Tabla 2.13 Frecuencia y frecuencia acumulada de la variable provincia de nacimiento 49
Tabla 2.14 Variables eliminadas por el criterio de concentración de información..... 49
Tabla 2.15 Categorías resultantes del árbol de decisión 51
Tabla 2.16 Reglas de Thumb..... 57
Tabla 2.17 Selección de variables regresoras dicotómicas 59
Tabla 2.18 Índice de estabilidad de variable antigüedad de residencia 61
Tabla 2.19 Recodificación de la variable antigüedad de residencia..... 62
Tabla 2.20 Análisis de multicolinealiadad de las variables dicotómicas..... 63
Tabla 2.21 Regresión logística preliminar..... 65
Tabla 2.22 Test Chi-Cuadrado del coeficiente de la variable antigüedad de vivienda 68
Tabla 2.23 Estimación de los coeficientes de la regresión final..... 69
Tabla 2.24 Descripción de las variables dicotómicas del modelo final..... 75
Tabla 2.25 Coeficientes R2..... 77

Tabla 2.26 Tabla de contingencia para el test Hosmer-Lemeshow 77
Tabla 2.27 Prueba de Hosmer-Lemeshow para la bondad de ajuste del modelo..... 78
Tabla 2.28 Resultados para el estadístico de Wald 78
Tabla 2.29 Prueba Omnibus para la significancia conjunta de las estimaciones..... 79
Tabla 2.30 Matriz de confusión 82
Tabla 2.31 Acierto en la predicción de clientes buenos y malos..... 82
Tabla 2.32 Cálculo del indicador KS y Gini 87
Tabla 2.33 Indicadores KS y GINI para la muestra de testing 89
Tabla 3.1 Partición de la muestra por disponibilidad del salario 105
Tabla 3.2 Indicador de pérdida por puntaje de score..... 106
Tabla 3.3 Perfil de riesgo – Modelo I 111
Tabla 3.4 Carga financiera máxima – Modelo I..... 112
Tabla 3.5 Endeudamiento en la competencia – Modelo I 114
Tabla 3.6 Variación de la deuda último semestre – Modelo I..... 115
Tabla 3.7 Perfil de riesgo – Modelo II 116
Tabla 3.8 Máximo endeudamiento último año – Modelo II..... 117
Tabla 3.9 Endeudamiento en la competencia – Modelo II 118
Tabla 3.10 Variación del endeudamiento último semestre – Modelo II..... 119
Tabla 3.11 Equivalencia entre valores lingüísticos y numéricos 122
Tabla 3.12 Ejemplo de equivalencia entre valores lingüísticos y numéricos..... 122
Tabla 3.13 Percentiles asociados a las reglas tipo Mandani..... 123
Tabla 3.14 Vector de características de un individuo – Modelo I 123
Tabla 3.15 Reglas tipo Mandani para un vector de características..... 124
Tabla 3.16 T-norma del minimo para el antecedente de las reglas Mandani..... 125
Tabla 3.17 Evaluación de resultados – Modelo 1 128
Tabla 3.18 Evaluación de resultados – Modelo II 130

RESUMEN

Este proyecto presenta la construcción de una metodología para la evaluación crediticia de potenciales clientes de una entidad financiera privada regulada por la SBS (Superintendencia de Bancos y Seguros de Ecuador). La propuesta se basa en un modelo de inferencia difuso que conjuga por un lado, el análisis del perfil de riesgo del solicitante y por otro, la forma de administrar su endeudamiento en la competencia; con el análisis de esta información se busca sugerir un cupo de crédito para que sea ofertado como pre aprobado.

El trabajo empieza con la construcción del perfil de riesgo de un potencial cliente por medio de una regresión logística binaria; el resultado se incluye como una variable del modelo difuso, la evaluación crediticia se complementa con las variables relacionadas al endeudamiento actual e histórico en la competencia. Lo anterior permite identificar la relación entre las variables de entrada y salida mediante la base de conocimiento que incluye: el conocimiento que tienen los expertos sobre el fenómeno, los objetivos de negocio que tiene la institución y el resultado del análisis estadístico de las variables.

Esto permite obtener una metodología analítica y estadística, que se convierte en una herramienta objetiva y estandarizada que ayuda a minimizar el riesgo de crédito debido al incumplimiento del pago de los clientes nuevos de la entidad y/o al sobre endeudamiento de ellos.

Palabras clave: Evaluación crediticia, regresión logística, modelo de inferencia difusa, segmentación, riesgo de crédito.

ABSTRACT

This project presents the construction of a methodology for the credit assessment of prospects of a private financial institution regulated by SBS (Superintendencia de Bancos y Seguros de Ecuador). The proposal is based on a fuzzy inference model which combines on the one hand, the analysis of the risk profile of the applicant and secondly, how to manage your debt in competition; with the analysis of this information it seeks to suggest a line of credit to be offered as pre-approved.

The work begins with the construction of the risk profile of a potential customer through a binary logistic regression; the result is included as a variable of the fuzzy model, the credit assessment is complemented by the variables related to current and historical debt in the competition. This identifies the relationship between input and output variables using the knowledge base which includes: knowledge among experts on the phenomenon, the business objectives of the institution and the result of statistical analysis of the variables.

This allows to get an analytical and statistical methodology, which becomes an objective and standardized tool that helps minimize credit risk due to nonpayment of new customers of the entity and / or over-indebtedness of them.

Keywords: Credit assessment, logistic regression, fuzzy inference model, segmentation, credit risk.

CAPÍTULO 1
LOS PROCESOS DE EVALUACIÓN CREDITICIA DE POTENCIALES CLIENTES.

Uno de los mayores retos para las instituciones privadas del sistema financiero es, la permanencia en el mercado, cuidando que sus indicadores de riesgo permanezcan estables y ello exige que sean más rápidas y sólidas [1]. Esto ha motivado a que organizaciones del sistema financiero ecuatoriano busquen constantemente nuevas herramientas metodológicas y tecnológicas que les permita mejorar su posición competitiva en el mercado. Es por esta razón que la financiera ABC, llamada así por motivos de confidencialidad, se encuentra en la búsqueda nuevas técnicas que ayuden a optimizar los procesos de evaluación, segmentación y atracción de clientes nuevos; esto con la finalidad de incrementar los volúmenes de crédito, particularmente de su cartera de consumo, que es el tema central del presente proyecto.

En este capítulo se realiza una descripción de los aspectos más relevantes que intervienen al momento de realizar la evaluación crediticia de potenciales clientes de la cartera de consumo de ABC; en ese sentido, se presentan los siguientes tópicos:

- Exposición de los aspectos relacionados a la evaluación crediticia, según lo dispuesto por la Superintendencia de Bancos y Seguros (SBS).
- El ciclo de crédito y sus principales implicaciones en los procesos internos de ABC.
- Aspectos principales que rigen en la gestión integral de riesgos, haciendo especial énfasis en la administración del riesgo de crédito, según las recomendaciones de la SBS.
- Segmentación de potenciales clientes, presentando principalmente las ventajas competitivas que adquiere ABC al focalizar la atracción de clientes nuevos a la institución.

- Descripción de los procedimientos actuales de ABC, donde básicamente se hace una introducción a los conceptos fundamentales que rigen en la investigación, además de un reconocimiento a los procedimientos que sigue ABC como parte de la avaluación crediticia de potenciales clientes.
- Atributos de los modelos de regresión logística binaria y de inferencia difusa, específicamente en la labor de identificar del nivel de riesgo de un mercado potencial de clientes y en la generación de cupos de crédito para el segmento deseado.

1.1 LA EVALUACIÓN CREDITICIA DENTRO DEL MARCO REGULATORIO ECUATORIANO.

En este proyecto se hace un recorrido por la normativa vigente emitida por la SBS, al respecto de la evaluación crediticia; pero antes es preciso tomar en cuenta que, la investigación se centra específicamente en la cartera de consumo de ABC; en ese sentido, es importante conocer la definición de un crédito de consumo.

> *Son créditos otorgados a personas naturales destinados al pago de bienes, servicios o gastos no relacionados con una actividad productiva, cuya fuente de pago es el ingreso neto mensual promedio del deudor, entendiéndose por éste el promedio de los ingresos brutos mensuales del núcleo familiar menos los gastos familiares estimados mensuales obtenidos de fuentes estables como: sueldos, salarios, honorarios, remesas y/o rentas promedios. Generalmente se amortizan en función de un sistema de cuotas periódicas.*[1]

En adelante, los temas que se presentan hacen referencia a este tipo de operaciones crediticias.

[1] Disposición emitida en el numeral 1.2.1 del artículo 5 en la sección II “Elementos de la calificación de activos de riesgo y su clasificación” en el capítulo II “Calificación de activos de riesgo y constitución del provisiones por parte se las instituciones controladas por la Superintendencia de Banco y Seguros”, del título IX “De los activos y de los límites del créditos”, del libro I “Sistema Financiero”, de la Codificación de Resoluciones de la Superintendencia de Bancos y Seguros y de la Junta Bancaria.

La normativa vigente de la SBS dispone también que, todas las operaciones efectuadas a través del sistema de tarjetas de crédito a personas naturales, se consideran como créditos de consumo, siempre y cuando el destino del crédito sea el pago de bienes, servicios o gastos no relacionados con una actividad productiva; sin embargo, en este proyecto no se incluyen en el proceso de evaluación crediticia a los potenciales clientes de la tarjeta de crédito que emite ABC.

Una vez que se ha definido el segmento de crédito sobre el que gira la investigación, nace la necesidad de revisar los lineamientos establecidos por la Superintendencia de Bancos y Seguros, al respecto de la evaluación crediticia de potenciales clientes de las instituciones financieras ecuatorianas. Por tratarse de aspectos regulatorios, ABC fija sus estrategias y políticas de administración de acuerdo a estos lineamientos.

La normativa vigente establece que, en el proceso de administración de créditos de consumo se deberá dar especial importancia a la política que la institución del sistema financiero aplique para la selección de los sujetos de crédito, a la determinación de la capacidad de pago del deudor y a la estabilidad de la fuente de sus recursos, provenientes de los sueldos, salarios, honorarios, remesas, rentas promedio u otras fuentes de ingreso redituables, adecuadamente verificados por la institución prestamista del sistema financiero. ABC debe responder a lo dispuesto por el órgano rector del sistema financiero; es decir, incluir dentro de su política interna instrumentos metodológicos que les permita realizar una selección adecuada de los sujetos de crédito, en particular sus potenciales clientes. Así mismo debe establecer mecanismos para la verificación de la información presentada por el solicitante, poniendo especial énfasis en la validación de las fuentes de ingreso, pues a partir de ella se establece la capacidad de pago.

En relación a la evaluación crediticia de potenciales clientes es importante también considerar que, las instituciones financieras están obligadas a desarrollar y utilizar metodologías o sistemas internos para la aprobación y desembolso de las operaciones crediticias. Para este efecto la SBS dispone que, las metodologías implantadas deben considerar la combinación de criterios cuantitativos y cualitativos, de acuerdo con la

experiencia y las políticas estratégicas de la entidad; deben permitir monitorear y controlar la exposición crediticia de los diferentes portafolios. Esta metodología debe ser evaluada periódicamente a fin de garantizar la idoneidad de la misma, al igual que la relevancia de las variables utilizadas.

Esto concuerda con criterios expresados por otros autores. En la evaluación de crédito se consideran variables cuantitativas y cualitativas. Históricamente los primeros se han estudiado de una manera sistemática, a través de instrumentos estadísticos y econométricos; los segundos, o aspectos cualitativos, han sido estudiados acorde a los criterios de las entidades financieras y por los expertos en el tema [1].

1.2 EL CICLO DE CRÉDITO.

El siguiente gráfico muestra las etapas del ciclo de crédito, de acuerdo a la visión de Equifax[2].

Figura 1.1 Ciclo de crédito según la visión de Equifax

[2] Una de las empresas líder en la calificación de activos de riesgo y servicios de buró de crédito.

El criterio de Equifax acerca de las etapas del ciclo de crédito concuerda con el de la SBS; adicionalmente, este organismo rector establece que las entidades del sistema financiero desarrollen criterios, metodologías y sistemas internos de evaluación crediticia para la selección y otorgamiento de créditos, que se ajusten al perfil de riesgo de la entidad, los que deben ser consistentes con la naturaleza, tamaño y complejidad de las operaciones de la institución controlada; por ejemplo, *credit scoring* para créditos de consumo a personas naturales.

En ese sentido, en cada una de las etapas del ciclo de crédito es posible construir un modelo *scoring* diferente, cada uno con propósitos específicos, tomando en cuenta que la información disponible para cada uno de ellos son de diferente naturaleza; esta investigación se ocupa de la etapa de preventa de ABC, en donde el objetivo del modelo *scoring* es, seleccionar a los potenciales clientes según su perfil de riesgo con el propósito de focalizar la concesión de créditos en los sujetos que representan menor riesgo para la institución y de esa manera incrementar los volúmenes de la cartera de crédito de consumo.

A continuación se describen cada una de las etapas del ciclo de crédito:

1.2.1 PREVENTA.

Es la etapa del ciclo en la que se realiza una evaluación previa de los clientes o potenciales clientes, con la finalidad de seleccionar las condiciones de crédito más adecuadas para cada segmento de individuos. El desarrollo de esta etapa permite determinar el tiempo de respuesta y los gastos de colocación.

Es importante mencionar que la investigación se enfoca únicamente en los potenciales clientes de ABC; es decir, individuos que no han tenido ninguna operación de crédito con la institución, pero que se encuentran dentro del mercado objetivo de la empresa y con los que ABC podría tener una oportunidad de negocio, dado que presentan condiciones favorables para el otorgamiento de un crédito.

De acuerdo al punto de vista y experiencia de la empresa Equifax, los beneficios de atender esta etapa del ciclo de crédito son:

- En la gestión de venta, permite identificar a los sujetos de crédito, seleccionando así los clientes o potenciales clientes que cumplen con el nivel de riesgo que fija la institución.
- En el control de sus clientes y atracción de nuevos clientes, permite la evaluación rápida y la clasificación de los mismos, para ejecutar acciones diferenciadas por el perfil de riesgo.
- Eficiencia y efectividad en los contactos a realizar en la preventa, mejorando los tiempos y el costo de captación de clientes nuevos.
- Aumentan el retorno sobre las inversiones realizadas con clientes nuevos de la institución.

La estrategia que ha ganado un gran espacio en el mercado financiero, para alcanzar los beneficios citados, es la generación de los modelos *scoring* de originación; con los cuales se busca evaluar el perfil de riesgo de los potenciales clientes, en base a fuentes de información externas a la institución. Estos modelos se componen fundamentalmente de variables socio-demográficas y del buró de crédito.

Adicional a la construcción de un modelo *scoring* de originación para ABC, en este proyecto se plantea la construcción de un modelo de inferencia difuso que permite evaluar el perfil de riesgo y variables de endeudamiento actual e histórico del sujeto en la competencia, con la finalidad estudiar la forma en que los sujetos administran su deuda con otras instituciones y a partir de ese entendimiento, sugerir un monto de crédito para que sea ofertado a los potenciales clientes de ABC; de esta forma, no sólo se analiza a quien ofertar, sino también que ofertar.

1.2.2 ANÁLISIS.

Una vez que se ha identificado, desde la etapa anterior, a los potenciales clientes y el producto que se desea ofertar para atraerlo a la institución, es importante cerrar el

negocio de manera ágil y sin dejar de atender los requerimientos de la SBS, respecto a la revisión de documentos y verificación de la información presentada por el futuro cliente, antes proceder con el desembolso del crédito.

Es así que, en esta etapa, la intervención de los asesores de crédito toma un rol muy importante, pues son los encargados de revisar y validar los datos que declaran los individuos en la solicitud de crédito. Cabe mencionar que tanto los clientes de ABC como los potenciales clientes deben someterse a esta etapa del ciclo de crédito; y, como se ha mencionado anteriormente, este proyecto de centra en el segundo grupo de individuos.

Para cumplir con el análisis de la documentación presentada por los solicitantes, se aplican estrategias diferenciadas entre clientes y futuros clientes de la institución; para los primeros, los controles son menos estrictos. Esto se debe básicamente a que la institución financiera almacena la información de los clientes en las bases de datos, misma que ha sido analizada previo al desembolso de créditos anteriores; en cambio, esto no sucede con los clientes nuevos, pues se debe verificar por primera vez la información presentada en la solicitud.

La documentación que deben adjuntar los futuros clientes de ABC va acorde a lo revisado anteriormente en la normativa de la SBS; los de mayor relevancia son:

a) Soporte de la fuente de los ingresos;
b) Dirección laboral y de domicilio;
c) Mecanizado del IESS;
d) Declaración patrimonial;
e) Referencias bancarias y deudas en el sistema financiero (reporte del buró);
f) Información del cónyuge y dependientes, según sea el caso; y,
g) Referencias personales.

La principal actividad del asesor de crédito, en esta etapa del ciclo de crédito, es validar la información presentada por el solicitante. La confirmación de la información se realiza, generalmente, por medio de llamadas telefónicas o verificaciones físicas,

según sea el caso; también resulta de utilidad el servicio de bases de datos públicas como el registro civil, seguro social, matriculación de vehículos y municipio.

La implementación y automatización de modelos matemáticos y estadísticos en la atapa de análisis del ciclo de crédito, según Equifax, trae los siguientes beneficios:

- Permite atender ágilmente a los potenciales clientes identificados con un menor perfil de riesgo.
- Objetividad en el análisis de la información externa a la institución del potencial cliente.

Los modelos estadísticos que comúnmente se construyen para estos efectos son los *scoring*, y como se dijo anteriormente, en este proyecto se propone también la construcción de un modelo de inferencia difusa.

1.2.3 SEGUIMIENTO.

El monitoreo sobre el movimiento de portafolio de clientes es fundamental para establecer estrategias comerciales futuras; esto permite a una institución financiera, contar con una alerta temprana sobre las informaciones negativas de sus clientes y de esta manera mantener el control del nivel de riesgo de su cartera de crédito. Es preciso indicar que en esta etapa del ciclo de crédito están involucrados todos los clientes, pues en el seguimiento ya no existe la concepción de cliente potencial.

Una empresa que dedique sus esfuerzos a cumplir con esta etapa del ciclo de crédito, obtiene beneficios, tales como:

- Selección de clientes con calificación excelente para fijar nuevas estrategias de mercado con este grupo de individuos.
- Fidelización de la cartera clientes, especialmente de aquellos que muestran ser excelentes en el cumplimiento de sus obligaciones financieras con la institución.
- Seguimiento a clientes con problemas potenciales para evitar, de manera temprana, un posible deterioro de los indicadores de riesgo de la institución.

Los modelos más adecuados para esta etapa del ciclo de crédito son los denominados *scoring* de seguimiento, los cuales toman información del comportamiento interno de pago del portafolio de clientes con la finalidad de estimar la probabilidad de que un cliente incumpla con sus obligaciones de pago en la institución financiera.

1.2.4 RECUPERACIÓN.

Consiste en el desarrollo de herramientas y metodologías que permitan la protección de la cartera de crédito, con la finalidad de mantener estables los indicadores de rendimiento de las instituciones financieras.

En esta etapa del ciclo de crédito es posible seleccionar a los clientes por nivel de morosidad y de esa manera elaborar estrategias diferenciadas de recuperación de la cartera, empezando por una gestión preventiva para aquellos clientes que presenten una probabilidad de incumplir con sus pagos futuros hasta los que mantienen niveles críticos de mora en la institución.

Es claro mirar que, esta etapa requiere de recursos humanos y económicos para que pueda llevarse a cabo; en ese sentido, una correcta segmentación de los clientes por medio de modelos *scoring* de cobranzas, permite optimizar la inversión de recursos, puesto que permiten elaborar estrategias diferenciadas de manera objetiva y minimizando el criterio subjetivo de los asesores de crédito que se encargan de asignar la intensidad de la cobranza.

Pese a que en esta etapa del ciclo de crédito intervienen exclusivamente los clientes y no los potenciales clientes, la evaluación debe considerar no únicamente el comportamiento de pago interno sino también en la competencia, puesto que ello podría alertar un posible incumplimiento dentro de la institución.

Los métodos más usuales dentro de la gestión de cobranzas son:

- **Llamadas telefónicas:** está orientada a la recuperación del cliente y posterior pago de la deuda, este tipo de gestión normalmente se utiliza cuando la intensidad de la cobranza es baja o preventiva.
- **Localización del deudor:** mantener una base de datos actualizada de clientes permite, en una gestión de cobranzas, ubicar a aquellos que se encuentran en mora; generalmente esta gestión se realiza con los clientes que muestran niveles críticos de morosidad en la institución, haciendo necesaria una alta intensidad en la cobranza.

1.3 GESTIÓN INTEGRAL Y CONTROL DE RIESGOS.

Este trabajo se basa fundamentalmente en los lineamientos de la SBS, al respecto de la gestión y administración de riesgos, esto se debe básicamente a que ABC es una institución que debe cumplir con las disposiciones emitidas por el órgano rector. Por este motivo, la sección que se va tratar a continuación, considera las definiciones planteados por la SBS, al respecto del riesgo, su clasificación y administración.

1.3.1 DEFINICIÓN DE RIESGO.

La SBS define al riesgo como la posibilidad de que se produzca un hecho generador de pérdidas que afecten el valor económico de las operaciones [7].

La definición plantea posibles pérdidas que pueden afectar al correcto desempeño de una institución financiera, es por eso que, la SBS emite recomendaciones para una correcta administración del riesgo; a la que define como: “El proceso mediante el cual las instituciones financieras identifican, miden, controlan / mitigan y monitorean los riesgos inherentes al negocio, con el objetivo de definir el perfil de riesgo, el grado de exposición que la institución está dispuesta a asumir en el desarrollo del negocio y los mecanismos de cobertura, para proteger los recursos propios y de terceros que se encuentran bajo su control y administración”.

Las etapas de la administración de riesgos se definen como:

Figura 1.2 Administración integral del riesgo

1. **Identificación.-** Es el levantamiento de los riesgos, a partir de un inventario estándar de instituciones, mediante entrevistas con personal clave.
2. **Medición.-** Análisis de los riesgos identificados, depuración, consolidación y agrupación, así como los controles asociados.
3. **Control o mitigación del riesgo.-** Determinación de controles, planes de mitigación y responsables.
4. **Monitoreo.-** Seguimiento de los planes de mitigación conjuntamente con los responsables.

1.3.2 CLASIFICACIÓN DEL RIESGO.

La normativa vigente emitida por la SBS[3], clasifica al riesgo de la siguiente manera:

1. **Riesgo de crédito.-** Es la posibilidad de pérdida debido al incumplimiento del prestatario o la contraparte en operaciones directas, indirectas que conlleva el no pago, el pago parcial o la falta de oportunidad en el pago de las obligaciones pactadas.

[3] Disposición emitida en el artículo 2 en la sección I “Alcance y definiciones” en el capítulo I “De la gestión integral y control de riesgos”, del título X “De la gestión y administración de riesgos”, del libro I “Normas generales para las instituciones del Sistema Financiero”, de la Codificación de Resoluciones de la Superintendencia de Bancos y Seguros.

2. **Riesgo de mercado.-** Es la contingencia de que una institución del sistema financiero incurra en pérdidas debido a variaciones en el precio de mercado de un activo financiero.

3. **Riesgo de tasa de interés.-** Es la posibilidad de que las instituciones del sistema financiero asuman pérdidas como consecuencia de movimientos adversos en las tasas de interés pactadas.

4. **Riesgo de tipo de cambio.-** Es el impacto sobre las utilidades y el patrimonio de la institución controlada, por variaciones en el tipo de cambio.

5. **Riesgo de liquidez.-** Es la contingencia de pérdida que se manifiesta por la incapacidad de la institución del sistema financiero para enfrentar una escasez de fondos y cumplir sus obligaciones, y que determina la necesidad de conseguir recursos alternativos.

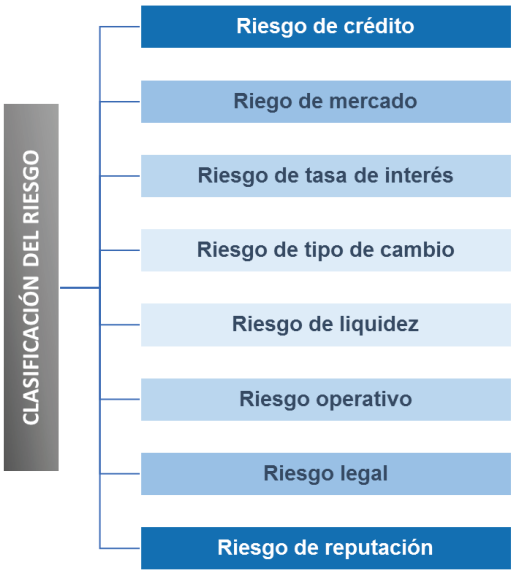

Figura 1.3 Clasificación del riesgo, según la SBS.

6. **Riesgo operativo.-** Es la posibilidad de que se produzcan pérdidas debido a eventos originados en fallas o insuficiencia de procesos, personas, sistemas internos, tecnología, y en la presencia de eventos externos imprevistos.

7. **Riesgo legal.-** Es la probabilidad de que una institución del sistema financiero sufra pérdidas directas o indirectas; de que sus activos se encuentren expuestos a situaciones de mayor vulnerabilidad; de que sus pasivos y contingentes puedan verse incrementados más allá de los niveles esperados, o de que el desarrollo de sus operaciones enfrente la eventualidad de ser afectado negativamente, debido a error, negligencia, impericia o imprudencia, que deriven de la inobservancia, incorrecta o inoportuna aplicación de disposiciones legales o normativas.

8. **Riesgo de reputación.-** Es la posibilidad de afectación del prestigio de una institución del sistema financiero por cualquier evento externo, fallas internas hechas públicas, o al estar involucrada en transacciones o relaciones con negocios ilícitos, que puedan generar pérdidas y ocasionar un deterioro de la situación de la entidad.

Los modelos que se proponen en este proyecto, buscan aportar en la administración del riesgo de crédito, en ese sentido, a continuación se presentan los aspectos más relevantes que giran alrededor de este tema.

1.3.3 ADMINISTRACIÓN DEL RIESGO DE CRÉDITO.

Como se ha mencionado anteriormente, para una correcta administración de riesgo de crédito es importante cumplir con las etapas de identificación, medición, control y monitoreo del riesgo dentro del ciclo de crédito. La SBS reconoce que cada institución financiera tiene su propio perfil de riesgo, según sus actividades y circunstancias

específicas; por lo tanto, al no existir un esquema de administración único, cada institución debe desarrollar un sistema propio para una correcta gestión.

La identificación del riesgo en cada una de etapas del ciclo de crédito está orientado reconocer y entender los riesgos existentes en cada operación efectuada, y aquellos que pueden surgir de nuevas iniciativas. La alta gerencia es la encargada de decidir, entre otros aspectos, su estrategia de negocio, políticas, segmento de mercado objetivo y productos a ofertar a sus clientes o potenciales clientes. Estos procedimientos permiten definir el nivel de riesgo considerado como aceptable para la institución, y partir del cual se empiezan a ejecutar acciones que lo lleven a la práctica.

Una vez identificados los riesgos, es importante cuantificarlos o medirlos, con la finalidad de determinar el cumplimiento de las políticas, los umbrales determinados para el nivel de riesgo y el impacto económico que representa para la institución; de esta manera es posible que los responsables de la administración instauren los controles o correctivos necesarios.

Los sistemas de control implantados en las instituciones financieras permiten constatar la validez del cumplimiento de las estrategias de negocio, políticas y límites de los niveles de riesgo establecidos durante la ejecución de las operaciones de la entidad. Para garantizar que la entidad está administrando adecuadamente el riesgo de crédito, conforme a la política de la institución, debe establecer controles de tipo administrativo y tecnológico (tecnología crediticia[4]) que permitan evidenciar la calidad de la gestión.

La unidad encargada de la administración del riesgo de crédito en la institución, es la responsable de realizar un seguimiento sistemático de las exposiciones de riesgo y de los resultados de las acciones adoptadas, lo que implica un monitoreo permanente a través de un sistema de información para la gestión del riesgo de crédito, que vaya acorde a las necesidades de la institución y que a la vez apoye en los procesos de

[4] Es el término utilizado para referirse al conjunto de procedimientos tecnológicos utilizados durante el ciclo de crédito, enmarcados dentro de la política de concesión crediticia, y encaminados a la minimización del riesgo crediticio. Brindan agilidad, sencillez y satisfacción por el servicio prestado por la entidad.

toma de decisiones, que permita la generación de informes permanentes, mensuales, oportunos, objetivos, relevantes y consistentes.

1.4 PROCESOS DE SEGMENTACIÓN DE POTENCIALES CLIENTES.

En el caso del mercado financiero ecuatoriano, la SBS define la segmentación como una opción de clasificación utilizada para definir, identificar y analizar adecuadamente los grupos de sus clientes y potenciales clientes, en relación con la gestión del riesgo de crédito.

El hecho de contar con una base de datos que contiene a los potenciales clientes de ABC, permite seleccionar a aquellos que forman parte del mercado objetivo de la institución y sobre ese grupo de personas aplicar estrategias comerciales diferenciadas por segmento identificado, permitiendo que la organización minimice el riesgo de originación y los costos relacionados con la captación de clientes nuevos.

1.4.1 PROCESAMIENTO DE INFORMACIÓN POR LOTES.

El desarrollo de las herramientas tecnológicas y metodológicas permite que las instituciones se mantengan en una constante búsqueda de métodos y procedimientos que ayuden a mejorar su posición competitiva en el mercado. Una de estas estrategias tecnológicas se denomina procesamiento de información por lotes (en inglés *batch proccesing*), también conocido como procesamiento en modo *batch*.

En el ámbito comercial y financiero, las herramientas tecnológicas se han convertido en un referente muy importante para buscar nuevos mercados potenciales; en particular, un proceso *batch* le permite a una institución, ejecutar programas complejos sin el control o la supervisión del usuario. Esto no quiere decir que se deja de lado la validación final por parte del mismo, para certificar que los resultados son los esperados; la decisión final al respecto de los resultados finales siempre queda a consideración del usuario y de los expertos en el tema.

Generalmente los procesos en modo *batch* se especifican en distintos pasos de ejecución en un script. Estos procesos se caracterizan por la flexibilidad para incluir reglas de negocios, políticas de la institución y estrategias comerciales; entre los programas más conocidos para la ejecución por lotes, tenemos: *Microsoft SQL Server, Oracle Database, R-Project, Excel*.

La ventaja fundamental de estos sistemas en el caso de la evaluación crediticia para la pre-aprobación de cupos de crédito es que, permite procesar masivamente al lote de potenciales clientes de los que se dispone de información en las bases de datos. Además, se minimiza el riesgo operativo que representa la interacción del usuario con el programa, pues al ser actividades repetitivas podrían generarse errores en alguna de ellas.

Entre las principales actividades de ejecución que se incluyen en un proceso *batch*, para la evaluación crediticia para potenciales clientes y gestión comercial, se citan las siguientes:

- Identificación de los sujetos de crédito, aquellos individuos que se encuentran dentro del mercado objetivo de la institución y cumplen con las políticas tanto internas como de la SBS, para la concesión de créditos.
- Evaluación del perfil de riesgo, por medio de los modelos *scoring* que se utilizan en las diferentes etapas del ciclo de crédito.
- Segmentación de potenciales clientes, tomando como dimensión principal el nivel de riesgo que representan para la institución.
- Estimación de un cupo de crédito, que en el caso de potenciales clientes se gestiona como oferta pre-aprobada, por medio de una plataforma comercial.
- Ubicación geográfica y telefónica, en el caso de que exista, de los potenciales clientes para que puedan ser contactados por los representantes de la plataforma comercial, a través de canales de venta como el *call center*.

1.4.2 CAPTACIÓN DE CLIENTES NUEVOS.

Las entidades financieras son empresas dedicadas a la prestación de servicios, en donde los clientes recurren a la misma para satisfacer sus necesidades de capital, y como toda empresa necesita de ellos para permanecer en el mercado.

Por muchos años, el marketing transaccional[5] ha sido el más común entre las empresas prestadoras de servicios, pero con el paso del tiempo se ha ido dejando de lado ya que es una estrategia muy alejada de la relación con el cliente y sus verdaderas necesidades en un mercado altamente competitivo; en su lugar, las instituciones financieras han optado por implantar el marketing relacional, en donde gracias a las nuevas tecnologías de la información y comunicación es posible mantener relaciones personalizadas de forma masiva, permanente y geográficamente dispersa.

En ese sentido, el *Customer Relationship Management* (CRM[6]) es una forma de administración basada en la relación con el cliente, altamente recomendada para el crecimiento de una entidad financiera, ya que no incurre en gastos de planes de captaciones y además crea una relación personal con los clientes de la institución; desarrollando un conjunto de acciones que den como resultado la fidelización del cliente.

El CRM es la estrategia de negocio enfocada a seleccionar y gestionar los clientes con el fin de optimizar su valor a largo plazo, el mismo que requiere de una filosofía de negocio centrada en el cliente y una cultura de empresa que apoye decididamente los procesos de marketing, ventas y servicio. Las aplicaciones CRM permiten implantar la gestión de la relación con los clientes cuando la empresa tiene el liderazgo, la estrategia y la cultura acertada.

---

[5] El marketing transaccional es aquel que se limita a generar productos y ofertarlos en el mercado, esperando captar clientes susceptibles de comprarlos.

[6] *Customer Relationship Management* es el proceso que consiste en identificar, captar, satisfacer, retener y potenciar (y cuando sea necesario, terminar) relaciones rentables con los mejores clientes de la institución.

El manejo de esta herramienta es un método altamente efectivo para la selección de potenciales clientes y fidelización de los mismos.

La Gerencia de la institución, entre otras actividades, se ha esforzado por lograr captar la mayoría del mercado disponible; es decir, se ha esforzado por satisfacer las necesidades de sus clientes nuevos. Esto lo realiza aportando una buena cantidad de recursos, no solamente monetarios sino también humano y de tiempo. Una entidad financiera también es una empresa que se esfuerza por captar nuevos clientes y satisfacer sus necesidades, que en este caso son de capital y financiamiento, pero qué tan factible es conseguir nuevos clientes desde el punto de vista económico y financiero. El esfuerzo de captar nuevos clientes demanda muchos más recursos que mantener a la cartera vigente de clientes satisfecha, ya que se debe destinar dichos recursos para tareas como:

- Exploración de campo para encontrar clientes potenciales.
- Investigación de mercado, en donde se denoten las necesidades y preferencias que tiene el cliente.
- Planificación estratégica para convertir a los potenciales clientes en clientes reales de la institución.
- Servicios pos desembolso.

Con estos antecedentes, varios autores afirman que el éxito de una organización no consta en captar nuevos clientes, sino en fidelizar la cartera de clientes que ya tiene, debido a que representa una inversión menor de recursos; sin embargo, con la propuesta de este proyecto y como se ha mencionado en secciones anteriores, una correcta segmentación y su automatización en un proceso *batch* ayudan a incrementar el volumen de captación de clientes nuevos; minimizando los costos y mejorando los estándares de segmentación de la institución.

1.5 PLANTEAMIENTO DEL PROBLEMA.

Con el desarrollo de las herramientas tecnológicas, ABC ha logrado almacenar información socio-demográfica y del buró de créditos de potenciales clientes. Estos recursos permiten el procesamiento de información por lotes o en modo *batch*, el cual tiene como propósito identificar a aquellas personas con las que ABC podría tener una oportunidad de negocio, es decir potenciales clientes con condiciones favorables para el otorgamiento de un crédito. Una vez que se ha determinado el segmento potencial de clientes, se procede a fijar un cupo de crédito para que sea ofertado por la plataforma comercial a través de estrategias como visitas personalizadas o llamadas por *call center*.

Como se ha mencionado anteriormente, por disposición de la SBS, en la evaluación crediticia de potenciales clientes, las instituciones financieras deben utilizar metodologías de aprobación y otorgamiento de créditos. Es así que, previo a la asignación de cupos y demás condiciones crediticias, el marco regulatorio exige evaluar el perfil de riesgo así como la capacidad financiera del solicitante, usando la información que consta en el reporte del buró de créditos y en la solicitud de crédito. La normativa muestra claramente que no es suficiente evaluar el nivel de riesgo de un solicitante sino que también es importante el análisis de la capacidad financiera para disminuir el riesgo de contrapartida debido a un posible sobre endeudamiento.

Para créditos de consumo, la metodología recomendada por la SBS para evaluar la capacidad financiera y asignar el cupo se basa fundamentalmente en el ingreso neto mensual promedio disponible, la normativa vigente señala:

> *La estimación del ingreso neto mensual promedio disponible se la realizará con las siguientes consideraciones: (incluido con resolución No. JB-2012-2217 de 22 de junio del 2012)*
>
> *1.2.2.1. Ingreso neto mensual = Ingreso promedio mensual – gasto mensual promedio – cuota mensual estimada promedio que consta en el reporte del buró de información crediticia.*

1.2.2.2. La institución del sistema financiero para estimar el ingreso mensual promedio de los potenciales clientes crediticios deberá requerir los documentos de soporte respectivos que evidencie el ingreso real mensual del cliente.

1.2.2.3. El gasto mensual promedio estará constituido al menos por los siguientes rubros: alimentación, vivienda, servicios básicos, vestimenta, transporte, salud, educación. Las instituciones del sistema financiero deberán solicitar la documentación que respalde el nivel de gastos o de ser el caso analizar la razonabilidad de los gastos declarados por el cliente. [7]

Es decir, la normativa vigente obliga a todas las instituciones del sistema financiero a atender dos frentes para evaluar la capacidad financiera de un solicitante:

a) Conocer como información fundamental el salario y gastos mensuales del solicitante para la aprobación del cupo de crédito.
b) Solicitar los documentos de soporte para verificar la fuente de los ingresos.

Ambas acciones se pueden llevar a cabo cuando empieza la negociación entre ABC y el cliente nuevo, a través de la solicitud de crédito y la revisión de la documentación presentada por el solicitante; pero en el proceso *batch* de ABC para la pre-aprobación de cupos de crédito, el salario es un dato incierto o desconocido dado que se tratan de no clientes de la institución; para solventar este inconveniente, ABC asigna el cupo de crédito con la información disponible dejando de lado a las personas que no la tienen. En cuanto a la verificación de la documentación se la realiza antes del desembolso, comprobando previamente que el cupo no ha sido sobre estimado en el proceso de pre-aprobación, en el caso de que el cupo estimado en el *batch* supere al valor planteado según la metodología recomendada por la SBS, el asesor de crédito procede a negociar con el cliente las condiciones del crédito tal como el plazo y la

[7] Disposición emitida en el numeral 1.2.2. Metodologías y/o sistemas internos de calificación de créditos de consumo, del artículo 5 en la sección II "Elementos de la calificación de activos de riesgo y su clasificación" en el capítulo II "Calificación de activos de riesgo y constitución del provisiones por parte se las instituciones controladas por la Superintendencia de Banco y Seguros", del título IX "De los activos y de los límites del créditos", del libro I "Sistema Financiero", de la Codificación de Resoluciones de la Superintendencia de Bancos y Seguros y de la Junta Bancaria.

cuota mensual, incentivando de esa forma a que la persona concrete el negocio con ABC.

La metodología utilizada actualmente por ABC para la asignación de cupos por proceso *batch* obedece a lo expuesto en el marco regulatorio:

> *Si las instituciones del sistema financiero no presentaren sus metodologías para ser evaluadas o si éstas no cumplieren con los requisitos que establezca la Superintendencia de Bancos y Seguros, deberá considerar como límite máximo de exposición en sus operaciones de financiamiento de consumo, que los dividendos o cuotas mensuales pactados por éstas, no sobrepasen el 50% del ingreso neto mensual promedio del deudor.*[8]

Es claro ver que la limitación fundamental en el proceso *batch* actual de ABC es que la pre-aprobación de cupos de crédito está orientada únicamente hacia los individuos que disponen información del salario, dejando de lado a aquellos que no la tienen; por tales motivos se ve la necesidad de crear nuevas metodologías que permitan optimizar el uso de los recursos de información que dispone ABC sobre potenciales clientes.

En este proyecto se plantea realizar una evaluación crediticia de los potenciales clientes de ABC, con la ayuda de modelos estadísticos y analíticos que permitan:

a) Evaluar el perfil de riesgo.
b) Asignar un cupo de crédito en la cartera de consumo.

Ambas acciones ejecutadas dentro del contexto de la pre-aprobación *batch* de ABC.

La propuesta consiste en evaluar el perfil de riesgo con la ayuda de un modelo de regresión logística, que obtiene como resultado la probabilidad de que un individuo incumpla con las obligaciones financieras que ha adquirido en ABC; en lo que se refiere a la asignación del cupo de crédito, se propone hacerlo mediante un modelo de

---

[8] Disposición emitida en el numeral 1.2.2. Metodologías y/o sistemas internos de calificación de créditos de consumo, del artículo 5 en la sección II "Elementos de la calificación de activos de riesgo y su clasificación" en el capítulo II "Calificación de activos de riesgo y constitución del provisiones por parte se las instituciones controladas por la Superintendencia de Banco y Seguros", del título IX "De los activos y de los límites del créditos", del libro I "Sistema Financiero", de la Codificación de Resoluciones de la Superintendencia de Bancos y Seguros y de la Junta Bancaria.

inferencia difusa que permita explicar la forma en que los potenciales clientes de ABC administran su deuda con la competencia y a partir de ello, estimar un cupo a ofertar por el *call center*, tomando en cuenta las limitaciones de la variable salario del proceso *batch*.

La metodología que se plantea en este proyecto necesariamente debe articularse al procedimiento actual que sigue ABC para el cumplimiento de la normativa; es decir, previo al desembolso del dinero se debe llenar la solicitud de crédito y los documentos de soporte para que el asesor de crédito verifique la validez de los documentos presentados. Con estas acciones, ABC completa las dos etapas del ciclo de crédito para potenciales clientes: preventa y análisis. En el proceso *batch* se implementan las metodologías estadísticas y matemáticas necesarias para una correcta segmentación de los individuos que se encuentran dentro del mercado objetivo de ABC, estas acciones son propias de la etapa de preventa; mientras que, como parte de la etapa de análisis se consideran los procedimientos y políticas de la institución para la verificación de los documentos de soporte presentados por el cliente nuevo.

Para una mejor comprensión del procedimiento que sigue ABC, para la pre-aprobación y concesión de créditos, el lector puede revisar el diagrama de flujo que se presenta en la figura 1.4. Un aspecto importante a resaltar en el diagrama de flujo es que, en la etapa de análisis se canalizan por el *call center* a las personas que cuentan con un cupo pre-aprobado; en cambio, a las que llegan por sus propios medios deben continuar por otro canal. La diferencia fundamental es que el tiempo de respuesta es menor para aquellos que cuenta con un cupo de crédito desde el proceso *batch*.

La implementación de las metodologías propuestas permite que ABC mejore su posición competitiva al utilizar herramientas técnicas que mejoran la administración de la información almacenada de potenciales clientes; a su vez, minimizan el tiempo de atención a los clientes nuevos que cuentan una cupo de crédito pre-aprobado desde el proceso *batch* y son contactados por el *call center*.

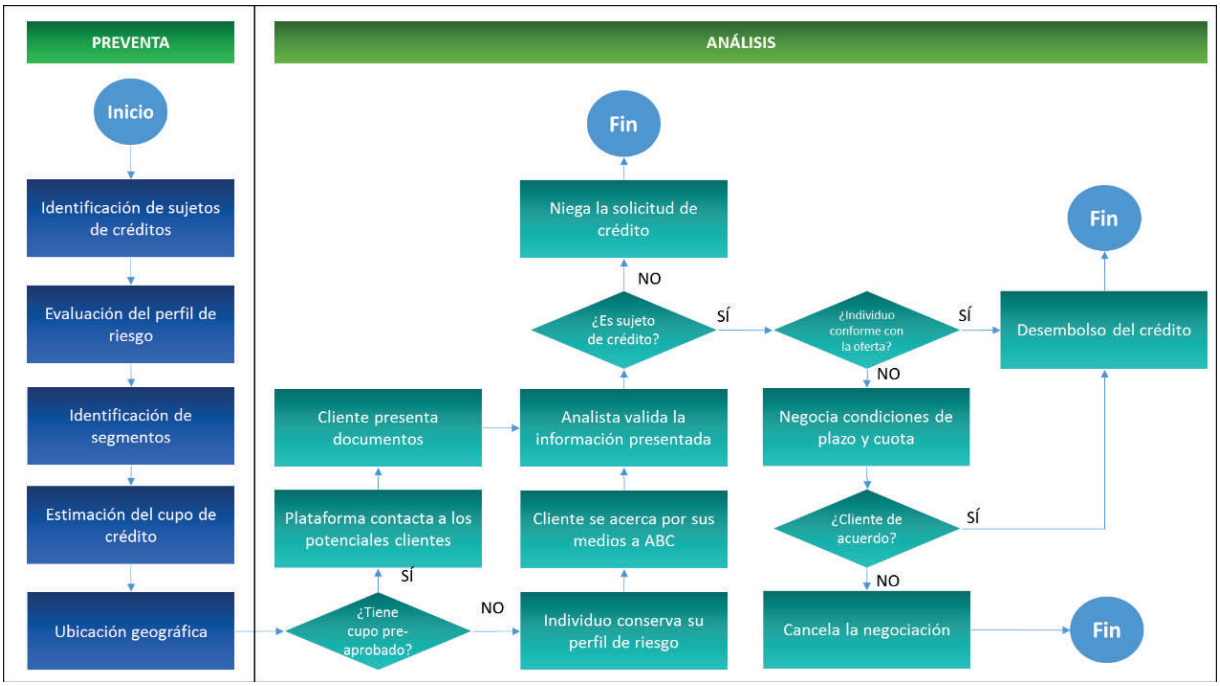

Figura 1.4 Etapas de preventa y análisis del ciclo de crédito de ABC

1.6 JUSTIFICACIÓN DEL TRABAJO.

En este trabajo se presenta una metodología que permite efectuar una evaluación crediticia de potenciales clientes de ABC, integrando al análisis del perfil de riesgo del individuo y la estimación de un cupo de crédito con el propósito de que sean gestionados como ofertas pre-aprobadas, siguiendo los procedimientos de análisis y otorgamiento implantados por ABC para el cumplimiento de la normativa emitida por la SBS. Es decir, la propuesta se basa en la integración de instrumentos estadísticos y matemáticos, así como el criterio de los expertos de la institución en el tema de concesión de créditos, en un modelo de inferencia difuso.

1.6.1 MODELOS SCORING DE ORIGINACIÓN.

Al definir una variable endógena binaria, que distingue entre clientes nuevos buenos y malos, la regresión logística se convierte en una herramienta estadística útil para la

evaluación del perfil de riesgo de un individuo antes del otorgamiento de su primera operación crediticia. La metodología permite obtener, para cada individuo de la muestra de estudio, la probabilidad de que el crédito concedido por una entidad bancaria será devuelto en su fecha de su vencimiento [2].

El trabajo consiste en tomar como referencia a los clientes de ABC que recibieron su primer crédito y a partir de ello, evaluar su comportamiento de pago y clasificarlos como buenos o malos para estimar el modelo de regresión logística. Es decir, en base al comportamiento que han tenido los clientes nuevos de la institución, se busca estimar la probabilidad de que los potenciales clientes que forman parte del mercado objetivo de ABC, cumplan o incumplan con los pagos del crédito concedido, o sea que en un periodo hacia delante puedan ser categorizados como buenos o malos clientes.

Con el resultado que se obtiene del modelo logístico es posible, segmentar el lote de potenciales clientes que dispone ABC en las bases de datos almacenadas, en función de su semejanza con los clientes actuales y el riesgo asociado a ellos; permitiendo fijar los límites de riesgo que asume la institución en la originación de sus clientes nuevos.

En un modelo de elección binaria se pretende explicar la preferencia entre dos alternativas posibles, es así que la variable dependiente puede tomar dos valores: $Y_i = \{malo(0), bueno(1)\}$. La elección entre los dos valores se puede explicar en función de:

a) Las variables que caracterizan a los individuos y que se denotan por el vector de dimensión K, $\boldsymbol{x}_i$; y,
b) Un término escalar de error, u_i, entre los valores observados de Y_i y sus valores previstos [2].

El modelo logit surge cuando se usa la función de distribución logística para representar la probabilidad de que un individuo sea calificado con la opción $Y_i = 1$. La fórmula que describe la función logística, se expresa por:

$$F(z) = \frac{e^z}{1 + e^z}, \quad -\infty < z < \infty \tag{1.1}$$

Entonces, la probabilidad de elegir la opción $Y_i = 1$, es igual a:

$$P_i = P(Y_i = 1) = F(z) = \frac{e^z}{1 + e^z} \tag{1.2}$$

Donde,

$$z = \boldsymbol{x}_i^T \boldsymbol{\beta} = \beta_0 + \sum_{k=1}^{K-1} x_{ik}\beta_k \tag{1.3}$$

El vector $\boldsymbol{\beta}$ contiene los K parámetros que se estiman en el modelo.

De este modo, la estimación de la probabilidad de escoger la opción $Y_i = 1$, se plantea como:

$$p_i = P_i + u_i = \frac{e^z}{1 + e^z} + u_i \tag{1.4}$$

Por obvias razones, la variable dependiente binaria de la regresión logística define una partición de la muestra en donde el número de observaciones en uno de los grupos es mucho más pequeño que en el otro. Por tal motivo, en la estimación del modelo logístico es preciso solventar el problema de muestras desproporcionadas, a fin de corregir el desplazamiento generado en la curva logística [3].

En el Capítulo 2, se presenta la metodología utilizada para estimar los parámetros del modelo y corregir el desplazamiento de la curva logística; así mismo se exponen las pruebas de consistencia para valorar la calidad de discriminación del modelo.

1.6.2 MODELO DE INFERENCIA DIFUSA.

El modelo de inferencia difuso, permite incluir el resultado del análisis del perfil de riesgo así como variables de endeudamiento actual e histórico para explicar la forma en que el cliente potencial del ABC administra su deuda con la competencia y a partir de ello sugerir un cupo de crédito a ofertar como pre-aprobado, siguiendo los lineamientos de la institución que orientan al cumplimiento de la normativa de la SBS.

Zadeh (1965) sienta las bases sobre la teoría formal de conjuntos difusos. La Lógica Difusa actualmente está relacionada y fundamentada en la teoría de los conjuntos difusos, según la cual, el grado de pertenencia de un elemento a un conjunto está determinado por una función de pertenencia que puede tomar todos los valores reales comprendidos en el intervalo (0,1) [4]. Años más tarde, Mandani (1977) extiende el concepto de conjunto difuso a sistemas de lógica difusa, los cuales son estudiados en este trabajo, como parte del proceso de evaluación crediticia de ABC.

Con la Lógica Difusa se abre la posibilidad de dar solución a problemas expresados desde la perspectiva humana y que, por ésta condición, no pueden tener una solución única desde lo “falso” o “verdadero” sino pueden tomar condiciones intermedias para dar soluciones satisfactorias a los problemas planteados [4]. En cambio, las técnicas de la Lógica Clásica utilizan herramientas para codificar las variables cualitativas y cuantitativas, los mismos asignan un umbral a partir del cual se define la pertenencia o no entre un conjunto y otro. A manera de ejemplo, se puede citar el caso de la variable edad, que por política de ABC, un individuo es calificado como sujeto de crédito si su edad es mayor o igual a 25 años; desde el punto de vista de la lógica clásica, los casos cercanos al umbral definido por la institución pasan de manera abrupta de pertenecer de un conjunto a otro; es decir una persona que en el mes entrante cumple los 25 años, hoy mismo no es un sujeto de crédito. En la Lógica Difusa, el pasar de un estado a otro es gradual; en términos lingüísticos el individuo del ejemplo debería estar más asociado al término “es sujeto de crédito” que a “no es sujeto de crédito”; de esta manera, se rompe con el principio de la no contradicción o del tercero excluido, pues el individuo objeto de análisis puede pertenecer a la vez a conjuntos contrarios en cierto grado.

La Lógica Difusa ha mantenido diversas aplicaciones, una de ellas es que el conocimiento experto permite la automatización de tareas [5]. En el caso de la evaluación crediticia de ABC, se plantea construir una base de conocimientos sobre la base de la opinión de expertos y políticas de la institución, con ello construir un modelo de inferencia difusa para introducirlo en el proceso *batch* de pre-aprobación de cupos de crédito.

Al respecto de la evaluación crediticia, varios autores han formulado modelos de inferencia difusa con la finalidad de analizar el nivel de solvencia de los solicitantes de créditos; es así que, una evaluación crediticia integral basada en técnicas estadísticas y analíticas genera beneficios para las instituciones financieras; tal como lo expone [4] en un estudio realizado para analizar el nivel de solvencia de los clientes de una cartera de microcrédito, los beneficios que expone son:

a) Proporcionan mucha más imparcialidad en la concesión de créditos a los clientes de la institución, pues incorpora al modelo las mejores prácticas anteriores de la institución en la evaluación crediticia.
b) Es una herramienta de apoyo a los especialistas, desarrollada de manera sencilla y considerando que el conocimiento previo de los expertos es una metodología que permite la minimización del riesgo de crédito.

De manera similar, [1] estudia la capacidad de pago de los clientes de una institución financiera, por medio de un modelo de inferencia difusa, a partir del cual permite citar algunas bondades en la evaluación de crediticia:

a) Permiten incluir el criterio de los expertos en el tema, durante la etapa de construcción del modelo. Esto se logra al momento de definir las relaciones entre las variables de entrada y salida, logrando definir la base de conocimiento que representa por una parte, el entendimiento que los expertos tienen del fenómeno y por otra, sus sistemas de razonamiento.
b) Permiten optimizar la gestión de la información almacenada en las bases de datos.
c) Permiten determinar la calidad financiera de los usuarios y su capacidad de cumplir con el contrato de pago.
d) Permiten tomar decisiones sobre las estrategias de otorgamiento de créditos a determinados segmentos de potenciales clientes, evaluando fundamentalmente el nivel de riesgo que asume la institución con ellos y el posicionamiento deseado dentro de un mercado objetivo; ambos criterios orientados a incrementar la rentabilidad del negocio.

Bajo el punto de vista de la lógica financiera, es natural suponer que si una persona mantiene una deuda con la competencia, cumple con los requerimientos exigidos por la SBS para ser calificado como un sujeto de crédito. Adicional, el comportamiento de pago y tanto el endeudamiento actual como histórico proporcionan información de la forma en que los clientes nuevos de ABC administran su deuda con la competencia, al momento en el que recibe su primer crédito en la institución. Partiendo de ese entendimiento, el modelo de inferencia difusa busca sugerir un monto de crédito para que ABC lo gestione a través de sus diferentes canales de venta.

En el capítulo 3 se explica la metodología desarrollada para obtener el modelo de inferencia difusa; cabe mencionar que el resultado obtenido con la regresión logística se incluirá como una variable del modelo, esto con la finalidad de elaborar estrategias comerciales diferenciadas por nivel de riesgo.

1.7 HIPÓTESIS DEL TRABAJO.

La construcción de un modelo de inferencia difuso permite interpretar la forma en que los potenciales clientes de ABC administran su deuda con la competencia; a través de la información de endeudamiento y cumplimiento de sus pagos que consta en el reporte del buró de créditos, lo que permite evaluar el nivel de riesgo y la capacidad de endeudamiento que ha demostrado históricamente.

Comprender la forma en que una persona administra su deuda en otras instituciones financieras, le permite a ABC, ofertar un cupo pre-aprobado a sus potenciales clientes por los canales comerciales adecuados, mismos que están sujetos a las políticas internas de otorgamiento de créditos y que a la vez conducen al cumplimiento de la normativa emitida por la Superintendencia de Bancos y Seguros.

1.8 OBJETIVOS DEL ESTUDIO.

Objetivo general.

Desarrollar una metodología para la asignación de cupos de crédito a potenciales clientes de ABC en un escenario de pre-aprobación, tomando en cuenta el perfil de riesgo e información histórica de endeudamiento en el sistema financiero

Objetivos específicos.

a) Asignar un perfil de riesgo a potenciales clientes de ABC.
b) Estimar cupos de crédito para que sean gestionados como pre-aprobados por el *call center*.
c) Comparar la metodología recomendada por la SBS para el cálculo de cupos de crédito con la propuesta, de tal manera que se pueda medir la bondad del modelo de inferencia difusa.

CAPÍTULO 2
ESTIMACIÓN DEL PERFIL DE RIESGO.

La metodología escogida en este proyecto para determinar el perfil de riesgo de un cliente nuevo es la regresión logística binaria, por su facilidad de interpretar la probabilidad de que un cliente potencial de ABC incumpla sus obligaciones financieras.

En este capítulo se desarrolla la metodología y se presentan los principales resultados que se han obtenido para discriminar los grupos de potenciales clientes, buenos y malos.

En ese sentido se plantean los siguientes puntos a desarrollar:

- Definición del universo de estudio para el proyecto y variables solicitadas en el requerimiento de información.
- Elección de la muestra de estudio.
- Definición del tipo de cliente (bueno o malo), que se convierte en la variable dependiente del estudio.
- Análisis descriptivo de las variables construidas para este proyecto.
- Identificación del poder discriminante de las variables regresoras.
- Estimación del modelo de regresión logística binaria.
- Elección del modelo óptimo.
- Valoración de la calidad de discriminación del modelo.

Previo al desarrollo de las actividades propuestas se ha realizado un requerimiento de información, el mismo que es trabajado por la Unidad de Sistemas de Información de ABC, mientras que la certificación de la validez y consistencia forman parte de las tareas de este proyecto de manera conjunta con la Unidad de Auditoría de ABC, esto con la finalidad de garantizar la confiabilidad de la información.

2.1 DEFINICIÓN DEL UNIVERSO DE INDIVIDUOS Y VARIABLES.

Para este estudio se ha escogido como universo de análisis a todos los clientes nuevos de ABC. Un individuo se denomina cliente nuevo cuando recibe por primera vez una operación crediticia en la institución. La Superintendencia de Bancos y Seguros del Ecuador (SBS) sugiere, como parte del marco regulatorio, que las metodologías para la medición del riesgo de crédito sean desarrolladas con datos históricos de mínimo 3 años[9]. En ese sentido, se ha tomado información de los individuos a quienes se otorgó su primer crédito durante el periodo comprendido entre los meses de abril 2012 y marzo 2015, a continuación el detalle de los registros obtenidos.

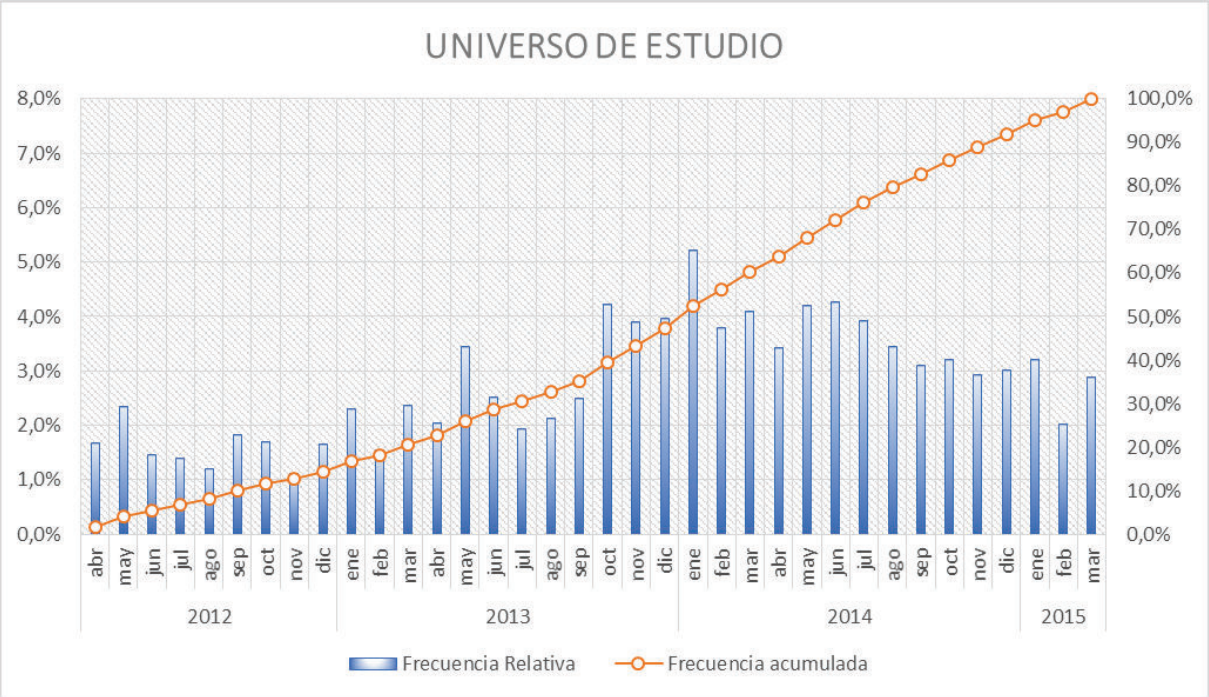

Figura 2.1 Frecuencia de clientes nuevos por mes de estudio

[9] Disposición emitida en el numeral 7.3 del artículo 7 en el capítulo II "De la administración de riesgo de crédito", del título X "De la gestión y administración de riesgos", del libro I "Sistema Financiero", de la Codificación de Resoluciones de la Superintendencia de Bancos y Seguros.

Tabla 2.1 Frecuencia de clientes nuevos de ABC, por mes de estudio

Año	Mes	Frecuencia	Frecuencia Relativa	Frecuencia acumulada
2012	abr	500	1,7%	1,7%
	may	703	2,3%	4,0%
	jun	437	1,5%	5,5%
	jul	418	1,4%	6,9%
	ago	362	1,2%	8,1%
	sep	545	1,8%	9,9%
	oct	510	1,7%	11,6%
	nov	335	1,1%	12,7%
	dic	494	1,6%	14,4%
2013	ene	690	2,3%	16,7%
	feb	461	1,5%	18,2%
	mar	713	2,4%	20,6%
	abr	615	2,1%	22,6%
	may	1.035	3,5%	26,1%
	jun	754	2,5%	28,6%
	jul	579	1,9%	30,5%
	ago	636	2,1%	32,6%
	sep	748	2,5%	35,1%
	oct	1.269	4,2%	39,4%
	nov	1.172	3,9%	43,3%
	dic	1.189	4,0%	47,3%
2014	ene	1.562	5,2%	52,5%
	feb	1.138	3,8%	56,3%
	mar	1.229	4,1%	60,4%
	abr	1.025	3,4%	63,8%
	may	1.258	4,2%	68,0%
	jun	1.278	4,3%	72,2%
	jul	1.178	3,9%	76,2%
	ago	1.034	3,4%	79,6%
	sep	930	3,1%	82,7%
	oct	964	3,2%	85,9%
	nov	878	2,9%	88,9%
	dic	901	3,0%	91,9%
2015	ene	963	3,2%	95,1%
	feb	605	2,0%	97,1%
	mar	868	2,9%	100,0%
TOTAL		**29.976**	**100,0%**	

En la tabla 2.1 y figura 2.1 se puede apreciar, por ejemplo, que en el mes de abril de 2012 se ha otorgado 500 operaciones a clientes nuevos de ABC y representan el 1,7% de los datos de análisis.

Con respecto a las variables solicitadas en el requerimiento de información, en total se definen 92, clasificadas en cuatro grupos:

1. Variables informativas de la matriz de datos, tales como fecha de otorgamiento de la operación, número de operación e identificación del cliente.
2. Variables relativas al comportamiento de pago del cliente nuevo de ABC, que básicamente son de utilidad al momento de definir la variable dependiente del modelo: atraso máximo y promedio de los 12 meses posteriores al desembolso, monto y provisión del crédito otorgado.
3. Variables socio demográficas, que describen información del cliente como edad, género, provincia de domicilio, etc. En total se contabilizan 27 variables dentro de este grupo.
4. Variables con información del buró de créditos, recogen el comportamiento del cliente en el sistema financiero previo a la relación con ABC. En este grupo se definen 58 variables.

En el anexo I se detalla la definición de todas las variables citadas anteriormente.

2.2 ELECCIÓN DE LA MUESTRA DE ESTUDIO.

Para determinar la muestra de estudio, sobre el universo de individuos identificado en el punto anterior se han tomado en cuenta criterios de madurez de la cartera y representatividad de la muestra.

2.2.1 MADUREZ DE LA CARTERA.

En este punto se debe garantizar que el tiempo transcurrido luego de la fecha de desembolso del crédito sea el suficiente para medir características de comportamiento de pago de un cliente nuevo de la institución, esto debido a que las mejores prácticas demuestran que los indicadores de pérdida y morosidad tienden a estabilizarse a partir de un determinado periodo posterior al otorgamiento.

ABC considera que la cartera es madura a partir de los 12 meses posteriores al desembolso de la operación. Aplicado el criterio sobre el universo de individuos de la

tabla 2.1, se selecciona como cartera madura al periodo comprendido desde el mes de abril 2012 hasta marzo 2014.

2.2.2 REPRESENTATIVIDAD DE LA MUESTRA.

La muestra de estudio seleccionada debe ser representativa, de tal manera que las inferencias realizadas a partir de ella tengan un mayor grado de confiabilidad y permitan caracterizar a la población de potenciales clientes nuevos de ABC.

En este proyecto no se ha aplicado una técnica de muestreo para seleccionar a los sujetos que intervienen en el estudio, sino que se ha decidido trabajar con todos los individuos identificados en la tabla 2.1. Esto con la finalidad de que las estimaciones e inferencias puedan hacerse con un mayor nivel de confianza, además que el número de elementos presentes en el estudio no representa un problema al momento de trabajar con los paquetes estadísticos.

En resumen, considerando los criterios de madurez y representatividad, la muestra de estudio seleccionada en este proyecto se compone de todos los clientes nuevos de ABC a los que se les otorgó un crédito entre los meses de abril 2012 y marzo 2014, lo que representa un total de 18.094 individuos que equivale al 60,4% del universo identificado en la tabla 2.1.

2.2.3 MUESTRAS DE APRENDIZAJE Y TESTING.

La muestra seleccionada a partir de los criterios revisados en 2.2.1 y 2.2.2 se clasifica en dos grupos: uno para la construcción del modelo al que se denomina muestra de aprendizaje, y otro que se usa para el control del modelo denominado muestra de testing.

Para la muestra de aprendizaje se toma aproximadamente el 60% de los 18.094 individuos a los que se otorgó una operación entre los meses abril 2012 y septiembre 2013; mientras que, la muestra de testing corresponde al 40% restante, entre los

meses octubre 2013 y marzo 2014. Bajo estas consideraciones, las muestras se conforman tal como se observa en la tabla 2.2.

Tabla 2.2 Muestras de aprendizaje y testing

Muestra	Periodo	Frecuencia	Frec. Relativa
Aprendizaje	Abr.2012 - Sep.2013	10.535	58,2%
Testing	Oct.2013 – Mar.2014	7.559	41,8%
Total		**18.094**	**100,0%**

2.3 ELECCIÓN DE LA VARIABLE DEPENDIENTE DEL MODELO.

En esta etapa se busca una partición sobre la muestra de aprendizaje, de tal manera que se pueda identificar a los clientes buenos y malos de la institución. A partir esta definición se obtiene la variable dependiente de la regresión logística, los resultados se extenderán a la muestra de testing.

2.3.1 INDICADOR DE PÉRDIDA POR OPERACIÓN.

Previo a identificar a los clientes buenos y malos se hace una introducción a la calificación de activos de riesgo[10] y constitución de provisiones[11], dispuesto por la SBS.

Para carteras de consumo, la SBS ha definido la calificación de los créditos en base al rango de morosidad que presentan, los parámetros se definen en la tabla 2.3.

[10] Disposición emitida en el numeral 1.2.1 del artículo 5 en la sección II “Elementos de la calificación de activos de riesgo y su clasificación” en el capítulo II “Calificación de activos de riesgo y constitución del provisiones por parte se las instituciones controladas por la Superintendencia de Banco y Seguros”, del título IX “De los activos y de los límites del créditos”, del libro I “Sistema Financiero”, de la Codificación de Resoluciones de la Superintendencia de Bancos y Seguros y de la Junta Bancaria.

[11] Disposición emitida el artículo 6 en la sección III “Constitución de provisiones” del mismo capítulo, título y libro de la Codificación de Resoluciones de la Superintendencia de Bancos y Seguros, mencionados en la nota anterior.

Tabla 2.3 Calificación de créditos de consumo según la SBS

CATEGORÍAS	DÍAS DE MOROSIDAD
A1	0
A2	1-8
A3	9-15
B1	16-30
B2	31-45
C1	46-70
C2	71-90
D	91-120
E	+120

A partir de la calificación asignada a cada operación, según lo explicado en la tabla 2.3, se procede a la constitución de provisiones del monto expuesto o saldo de la operación al momento de realizar la calificación del activo. En la tabla 2.4 se presentan los porcentajes de pérdida sobre el monto expuesto que establece la normativa de la SBS.

Tabla 2.4 Constitución de provisiones de créditos de consumo

CATEGORÍAS	% PROVISIÓN	
	Mínimo	Máximo
A1	1%	
A2	2%	
A3	3%	5%
B1	6%	9%
B2	10%	19%
C1	20%	39%
C2	40%	59%
D	60%	99%
E	100%	

Apalancado en las definiciones de las tablas 2.3 y 2.4, ABC construye su indicador de pérdida por operación como la relación entre la provisión máxima y el monto total del crédito otorgado a un cliente nuevo, matemáticamente se expresa por:

$$Pérdida = \frac{Provisión}{Monto} \tag{2.1}$$

2.3.2 DEFINICIÓN DE LA VARIABLE DEPENDIENTE.

Una vez definido el indicador de pérdida, se procede al análisis de los siguientes aspectos:

a) La peor calificación asignada al crédito del cliente nuevo durante los 12 meses posteriores al desembolso; y,
b) El promedio de los días de mora que ha tenido la operación en la misma ventana de tiempo considerada en el punto a.

En la tabla 2.5 se observa un resumen descriptivo de la variable peor calificación de los créditos otorgados a los clientes nuevos de ABC, conforme a la información de la muestra de aprendizaje. El porcentaje de pérdida se calcula a partir de la relación (2.1), aplicada sobre los créditos contenidos en cada calificación; de igual menara para conocer la pérdida general de ABC, la relación (2.1) se aplica sobre el total de créditos considerados en la muestra, es por esta razón que la suma del porcentaje de pérdida no es 100%.

En ese sentido, se puede apreciar que la pérdida global de la cartera de consumo de ABC, considerando la información de la muestra de aprendizaje, es 6,9%. Además, se puede apreciar que desde la calificación A1 hasta B2 no superan la pérdida global de ABC y concentran el 82,0% de los clientes nuevos, por lo que podría considerarse un punto de corte para la variable dependiente.

Tabla 2.5 Peor calificación de clientes nuevos en la muestra de aprendizaje

Peor calificación	Frecuencia	%	% Acumulado	% Pérdida
A1	1.784	16,9%	16,9%	0,2%
A2	3.482	33,1%	50,0%	0,2%
A3	1.440	13,7%	63,7%	0,7%
B1	1.280	12,1%	75,8%	1,6%
B2	652	6,2%	82,0%	3,5%
C1	550	5,2%	87,2%	9,3%
C2	320	3,0%	90,3%	18,8%
D	314	3,0%	93,2%	30,3%
E	713	6,8%	100,0%	61,3%
Total general	**10.535**	**100,00%**		**6,9%**

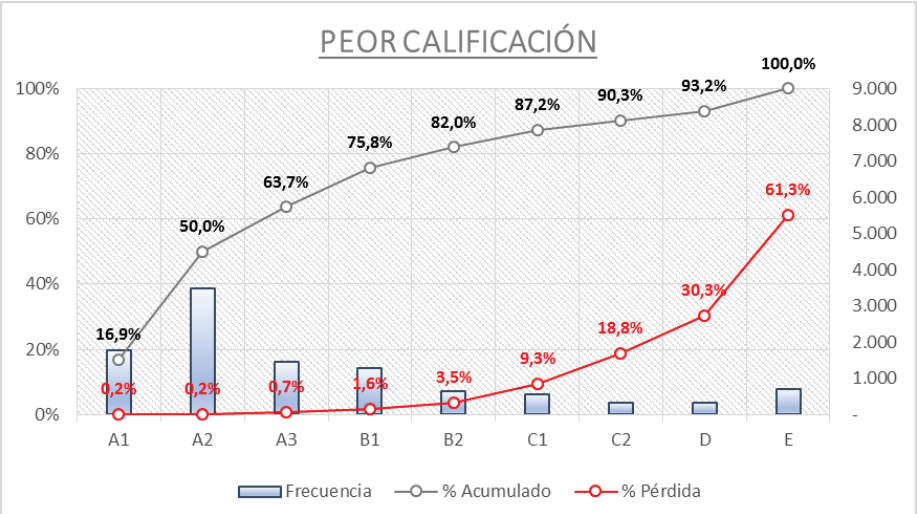

Figura 2.2 Frecuencia y pérdida de la variable peor calificación de clientes nuevos

La figura 2.2 evidencia que las calificaciones que representan la menor morosidad tienen la menor pérdida y a su vez concentran la mayor parte de la información, en cambio para las calificaciones que representan morosidades mayores sucede lo contrario. A manera de ejemplo, las categorías A1 y A2 tienen una pérdida de 0,2% y juntas representan el 50% de los datos.

Tabla 2.6 Promedio de los días de mora de los créditos otorgados a los clientes nuevos

RangoAtrasoProm12M	Frecuencia	%	% Acumulado	% Pérdida
(1) 0	3.814	36,2%	36,2%	0,2%
(2) 1-5	3.496	33,2%	69,4%	0,9%
(3) 6-10	903	8,6%	78,0%	4,2%
(4) 11-15	539	5,1%	83,1%	14,8%
(5) 16-20	363	3,4%	86,5%	18,7%
(6) 21-25	268	2,5%	89,1%	18,0%
(7) 26-30	191	1,8%	90,9%	24,0%
(8) 31-40	315	3,0%	93,9%	28,9%
(9) 41-50	193	1,8%	95,7%	36,7%
(10) 51-60	136	1,3%	97,0%	37,9%
(11) 61-90	188	1,8%	98,8%	47,3%
(12) 91-120	71	0,7%	99,4%	63,2%
(13) >120	58	0,6%	100,0%	76,7%
Total general	**10.535**	**100,00%**		**6,9%**

Con la finalidad de ampliar el análisis del comportamiento de pago del cliente para definir la variable dependiente, en la tabla 2.6 se presenta un nuevo resumen descriptivo, tomando en cuenta la variable promedio de los días de mora.

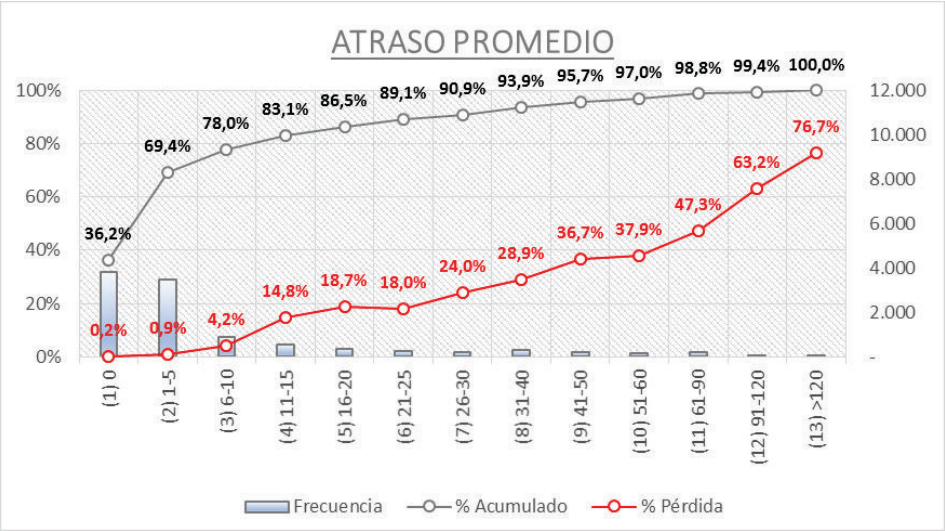

Figura 2.3 Promedio de los días de mora de los créditos otorgados a los clientes nuevos

La tabla 2.6 ratifica que la pérdida general de ABC es 6,9% y hasta el rango 6 a 10 días de mora no se supera ese valor, por lo que podemos asumir que es un nuevo punto de corte para definir la variable dependiente del modelo.

Con los resultados de las tablas 2.5 y 2.6 se puede afirmar que un buen cliente es aquel individuo cuya peor calificación es mejor o igual a B2 y el promedio de los días de mora es menor o igual a 10; sin embargo al revisar el nivel de pérdida por medio de una tabla de doble entrada que contiene las variables peor calificación asignada y promedio de los días de mora se modifica la definición.

La tabla 2.7 muestra la definición final de la variable dependiente del modelo, para lo cual se han realizado las siguientes consideraciones:

- El umbral de la peor calificación permitida para el cliente bueno ha sido reducido a B1, pues las prácticas anteriores de la institución no permiten aceptar moras

superiores a 30 días para calificarlos como buenos, pese a que el grupo con peor calificación B2 no supera la pérdida global de ABC.

- Los clientes inicialmente identificados como buenos y cuya peor calificación es B2 se consideran como indeterminados, pues su nivel de pérdida es pequeño pero las políticas de la institución no permiten categorizarlos como buenos.
- El límite para el promedio de los días de mora se mantiene en 10. La condición permite restringir el carácter de buen cliente a aquellos individuos cuya peor calificación es A3 o B1 (mora entre 8 y 31 días); pues si son reincidentes en moras superiores a 10 días durante los 12 meses posteriores al desembolso, según sea el caso, adquieren la característica de cliente indeterminado.

Tabla 2.7 Pérdida de ABC respecto a la peor calificación y promedio de los días de mora

RangoAtrasoProm12M	Peor calificación									
	A1	A2	A3	B1	B2	C1	C2	D	E	Total general
(1) 0	0,2%	0,2%	0,5%	2,3%						0,2%
(2) 1-5		0,3%	0,8%	1,6%	3,6%	18,1%	20,5%			0,9%
(3) 6-10		0,4%	0,7%	1,5%	3,1%	12,0%	22,8%	41,2%		4,2%
(4) 11-15			0,4%	1,6%	3,5%	9,1%	21,8%	30,8%	68,2%	14,8%
(5) 16-20				1,2%	3,4%	8,0%	21,6%	29,2%	64,2%	18,7%
(6) 21-25				1,5%	3,9%	5,9%	15,5%	31,1%	61,5%	18,0%
(7) 26-30				2,3%	3,9%	8,0%	19,5%	32,1%	54,8%	24,0%
(8) 31-40					4,6%	9,5%	16,1%	30,0%	57,1%	28,9%
(9) 41-50					3,0%	9,9%	13,8%	27,4%	57,9%	36,7%
(10) 51-60						13,6%	17,2%	28,9%	56,0%	37,9%
(11) 61-90							18,5%	30,3%	58,2%	47,3%
(12) 91-120								31,2%	66,8%	63,2%
(13) >120									76,7%	76,7%
Total general	**0,2%**	**0,2%**	**0,7%**	**1,6%**	**3,5%**	**9,3%**	**18,8%**	**30,3%**	**61,3%**	**6,9%**

Todos los razonamientos anteriores conllevan a definir la variable dependiente del modelo como sigue:

Cliente bueno: si su peor calificación en los 12 meses posteriores al desembolso es menor o igual a B1 y el promedio de los días de mora es menor o igual 10. Corresponde a la zona de color verde de la tabla 2.7.

Cliente malo: si su peor calificación es mayor o igual a C1; el promedio de los días de mora es indiferente. La zona de color anaranjado de la tabla 2.7 muestra la definición.

Cliente indeterminado: si no se ha categorizado como cliente bueno o malo, es decir no existe suficiente evidencia para asignarlos dentro de uno de esos grupos. Ver la zona de color celeste de la tabla 2.7.

El objetivo de esta etapa del proyecto es discriminar la población de clientes buenos y malos, de tal manera que en el estudio se incluyen los clientes nuevos etiquetados con esas marcas. Los indeterminados se dejan de lado, pues representan el 8% de los datos de la muestra de aprendizaje y además la metodología seleccionada en este trabajo, la regresión logística binaria, exige que la variable dependiente tenga únicamente dos categorías.

La frecuencia de las categorías obtenidas para la variable dependiente se muestra en la tabla 2.8. En adelante, para estimar los parámetros del modelo se trabaja únicamente con los clientes nuevos que han sido calificados como buenos o malos, 9.691 en total.

Tabla 2.8 Frecuencia de la variable dependiente sobre la muestra de aprendizaje

Tipo cliente	Frecuencia	Frec. Relativa	%Pérdida
Malo	2.549	24,20%	25,99%
Bueno	7.142	67,79%	0,38%
Indeterminado	844	8,01%	1,71%
Total aprendizaje	**10.535**	**100,00%**	**6,88%**

La definición de la variable dependiente diseñada sobre la muestra de aprendizaje se extiende a la de testing y será de utilidad al momento de realizar la validación y consistencia del modelo en la sección 2.8.

2.3.3 ESTABILIDAD DE LA DEFINICIÓN.

El objetivo es, demostrar que la definición de la variable dependiente se mantiene a lo largo de la muestra de aprendizaje. Para ello, durante los meses de estudio se va a evaluar el progreso de la colocación (otorgamiento de créditos) en clientes nuevos de ABC, tomando como referencia las operaciones otorgadas en abril 2012 y se monitorea el resto de meses de la ventana de análisis (aprendizaje y testing) para demostrar que la proporción de clientes buenos y malos es estable.

Utilizando el índice de estabilidad de la población (IPS), según lo menciona [8], se va a garantizar que, la definición del tipo de cliente guarda estabilidad y se ajusta a la realidad de la institución.

Para cada mes de análisis (desde mayo 2012) se calcula el IPS, matemáticamente definido por:

$$IPS = \sum_{i=1}^{n} (O_i - E_i) * ln\left(\frac{O_i}{E_i}\right) \tag{2.2}$$

Donde,

O_i: es la proporción observada de la categoría i de la variable de análisis.

E_i: es la proporción esperada de la categoría i de la variable de análisis.

En el caso de la variable dependiente del modelo las categorías son únicamente dos, clientes buenos y malos. El término “proporción esperada” se refiere al porcentaje de buenos y malos de abril 2012; mientras que la “proporción observada” es el valor correspondiente al mes de evaluación. Bajo esas consideraciones, el IPS para cada mes de análisis se presenta en la tabla 2.9.

Un IPS menor a 10% indica que la población comprendida entre los meses de mayo 2012 y marzo 2014 es similar a la estudiada en abril 2012. Los resultados de la tabla

2.9 demuestran entonces que la definición de la variable dependiente es estable, de hecho, al realizar un gráfico de la proporción de malos (figura 2.4) se puede evidenciar una tendencia constante en la ventana de análisis.

Tabla 2.9 Proporción de malos por mes en la muestra de aprendizaje

						Índice de estabilidad de la población		
Muestra	**Años**	**FechaCorte**	**% Malos**	**% Buenos**	**Total**	**i = malos**	**i = buenos**	**IPS**
APRENDIZAJE	2012	abr	22,94%	77,06%	100,00%			
		may	21,25%	78,75%	100,00%	0,13%	0,04%	0,17%
		jun	30,27%	69,73%	100,00%	2,03%	0,73%	2,76%
		jul	22,56%	77,44%	100,00%	0,01%	0,00%	0,01%
		ago	25,53%	74,47%	100,00%	0,28%	0,09%	0,37%
		sep	28,88%	71,12%	100,00%	1,37%	0,48%	1,84%
		oct	25,00%	75,00%	100,00%	0,18%	0,06%	0,23%
		nov	24,14%	75,86%	100,00%	0,06%	0,02%	0,08%
		dic	26,80%	73,20%	100,00%	0,60%	0,20%	0,80%
	2013	ene	30,14%	69,86%	100,00%	1,97%	0,71%	2,67%
		feb	23,82%	76,18%	100,00%	0,03%	0,01%	0,04%
		mar	29,22%	70,78%	100,00%	1,52%	0,53%	2,05%
		abr	26,08%	73,92%	100,00%	0,40%	0,13%	0,53%
		may	29,05%	70,95%	100,00%	1,44%	0,50%	1,94%
		jun	26,72%	73,28%	100,00%	0,57%	0,19%	0,76%
		jul	27,65%	72,35%	100,00%	0,88%	0,30%	1,18%
		ago	22,92%	77,08%	100,00%	0,00%	0,00%	0,00%
		sep	26,27%	73,73%	100,00%	0,45%	0,15%	0,60%
TESTING		oct	22,56%	77,44%	100,00%	0,01%	0,00%	0,01%
		nov	21,25%	78,75%	100,00%	0,13%	0,04%	0,17%
		dic	26,07%	73,93%	100,00%	0,40%	0,13%	0,53%
	2014	ene	22,14%	77,86%	100,00%	0,03%	0,01%	0,04%
		feb	21,82%	78,18%	100,00%	0,06%	0,02%	0,07%
		mar	22,92%	77,08%	100,00%	0,00%	0,00%	0,00%

Una forma alternativa de demostrar la estabilidad de la definición del tipo de cliente es, aplicar el test de Dickey Fuller, el cual permite determinar si la serie temporal que describe el porcentaje de malos es estacionaria, o sea que no existe suficiente evidencia estadística para afirmar que existe una tendencia creciente o decreciente de

la serie de datos. Esta prueba se realiza en el paquete estadístico Eviews 7 y se extiende a las muestras de aprendizaje y testing.

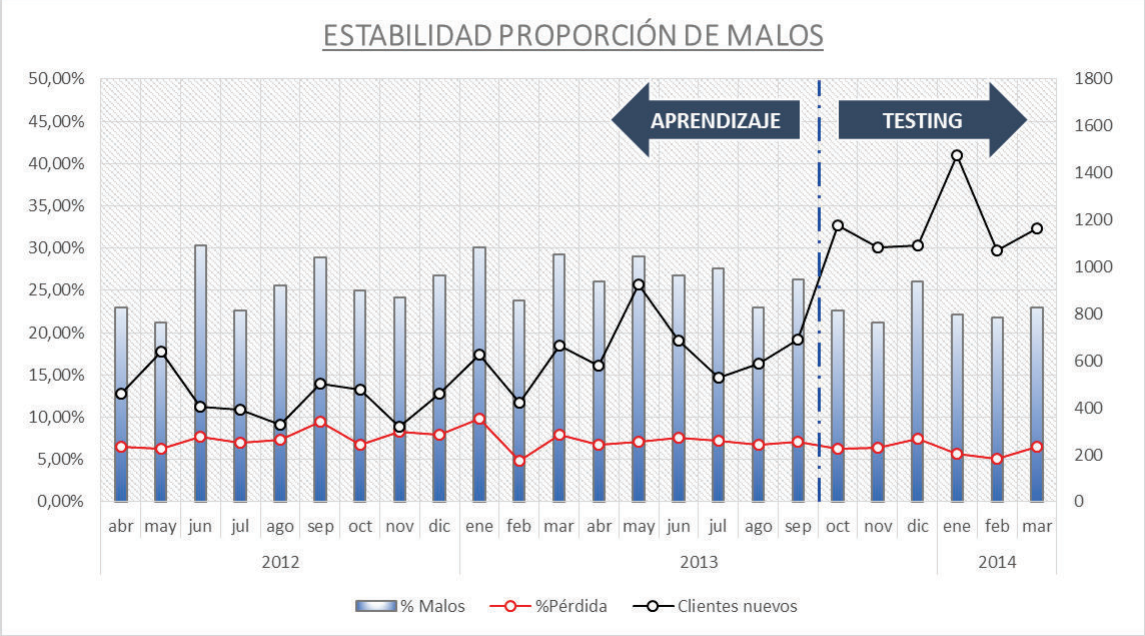

Figura 2.4 Proporción de malos en la muestra de aprendizaje

El test contrasta las hipótesis nula y alternativa:

$$H_o: La\ serie\ de\ tiempo\ no\ es\ estacionaria\ (presenta\ raíz\ unitaria)$$

$$H_a: La\ serie\ de\ tiempo\ es\ estacionaria\ (no\ presenta\ raíz\ unitaria)$$

La hipótesis nula se acepta si el valor observado del estadístico t de la tabla 2.10 es mayor al valor crítico correspondiente. Se puede observar que para los valores críticos 1%, 5% y 10%, el test rechaza la hipótesis nula a favor de la alternativa; es decir la serie de tiempo no presenta una raíz unitaria y por tanto es estacionaria. De esta manera se demuestra, estadísticamente, que la definición de la variable dependiente

el modelo es estable en la ventana de análisis y se ajusta a la realidad del proceso de otorgamiento de créditos a clientes nuevos de la institución.

Cabe mencionar que el estadístico t no sigue una distribución *t-student* convencional (bajo la hipótesis inicial), Dickey y Fuller calcularon los niveles críticos para diferentes tamaños de muestra en base a simulaciones de Monte Carlo [2]; sin embargo, estas tablas no resultaban ser del todo adecuadas por lo cual fueron extendidas por MacKinnon (1996) y actualmente son las que utiliza Eviews para el contraste de raíces unitarias. El software es capaz de trabajar con muestras de 20 observaciones en adelante, en el caso de utilizar muestras con un menor número de observaciones se alerta al usuario que los resultados podrían son poco confiables. Entre la muestra de aprendizaje y testing se obtienen 24 observaciones, con lo que se puede garantizar la confiabilidad de los resultados obtenidos en Eviews.

Tabla 2.10 Resultados del test de Dickey Fuller

Null Hypothesis: MALOS has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=5)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.664669	0.0013
Test critical values:	1% level	-3.752946	
	5% level	-2.998064	
	10% level	-2.638752	

*MacKinnon (1996) one-sided p-values.

2.4 ANÁLISIS DESCRIPTIVO DE LAS VARIABLES.

En esta sección se realiza un tratamiento previo de los datos con la finalidad identificar datos nulos y concentrados. El análisis se realiza sobre las variables con información socio demográfica y del buró de créditos que se definieron en la sección 2.1, pues a partir de ellas se elegirán los regresores del modelo.

La metodología estadística escogida para identificar la representatividad de los valores válidos y nulos de cada variable, es el análisis de la frecuencia y frecuencia relativa. Para determinar la concentración de información de una variable, en cambio, se utilizan los conceptos de distribución acumulada y acumulada relativa para variables aleatorias cualitativas, cuantitativas discretas y en el caso de las cuantitativas continuas se determinan intervalos de clase previo a analizar la distribución acumulada.

2.4.1 IDENTIFICACIÓN DE VALORES NULOS.

Para ilustrar la metodología de identificación de datos nulos se presenta un ejemplo con variables socio demográficas y del sistema financiero.

Tabla 2.11 Frecuencia y frecuencia relativa de datos válidos y nulos

Variable	Válidos	Nulos	Total	% Válidos	% Nulos	% Total
ProvinciaNacimiento	7.004	2.687	9.691	72,3%	27,7%	100,0%
NumCalificacionCDE_12UM	9.691	-	9.691	100,0%	0,0%	100,0%
DiasMoraCorte	9.691	-	9.691	100,0%	0,0%	100,0%

En la tabla 2.11 puede observarse que la variable provincia de nacimiento del cliente nuevo de ABC se compone del 27,7% de valores nulos, mientras que en las otras variables del ejemplo la información es completa.

En términos globales, 16 de las 85 variables presentan problemas de valores nulos de manera similar al caso de la variable provincia de nacimiento; este grupo de variables representan el 18,8% del total. En la tabla 2.12 se evidencian los resultados, los porcentajes en paréntesis representan la proporción de variables que presentan valores nulos o información completa, según sea el caso.

El sentido de realizar este análisis es tomar una decisión al respecto de las variables que presentan valores nulos. A partir de los resultados obtenidos, de las 16 variables

identificadas se han eliminado del estudio 2 de ellas, las cuales son parroquia y ciudad de nacimiento cuyo porcentaje de valores nulos es de 50% y 52%, respectivamente. A pesar de que la parroquia y ciudad de nacimiento se eliminan, se conserva en el estudio la provincia por considerar que es una variable demográfica relevante en el estudio y tan solo presenta un 27,7% de valores nulos.

Tabla 2.12 Resumen del análisis de los valores nulos

Grupo de variables	Presencia de nulos	Información completa	Total de variables
Socio demográficas	12 (46,2%)	15 (53,8%)	27 (100,0%)
Buró de créditos	4 (80,0%)	54 (20,0%)	58 (100,0%)
TOTAL	**16 (18,8%)**	**69 (80,2%)**	**85 (100,0%)**

Las otras variables que presentan valores nulos (14 restantes según la tabla 2.12) se mantienen en el estudio debido a que el porcentaje de nulos para todas ellas oscila entre el 1% y 29%, y se considera que no es un argumento suficiente para eliminarlas.

2.4.2 ANÁLISIS DE LA CONCENTRACIÓN DE INFORMACIÓN.

Ahora, para determinar la concentración de información, en este proyecto se estudia la distribución acumulada de las variables. Para el caso de las variables aleatorias cuantitativas continuas se utiliza el criterio de intervalos de clase para agrupar los datos, previo al análisis de la concentración de información.

Existen diversas metodologías que se recomiendan para la construcción de intervalos de clase, en este caso se lo hace con la ayuda de la regla de Sturges para la elección del número de intervalos. Matemáticamente la fórmula se describe por:

$$k = 1 + log_2(n) \tag{2.3}$$

Donde, k es el número de intervalos de clase y n es el tamaño de la muestra. Para todas las variables cuantitativas continuas del estudio, de acuerdo con la fórmula de Sturges, el número de intervalos de clase calculados es 14.

Se trata de una variable cualitativa nominal y para cada categoría se calcula la frecuencia acumulada. En términos generales, el criterio que se ha adoptado para eliminar del estudio variables con problemas de concentración de información es que no debe existir un valor, categoría o intervalo de clase que concentre más del 95% de información. En el ejemplo de la tabla 2.13 se puede notar que no existen categorías en las que se concentra información, pero sí existen varias en las que el porcentaje de representatividad es menor a 5%, este último problema se solventa en la sección 2.5 donde se seleccionan a las variables regresoras del modelo.

Bajo estas consideraciones se han eliminado del estudio las variables que se resumen en la tabla 2.14, donde a manera de ejemplo se puede observar que el 98% de los individuos considerados en el estudio vive en una zona urbana, o que el 97% no registran calificación “B”, “C” o “D” en los 12 últimos meses (el 97% de la información se concentra en el valor 0).

En resumen, los resultados de las secciones 2.4.1 y 2.4.2 indican que es necesario eliminar del estudio 16 de las 85 variables iniciales, de tal manera que en las secciones siguientes se trabajará únicamente con las 69 variables resultantes de este primer análisis.

Tabla 2.13 Frecuencia y frecuencia acumulada de la variable provincia de nacimiento

Nro.	Categoría	Frecuencia	% Frecuencia	Frec. acumulada	% Frec. acumulada
1	NULOS	2.687	27,73%	2.687	27,73%
2	PICHINCHA	2.281	23,54%	4.968	51,26%
3	GUAYAS	1.468	15,15%	6.436	66,41%
4	MANABI	883	9,11%	7.319	75,52%
5	IMBABURA	315	3,25%	7.634	78,77%
6	LOS RIOS	251	2,59%	7.885	81,36%
7	COTOPAXI	234	2,41%	8.119	83,78%
8	LOJA	202	2,08%	8.321	85,86%
9	ESMERALDAS	199	2,05%	8.520	87,92%
10	CHIMBORAZO	195	2,01%	8.715	89,93%
11	EL ORO	195	2,01%	8.910	91,94%
12	TUNGURAHUA	168	1,73%	9.078	93,67%
13	CARCHI	154	1,59%	9.232	95,26%
14	BOLIVAR	127	1,31%	9.359	96,57%
15	SANTO DOMINGO	94	0,97%	9.453	97,54%
16	SANTA ELENA	81	0,84%	9.534	98,38%
17	AZUAY	70	0,72%	9.604	99,10%
18	CAÑAR	25	0,26%	9.629	99,36%
19	SUCUMBIOS	15	0,15%	9.644	99,52%
20	MORONA SANTIAGO	13	0,13%	9.657	99,65%
21	NAPO	11	0,11%	9.668	99,76%
22	GALAPAGOS	8	0,08%	9.676	99,85%
23	ORELLANA	8	0,08%	9.684	99,93%
24	PASTAZA	4	0,04%	9.688	99,97%
25	ZAMORA CHINCHIPE	3	0,03%	9.691	100,00%

Tabla 2.14 Variables eliminadas por el criterio de concentración de información

Variable	Valor/clase concentración	% Concentración
ZonaVivienda	Urbana	98,0%
NumCalificacionCDE_12UM	0	97,0%
MaxMontoHistorico_Comercial	0	97,0%
MaxMontoHistorico_Vivienda	0	98,0%
MaxMontoHistorico_Olla	0	100,0%
TieneCreditoComercial	0	98,0%
TieneCreditoOlla	0	100,0%
TieneCreditoVivienda	0	99,0%
ValorDemandaJudicialCorte	0	100,0%
CarteraCastigadaCorte	0	100,0%
OperacionesDemandasHis	0	100,0%
OperacionesCastigoHis	0	99,0%
MaxValorDemandaJudicialHis	0	100,0%
Bancarizado_36UM	1	100,0%

2.5 SELECCIÓN DE VARIABLES REGRESORAS DEL MODELO.

La finalidad de esta sección es, determinar las variables que discriminan de mejor manera a la población de los clientes buenos y malos.

A las variables descritas en la sección 2.1 se han realizado transformaciones que ayuden a explicar mejor el comportamiento de la variable dependiente, en ese sentido se han construido los siguientes grupos:

a) Variables binarias, indican la presencia o ausencia de una característica del individuo objeto de análisis.

b) Variables recodificadas, resultan de asignar el valor del coeficiente WOE (detallado más adelante) a los individuos que pertenecen a una determinada categoría de las diferentes variables de análisis.

c) Variables logarítmicas, básicamente son el resultado de calcular el logaritmo natural sobre las variables cuantitativas continuas.

Con el fin de seleccionar las variables que se incluyen en el modelo, se analiza la relación entre la variable dependiente del estudio con las posibles variables predictoras. Los criterios estadísticos usados para cumplir con ese objetivo son: a) test Chi Cuadrado de bondad de ajuste, b) índice de estabilidad de la población, c) valor de la información; y, d) factor de inflación de la varianza.

2.5.1 CATEGORIZACIÓN DE VARIABLES.

Una vez que se dispone de un conjunto de variables en las que se han eliminado los problemas de concentración de información y presencia de valores nulos, por medio de árboles de decisión, se busca: a) definir puntos de corte de las variables cuantitativas para transformarlas en variables categóricas, b) reagrupar las categorías de las variables cualitativas, y c) crear combinaciones entre las variables.

El árbol de decisión es un algoritmo de particionamiento recursivo, usado en minería de datos para resolver problemas de clasificación y predicción. El propósito con esta

metodología es establecer una agrupación de cada una de las variables del estudio o crear combinaciones entre ellas, de tal forma que expliquen mejor el comportamiento de la variable dependiente y así se conviertan en relevantes para el modelo.

Entre los diferentes algoritmos existentes para la construcción de árboles de decisión, se ha escogido el método CHAID (*Chi-squared automatic interaction detector*), desarrollado y publicado por G. V. Kass (1980)[12]. Para definir los valores de corte de la potencial variable regresora, el método CHAID, valida el test Chi-cuadrado de independencia entre el tipo de cliente y la variable que ha sido categorizada con el algoritmo; en el caso de que la variable presente valores nulos, son tratados como una categoría válida adicional.

La metodología de árboles de decisión se replica sobre las sesenta y nueve variables resultantes de la sección 2.4, con la ayuda del programa *Answer Tree* 3.1.

A manera de ejemplo, se puede observar la variable saldo por vencer histórico (X) de la figura 2.5; el estadístico Chi-cuadrado obtenido es 46,3214 y el p-valor asociado es menor al 5% (p-value=0,00), por lo tanto, se rechaza la hipótesis de independencia a un nivel de confianza del 95% y se concluye que el tipo de cliente depende de X.

En conclusión, con las categorías construidas a partir del árbol de decisión, X es candidata a considerarse como regresora del modelo. El resumen descriptivo de la variable obtenida se presenta en la tabla 2.15.

Tabla 2.15 Categorías resultantes del árbol de decisión

Saldo por vencer histórico				
Categoría	**% Malos: 0**	**% Buenos: 1**	**Frecuencia**	**Frec. Relativa**
<= 21.134,28	28,58%	71,42%	5.815	60%
(21.134,28 ; 133.083,44]	21,78%	78,22%	2.907	30%
> 133.083,44	26,21%	73,79%	969	10%
Total			**9.691**	**100%**

[12] G. V. Kass: An Exploratory Technique for Investigating Large Quantities of Categorical Data, Applied Statistics, Vol. 29, No. 2 (1980), pp. 119-127.

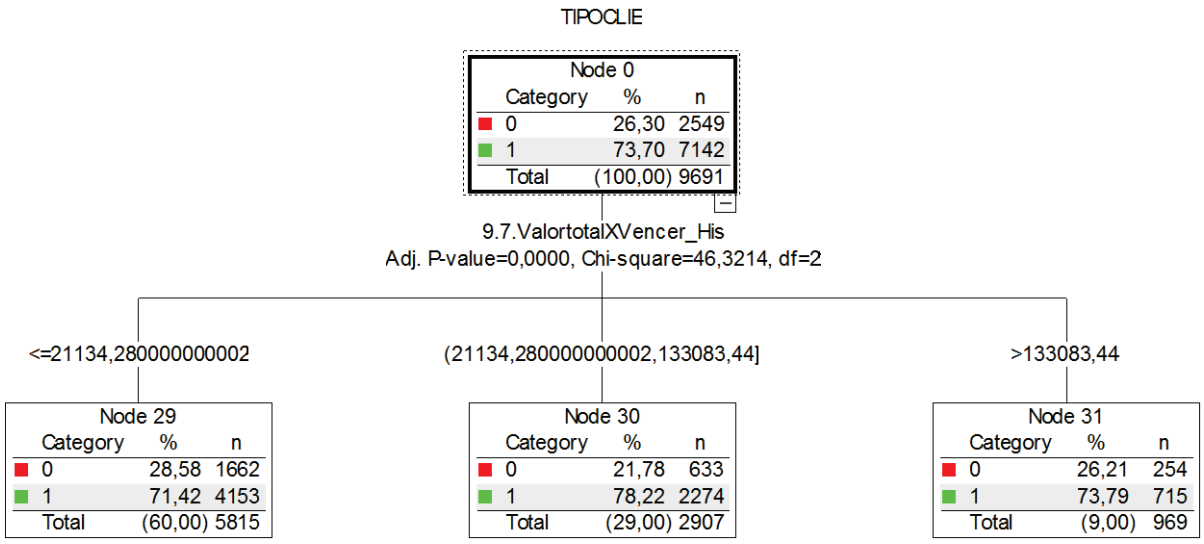

Figura 2.5 Punto de corte de la variable saldo por vencer histórico

La figura 2.6, en cambio, muestra que no es posible determinar una agrupación de las categorías del tipo de vivienda, pues a un nivel de confianza del 95% se acepta la hipótesis de independencia entre el tipo de cliente y la variable estudiada, por lo tanto no se la considera dentro del grupo de variables candidatas a ser regresoras del modelo.

En el árbol de decisión de la figura 2.7, se han combinado las variables antigüedad y tipo de vivienda. El p-valor del nodo 1 sugiere que si la antigüedad de residencia es menor o igual a 99 meses no se debe combinar con el tipo de vivienda, por el contrario, si es mayor que 99, la variable resultante de esa combinación puede considerarse como candidata a regresora del modelo.

Nótese que el tipo de vivienda por sí solo no es capaz de explicar el comportamiento de la variable dependiente, sin embargo al ser combinada con la antigüedad de residencia podría integrarse en la regresión.

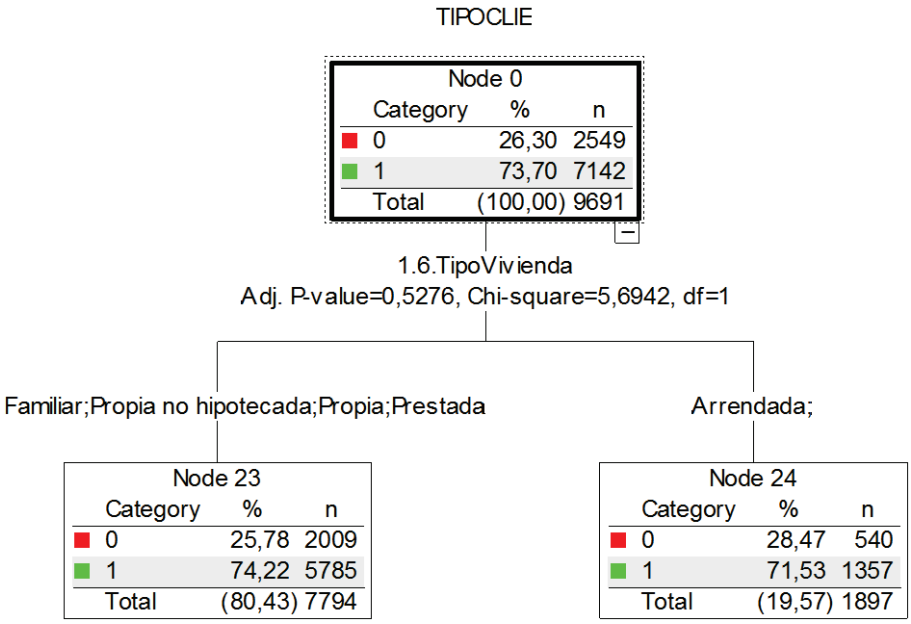

Figura 2.6 Reagrupación de las categorías de la variable tipo de vivienda

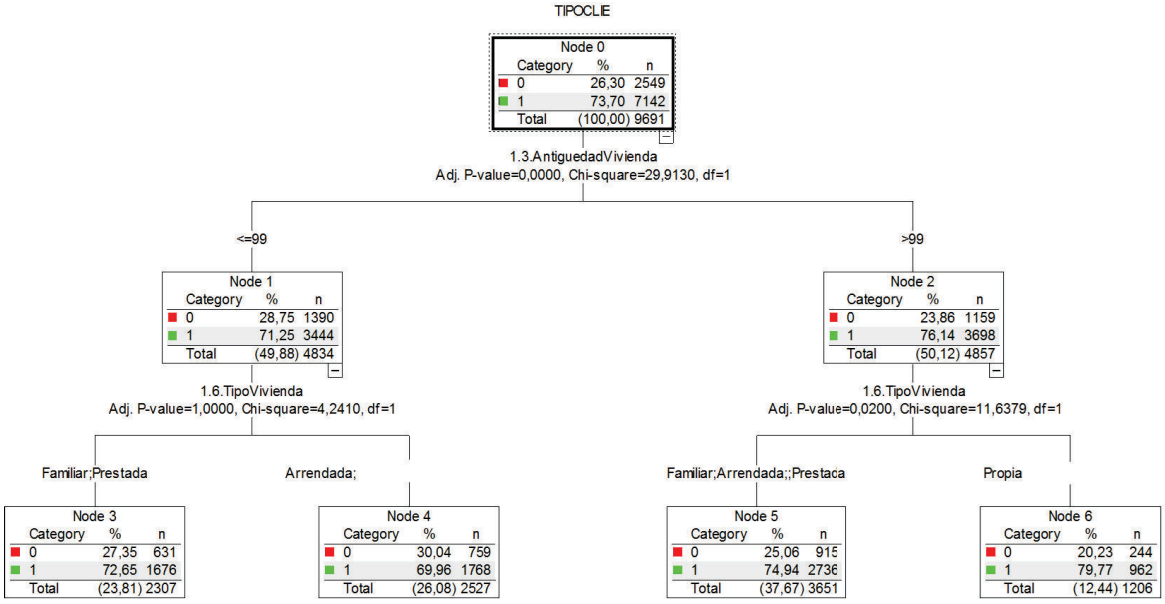

Figura 2.7 Combinación de las variables antigüedad de residencia y tipo de vivienda

2.5.2 CRITERIOS PARA LA SELECCIÓN DE VARIABLES REGRESORAS.

En este proyecto se han escogido cuatro criterios estadísticos para la selección de las variables regresoras del modelo, los mismos que son aplicados de acuerdo al tipo de variable que se va tratar.

i) Test Chi-Cuadrado de bondad de ajuste.

El test permite verificar que la muestra proviene de una población con una distribución teórica dada, $F_o(x)$, [8].

Las hipótesis nula y alternativa son, respectivamente:

$$\begin{cases} H_o: F(x) = F_o(x) \\ H_a: F(x) \neq F_o(x) \end{cases} \tag{2.4}$$

Donde $F(x)$ es la distribución empírica de la muestra.

La prueba considera una muestra de tamaño n, cuyas observaciones están agrupadas en k categorías o valores individuales, dependiendo si la variable de estudio es cualitativa o cuantitativa discreta, respectivamente. Se denota por o_i a la cantidad de observaciones en cada clase o valor $i = 1,2, \dots, k$

A partir de la distribución teórica $F_o(x)$ se puede calcular la probabilidad p_i de que un dato cualquiera pertenezca a una clase i o que sea igual a un dato individual.

Con el valor de probabilidad es posible calcular la frecuencia esperada e_i para la clase o dato individual i, según sea el caso; es decir se procede al cálculo del número de observaciones que deberían ser incluidas en la clase o valor i según la distribución teórica dada.

$$e_i = np_i, \qquad i = 1,2, \dots, k \tag{2.5}$$

Entonces, se han identificado dos valores de frecuencia cada clase o dato individual i: a) o_i la frecuencia observada, y b) e_i la frecuencia esperada o teórica.

El estadístico usado por el test es:

$$\chi_c^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \tag{2.6}$$

Que sigue una distribución Chi-Cuadrado con $k - r - 1$ grados de libertad, donde r es la cantidad de parámetros de la distribución que deben estimarse a partir de la muestra. Una condición necesaria para aplicar el test que,

$$Porcentaje\ e_i \geq 5\%, \qquad \forall e_i \tag{2.7}$$

Para un nivel de significancia α, se define un valor crítico χ_α^2 (fractil de orden α de la distribución Chi-Cuadrado de cola derecha con $k - r - 1$ grados de libertad) para evaluar la hipótesis inicial.

Si las frecuencias observadas no difieren significativamente de las esperadas, entonces el valor del estadístico χ_c^2 será cercano a cero y se acepta la hipótesis H_o, caso contrario es rechazada. Es decir, se rechaza la hipótesis nula si y solo si $\chi_c^2 > \chi_\alpha^2$.

La bondad de ajuste se realiza entre la muestra de aprendizaje y testing, tomando la primera como teórica y segunda como observada, de esta forma se busca probar que las proporciones de clientes malos de cada categoría de las variables regresoras, son similares en los dos periodos.

ii) Índice de estabilidad de la población (IPS).

La metodología de cálculo y aplicación del IPS se explicó en la sección 2.3.3; la diferencia es que en la selección de variables regresoras se busca que la frecuencia relativa de las categorías sea similar entre el periodo de aprendizaje y testing, garantizando se esa manera la estabilidad de la información.

iii) Valor de la información (IV).

El estadístico IV permite seleccionar las variables predictoras, para una variable X con K categorías representadas por x_k, el estadístico se define matemáticamente por:

$$IV = \sum_{k=1}^{K} (b_k - m_k) \ln\left(\frac{b_k}{m_k}\right) \tag{2.8}$$

Donde:

b_k: es la proporción de clientes buenos de la categoría x_k, $k = 1,2, \dots, K$

m_k: es la proporción de clientes malos de la categoría x_k, $k = 1,2, \dots, K$

Los valores b_k y m_k son tales que:

$$\sum_{k=1}^{K} b_k = 1, \;\; y \;\; \sum_{k=1}^{K} m_k = 1 \tag{2.9}$$

Es decir, el estadístico IV busca que la composición de buenos y malos sea diferente en cada una de las K categorías de las variables regresoras.

Para evaluar la fuerza de un predictor, se recurre a las reglas de Thumb[13]:

Tabla 2.16 Reglas de Thumb

Rango	Criterio
< 0,02	NO PREDICTIVO
(0,02 ; 0,10]	DÉBIL
(0,10 ; 0,30]	MEDIO
> 0,30	ALTO

En este estudio se escogerán como regresoras a aquellas variables que sean diferentes de “no predictiva”, además el indicador se calcula tanto para la muestra de aprendizaje como para el testing, pues se busca que en los dos periodos sean diferentes de “no predictoras”.

iv) Factor de inflación de varianza.

El propósito es, analizar el problema de la multicolinealidad entre las variables regresoras del modelo.

La situación de multicolinealidad se presenta cuando las variables predictoras de la regresión mantienen un alto nivel de correlación entre ellas. Los indicios para detectar su presencia básicamente son:

- Pequeños cambios en los datos provocan grandes cambios en la estimación de los parámetros del modelo.
- Las estimaciones de los coeficientes suelen presentar signos y magnitudes diferentes a los esperados.

En este proyecto se ha escogido trabajar con el factor de inflación de la varianza para detectar la presencia de multicolinealidad. La metodología consiste en

[13] Datos tomados de: Information Value Statistic. Lund, B. Brotherton, D.

realizar una regresión de cada una de las variables predictoras y analizar sus coeficientes de determinación R^2. Un valor alto de R^2 indica la posible existencia de multicolinealidad, a partir de esta afirmación es posible calcular el factor de inflación de la varianza (VIF) como:

$$VIF_j = \frac{1}{(1 - R_j^2)} \tag{2.10}$$

Siendo R_j^2 el coeficiente de determinación de la regresión de la variable X_j.

Fox y Monnette (1992)[14] generalizan la noción de inflación de la varianza, pues el VIF no es adecuado ante la presencia de variables categóricas como predictoras del modelo; obtuvieron una medida llamada factor de inflación de la varianza generalizado (GVIF), matemáticamente el GVIF se define por:

$$GVIF = \frac{det(R_{11}) \times det(R_{22})}{det(R)} \tag{2.11}$$

Donde R_{11} es la matriz de correlaciones del conjunto de las variables categóricas del modelo, R_{22} es la matriz de correlaciones del resto de regresores y R es la matriz de correlaciones de todas las variables de todo el modelo.

En este proyecto se utiliza, para el cálculo del GVIF, el paquete estadístico R; puesto que en la librería CAR ya se encuentra implementada la fórmula.

Con un valor de GVIF mayor a 10 concluiremos que la multicolinealidad de la variable analizada con el resto de regresores es alta y por lo tanto la decisión será quitar a esa variable del modelo.

[14] Publicación original: Fox J. y Monnette G., Generalized Collinearity Diagnostics, American Statistical Associaton, Vol. 87, No. 417 (1992), pp. 178-182.

Es preciso indicar que todos los criterios descritos en este punto, serán utilizados más adelante de acuerdo al tipo de variable a ser incluida en el modelo, es decir no se usarán todos al mismo tiempo.

2.5.3 SELECCIÓN DE VARIABLES REGRESORAS DICOTÓMICAS.

A partir de las categorías construidas con los árboles de decisión de la sección 2.5.1, se construyen variables dicotómicas que indican la presencia o ausencia de la característica definida en cada categoría.

Por ejemplo, partiendo de los resultados de la tabla 2.17 se construyen tres variables dicotómicas, específicamente son:

- *VD_1:* Indica 1 si el saldo total por vencer histórico del individuo analizado es menor o igual a 21.134,28 dólares, caso contrario el valor de la variable es 0.
- *VD_2:* Indica 1 siempre y cuando el saldo total por vencer histórico del individuo se encuentra entre 21.134,28 y 133.083,44 dólares, caso contrario el valor de la variable es 0.
- *VD_3:* indica 1 si el individuo registra un saldo total por vencer histórico mayor a 133.083,44, caso contrario el valor de la variable es cero.

Para seleccionar las variables dicotómicas a ser incluidas en el modelo, se utilizan los criterios citados en 2.5.2. A manera de ejemplo, se toman las mismas variables *VD_1, VD_2 y VD_3* para mostrar la metodología empleada.

Tabla 2.17 Selección de variables regresoras dicotómicas

Variable	ChiCuadrado	IPS	IV-Aprendizaje	IV-Testing	Conclusión
VD_1	Acepta Ho	9,4%	2,1%	9,9%	REGRESOR
VD_2	Acepta Ho	1,0%	2,4%	2,9%	REGRESOR
VD_3	Rechaza Ho	9,0%	0,0%	4,1%	

La tabla 2.17 muestra que las variables dicotómicas VD_1 y VD2 pueden ser incluidas en el modelo, por las siguientes razones:

i) Acepta la hipótesis nula del test Chi-Cuadrado de bondad de ajuste, realizado entre las muestras de aprendizaje y testing usando la proporción de clientes malos, eso quiere decir que la población de malos es similar en ambos periodos.
ii) El IPS es menor que 10%, por lo tanto la frecuencia relativa de las categorías 0 y 1 de la muestra de testing es similar a la estudiada en la muestra de aprendizaje.
iii) El IV en la muestra de aprendizaje y testing es mayor que 2%, es decir las variables dicotómicas son diferentes de “no predictivas” en ambos periodos.

Los criterios analizados en i, ii y iii ayudan a seleccionar variables dicotómicas que explican el comportamiento de la variable dependiente del modelo, tanto con la muestra de aprendizaje como con la de testing.

La razón por la que se utilizan estos criterios es que algunas de las variables del estudio se relacionan con la administración de políticas de la institución, por tal motivo la distribución de las variables en la línea del tiempo podría cambiar y provocar que el modelo sea eficiente únicamente con los datos usados en la etapa de construcción del mismo, pero no con ventanas de datos diferentes. Por ejemplo, desde un mes en particular la gestión de colocación podría enfocarse en potenciales clientes nuevos mayores a 30 años de edad, pero desde otro mes en aquellos cuya edad oscila entre 35 y 40 años, provocando de esta manera que la distribución de la variable edad cambie en esos meses.

2.5.4 SELECCIÓN DE VARIABLES REGRESORAS RECODIFICADAS.

Con la finalidad de construir nuevas variables que ayuden a describir el comportamiento de la variable dependiente del modelo, se analiza la relación entre la

proporción de clientes buenos y malos dentro de cada categoría de las variables construidas con los árboles de decisión.

Para ello, consideramos una variable X con K categorías para la cual se define el indicador WOE (*weight of evidence*) como sigue:

$$WOE = LN\left(\frac{\%\ buenos\ en\ la\ categoría\ k}{\%\ malos\ en\ la\ categoría\ k}\right),\quad k = 1,2,\dots,K \tag{2.12}$$

Es decir, se busca evaluar la relación buenos / malos en cada categoría. Cuando el porcentaje de clientes buenos y malos es igual, el WOE es cero. Si el porcentaje de buenos en una categoría es mayor entonces el WOE es positivo, por el contrario si el porcentaje de clientes malos es mayor entonces el WOE es negativo.

Para seleccionar las variables que ingresan al modelo con el coeficiente WOE, primero se debe garantizar, mediante el índice de estabilidad de la población, que la proporción de clientes malos se mantiene estable entre los periodos de aprendizaje y testing. Para ejemplificar la metodología se toma la variable antigüedad de residencia, en la tabla 2.18 puede observarse que el valor del IPS es 4,3% y por ser menor a 10% se concluye que es estable, por lo tanto la variable es incluida en la regresión logística.

Tabla 2.18 Índice de estabilidad de la variable antigüedad de residencia

Antigüedad de residencia					
Categoría	**Frec. Aprendizaje**	**Frec. Testing**	**Frec. Rel. Modelo**	**Frec. Rel. Testing**	**IPS**
<= 99	1.390	882	54,53%	61,85%	0,9%
(99 ; 164]	497	175	19,50%	12,27%	3,3%
> 164	662	369	25,97%	25,88%	0,0%
Suma	**2.549**	**1.426**	**100,00%**	**100,00%**	**4,3%**

Una vez que se ha demostrado estadísticamente que la proporción de clientes malos es estable entre los dos periodos, se procede a recodificar la variable con los valores

del WOE correspondientes a cada categoría, según se muestra en la tabla 2.19. En el cálculo se incluyen los individuos de la muestra de aprendizaje.

Tabla 2.19 Recodificación de la variable antigüedad de residencia

1.3.AntiguedadVivienda					
Categoría	**Malos**	**Buenos**	**Proporción malos**	**Proporción buenos**	**WOE**
<= 99	1.390	3.444	54,53%	48,22%	-0,1230
(99 ; 164]	497	1.460	19,50%	20,44%	0,0473
> 164	662	2.238	26,97%	31,34%	0,1878
Suma	**2.549**	**7.142**	**100,0%**	**100,0%**	

2.5.5 SELECCIÓN DE VARIABLES REGRESORAS LOGARÍTMICAS.

Con el fin de explicar mejor el comportamiento de la variable dependiente, se han incluido en el modelo variables que se relacionan con el logaritmo natural de las variables cuantitativas del estudio.

Al momento de incluir las variables se ha verificado previamente que,

a) Aporten al ajuste del modelo, y
b) No contribuyan al problema de multicolinealidad.

Si alguna de las dos condiciones no se cumple, la variable no se incluye en el modelo.

2.5.6 ESTUDIO DEL PROBLEMA DE MULTICOLINEALIDAD EN LAS VARIABLES PREDICTORAS.

Con las variables tratadas en las secciones precedentes, se da paso a la validación del factor de inflación de la varianza (VIF) para resolver el problema de correlación que pudiera existir dentro de cada grupo: dicotómicas, recodificadas o logarítmicas.

El procedimiento de verificación considera los siguientes pasos:

1. Calcular el VIF para cada variable dentro de los grupos identificados anteriormente.
2. Eliminar la variable que presentan el mayor valor del VIF siempre y cuando sea superior a 10.
3. Luego de eliminar la variable en el punto 2, verificar si existe al menos una variable cuyo valor del VIF supera 10; se desprenden dos opciones:
    a. Si se cumple la condición entonces volver al punto 1.
    b. Si no se cumple la condición, es decir si no existe variables con VIF superior a 10, termina la validación del problema de multicolinealidad.

A manera de ejemplo, en la tabla 2.20 se toman seis variables dicotómicas que han sido seleccionadas como regresoras y se aplica el procedimiento planteado.

Tabla 2.20 Análisis de multicolinealiadad de las variables dicotómicas

Nro.	Variable	VIF1	VIF2	VIF3
1	X.NumCalificacionB_12UM11	4,37	4,37	4,01
2	X.NumCalificacionBlanco_12UM31	80,96	80,97	Eliminada
3	X.C1.NumCalificacionB_A_12UM21	201,17	Eliminada	Eliminada
4	X.C1.NumCalificacionB_A_12UM41	4,61	4,61	4,22
5	X.C2.NumCalificacionA_Antiguedad12UM11	79,20	79,21	2,30
6	X.C4.NumCalificacionB_Antiguedad12UM21	199,68	2,49	2,45

Se puede observar que al tomar las 6 variables en conjunto el valor del VIF es superior a 10 para 4 de ellas; de acuerdo con el procedimiento planteado se elimina la variable 3 (VIF=201,17). Una vez que se ha eliminado esa variable, se vuelve a calcular el indicador con las 5 restantes, con lo cual las variables 2 y 5 vuelven a presentar un VIF superior a 10, por lo tanto se elimina la que presenta el mayor valor entre ellas. Finalmente, una vez eliminadas las dos variables se puede evidenciar que ninguna de las restantes supera un VIF de 10, por lo tanto serán incluidas en el modelo.

Esta metodología se aplica sobre todas las variables que conforman cada uno de los conjuntos: dicotómicas, recodificadas y logarítmicas.

El resultado de toda la sección 2.5 es un conjunto de variables que pueden ser parte de la regresión logística. En lo posterior se trabajará con ellas para estimar los coeficientes del modelo y de esa forma obtener la regresión final.

2.6 ESTIMACIÓN DEL MODELO DE REGRESIÓN LOGÍSTICA.

Una vez que se han seleccionado las variables que cumplen con las condiciones establecidas para ser consideradas como regresoras, se las incluye en una regresión preliminar, verificando previamente que los coeficientes estimados son significativos que no existen problemas de multicolinealidad.

Para obtener la regresión preliminar, se plantea el siguiente procedimiento:

1. Realizar una regresión con todas las variables resultantes de la sección 2.5.
2. Seleccionar las variables cuyo coeficiente estimado en la regresión es estadísticamente cero al 95% de confianza, es decir elegir aquellas cuyo p-valor supera el 5% y eliminar la de mayor valor. Con las variables restantes realizar una nueva regresión y verificar nuevamente la significancia de los coeficientes. El procedimiento se repite hasta obtener una regresión en donde todos los coeficientes estimados sean estadísticamente diferentes de cero, es decir que el p-valor asociado sea menor al 5%.
3. Con las variables que se incluyen en la regresión obtenida en el punto 2, se verifica el VIF para cada una de ellas. Al igual que en secciones anteriores se busca que los valores sean menores que 10. En el caso de que no se cumpla con esta condición es preciso quitar del modelo, una a una, las variables con VIF superior a 10, de la misma forma que se realizó en la sección 2.5.6.

El resultado del procedimiento descrito, da como resultado la regresión preliminar buscada y es presentada la tabla 2.21. Se denomina preliminar porque luego se aplicarán técnicas de *bootstrap* (remuestreo) para garantizar la significancia

estadística de la estimación de los coeficientes, y con esos resultados se obtendrá la regresión final.

Tabla 2.21 Regresión logística preliminar

Nro.	Variable	B	Sig.	Exp(B)
1	AntiguedadVivienda_WOE	0,52429	0,78%	1,689
2	MaxMontoHistoricoMicro_WOE	0,75997	0,00%	2,138
3	PorEndAntes12UM_WOE	0,67974	0,08%	1,973
4	DiasMora12_24_LN	-0,09768	0,00%	0,907
5	CR_36UM_LN	-0,06357	1,77%	0,938
6	PorEnd_12UM_LN	-0,03694	1,91%	0,964
7	EntVen36_LN	-0,05697	0,42%	0,945
8	NumeroHijos1(1)	0,19920	2,28%	1,220
9	NumeroHijos2(1)	0,37021	0,00%	1,448
10	Salario1(1)	-0,34550	0,00%	0,708
11	Salario2(1)	-0,18336	0,42%	0,832
12	Edad1(1)	-0,29458	0,24%	0,745
13	Edad2(1)	-0,14174	4,29%	0,868
14	Edad4(1)	0,22701	0,51%	1,255
15	Edad5(1)	0,33899	0,00%	1,404
16	AntiguedadRelacionDependencia(1)	-0,19551	0,07%	0,822
17	ProvinciaNacimiento1(1)	0,42571	0,00%	1,531
18	ProvinciaNacimiento2(1)	0,40156	0,00%	1,494
19	ProvinciaNacimiento3(1)	0,24248	0,54%	1,274
20	AntiguedadViv_TipoVivienda1(1)	0,29809	0,06%	1,347
21	Antiguedad_ProvDomicilio2(1)	-0,31990	0,02%	0,726
22	DiasMoraPromedio12UM_MoraCorte2(1)	0,71722	0,00%	2,049
23	DiasMoraPromedio12UM_MoraCorte3(1)	-0,47701	0,00%	0,621
24	MaxMontoHistoricoConsumo1(1)	-0,17812	3,67%	0,837
25	AntiguedadSF_MaxMontoHisConsumo3(1)	0,28262	0,72%	1,327
26	OperacionesAlDia_Vigentes1(1)	-0,23787	0,20%	0,788
27	ValortotalXVencer_His2(1)	0,18763	0,29%	1,206
28	Maxvencido_VecesAlDia12UM2(1)	0,31949	0,00%	1,376
29	Maxvencido_VecesAlDia12UM3(1)	-0,41601	0,00%	0,660
30	Antiguedad12UM_PorEndCorte3(1)	-0,18119	1,91%	0,834
31	EndPromSF_12UM1(1)	0,26419	0,20%	1,302
32	Constant	-0,67820	0,01%	0,508

2.6.1 BOOTSTRAP PARA ESTIMAR INTERVALOS DE CONFIANZA DE LOS COEFICIENTES DEL MODELO.

En el manual de IBM SPSS Bootstrapping 20, se define al *bootstrap* o remuestreo como un método para derivar estimaciones robustas de errores típicos e intervalos de confianza para estimaciones como la media, mediana, proporción, razón de ventajas, coeficiente de correlación o coeficientes de regresión. También se puede utilizar para realizar pruebas hipotéticas. *Bootstrap* es más útil como alternativa a estimaciones paramétricas en caso que el supuesto de esos métodos sea dudoso (como en el caso de los modelos con residuos heteroscedásticos que se ajustan a muestras pequeñas), o si la inferencia paramétrica no es posible o requiere fórmulas muy complicadas para el cálculo de errores típicos (como en el caso de cálculo de intervalos de confianza de mediana, cuartiles y otros percentiles).

En la práctica, para estimar los intervalos de confianza mencionados por IBM, el *bootstrap* trata a la muestra original como si fuera la población y a partir de ella se extraen, mediante muestreo con reposición, un elevado número de muestras para obtener la distribución de los parámetros y en base a ella estimar sus intervalos de confianza.

En este proyecto, el procedimiento del *bootstrap* se replica para cada uno de los coeficientes estimados en la regresión de la tabla 2.21 tomando 1000 muestras aleatorias de 5130 individuos cada una. El tamaño de la muestra ha sido calculado por medio de la siguiente fórmula:

$$n = \frac{Npq}{(N-1)D + pq}, \quad D = \left(\frac{B}{z_{\alpha/2}}\right)^2 \tag{2.13}$$

Donde,

N: Tamaño de la población, en el *bootstrap* se toma el tamaño de la muestra de aprendizaje, 9.691.

p: La proporción de clientes buenos de la muestra de aprendizaje, 0,6779.

q: La proporción de clientes malos de la muestra de aprendizaje, 0,2420.

B: Tamaño del error, 1%.

$z_{\alpha/2=0,005} = 2,58$: Es el fractil de la distribución normal estándar para un nivel de confianza del 99%.

Como siguiente paso, se realiza una regresión con cada una de las muestras considerando las variables del modelo preliminar; con ello es posible construir un histograma para todos los coeficientes estimados y verificar, mediante un test de bondad de ajuste, su distribución teórica. Por ejemplo, la figura 2.8 muestra el histograma del coeficiente de la variable antigüedad de residencia (WOE), el mismo que sugiere una distribución normal, para demostrarlo estadísticamente se utiliza el test Chi-Cuadrado descrito en la sección 2.5.2. Los resultados del test se muestran en la tabla 2.22.

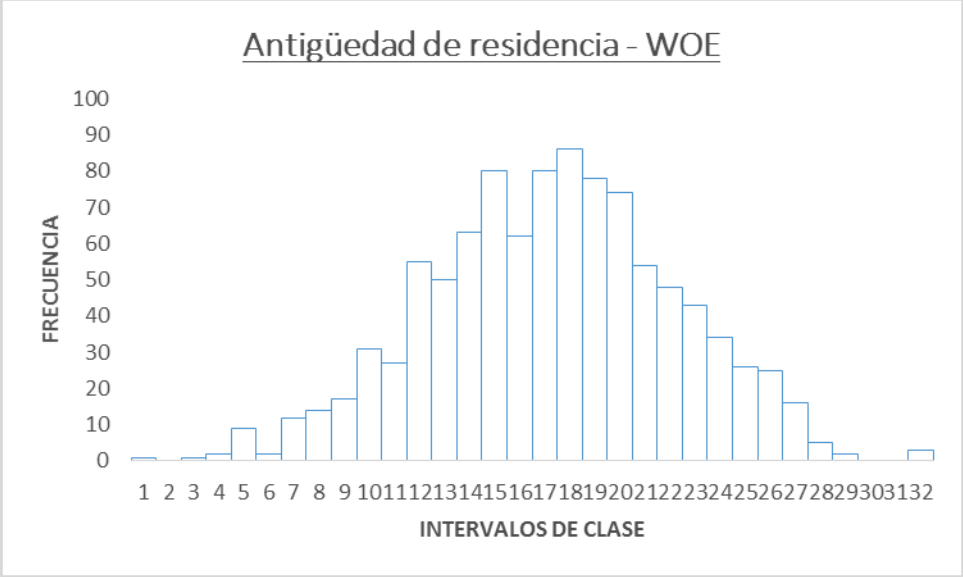

Figura 2.8 Histograma del coeficiente de la variable antigüedad de residencia (WOE)

Para un nivel de confianza del 95% se acepta la hipótesis nula pues el valor del estadístico $\chi_c^2 = 36,6$ es menor que el valor crítico $\chi_\alpha^2 = 42,56$, por lo tanto la distribución obtenida por *bootstrap* se comporta como una normal con media 0,53 y

desviación estándar 0,21. A partir de esa distribución se obtiene el intervalo de confianza para la estimación del coeficiente, como sigue:

$$\left[\bar{B} - t_{n-1}(1-\alpha/2)\frac{s}{\sqrt{n-1}}\ ; \bar{B} + t_{n-1}(1-\alpha/2)\frac{s}{\sqrt{n-1}}\right]$$

Donde, $\bar{B}$ es la media de los 1000 valores obtenidos por *bootstrap* para el coeficiente de la variable antigüedad de residencia (WOE), s es la desviación estándar de los mismos valores, y $t_{n-1}(1-\alpha/2)$ es el fractil de orden $1-\alpha/2$ de la ley t de Student con $n-1$ grados de libertad.

Tabla 2.22 Test Chi-Cuadrado del coeficiente de la variable antigüedad de vivienda

Intervalos de clase		Frec. Observ.	Frec. Teórica	Chi-Cuadrado
0	-0,14	1	0,7131	0,1154
-0,14	-0,10	0	0,6938	0,6938
-0,09	-0,05	1	1,2657	0,0558
-0,05	-0,01	2	2,2171	0,0213
-0,01	0,03	9	3,7289	7,4511
0,03	0,08	2	6,0218	2,6861
0,08	0,12	12	9,3375	0,7592
0,12	0,16	14	13,9021	0,0007
0,16	0,20	17	19,8740	0,4156
0,20	0,25	31	27,2798	0,5073
0,25	0,29	27	35,9541	2,2300
0,29	0,33	55	45,4997	1,9836
0,33	0,37	50	55,2868	0,5056
0,38	0,42	63	64,5040	0,0351
0,42	0,46	80	72,2609	0,8288
0,46	0,50	62	77,7272	3,1822
0,50	0,54	80	80,2778	0,0010
0,55	0,59	86	79,6104	0,5128
0,59	0,63	78	75,8048	0,0636
0,63	0,67	74	69,3069	0,3178
0,67	0,72	54	60,8427	0,7696
0,72	0,76	48	51,2852	0,2104
0,76	0,80	43	41,5077	0,0537
0,80	0,84	34	32,2565	0,0942
0,84	0,89	26	24,0690	0,1549
0,89	0,93	25	17,2445	3,4879
0,93	0,97	16	11,8631	1,4426
0,97	1,01	5	7,8360	1,0264
1,02	1,06	2	4,9698	1,7747
1,06	1,10	0	3,0265	3,0265
1,10	1,14	0	1,7697	1,7697
1,14	10,00	3	2,0626	0,4260
Suma		**1000**	**1000,00**	**36,6034**

Para la variable en cuestión, el intervalo de confianza es:

$$[0{,}5170; 0{,}5430]$$

Dado que el cero no se encuentra dentro del intervalo de confianza, se concluye que el coeficiente estimado de la variable antigüedad de residencia (WOE) es estadísticamente diferente se cero.

Así mismo se ha procedido con el resto de coeficientes de la regresión preliminar, concluyendo que las variables 8, 13 y 24 (de la tabla 2.21) deben eliminarse del modelo, pues los intervalos de confianza contienen el cero. Por lo tanto, la regresión final se compone de todas las variables que se muestran en la tabla 2.23.

Tabla 2.23 Estimación de los coeficientes de la regresión final

Nro. Variable	Variables	B	Sig.	Exp(B)	95% C.I.for EXP(B)	
					Lower	Upper
V1	AntiguedadResidencia_WOE	0,55005	0,52%	1,73	1,18	2,55
V2	MaxMontoHistoricoMicro_WOE	0,97724	0,00%	2,66	2,02	3,50
V3	PorEndAntes12UM_WOE	0,71217	0,04%	2,04	1,37	3,02
V4	DiasMora12_24_LN	-0,09758	0,00%	0,91	0,88	0,94
V5	CR_24UM_LN	-0,06641	1,32%	0,94	0,89	0,99
V6	PorEnd_U12UM_LN	-0,03675	1,96%	0,96	0,93	0,99
V7	EntVen24_LN	-0,05365	0,69%	0,95	0,91	0,99
V8	NumeroHijos2(1)	-0,18532	2,77%	0,83	0,70	0,98
V9	Salario1(1)	-0,34687	0,00%	0,71	0,62	0,81
V10	Salario2(1)	-0,18669	0,35%	0,83	0,73	0,94
V11	Edad1(1)	-0,18532	2,77%	0,83	0,70	0,98
V12	Edad4(1)	0,28705	0,00%	1,33	1,17	1,52
V13	Edad5(1)	0,41753	0,00%	1,52	1,32	1,74
V14	AntiguedadRelacionDependencia(1)	-0,20545	0,03%	0,81	0,73	0,91
V15	ProvinciaNacimiento1(1)	0,41917	0,00%	1,52	1,34	1,72
V16	ProvinciaNacimiento2(1)	0,39425	0,00%	1,48	1,29	1,70
V17	ProvinciaNacimiento3(1)	0,23002	0,82%	1,26	1,06	1,49
V18	AntiguedadViv_TipoVivienda1(1)	0,29579	0,06%	1,34	1,13	1,59
V19	Antiguedad_ProvDomicilio2(1)	-0,32062	0,02%	0,73	0,61	0,86
V20	DiasMoraPromedio12UM_MoraCorte2(1)	0,71274	0,00%	2,04	1,61	2,59
V21	DiasMoraPromedio12UM(1)	-0,48008	0,00%	0,62	0,54	0,71
V22	AntiguedadSF_MaxMontoHisConsumo3(1)	0,30757	0,33%	1,36	1,11	1,67
V23	OperacionesVigentes1(1)	-0,25634	0,08%	0,77	0,67	0,90
V24	ValortotalXVencer_His2(1)	0,19386	0,21%	1,21	1,07	1,37
V25	Maxvencido_VecesAlDia12UM2(1)	0,32131	0,00%	1,38	1,23	1,54
V26	Maxvencido_VecesAlDia12UM3(1)	-0,42154	0,00%	0,66	0,55	0,79
V27	Antiguedad12UM_PorEndCorte3(1)	-0,18414	1,71%	0,83	0,71	0,97
V28	EndPromSF_12UM1(1)	0,27108	0,15%	1,31	1,11	1,55
V29	Constant	-0,60919	0,01%	0,54		

2.6.2 DESCRIPCIÓN DE LAS VARIABLES DEL MODELO.

En este punto se describe la definición de cada una de las variables que se incluyeron en el modelo final, además se revisa la consistencia de los signos de cada coeficiente estimado en la regresión.

1. **AntiguedadResidencia_WOE (V1):** Variable que indica el número de meses que la persona evaluada reside en su vivienda, a cada categoría de la variable se ha asociado el valor del WOE, de la siguiente manera.

$$V1 = \begin{cases} -0{,}1229 & si\ AntigResid \leq 99 \\ 0{,}0473 & si\ 99 < AntigResid \leq 164 \\ 0{,}1877 & si\ AntigResid > 164 \end{cases}$$

   El coeficiente estimado es positivo lo cual significa que, para las personas con antigüedad de residencia menor a 99 meses, la probabilidad de ser buen cliente disminuye; en cambio si su antigüedad es mayor a 99 y menor igual a 164 la probabilidad aumenta. El incremento en la probabilidad de ser buen cliente es mayor para las personas cuya antigüedad es mayor a 164 meses.

2. **MaxMontoHistoricoMicro_WOE (V2):** Indica el máximo monto otorgado a una persona por microcréditos en el sistema financiero, a cada categoría de la variable se ha asignado el valor del WOE, de la siguiente manera:

$$V2 = \begin{cases} 0{,}0820 & si\ Monto = 0 \\ -0{,}4417 & si\ 0 < Monto \leq 1510{,}66 \\ -0{,}0165 & si\ Monto > 1510{,}66 \end{cases}$$

   El coeficiente estimado es positivo lo cual significa que, para las personas con máximo monto histórico en microcrédito igual a cero, la probabilidad de ser buen cliente aumenta, en cambio si el monto es mayor a 0 y menor igual a 1510,66 la probabilidad disminuye; la penalización en la probabilidad de ser buen cliente es menor para las personas cuyo monto supera 1510,66 dólares.

3. **PorEndAntes12UM_WOE (V3):** Es la relación del endeudamiento promedio entre los meses 12 y 24 antes del mes de evaluación, y el ingreso del individuo, se calcula por la relación:

$$PorEndAntes12UM = \frac{EndeuProm12_24}{Ingreso}$$

Donde $EndeuProm12_24$ el promedio de saldos reportados por el buró de créditos entre los meses 12 y 24 previos al mes de evaluación del individuo.
Al resultado de la variable $PorEndAntes12UM$ se ha asignado el valor del WOE asociado a cada categoría construida con el árbol de decisión, de la siguiente manera:

$$V3 = \begin{cases} -0{,}2922 \; si \; PorEndAntes12UM = 0 \\ 0{,}0909 \quad si \; PorEndAntes12UM > 0 \end{cases}$$

El coeficiente estimado es positivo lo cual significa que, para las personas con endeudamiento promedio igual a cero, la probabilidad de ser buen cliente disminuye, en cambio si el endeudamiento es mayor a 0 la probabilidad aumenta.

4. **DiasMora12_24_LN (V4):** Es el logaritmo natural del indicador de mora, definido por la relación:

$$IndicadorMora = \frac{DíasMora12UM}{DíasMora24UM}$$

Donde $DíasMora12UM$ es el máximo de días de mora reportados por el buró de créditos en los 12 últimos meses, contados a partir de la fecha de evaluación; $DíasMora24UM$ se define de manera similar pero en una ventana de 24 meses.

Es decir, la variable busca evaluar la recuperación, en términos de morosidad, del cliente nuevo de ABC en los últimos 12 meses. La construcción del logaritmo se realiza como sigue.

$$V4 = \begin{cases} DiasMora12UM & si\ DiasMora24UM = 0 \\ ln(IndicadorMora + 0{,}001) & si\ \ DiasMora24UM \neq 0 \end{cases}$$

El coeficiente es negativo lo cual significa que, para valores menores a uno del indicador de mora (disminuye la morosidad en los 12 últimos meses respecto a la ventana de 24 meses), incrementa la probabilidad de ser buen cliente pues el logaritmo en estos casos es negativo. Por el contrario, si el indicador de mora es mayor o igual a uno la probabilidad disminuye.

5. **CR_24UM_LN (V5):** Mide la cartera en riesgo de los últimos 24 meses, previo al cálculo del logaritmo se define la cartera en riesgo como:

$$CR_24UM = \frac{MayorValorVencido + MayorValorNDI}{MayorDeuda24UM}$$

Donde $MayorValorVencido$ es el máximo saldo vencido reportado por el buró de créditos, $MayorValorNDI$ es el máximo saldo por vencer que no genera interés (NDI: no devenga interés) y $MayorDeuda24UM$ es el máximo saldo que reporta del cliente nuevo de ABC en el sistema financiero; las tres variables se calculan dentro de una ventana de 24 meses.

La variable V5 busca evaluar el incumplimiento en los pagos de las cuotas pactadas por el cliente nuevo en instituciones diferentes a ABC, en términos de saldo vencido y NDI; se construye a partir de la siguiente definición:

$$V5 = ln(CR_24UM + 0{,}001)$$

El coeficiente estimado es negativo, lo cual indica que para valores del CR_24UM entre cero y uno o equivalentemente V5<0 (saldos pequeños por los que el individuo registra mora), la probabilidad de ser buen cliente aumenta; mientras que para valores de CR_24UM mayores que uno la probabilidad disminuye (saldos grandes por los que el individuo registra mora).

6. **PorEnd_U12UM_LN (V6):** Es la relación entre el endeudamiento promedio de los 12 últimos meses y el ingreso del individuo, se calcula por la relación:

$$PorEnd12UM = \frac{EndeuProm12UM}{Ingreso}$$

Donde $EndeuProm12UM$ el promedio de saldos reportados por el buró de créditos en los 12 últimos meses, contados hacia atrás a partir de la fecha de evaluación.

Con el resultado de la variable $\mathrm{PorEnd12UM}$ se ha calculado el logaritmo natural como sigue.

$$V6 = ln(PorEnd12UM + 0{,}001)$$

El coeficiente estimado es negativo lo cual significa que, para valores entre cero y uno de $PorEnd12UM$ o equivalentemente V6<0 (cuando el endeudamiento promedio no supera el salario mensual), la probabilidad de ser buen cliente aumenta; en cambio para valores mayores a uno de $PorEnd12UM$ (cuando el endeudamiento promedio supera el salario mensual), la probabilidad disminuye.

7. **EntVen24_LN (V7):** Es el logaritmo natural del número de entidades en las que el individuo ha registrado al menos un día de morosidad durante los últimos 24 meses, antes de empezar su relación con ABC; matemáticamente se expresa por:

$$V7 = ln(EntVen24UM + 0{,}001)$$

El coeficiente estimado es negativo lo cual indica que entre mayor es el número de entidades en las que el individuo ha incumplido con sus pagos, menor es la probabilidad de ser buen cliente.

Las variables que se describen en la tabla 2.24 son dicotómicas (desde V8 hasta V28). Para todas ellas, el signo positivo del coeficiente indicará que la probabilidad de ser buen cliente aumenta siempre y cuando el individuo cumpla con la propiedad que señala la variable dicotómica, es decir que el valor sea 1. En cambio, el signo negativo indicará que el individuo es más propenso a ser un mal cliente en ABC si cumple con la propiedad sugerida por la variable dicotómica.

En la práctica, cuando se trabaja con variables dicotómicas en una regresión logística y los coeficientes estimados son positivos se dice que premian la probabilidad de ser buen cliente; por el contrario, cuando el coeficiente es negativo se dice la variable castiga la probabilidad se ser buen cliente, dependiendo de la codificación de la variable.

29. **Constante (V29):** Por default, en ABC la probabilidad de ser buen cliente para una persona de la que no se conoce información alguna es igual a 0,352. El valor se calcula con la constante del modelo, asumiendo que el resto de variables es cero, o sea:

$$P(buen\ cliente) = \frac{e^{constante}}{1 + e^{constante}} = \frac{e^{-0{,}609}}{1 + e^{-0{,}609}} = 0{,}352 \qquad (2.14)$$

Tabla 2.24 Descripción de las variables dicotómicas del modelo final

Nro.	Variable dicotómica	Criterio	Propiedad
V8	NumeroHijos2	Castiga	Tiene uno o dos hijos.
V9	Salario1	Castiga	Salario menor a 445,25 dólares.
V10	Salario2	Castiga	Salario mayor a 445,35 y menor o igual que 777,5 dólares.
V11	Edad1	Castiga	Edad menor o igual a 23 años.
V12	Edad4	Premia	Edad mayor igual 38 y menor igual a 46 años
V13	Edad5	Premia	Edad mayor igual a 47 años
V14	AntiguedadRelacionDependencia	Castiga	Antigüedad en relación de dependencia menor o igual a 48 meses.
V15	ProvinciaNacimiento1	Premia	Provincia de nacimiento CARCHI, CHIMBORAZO, COTOPAXI, LOJA.
V16	ProvinciaNacimiento2	Premia	Provincia de nacimiento IMBABURA, PICHINCHA, SANTO DOMINGO.
V17	ProvinciaNacimiento3	Premia	Provincia de nacimiento GUAYAS.
V18	AntiguedadViv_TipoVivienda1	Premia	Antigüedad de residencia mayor a 99 meses y tipo de vivienda Propia.
V19	AntiguedadSF_ProvDomicilio2	Castiga	Antigüedad en el sistema financiero menor igual a 9 meses y provincia de domicilio diferente de PICHINCHA, IMBABURA, TUNGURAHUA, COTOPAXI, AZUAY Y CHIMBORAZO.
V20	DiasMoraPromedio12UM_MoraCorte2	Premia	Promedio de días de mora en los 12 últimos meses mayor a cero y mora del último mes igual a cero.
V21	DiasMoraPromedio12UM	Castiga	Promedio de días de mora en los 12 últimos meses mayor a cero
V22	AntiguedadSF_MaxMontoHisConsumo3	Premia	Antigüedad en el sistema financiero mayor a 23 meses y máximo monto histórico de consumo entre 338,58 y 4500.
V23	OperacionesVigentes1	Castiga	No tiene operaciones vigentes en el mes de evaluación.
V24	ValortotalXVencer_His2	Premia	Saldo total por vencer histórico entre 21.134,28 y 133.033,44
V25	Maxvencido_VecesAlDia12UM2	Premia	El máximo saldo vencido histórico es menor igual a 27,19 dólares y durante los 12 últimos meses no ha tenido operaciones vencidas.
V26	Maxvencido_VecesAlDia12UM3	Castiga	El máximo saldo vencido histórico es mayor a 27,19 dólares y en 4 de los 12 últimos meses registra moras mayores a cero.
V27	Antiguedad12UM_PorEndCorte3	Castiga	El número de veces en que ha sido reportado el cliente en los 12 últimos meses es menor igual a 9 y el porcentaje de endeudamiento en el mes de evaluación es mayor a 0,6 (el porcentaje de endeudamiento es el resultado de dividir el saldo total por el ingreso en el mes de evaluación)
V28	EndPromSF_12UM1(1)	Premia	El endeudamiento promedio en el sistema financiero durante los 12 últimos meses es menor a 360,64. El endeudamiento se calcula como el promedio de los saldos totales registrados en los 12 últimos periodos.

2.7 MEDIDAS DE BONDAD DE AJUSTE DEL MODELO.

El objetivo es, estimar el ajuste de los datos que se ha alcanzado con el modelo de regresión logística descrito en este proyecto. En la práctica, no es posible explicar completamente el comportamiento de la variable dependiente, tipo de cliente; básicamente se debe a que el individuo evaluado es nuevo en ABC y apenas se conoce su comportamiento de pago con la competencia. A medida que se desarrolla el ciclo de crédito en la institución se tiene mejor información del cliente y eso ayuda a construir modelos con mayor poder de predicción. Al respecto, en la figura 2.9 se muestra el criterio de la calificadora de activos de riesgo Equifax, que respalda la afirmación citada en este apartado.

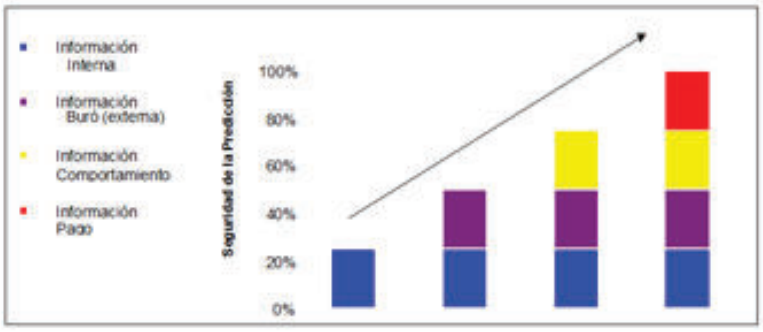

Figura 2.9 Poder predictivo de modelos scoring. Fuente: Equifax

2.7.1 COEFICIENTES R2.

La regresión múltiple por mínimos cuadrados ofrece el coeficiente de determinación R^2 como un indicador de ajuste del modelo; básicamente es una medida de la bondad del ajuste que compara la suma de cuadrados de los residuos. Sin embargo, en el modelo de regresión logística binaria no conviene usar directamente ese valor, pero es posible determinar medidas alternativas como el R^2 de Cox y Snell que se utiliza de la misma manera, los valores más grandes indican un mejor ajuste del modelo. Negelkerke propuso una modificación a la medida de Cox y Snell, en la tabla 2.25 se muestran dichas medidas del ajuste del modelo.

Tabla 2.25 Coeficientes R2

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	10142,285	,105	,153

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Si bien los coeficientes R^2 de Cox & Snell y Negelkerke no son muy grandes, se ha seleccionado el modelo con los mayores valores de R^2, además la figura 2.9 ayuda a justificar los resultados obtenidos.

2.7.2 BONDAD DE AJUSTE DE HOSMER-LEMESHOW.

Hosmer-Lemeshow proponen crear 10 grupos sobre las probabilidades estimadas por el modelo y en base a ellas construir una tabla de contingencia que compara las frecuencias observadas y esperadas del tipo de cliente. El test se basa en el estadístico Chi-Cuadrado de bondad de ajuste y contrasta la hipótesis inicial:

$$H_o: El\ modelo\ ajusta\ bien\ a\ los\ datos$$

El paquete estadístico SPSS obtiene los intervalos de clase directamente, así como los valores de la frecuencia observada y esperada; la tabla 2.26 muestra esta situación.

Tabla 2.26 Tabla de contingencia para el test Hosmer-Lemeshow

Intervalos de clase	TipoCliente = 0		TipoCliente = 1		Total
	Observado	Esperado	Observado	Esperado	
1	545	555,876	424	413,124	969
2	415	402,750	554	566,250	969
3	324	331,428	645	637,572	969
4	313	279,809	656	689,191	969
5	222	240,473	747	728,527	969
6	217	206,952	752	762,048	969
7	171	177,904	798	791,096	969
8	146	149,319	823	819,681	969
9	114	120,793	855	848,207	969
10	82	83,696	888	886,304	970

El resultado del test Chi-Cuadrado acepta la hipótesis nula al 95% de confianza, es decir el modelo está bien ajustado.

Tabla 2.27 Prueba de Hosmer-Lemeshow para la bondad de ajuste del modelo

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	10,323	8	,243

2.7.3 ESTADÍSTICO DE WALD.

Contrasta la hipótesis nula:

$$H_o: \beta_k = 0, \quad k = \{1,2, \dots ,29\}$$

Es decir, se busca verificar que la estimación de los parámetros, uno a uno, es estadísticamente igual a cero. Este contraste está basado en la normalidad asintótica de los estimadores. En el programa SPSS es posible mirar los resultados para cada uno de los coeficientes; para ejemplificar esta situación, en la tabla 2.28 se puede apreciar los resultados del estadístico de Wald para las estimaciones de los coeficientes de las cinco primeras variables de la regresión final así como el p-valor asociado.

Tabla 2.28 Resultados para el estadístico de Wald

Nro.	Variables	B	S.E.	Wald	df	Sig.
V1	AntiguedadVivienda_WOE	0,55005	0,19667	7,82254	1	0,52%
V2	MaxMontoHistoricoMicro_WOE	0,97724	0,13987	48,81577	1	0,00%
V3	PorEndAntes12UM_WOE	0,71217	0,20095	12,55938	1	0,04%
V4	DiasMora12_24_LN	-0,09758	0,01831	28,41553	1	0,00%
V5	CR_36UM_LN	-0,06641	0,02679	6,14684	1	1,32%

Para las variables del ejemplo y para todas las que conforman la regresión final, el estadístico de Wald se calcula como:

$$Wald_{V1} = \left(\frac{\widehat{\beta_{V1}}}{\widehat{SE_{V1}}}\right)^2 = \left(\frac{0{,}55005}{0{,}19667}\right)^2 = 7{,}82254$$

Además, el p-valor asociado es menor al 5% para todas las estimaciones, esto quiere decir que al 95% de confianza se rechaza la hipótesis nula; por lo tanto, se concluye que los coeficientes estimados son estadísticamente diferentes de cero.

2.7.4 PRUEBA DE OMNIBUS.

El test contrasta la significancia conjunta de las estimaciones del modelo, es decir la hipótesis nula y alternativa son, respectivamente:

$$H_o: \beta_1 = \beta_2 = \cdots = \beta_{29} = 0$$

$$H_o: \beta_k \neq 0, k = \{1,2, \ldots ,29\}$$

Tabla 2.29 Prueba Omnibus para la significancia conjunta de las estimaciones

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	1021,449	28	,000
	Block	1021,449	28	,000
	Model	1021,449	28	,000

Los resultados de la tabla 2.29 indican que se rechaza la hipótesis nula al 95% de confianza, pues el p–valor es menor que el 5%, por lo tanto la estimación de los coeficientes son estadísticamente significativos (diferentes de cero).

2.8 CORRECCIÓN DE LA CONSTANTE DEL MODELO.

Debido a que en el modelo se han usado muestras desproporcionadas de clientes buenos y malos (aproximadamente 76% vs. 24%), es preciso realizar la corrección de la constante de la regresión, el tema ha sido propuesto por Maddala[15]. Al no usar muestras balanceadas, con proporción 50% de buenos y 50% de malos, el punto de corte de la curva logística con el eje de las abscisas se desplaza como se puede apreciar en la figura 2.10.

Para explicar el efecto del desplazamiento de la regresión estimada, en este proyecto, se toma la ecuación general de regresión logística en términos del logaritmo natural.

$$ln\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{k=1}^{K} \beta_k$$

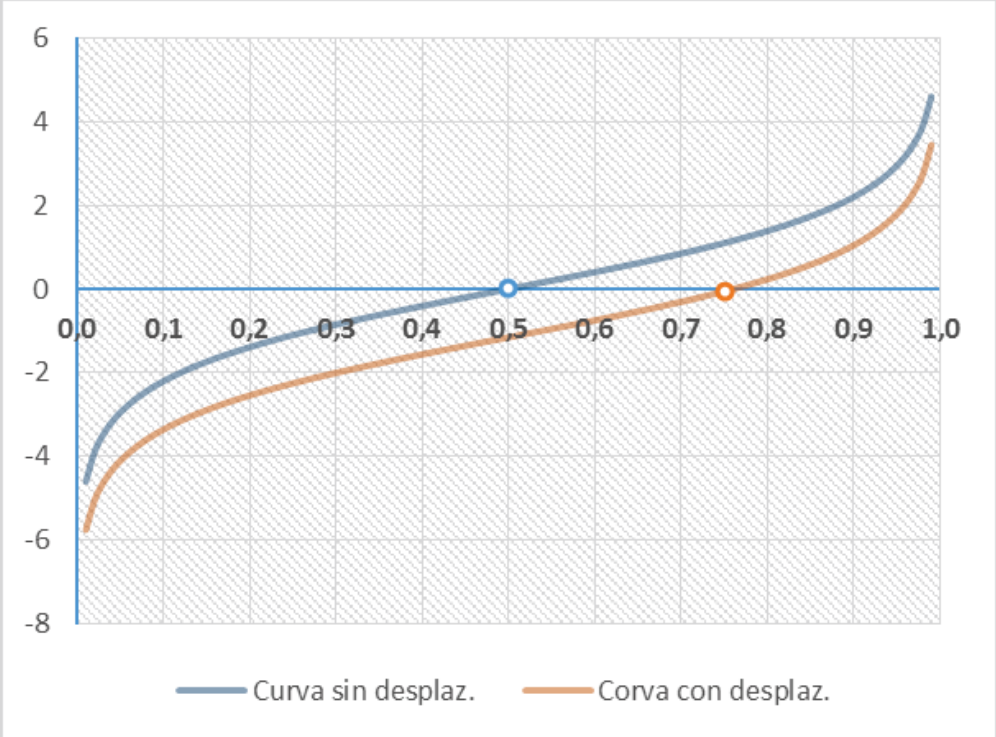

Figura 2.10 Desplazamiento de la curva logística

[15] G.S. Maddala, Introducción a la Econometría, segunda edición, Prentice-Hall Hispanoamericana, México, 1996, págs. 378-379.

Donde p es la probabilidad de ser buen cliente. Si se consideran las proporciones de clientes buenos y malos, p_1 y p_2, respectivamente; el desplazamiento de la curva logística se vería afectada por:

$$ln\left(\frac{p}{1-p}\times\frac{p_2}{p_1}\right)$$

Para este proyecto se tiene que $p_1 = 0{,}76$ y $p_2 = 0{,}24$. Por lo tanto, gráficamente el desplazamiento de la curva se ve afectado como la línea naranja de la figura 2.10. El problema influye básicamente cuando se hace una clasificación de clientes buenos y malos a partir de la probabilidad estimada por la regresión; bajo el esquema de muestras balanceadas, se debe tomar como punto de corte 0,50; pero en el caso de muestras desproporcionadas es un error mantener ese punto para predecir el tipo de cliente, bueno o malo.

Maddala demuestra que basta con modificar la constante del modelo para compensar el desplazamiento de la curva logística provocado por trabajar con muestras desproporcionadas. Para corregir el valor de la constante del modelo basta con restar el valor,

$$ln(p_1) - ln(p_2)$$

Es decir, a la constante de la regresión final obtenida en este estudio se resta la cantidad:

$$ln(0{,}76) - ln(0{,}24) = 1{,}1673$$

De manera que la constante corregida del modelo es,

$$-0{,}6092 - 1{,}1673 = -1{,}7766$$

2.8.1 EFECTOS DE CORREGIR LA CONSTANTE DEL MODELO.

El efecto de la modificación de la constante es notorio, como se dijo en la sección anterior, cuando se busca predecir la variable dependiente del modelo (tipo de cliente) y en base a ello construir la llamada matriz de confusión.

La matriz de confusión es una tabla de contingencia que presenta el tipo de cliente real y su pronóstico, el primero corresponde a la variable dependiente usada en la construcción del modelo y el segundo resulta de fijar un punto de corte sobre la probabilidad que se ha estimado con la regresión logística.

La tabla 2.30 muestra la matriz de confusión para el modelo, a partir de ella se define la asertividad de la clasificación como sigue:

Acierto clientes buenos: $\frac{a}{a+c}$

Acierto clientes malos: $\frac{d}{b+d}$

Acierto general del modelo: $\frac{a+d}{a+b+c+d}$

Tabla 2.30 Matriz de confusión

	Real		
Predicción	**Buenos**	**Malos**	**TOTAL**
Buenos	a	b	**a+b**
Malos	c	d	**c+d**
TOTAL	**a+c**	**b+d**	**a+b+c+d**

Tomando diferentes puntos de corte sobre la probabilidad estimada con la regresión logística que contiene la constante corregida, se pueden construir los indicadores de acierto de clientes buenos y malos, tal como se puede ver en la tabla 2.31.

Tabla 2.31 Acierto en la predicción de clientes buenos y malos

Pto. Corte	Acierto Buenos	Acierto Malos
0,385	79,9%	47,0%
0,425	73,8%	55,4%
0,465	66,3%	64,3%
0,505	57,8%	71,1%
0,545	48,9%	78,6%

Las líneas verticales de color azul de la figura 2.11 muestran gráficamente los puntos de corte considerados en la tabla 2.31. En general, se pueden calcular varios puntos de corte así como los respectivos valores de asertividad de clientes buenos y malos conocidos también como sensibilidad y especificidad, respectivamente. El paquete estadístico SPSS permite determinar dichos valores y a partir de ellos es posible obtener las curvas de la figura 2.11. El punto de intersección es (0,4745; 0,653); es decir, cuando el punto de corte de la probabilidad estimada por el modelo es igual a 0,4745 el acierto de buenos y malos es igual al 65,30%.

La figura 2.12, en cambio, muestra las curvas de acierto en la predicción de clientes buenos y malos, construidos a partir de la probabilidad estimada antes de corregir la constante del modelo. Las curvas se intersecan en el punto (0,744; 0,653), eso significa que al tomar como punto de corte 0,744 el acierto de buenos y malos, a la vez, es 65,30%.

En conclusión, el valor de asertividad es el mismo tanto en la figura 2.11 como en 2.12, pero es evidente que el punto de corte en el eje de las abscisas es diferente cuando se ha corregido la constante (figura 2.11) que cuando se trabaja con la constante original del modelo (figura 2.12). Entonces, el efecto de modificar la constante ha sido corregir el desplazamiento de la curva logística descrito en la figura 2.10. Es así que, las observaciones correctamente predichas, a partir del modelo que se obtiene luego de corregir la constante, se resumen en la matriz de confusión de la tabla 2.32.

En términos de porcentaje, las observaciones correctamente predichas para la población de clientes buenos, malos y global representan el 65,30%.

Tabla 2.32 Porcentaje de observaciones correctamente predichas

	Real		
Predicción	**Buenos**	**Malos**	**TOTAL**
Buenos	4.664	886	**5.550**
Malos	2.478	1.663	**4.141**
TOTAL	**7.142**	**2.549**	**9.691**

Es importante mencionar que en este proyecto el resultado de mayor contribución es la probabilidad estimada que se obtiene luego de corregir la constante de la regresión logística y no la predicción del tipo de cliente, pues con base a la probabilidad se estima un perfil de riesgo que sirve de *input* para calcular el monto a ofertar a un potencial cliente nuevo de ABC, mismo que se construye con la ayuda del modelo de inferencia difuso.

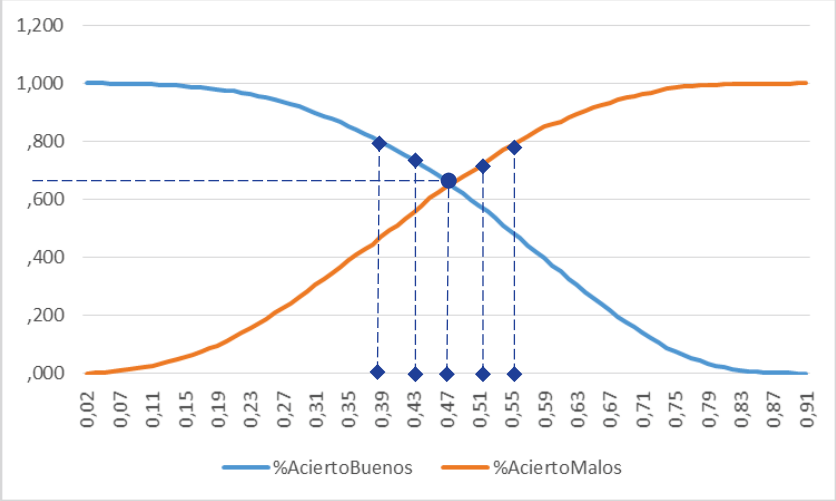

Figura 2.11 Curva de aciertos en la predicción de clientes buenos y malos luego de corregir la constante del modelo

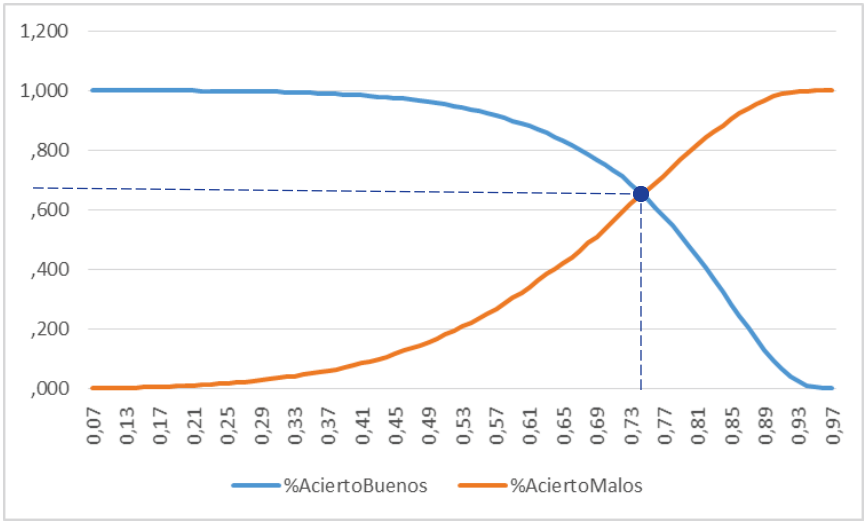

Figura 2.12 Curva de aciertos en la predicción de clientes buenos y malos antes de corregir la constante del modelo

2.9 VALORACIÓN DE LA CALIDAD DE DISCRIMINACIÓN.

En esta etapa se busca evaluar la eficiencia del modelo para discriminar la población de buenos y malos, para tal efecto se utiliza el test de Kolmogorov-Smirnov y el coeficiente de Gini.

2.9.1 TEST DE KOLMOGOROV-SMIRNOV (KS).

Indica la mayor diferencia absoluta entre las distribuciones empíricas de la probabilidad estimada para individuos buenos y malos. Al respecto de este indicador la SBS recomienda el KS de los modelos de originación no sea menor al 25%.

Matemáticamente, el KS se calcula con la relación:

$$KS = \max_i |P(B \leq b_i) - P(M \leq m_i)| \tag{2.15}$$

Donde $P(B \leq b_i)$ y $P(M \leq m_i)$ son las distribuciones de acumulación de buenos y malos hasta el decil i, respectivamente.

Para el modelo estimado en este proyecto se ha obtenido un KS igual a 30,8%. Con este valor se puede concluir que el modelo es capaz de discriminar entre buenos y malos a la población de potenciales clientes nuevos de ABC; además el indicador se encuentra sobre el nivel exigido por la SBS. En la figura 2.13 se muestra gráficamente el resultado del KS, mientras que en la tabla 2.33 se puede observar un resumen analítico del indicador.

2.9.2 COEFICIENTE DE GINI.

El coeficiente de Gini es un indicador que mide de igualdad o desigualdad distributiva de una población; tomando la distribución acumulada por deciles para clientes malos

y buenos de la tabla 2.33, se obtiene la curva de la figura 2.14. Por ejemplo, en el decil 1 el porcentaje acumulado de clientes malos es mayor que el de buenos, 21,4% y 6,0% respectivamente; en el decil 9 sucede algo similar, el porcentaje acumulado de malos es 87,7% y de buenos es 96,8%. Este resultado es el deseado, pues es natural pensar que cuando la probabilidad estimada de ser buen cliente aumenta, el porcentaje de clientes buenos también debe ser mayor en cada decil, en cambio el porcentaje de malos debe disminuir. En la tabla 2.33 es posible apreciar numéricamente la afirmación planteada.

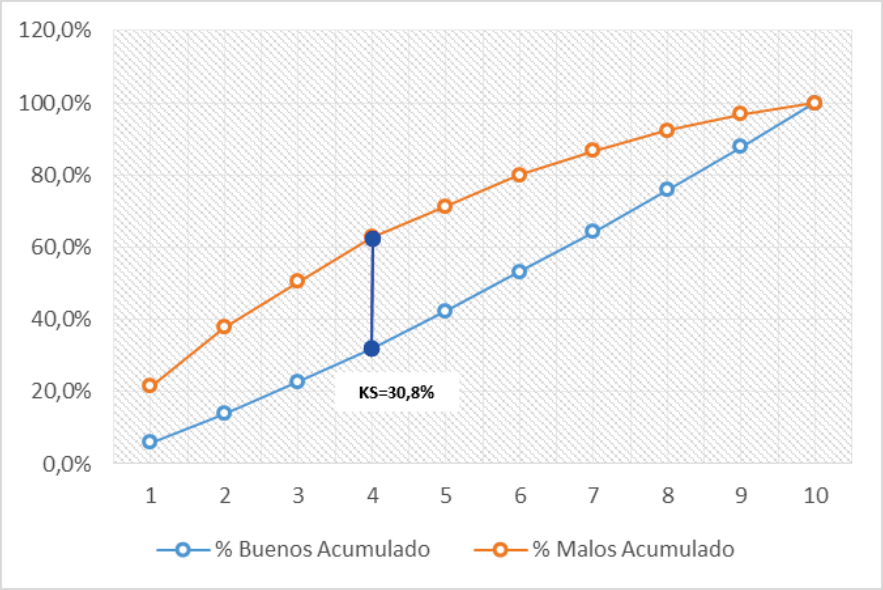

Figura 2.13 Indicador KS del modelo

Con base a las áreas A y B de la figura 2.14, el índice de Gini se define como,

$$GINI = \frac{A}{A + B}$$

En una situación de perfecta igualdad o equidad de distribución el indicador es cero, pues el área A es cero; mientras que en una situación de perfecta desigualdad el

coeficiente es 1, pues el área B es cero. En los modelos estadísticos discriminantes, se busca que el coeficiente Gini sea lo más cercano a 1.

En la práctica, el coeficiente de Gini se calcula con la fórmula de Brown.

$$GINI = \left| 1 - \sum_{k=1}^{K-1} (M_{k+1} - M_k)\,(B_{k+1} - B_k) \right|$$

Donde M_k y B_k son las proporciones acumuladas hasta el decil k de clientes malos y buenos, respectivamente.

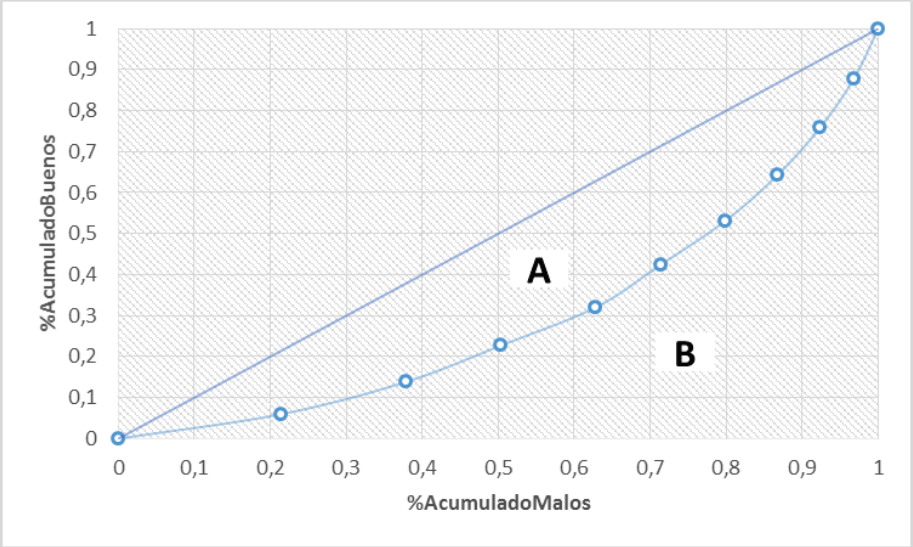

Figura 2.14 Coeficiente de Gini

Tabla 2.33 Cálculo del indicador KS y Gini

Percentil Score	# Buenos	% Buenos	% Buenos Acumulado	# Malos	% Malos	% Malos Acumulado	K - S	Variación Buenos	Variación malos	GINI
D1	425	6,0%	6,0%	545	21,4%	21,4%	15,4%			
D2	562	7,9%	13,8%	419	16,4%	37,8%	24,0%	20%	16%	3,25%
D3	638	8,9%	22,8%	321	12,6%	50,4%	27,7%	37%	13%	4,61%
D4	664	9,3%	32,0%	317	12,4%	62,8%	30,8%	55%	12%	6,82%
D5	739	10,3%	42,4%	217	8,5%	71,4%	29,0%	74%	9%	6,34%
D6	767	10,7%	53,1%	218	8,6%	79,9%	26,8%	96%	9%	8,17%
D7	791	11,1%	64,2%	173	6,8%	86,7%	22,5%	117%	7%	7,96%
D8	831	11,6%	75,8%	144	5,6%	92,3%	16,5%	140%	6%	7,91%
D9	847	11,9%	87,7%	113	4,4%	96,8%	9,1%	164%	4%	7,25%
D10	878	12,3%	100,0%	82	3,2%	100,0%	0,0%	188%	3%	6,04%
	7142	100,0%		**2549**	100,0%		**30,8%**			**41,66%**

En el modelo se ha obtenido un índice de Gini igual a 41,66% que es superior al mínimo establecido por la SBS (35%) para modelos de originación. El resultado se evidencia gráficamente en la figura 2.14, mientras que en la tabla 2.33 se aprecia analíticamente.

2.10 INTERVALOS DE CONFIANZA PARA LOS INDICADORES KS Y GINI.

En esta sección se estiman intervalos de confianza para la media de los indicadores KS y GINI, por medio de la metodología *bootstrap* que ya ha sido tratada en secciones anteriores.

Los intervalos servirán para compararlos con los valores de KS y GINI que se calculan con la muestra de testing. El objetivo es, verificar el poder discriminante del modelo usando una ventana de datos diferente a la de aprendizaje y de esta manera descartar la posibilidad de que el modelo presente buen nivel de discriminación únicamente con la muestra usada para su construcción.

Es común calcular los indicadores KS y GINI con los datos de la ventana de testing y compararlos con los obtenidos en la muestra de aprendizaje, sin embargo en este proyecto se verifica que los valores obtenidos con la muestra de testing se encuentren dentro de los intervalos de confianza estimados con la metodología del *bootstrap*; de esa forma se podrá concluir que el modelo conserva su calidad de discriminación.

Tomando 1000 muestras aleatorias (del periodo de aprendizaje) de tamaño 500 para cada una de ellas, la distribución acumulada empírica del indicador KS sigue una distribución normal con media 30,88% y desviación estándar 1,18%. El estadístico Chi-Cuadrado calculado es $\chi_c^2 = 18{,}34$ y el valor teórico es $\chi^2 = 19{,}67$, por lo tanto el test de bondad de ajuste asegura que la distribución del indicador KS se comporta como una curva normal con los parámetros descritos. Una situación similar sucede con el coeficiente GINI.

Los intervalos de confianza que se obtienen para ambos indicadores son,

$$KS = [28{,}45\%; 33{,}32\%]$$

$$GINI = [39{,}22\%; 44{,}28\%]$$

Nótese que los extremos superior e inferior de los intervalos de confianza se encuentran por encima del umbral que exige la SBS para los indicadores KS y GINI.

Para la muestra de testing los indicadores KS y GINI son 28,53% y 40,17% respectivamente, tal como se indica en la tabla 2.33. Los valores se encuentran dentro de los intervalos de confianza, por lo tanto se puede concluir que la calidad discriminante del modelo se mantiene si se usa una ventana de datos diferente a la considerada en la etapa de construcción.

Tabla 2.34 Indicadores KS y GINI para la muestra de testing

Decil	# Buenos	% Buenos	% Buenos Acumulado	# Malos	% Malos	% Malos Acumulado	K - S	Variación Buenos	Variación malos	GINI
D1	389	6,9%	6,9%	319	23,2%	23,2%	16,3%			
D2	485	8,6%	15,5%	205	14,9%	38,1%	22,6%	22,43%	15%	3,34%
D3	523	9,3%	24,8%	194	14,1%	52,2%	27,4%	40,32%	14%	5,69%
D4	549	9,7%	34,6%	150	10,9%	63,1%	28,5%	59,36%	11%	6,47%
D5	559	9,9%	44,5%	113	8,2%	71,3%	26,8%	79,03%	8%	6,49%
D6	597	10,6%	55,1%	121	8,8%	80,1%	25,0%	99,56%	9%	8,75%
D7	591	10,5%	65,6%	103	7,5%	87,6%	22,0%	120,65%	7%	9,03%
D8	633	11,2%	76,8%	73	5,3%	92,9%	16,1%	142,38%	5%	7,55%
D9	647	11,5%	88,3%	54	3,9%	96,8%	8,5%	165,11%	4%	6,48%
D10	659	11,7%	100,0%	44	3,2%	100,0%	0,0%	188,30%	3%	6,02%
	5.632	100,0%		**1.376**	100,0%		**28,53%**			**40,17%**

CAPÍTULO 3
CONSTRUCCIÓN DEL MODELO DE INFERENCIA DIFUSA.

El propósito de construir el modelo difuso es, estimar el cupo de crédito considerando el escenario de pre-aprobación *batch* de ABC; tal como se menciona en la sección 1.5, la SBS recomienda una metodología para el cálculo de cupos de créditos para clientes y potenciales clientes de las instituciones financieras reguladas, misma que utiliza fundamentalmente la información del salario; pero en un escenario de pre-aprobación *batch* esa variable es incierta para unos casos y para otros incluso desconocida, pues se dispone de una base de datos que corresponde a no clientes de ABC. Esto se convierte en una limitación para la estimación del cupo de crédito para potenciales clientes, de ahí que la construcción del modelo de inferencia difuso busca optimizar el uso de la información existente.

El modelo difuso propone estimar el cupo de crédito para un potencial cliente de ABC, a partir de la información actual e histórica de endeudamiento en la competencia, la misma que se publica en el reporte del buró de créditos; por otro lado, se desea incluir la probabilidad de incumplimiento que se obtuvo con la regresión logística como una variable del modelo, pues permite generar ofertas diferenciadas por perfil de riesgo. En ese sentido, el esquema de trabajo propuesto para la construcción del modelo y discusión de los resultados, es el siguiente:

- Fundamentos teóricos de un modelo de inferencia difusa.
- Principios que motivan a la construcción del modelo de inferencia difusa.
- Construcción del modelo de inferencia difusa.
- Comparación entre los cupos de crédito que realmente fueron otorgados y la estimación propuesta en este proyecto.

3.1 ETAPAS DE UN MODELO DE INFERENCIA DIFUSA.

La lógica difusa permite a un ordenador razonar en términos lingüísticos y reglas, de forma similar a como lo realizan los seres humanos [5]; es así que, la construcción de un modelo de inferencia difusa permite estimar el cupo de crédito para potenciales clientes de ABC, basándose en términos lingüísticos y reglas que permiten describir la forma de administrar su deuda en el sistema financiero. En ese sentido, se hace necesario presentar una breve introducción de los principales conceptos involucrados en las etapas de un sistema de lógica difusa (FIS).

En términos generales, un sistema de lógica difusa define una correspondencia no lineal entre variables de entrada y una de salida, proporcionando una base de conocimiento que permite tomar decisiones y definir patrones [1]. Para la construcción de un FIS, se deben llevar a cabo las etapas que se describen en la figura 3.1.

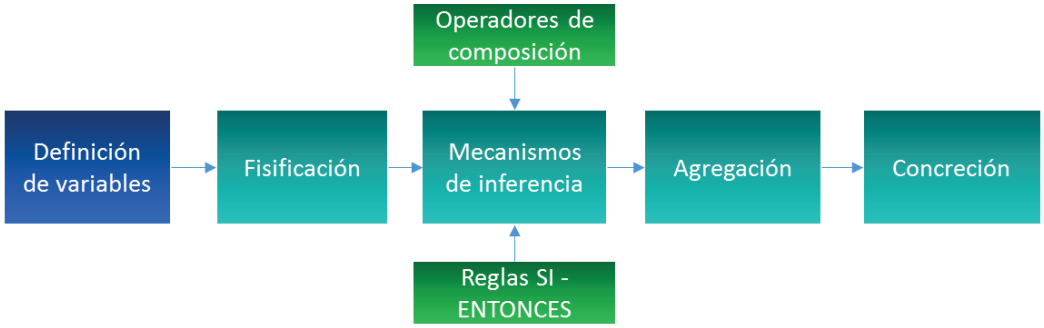

Figura 3.1 Etapas de un Sistema de Inferencia Difusa. Fuente: [1] Medina & Paniagua

3.1.1 DEFINICIÓN DE VARIABLES.

En la lógica clásica, la pertenencia o no de un elemento x conjunto A se puede definir en base a una función de pertenencia μ_A, de la siguiente manera:

$$\mu_A(x) = \begin{cases} 0 \; si \; x \notin A \\ 1 \; si \; x \in A \end{cases}$$

En la lógica difusa se relaja la condición de que la función de pertenencia tome únicamente los valores de 0 ó 1, y se admite que tome valores en el intervalo $[0,1]$; es así que, la concepción de un conjunto difuso es que el elemento x forma parte del conjunto A con un determinado grado de pertenencia.

A manera de ejemplo, se toma la variable edad que forma parte de la política institucional de ABC para determinar si un individuo es sujeto de crédito. Entre otras condiciones, la política de la institución requiere que la edad de un individuo sea mayor o igual a 25 años para ser reconocido como un sujeto de crédito. Bajo el punto de vista de la lógica clásica, los casos cercanos al umbral definido por la institución pasan de manera abrupta de pertenecer de un conjunto a otro (ver la figura 3.2); es decir, una persona que en el mes entrante cumple 25 años, hoy mismo no es un sujeto de crédito. En la lógica difusa, el pasar de un estado a otro es gradual (ver la figura 3.3); en términos lingüísticos el individuo del ejemplo debería estar más asociado al término “es sujeto de crédito” que a “no es sujeto de crédito”; de esta manera, se rompe con el principio de la no contradicción o del tercero excluido, pues el individuo objeto de análisis puede pertenecer a la vez a conjuntos contrarios en cierto grado.

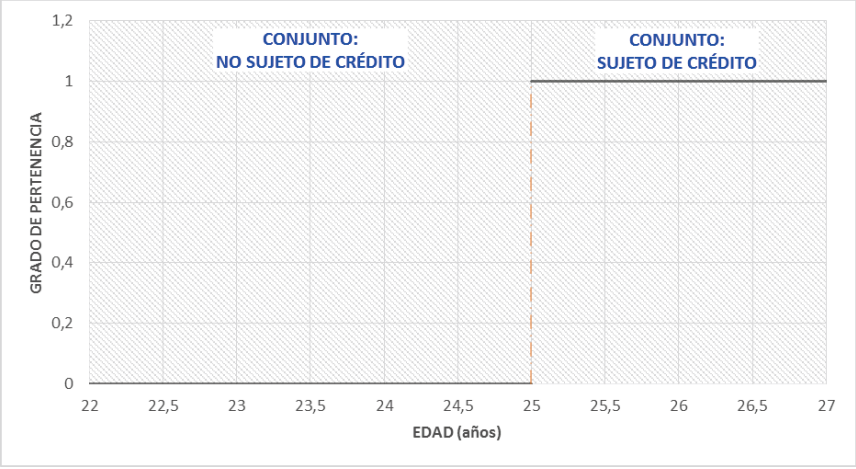

Figura 3.2 Función de pertenencia en la Lógica Clásica

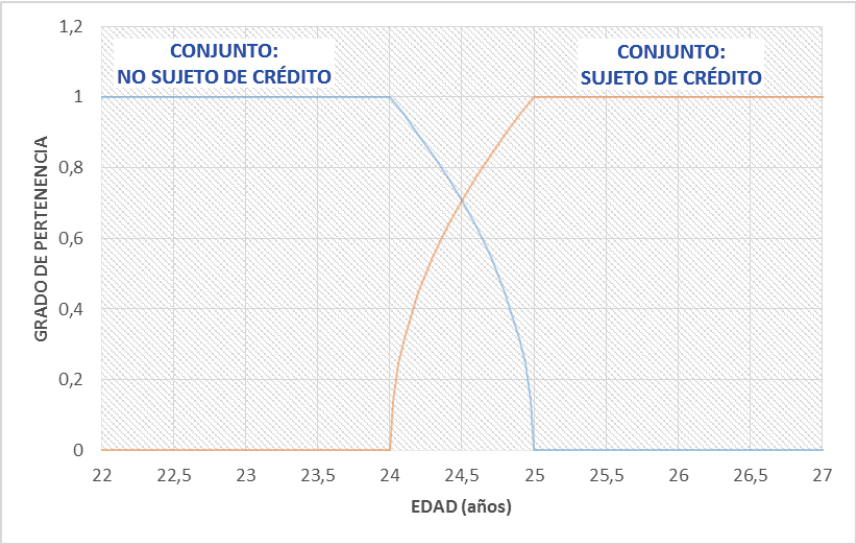

Figura 3.3 Función de pertenencia en la Lógica Difusa

En el contexto de la lógica difusa, la variable edad recibe el nombre de variable lingüística, los términos “sujeto de crédito” y “no sujeto de crédito” se denominan valores lingüísticos; al respecto se plantean las siguientes definiciones:

1. **Variable lingüística.** Es aquella noción o concepto que se califica de manera difusa, por ejemplo las variables edad o endeudamiento de un potencial cliente en la competencia. El nombre lingüístico se debe a que, en lo posterior, se definen las características de las variables mediante el lenguaje hablado.

2. **Valor lingüístico.** Tienen un significado semántico y pueden ser representados por funciones de pertenencia. Por ejemplo, la edad de un individuo lo define como “sujeto de crédito en ABC”, siendo este uno de los valores lingüísticos de la variable edad.

Considerando lo citado anteriormente, la etapa de definición de variables consiste en identificar aquellas que describen de mejor manera el fenómeno de estudio; esta etapa incluye la elección de las variables lingüísticas de entrada y de salida del modelo. En este trabajo se han seleccionado las variables en función del criterio de los expertos y los representantes de ABC, el mismo permite incluir en el modelo aquellas variables e

indicadores que describen la forma en que los potenciales clientes de ABC administran su deuda en la competencia hasta el momento en que reciben su primer crédito en la institución.

3.1.2 FUSIFICACIÓN DE VARIABLES.

Esta etapa también se conoce como parametrización de variables y consiste en definir los valores lingüísticos y sus funciones de pertenencia, a partir de lo cual es posible determinar los conjuntos difusos para el modelo.

Formalmente, un conjunto difuso se define como de la siguiente manera.

Sean:

X, el universo de valores permitidos para una variable lingüística.

x, un elemento cualquiera del X.

$A \subset X$, colección de elementos x que pertenecen a X.

El conjunto difuso A' en X se define como el conjunto de pares ordenados:

$$A' = \{(x, \mu_A(x))/x \in X\} \tag{3.1}$$

Donde:

$\mu_A(x)$, es la función de pertenencia del conjunto difuso A'. Para cada elemento x de A, asigna un valor del intervalo $[0,1]$, que se denomina valor de membresía o pertenencia.

Es decir, un conjunto difuso queda bien definido cuando se asocia una función de pertenencia a cada valor lingüístico de la variable. A manera de ejemplo, en la figura 3.3 se pueden identificar los siguientes elementos:

- Variable lingüística: Edad.
- Valores lingüísticos: “Es sujeto de crédito”, “No es sujeto de crédito”
- Funciones de pertenencia: curva que describe la línea de color azul y curva que describe la línea anaranjada, respectivamente para cada valor lingüístico.

3.1.3 MECANISMOS DE INFERENCIA.

Una vez que se han definido los conjuntos difusos para cada variable lingüística, se debe establecer una relación entre las variables de entrada y la de salida.

Previo a definir esas reglas es importante reconocer a una proposición simple dentro del contexto de la lógica difusa, la cual asigna un valor lingüístico a una determinada situación, por ejemplo: “el endeudamiento del prospecto es moderado”; es decir, una proposición difusa tiene asociada un valor lingüístico (en este caso “moderado”) y su respectiva función de pertenencia. Una proposición difusa compuesta, en cambio, resulta de agrupar una o más proposiciones simples. La agrupación se realiza mediante los conectores lógicos difusos “Y” y “O”, de la misma forma que en la lógica clásica; por ejemplo: “el endeudamiento del prospecto es moderado” Y “la variación del endeudamiento en el último semestre es pequeña”.

Con estos antecedentes es posible definir las reglas tipo Mandani “SI-ENTONCES” como, proposiciones difusas compuestas que toman la forma:

$$Si\ "X_1\ es\ A_1"\ y\ "X_2\ es\ A_2"\ y\ \dots\ y\ "X_k\ es\ A_k",\ entonces\ "Y\ es\ B" \tag{3.2}$$

Donde $A_1, A_2, \dots, A_k, B$ son valores lingüísticos definidos mediante conjuntos difusos para las variables lingüísticas $X_1, X_2, \dots, X_k, Y$, respectivamente.

La proposición difusa compuesta "$X_1\ es\ A_1$" y "$X_2\ es\ A_2$" $y\ \dots\ y$ "$X_k\ es\ A_k$" se denomina antecedente o premisa, mientras que el término "$Y\ es\ B$" es llamado consecuente o conclusión, de manera similar a la lógica clásica.

Las reglas tipo Mandani son las que permiten establecer relaciones entre las variables de entrada con la de salida, a este conjunto de reglas se denomina la base de conocimiento del sistema de inferencia difuso. Por ejemplo, en el desarrollo de una estrategia de mercado diferenciada por perfil de riesgo de potenciales clientes, para la promoción de cupos de crédito, se puede establecer la siguiente regla: “si el riesgo es

pequeño, entonces ofertar un cupo de crédito alto". La base de conocimientos se construye a partir del criterio de los expertos y con la ayuda herramientas matemáticas y estadísticas. Esta última consideración hace que un sistema de inferencia difuso permita conocer la forma en que un potencial cliente administra su deuda en la competencia y en función de ese resultado sugerir un cupo de crédito a ser ofertado por ABC.

La función de pertenencia asociada a una regla tipo Mandani, se expresa de la siguiente forma:

$$\begin{aligned} \mu_{p\to q}&: U \times V \to [0,1] \\ &(u,v) \longmapsto \mu_{p\to q}(u,v) \end{aligned} \tag{3.3}$$

Donde, p y q son proposiciones en las que intervienen los conjuntos difusos A y B, respectivamente, con conjuntos de la valores admisibles U y V.

Una función de pertenencia indica que se evalúa la activación de una regla (activación del consecuente) en función del grado de cumplimiento del antecedente; para ello, se utilizan las propiedades de los conectores lógicos difusos y operadores de conjuntos difusos, en la siguiente sección se hace una explicación más detallada sobre estos elementos de la lógica difusa.

3.1.4 AGREGACIÓN.

Esta etapa consiste en combinar cada una de las reglas definidas en la etapa anterior con la finalidad de obtener un único conjunto difuso como salida; para ello es importante conocer las definiciones y propiedades de los conectores lógicos difusos y operadores de los conjuntos difusos. Para una revisión formal sobre los fundamentos matemáticos de los conjuntos difusos, el lector puede consultar Kaufman 1990.

OPERACIONES ELEMENTALES CON CONJUNTOS DIFUSOS.

Al igual que en la lógica clásica, sobre la lógica difusa se definen las operaciones entre conjuntos, de la siguiente manera.

- **Complemento.** Sea el conjunto difuso A con función de pertenencia μ_A, el complemento se denota por $\bar{A}$ y está definido por una función de pertenencia que se calcula a partir de μ_A, de la siguiente manera:

$$\begin{aligned} \mu_{\bar{A}}&: [0,1] \to [0,1] \\ &\alpha \longmapsto \mu_{\bar{A}}(\alpha) = c(\alpha) \end{aligned} \tag{3.4}$$

  Donde, $a = \mu_A(x)$ es el valor de pertenencia de x al conjunto A.

  La función c debe verificar las siguientes condiciones:

  i) Concordancia con los conjuntos clásicos.

$$c(0) = 1,\ c(1) = 0.$$

  ii) Estrictamente decreciente.

$$\forall \alpha, \beta \in [0,1], \alpha > \beta \Rightarrow c(\alpha) < c(\beta).$$

  iii) Involución.

$$\forall \alpha [0,1], c\big(c(\alpha)\big) = \alpha.$$

  Algunas funciones que cumplen con las propiedades antes mencionadas son:

$$c(\alpha) = 1 - \alpha$$

$$c_w(\alpha) = (1 - \alpha^w)^{1/w}, w \in [0, \infty]$$

- **Intersección.** Sean los conjuntos difusos A y B con sus respectivas funciones de pertenencia μ_A y μ_B. El grado de pertenencia al conjunto intersección, $A \cap B$, conocido el grado de pertenencia a cada uno de los conjuntos originales, se define por:

$$\begin{aligned} \mu_{A\cap B}&: [0,1] \times [0,1] \to [0,1] \\ &(\alpha, \beta) \longmapsto \mu_{A\cap B}(\alpha, \beta) = i(\alpha, \beta) \end{aligned} \tag{3.5}$$

  Donde, $\alpha = \mu_A(x)$ y $\beta = \mu_B(z)$.

La función i es tal que:

i) Concordancia con los conjuntos clásicos.

$$i(0,1) = i(0,0) = i(1,0) = 0, i(1,1) = 1$$

ii) Conmutatividad.

$$\forall \alpha, \beta \in [0,1], i(\alpha, \beta) = i(\beta, \alpha)$$

iii) Asociatividad.

$$\forall \alpha, \beta, \gamma \in [0,1], i\big(\alpha, i(\beta, \gamma)\big) = i(i(\alpha, \beta), \gamma)$$

iv) Identidad.

$$\forall \alpha \in [0,1], i(\alpha, 1) = \alpha$$

v) Monotonía.

$$\forall \alpha, \beta \in [0,1], si\ \alpha \leq \alpha' y\ \beta \leq \beta', entonces\ i(\alpha, \beta) \leq i(\alpha', \beta')$$

El conjunto $([0,1], i)$ tiene la estructura de un grupo abeliano. En el contexto de la teoría de conjuntos difusos, las funciones que cumplen con las cinco propiedades anteriores se denominan normas triangulares o t-normas, las más usuales son las siguientes:

i) T-norma del mínimo: $i(\alpha, \beta) = \min(\alpha, \beta)$

ii) T-norma del producto: $i(\alpha, \beta) = \alpha * \beta$

iii) T-norma del producto drástico: $i(\alpha, \beta) = \begin{cases} \alpha \text{ si } \beta = 1 \\ \beta \text{ si } \alpha = 1 \\ 0 \text{ caso contrario} \end{cases}$

- **Unión.** Sean los conjuntos difusos A y B con sus respectivas funciones de pertenencia μ_A y μ_B. El grado de pertenencia al conjunto unión, $A \cup B$, conocido el grado de pertenencia a cada uno de los conjuntos originales, se define por:

$$\begin{aligned} \mu_{A \cup B} &: [0,1] \times [0,1] \to [0,1] \\ &(\alpha, \beta) \longmapsto \mu_{A \cup B}(\alpha, \beta) = u(\alpha, \beta) \end{aligned} \tag{3.6}$$

Donde $\alpha = \mu_A(x)$ y $\beta = \mu_B(z)$.

La función u es tal que:

i) Concordancia con los conjuntos clásicos.

$$u(0,1) = u(1,1) = u(1,0) = 1, u(0,0) = 0$$

ii) Conmutatividad.

$$\forall \alpha, \beta \in [0,1], u(\alpha, \beta) = u(\beta, \alpha)$$

iii) Asociatividad.

$$\forall \alpha, \beta, \gamma \in [0,1], u\big(\alpha, u(\beta, \gamma)\big) = u(u(\alpha, \beta), \gamma)$$

iv) Identidad.

$$\forall \alpha \in [0,1], u(\alpha, 0) = \alpha$$

v) Monotonía.

$$\forall \alpha, \beta \in [0,1], si\ \alpha \leq \alpha' y\ \beta \leq \beta', entonces\ u(\alpha, \beta) \leq u(\alpha', \beta')$$

vi) Leyes de De Morgan.

$$\forall \alpha, \beta \in [0,1], u(\alpha, \beta) = c(i(c(\alpha), c(\beta)))$$
$$\forall \alpha, \beta \in [0,1], i(\alpha, \beta) = c(u(c(\alpha), c(\beta)))$$

Las funciones que cumplen con las seis propiedades anteriores se denominan conormas triangulares o t-conormas, las más usuales son las siguientes:

iv) T-conorma del máximo: $u(\alpha, \beta) = \max(\alpha, \beta)$

v) T-conorma de la suma: $u(\alpha, \beta) = \alpha + \beta - (\alpha * \beta)$

vi) T-conorma de la suma drástica: $u(\alpha, \beta) = \begin{cases} \alpha \text{ si } \beta = 0 \\ \beta \text{ si } \alpha = 0 \\ 1 \text{ caso contrario} \end{cases}$

REGLAS TIPO MANDANI EN LA ETAPA DE AGREGACIÓN.

Los operadores lógicos difusos se definen de forma análoga a la teoría clásica de conjuntos, sean:

$p\ y\ q$, proposiciones difusas.

$A\ y\ B$, los conjuntos difusos que intervienen en ellas, respectivamente.

$\mu_A\ y\ \mu_B$, las funciones de pertenencia asociadas a ellas.

Los operadores lógicos se definen de la siguiente manera:

- NO $(\neg p)$: se asocia la función de pertenencia del complemento de A, por ejemplo $\mu_{\neg p}(x) = 1 - \mu_A(x)$.
- Y $(p \wedge q)$: se asocia una función de pertenencia de intersección, por ejemplo $\mu_{p \wedge q}(x, y) = \min(\mu_A(x), \mu_B(y))$.
- O $(p \vee q)$: se asocia una función de pertenencia de unión, por ejemplo $\mu_{p \vee q}(x, y) = \max(\mu_A(x), \mu_B(y))$.

De esta manera, la función de pertenencia asociada a la regla tipo Mandani de la ecuación (3.3), se expresa de la siguiente manera:

$$\mu_{p \to q}(u, v) = min(\mu_A(u), \mu_B(v)) \tag{3.7}$$

Es decir, Mandani propone que una condición tan sólo resulta cierta cuando el antecedente y el consecuente son ciertos, es decir $p \to q \equiv p \wedge q$.

A partir de esta definición es posible completar la etapa de agregación. Existen dos tipos de problemas a tratar en esta etapa: a) el antecedente de la implicación son conjuntos difusos; y, b) el antecedente es un número real, también conocido como antecedente nítido. En este proyecto se da un tratamiento al problema planteado en el literal b, es decir, se tratan situaciones como la siguiente:

Regla: Si el potencial cliente es riesgo bajo, entonces ofertar un cupo de crédito alto.

Antecedente nítido: 750/100

Problema: ¿Qué cupo de crédito se debe ofertar bajo estas condiciones del perfil de riesgo?

Es decir, la etapa de agregación parte de un conjunto de reglas tipo Mandani (definidas en la sección 3.1.3) y un conjunto de valores de las variables de entrada (valores

nítidos) del modelo y busca una conclusión que será uno o varios conjuntos difusos de la variable de salida.

En línea con el ejemplo anterior, supongamos que las funciones de pertenencia para el perfil de riesgo bajo y cupo de crédito alto son los de la figura 3.4.

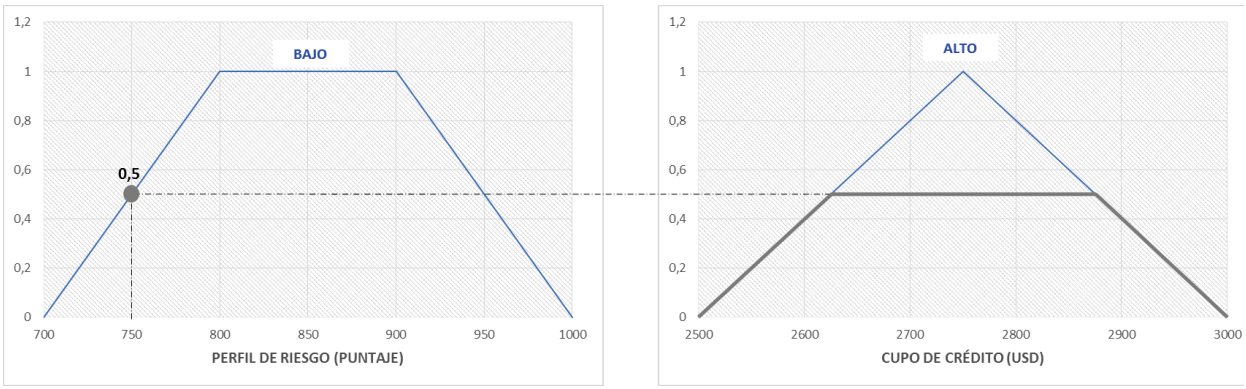

Figura 3.4 Funciones de pertenencia de perfil de riesgo y cupo de crédito

Gráficamente, el resultado de aplicar la regla tipo Mandani con antecedente nítido es el trapecio regular de la figura 3.4 (función truncada del lado derecho), es decir es el resultado de $\mu_{p\to q}(750, x) = \min\left(\mu_{BAJO}(750), \mu_{ALTO}(x)\right) = \min\left(0.5, \mu_{ALTO}(x)\right)$.

Cuando existen más reglas tipo Mandani, se obtienen varios conjuntos difusos como resultado en la variable de salida del modelo, los cuales se unen o agregan para formar uno solo, de ahí el nombre “etapa de agregación”.

3.1.5 CONCRECIÓN.

La etapa de concreción consiste en transformar de un conjunto difuso en un valor nítido, que permita realizar una acción a partir de ese resultado.

En línea con el ejemplo presentado en la figura 3.4, resulta de interés conocer un valor concreto para ofertar un cupo crédito, sabiendo que el perfil de riesgo es bajo. Existen

varios métodos que permiten convertir un valor difuso en nítido, los más empleados en el desarrollo de los modelos difusos, son los siguientes:

i) **MOM (Middle of maximum)**. Es el valor medio de los máximos que se obtienen en el conjunto o conjuntos difusos de la variable de salida, los mismos que resultan luego de aplicar el proceso de agregación.

   En el ejemplo anterior, la función de pertenencia (trapecio regular) tiene como máximo el valor 0,5 y sus correspondientes valores comprendidos entre 2625 y 2875 en el eje las abscisas. El método MOM toma el valor medio entre estos dos, 2750. La solución gráfica se puede observar en la figura 3.5.

ii) **COG (Center of gravity).** También llamado centroide o centro de gravedad, se define por:

$$y_{COG} = \frac{\sum_{x \in X}[x \times \mu_A(x)]}{\sum_{x \in X}[\mu_A(x)]} \tag{3.8}$$

En el ejemplo, el valor de COG es $\frac{(2500*0)+(2625*1)+(2875*1)+(3000*0)}{0+1+1+0} = 2750.$

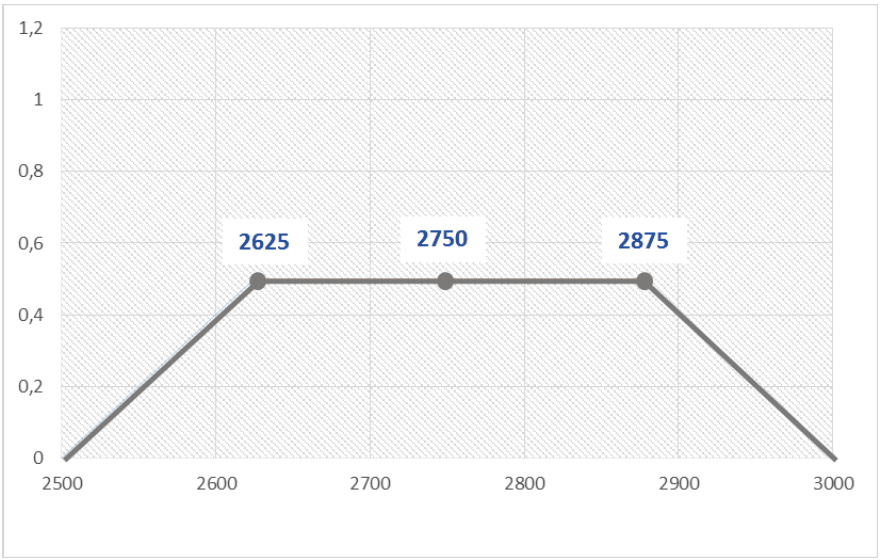

Figura 3.5 Concreción de conjuntos difusos

3.2 MOTIVACIÓN PARA LA CONSTRUCCIÓN DE UN MODELO DE INFERENCIA DIFUSA.

Partiendo de un punto de vista financiero, es natural suponer que si una persona mantiene una deuda con la competencia, cumple con los requerimientos exigidos por la SBS para ser calificado como un sujeto de crédito. Adicional, el comportamiento de pago y tanto el endeudamiento actual como histórico, proporcionan información de la forma en que los clientes nuevos de ABC administran su deuda con la competencia, hasta el momento en que reciben su primer crédito en la institución. Partiendo de ese entendimiento, el modelo de inferencia difusa busca sugerir un cupo de crédito para que ABC lo gestione a través de sus diferentes canales de venta.

La figura 3.6 muestra la forma en que una persona administra su deuda en la competencia antes y después de recibir su primer crédito en ABC; la dinámica de los créditos de consumo permite evidenciar una curva de endeudamiento lineal decreciente, siempre y cuando el individuo pague a tiempo las cuotas pactadas en la tabla de amortización, esto se debe a que en el caso de los crédito de consumo (sin considerar la tarjeta de crédito) el capital se amortiza mensualmente. Cuando la persona recibe su primera operación en ABC, como en el ejemplo, el endeudamiento total vuelve a ser similar al que tuvo en la competencia en el mes de marzo del 2013; es decir, el endeudamiento total es similar al que tuvo hace 11 meses. Esto sugiere que luego de un determinado periodo y bajo determinadas condiciones, el individuo tiene la capacidad de cumplir nuevas obligaciones de pago, en particular con ABC.

El presente proyecto busca generalizar esta idea por medio de un modelo de inferencia difusa en la muestra de clientes nuevos de ABC de un determinado periodo, de tal manera que permita sugerir un cupo de crédito para los potenciales clientes de la institución.

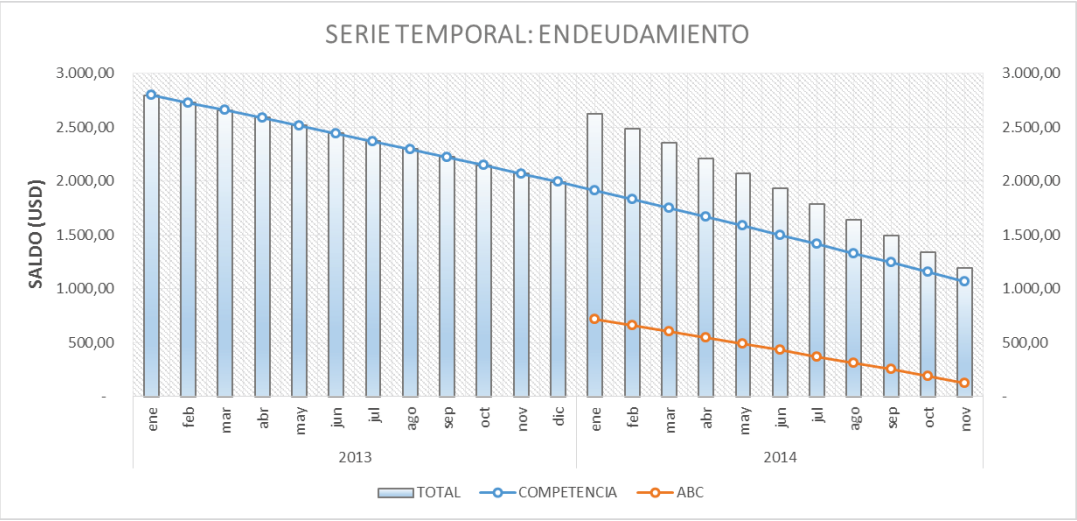

Figura 3.6 Serie temporal de endeudamiento de un potencial cliente de ABC

3.3 CONSTRUCCIÓN DEL MODELO DE INFERENCIA DIFUSA.

En esta sección se presentan los resultados obtenidos en la construcción del modelo de inferencia difusa para la evaluación crediticia de potenciales clientes de ABC, para conseguir tal propósito se han seguido los lineamientos presentados en las dos secciones precedentes.

3.3.1 DEFINICIÓN DEL UNIVERSO DE INDIVIDUOS.

El propósito de este apartado es, definir el universo de individuos con el que se va a construir el modelo de inferencia difusa. En la sección 1.5 se ha descrito la situación actual de ABC y se ha señalado la recomendación de la SBS para el cálculo del cupo de crédito para potenciales clientes, la misma que utiliza fundamentalmente la información del salario; por otro lado, también se ha mencionado que en el proceso de pre-aprobación *batch* de ABC, la variable salario es imprecisa e incluso desconocida para una porción del total de individuos. Esta situación conlleva a dividir la muestra de estudio en dos grupos: a) potenciales clientes que disponen de la información de salario; y, b) potenciales clientes para los que no es posible conocer el salario dentro

de las bases de datos de ABC. Es importante mencionar que cuando el individuo recibe su primer crédito en ABC, como parte de la etapa de verificación del ciclo de crédito, se procede a la verificación de los documentos que justifican sus ingresos, tal como se mencionó en la sección 1.2.2; por esta razón, la variable salario se recrea bajo las condiciones en las que se otorgó el primer crédito, pues ahí es donde se puede reconocer si se disponía o no de esa información para el proceso *batch* de pre-aprobación.

La partición de la muestra permite optimizar los recursos de información de la institución, pues el cálculo de cupos de crédito para potenciales clientes abarca los dos escenarios citados anteriormente. Los individuos seleccionados para construir el modelo de inferencia difusa son los que forman parte del periodo de *testing* del modelo *scoring*, esto por tratarse de la muestra más reciente dentro de la base de datos suministrada por el área de Sistemas de Información de ABC; es decir, se va trabajar con los clientes nuevos que han recibido su primer crédito en la institución entre los meses de octubre 2013 y marzo 2014. Cabe mencionar que, en la construcción del indicador KS de la ventana de *testing* se utilizó únicamente la población de buenos y malos, dejando de lado a los indeterminados (ver tabla 2.28); para el modelo difuso se incluyen las tres categorías, debido a que en la estimación de montos resulta poco relevante esa distinción.

Tabla 3.1 Partición de la muestra por disponibilidad del salario

Criterio	Frecuencia	Frec. relativa	Pérdida
Con salario	6.505	86%	4,9%
Sin salario	1.054	14%	9,1%
TOTAL	**7.559**	**100,00%**	**5,4%**

En la tabla 3.1 se muestra la partición de la muestra de *testing* del modelo *scoring* según la disponibilidad de la información, mostrando dos aspectos fundamentales: a) los individuos que disponían de la información de salario son mayoría y representan el 86%; y, b) el indicador de pérdida es mayor en el grupo de los clientes nuevos de ABC

que no disponían de la información de salario en la base de datos, antes de recibir su primer crédito en ABC, es decir la relación $provisión/monto$ es mayor en este grupo, eso significa que las operaciones son de menor calidad en términos de riesgo.

Para obtener la muestra definitiva es importante incluir los criterios descritos en 1.3.3 respecto a la administración del riesgo de crédito, en lo que se refiere a la identificación del nivel de riesgo considerado como aceptable para la institución; es así que, al revisar el indicador de pérdida por puntaje de *score*, se ha decidido incluir en la estimación del cupo de crédito pre-aprobado únicamente a los individuos que pertenecen a los deciles 5, 6, 7, 8, 9 y 10; o sea individuos cuyo puntaje supere 547.

Es decir, se incluyen los individuos que se encuentran en los grupos que no representan una pérdida superior a la global de 5,4%. En la práctica, aquellas personas con un puntaje de *score* menor al umbral fijado, deben someterse a un análisis más riguroso por parte de un asesor o un comité de crédito, según sea el caso. Como parte del análisis se pueden solicitar documentos o garantías que sustenten el otorgamiento de un crédito; caso contrario, la solicitud es negada.

Tabla 3.2 Indicador de pérdida por puntaje de score

Deciles	Puntaje score	Con salario		Sin salario		Global	
		Frecuencia	Pérdida	Frecuencia	Pérdida	Frecuencia	Pérdida
1	340	506	12,5%	307	13,6%	813	12,9%
2	433	573	7,9%	211	9,2%	784	8,2%
3	496	626	7,0%	146	8,5%	772	7,3%
4	547	659	6,4%	106	8,1%	765	6,7%
5	597	657	5,1%	96	5,2%	753	5,2%
6	640	688	3,9%	73	4,7%	761	3,9%
7	681	683	3,6%	41	4,1%	724	3,6%
8	725	697	2,5%	44	3,0%	741	2,5%
9	778	708	2,1%	15	2,7%	723	2,2%
10	926	708	2,0%	15	3,3%	723	2,0%
TOTAL		**6505**	**4,9%**	**1054**	**9,1%**	**7559**	**5,4%**

En adelante, se llamará modelo I al que se construye con la muestra de individuos que disponen de la información de salario y modelo II al que se construye con la muestra restante.

3.3.2 SELECCIÓN DE LAS VARIABLES DEL MODELO.

A partir de la muestra seleccionada en el punto anterior, se procede a analizar las variables que se van a incluir en los modelos I y II, para ello se utiliza fundamentalmente el criterio de los expertos, con el propósito de incluir las mejores prácticas anteriores y de esa manera cumplir con los objetivos plantados como institución.

Las variables que deben formar parte del estudio, son aquellas que describen el nivel de endeudamiento actual e histórico de los potenciales clientes de ABC en la competencia. En ese sentido, las variables seleccionadas para construir los modelos son las que se describen a continuación.

- **Carga financiera máxima.** Indica el nivel de solvencia del cliente, pues recoge la información de la máxima cuota que dispone una persona para cubrir sus compromisos con las instituciones del sistema financiero regulado; su forma de cálculo responde a lo mencionado por la normativa de la SBS, bajo las siguientes consideraciones:

  a) La política interna de ABC destina el 30% del salario al gasto promedio mensual de una persona, de esta manera se obtiene el ingreso neto mensual del solicitante; esto aplica cuando no es posible conocer la información de gastos del individuo que objeto de evaluación, tal como sucede en el proceso *batch* de pre-aprobación de cupos de crédito.

  b) En cumplimiento con la normativa de la SBS, del ingreso neto mensual se destina únicamente el 50% al pago de deudas con entidades del sistema financiero, a lo que se ha denominado la carga financiera máxima del potencial cliente.

Es importante recalcar que, el valor que se destina a los gastos de cada persona (30% del salario) responde a las mejores prácticas anteriores de ABC en procesos *batch* para la estimación de cupos de crédito. El lector debe recordar que este tipo de información (salario y gastos mensuales) es verificada por el asesor de crédito y registrada en el sistema *core* del banco, previo al desembolso del cupo de crédito pre aprobado.

- **Saldo que adeuda en la competencia.** Indica el nivel de endeudamiento del potencial cliente en la competencia. La experiencia de ABC en el mercado financiero permite concluir que a menor endeudamiento del individuo, aumenta la probabilidad de consolidar una negociación.

- **Máximo endeudamiento en el último año.** Indica el mayor saldo adeudado en la competencia durante el último año, antes de obtener su primer crédito en ABC. En la estimación de cupos de crédito se busca que el valor ofertado por ABC sea similar al que tuvo en la competencia durante el último año. Esta situación se observa en la figura 3.6, en el mes de enero 2014 el individuo obtiene su primer crédito en ABC, y su deuda total (incluyendo ABC y la competencia) es similar a la que tuvo hace un año. La razón por la que se toma una variable con temporalidad de 12 meses es que los cupos del proceso *batch* también se ofertan con ese plazo, de tal manera que el modelo busca retomar la deuda del individuo de hace un año, tal como se muestra en el ejemplo de la figura 3.6.

- **Variación de la deuda en el último semestre.** Indica si el saldo adeudado por un individuo en la competencia crece o decrece durante el último semestre, antes de obtener su primer crédito en ABC. A manera de ejemplo, en la figura 3.6 se muestra que la variación del endeudamiento en el último semestre decrece constantemente como consecuencia de los pagos regulares que realiza el individuo, permitiendo concluir que los potenciales clientes con este comportamiento pueden acceder a un nuevo compromiso con las instituciones financieras, en particular con ABC.

- **Perfil de riesgo.** Variable relativa al puntaje del *score* que se obtuvo con el modelo logístico. Esta variable no describe explícitamente la forma en que un potencial cliente administra una deuda con la competencia; pero permite elaborar estrategias comerciales diferenciadas por perfil de riesgo, de tal manera que a los individuos que representan un nivel de riesgo menor para ABC son sujetos de ofertas mayores que aquellos de mayor riesgo.

Las variables presentadas se utilizan en cada modelo, de acuerdo al siguiente detalle:

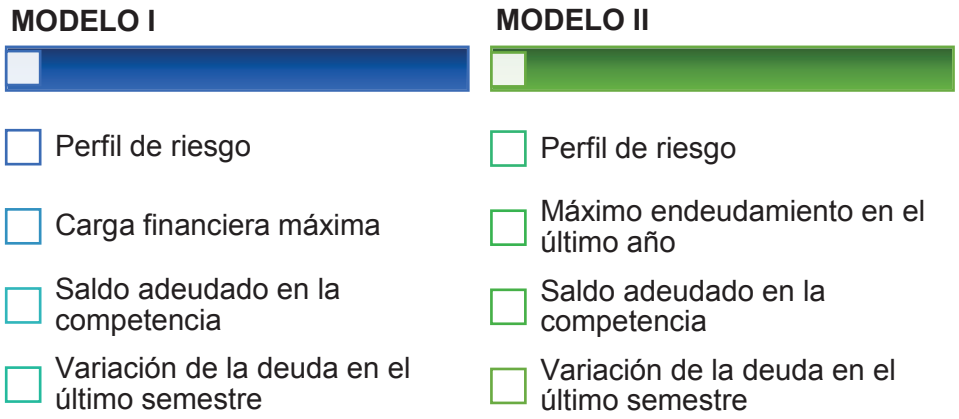

Figura 3.7 Variables seleccionadas para los modelos I y II

3.3.3 DEFINICIÓN DE CONJUNTOS DIFUSOS.

El propósito de presentar esta sección es, identificar los valores lingüísticos y sus respectivas funciones de pertenencia para las variables de entrada que se han descrito en la sección anterior y para la variable de salida.

MODELO I.

1. **Perfil de riesgo.**

   Se definen tres conjunto difusos: A, AA y AAA. Los individuos del conjunto AAA son los que representan un nivel de riesgo menor para la institución, es el mejor perfil para los potenciales clientes. Con la finalidad de realizar un análisis

exploratorio de los datos, se construyen intervalos de clase con la ayuda de la regla de *Sturges* presentada en la ecuación 2.3; a partir de lo cual se puede observar la distribución de los datos en la tabla 3.3.

Para determinar las funciones de pertenencia se procede de la siguiente manera:

a) Calcular los percentiles sobre la muestra de 4.141 observaciones. En cada grupo formado asociar el indicador de pérdida, de esta manera se construyen grupos homogéneos en términos del número de individuos y nivel de pérdida que representan para la institución.

b) Determinar los intervalos donde las funciones de pertenencia toman el valor de 1, esta consideración permite identificar la pertenencia absoluta de los elementos a un determinado conjunto. Para fijar los límites de cada intervalo se analiza el nivel de pérdida asociado a los clientes nuevos de la muestra de 4.141 individuos. A manera de ejemplo, en la figura 3.8 (cuadro derecho) se puede observar que los potenciales clientes pertenecen al conjunto difuso A con grado 1, siempre y cuando su puntaje de *score* no supere 621 puntos. El nivel de pérdida asociado a este grupo de personas es mayor a cualquier otro.

c) Determinar los intervalos donde las funciones de pertenencia se solapan. Esto permite identificar el grado de pertenencia de un elemento a diferentes conjuntos difusos a la vez. Por ejemplo, una persona con un puntaje igual a 550 pertenece al conjunto A con grado de membresía 0,54; al mismo tiempo esta persona pertenece al conjunto AA, pero con grado 0,46.

d) Sobre los intervalos identificados en el punto anterior, se construyen las funciones que describen la pertenencia a los conjuntos difusos, para ello se toman los siguientes criterios:

    - Inicialmente todas las funciones son lineales o toman la forma determinada por los expertos.

- Luego, en cada simulación del cupo de crédito se modifican, hasta alcanzar el objetivo primordial sobre la variable de salida, que es mantener el cupo de crédito promedio estimado con la metodología vigente de ABC; de esta manera las funciones pueden dejar de ser lineales y convertirse en cuadráticas o de cualquier otro tipo.

Tabla 3.3 Perfil de riesgo – Modelo I

Nro.	Intervalo	Frecuencia	Frec. Relativa
1	[548,00; 577,08)	365	8,81%
2	[577,08; 606,16)	425	10,26%
3	[606,16; 635,24)	469	11,33%
4	[635,24; 664,32)	486	11,74%
5	[664,32; 693,40)	477	11,52%
6	[693,40; 722,48)	452	10,92%
7	[722,48; 751,56)	455	10,99%
8	[751,56; 780,64)	329	7,94%
9	[780,64; 809,72)	324	7,82%
10	[809,72; 838,80)	198	4,78%
11	[838,80; 867,88)	102	2,46%
12	[867,88; 896,96)	44	1,06%
13	[896,96; 926,04)	15	0,36%
	Total	**4.141**	**100,00%**

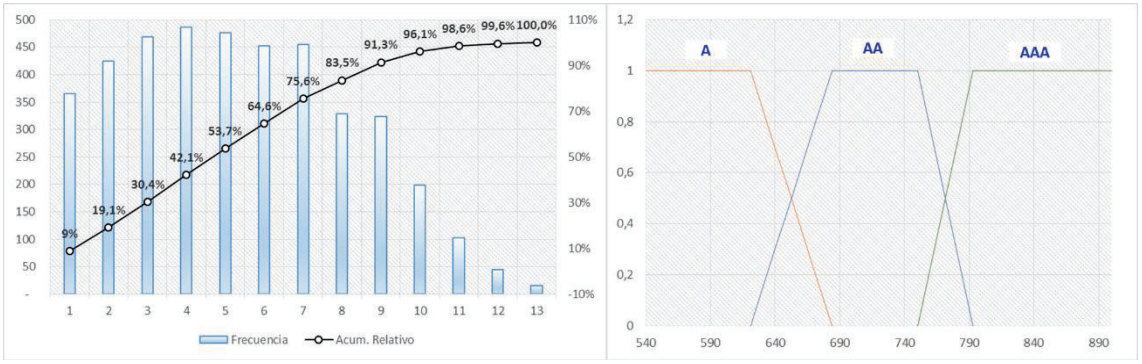

Figura 3.8 Histograma y función de pertenencia del perfil de riesgo – Modelo I

Las funciones de pertenencia finales para el perfil de riesgo, se muestran en el cuadro derecho de la figura 3.8, las ecuaciones respectivas son:

$$A = \begin{cases} 1 & si\ 540 \leq x \leq 621 \\ -\frac{1}{63}(x - 684) & si\ 621 \leq x \leq 684 \end{cases}$$

$$AA = \begin{cases} \frac{1}{63}(x - 621) & si\ 621 \leq x \leq 684 \\ 1 & si\ 684 \leq x \leq 750 \\ -\frac{1}{43}(x - 793) & s\ 750 \leq x \leq 793 \end{cases} \quad (3.9)$$

$$AAA = \begin{cases} \frac{1}{43}(x - 750) & si\ 750 \leq x \leq 793 \\ 1 & si\ 793 \leq x \leq 1000 \end{cases}$$

2. Carga financiera máxima (CFM).

De forma similar a la variable perfil de riesgo, luego de analizar el histograma de frecuencias, se procede a construir los conjuntos difusos con la ayuda de percentiles y desde luego incluyendo el criterio de profesionales expertos.

Los conjuntos difusos definidos para la variable carga financiera máxima son: PEQUEÑO, MODERADO y GRANDE. Los nombres hacen referencia al valor monetario de la carga financiera máxima, por ejemplo, se considera como CFM pequeña aquella que no supera 210 dólares.

Tabla 3.4 Carga financiera máxima – Modelo I

Nro.	Intervalo	Frecuencia	Frec. Relativa
1	[140,00; 263,85)	1575	38,03%
2	[263,85; 387,70)	1104	26,66%
3	[387,70; 511,55)	609	14,71%
4	[511,55; 635,40)	368	8,89%
5	[635,40; 759,25)	208	5,02%
6	[759,25; 883,10)	111	2,68%
7	[883,10; 1006,95)	52	1,26%
8	[1006,91; 1130,80)	42	1,01%
9	[1130,80; 1254,65)	23	0,56%
10	[1254,65; 1378,50)	22	0,53%
11	[1378,50; 1502,35)	11	0,27%
12	[1502,35; 1626,20)	9	0,22%
13	[1626,20; 1750,05)	7	0,17%
	Total	**4.141**	**100,00%**

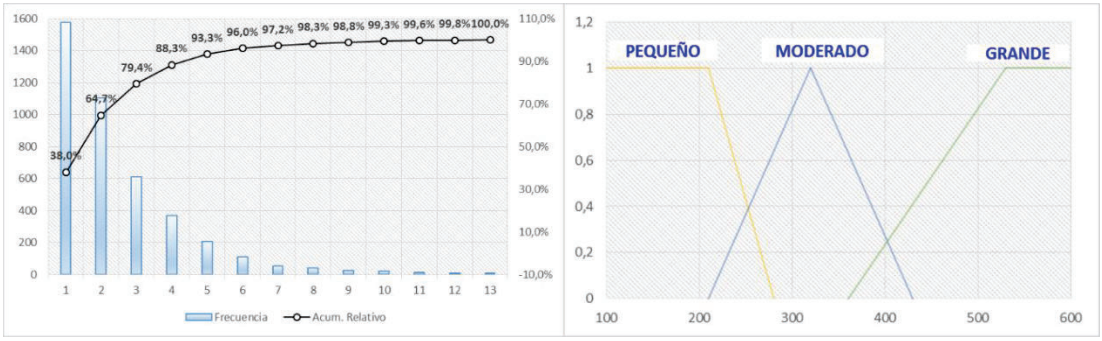

Figura 3.9 Histograma y función de pertenencia de carga financiera máxima – Modelo II

$$PEQ = \begin{cases} 1\ si\ 0 \leq x \leq 210 \\ -\frac{1}{70}(x-210) si\ 210 < x \leq 280 \end{cases}$$

$$MOD = \begin{cases} \frac{1}{110}(x-210)\ \ si\ 210 \leq x \leq 320 \\ -\frac{1}{110}(x-430) si\ 320 < x \leq 430 \end{cases} \qquad (3.10)$$

$$GRAN = \begin{cases} \frac{1}{170}(x-360) si\ 360 \leq x \leq 530 \\ 1\ si\ x > 530 \end{cases}$$

3. **Saldo adeudado en la competencia.**

   Los conjuntos difusos definidos para esta variable son, al igual que la carga financiera: PEQUEÑO, MODERADO y GRANDE. La construcción de las funciones de pertenencia sigue una lógica similar a las dos variables anteriores Los resultados se muestran a continuación.

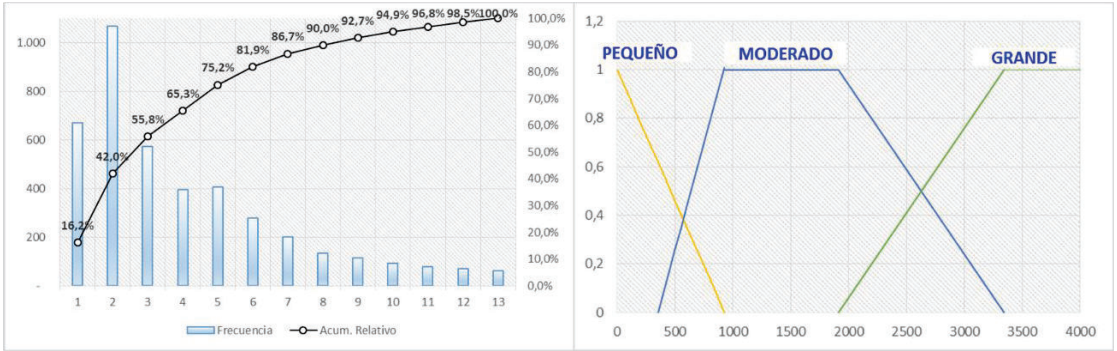

Figura 3.10 Histograma y función de pertenencia de endeudamiento – Modelo I

Tabla 3.5 Endeudamiento en la competencia – Modelo I

Nro.	Intervalo	Frecuencia	Frec. Relativa
1	Cero	670	16,18%
2	[1; 1251)	1.068	25,79%
3	[1251; 2501)	574	13,86%
4	[2501; 3751)	394	9,51%
5	[3751; 5001)	406	9,80%
6	[5001; 6251)	278	6,71%
7	[6251; 7501)	202	4,88%
8	[7501; 8751)	133	3,21%
9	[8751; 10001)	114	2,75%
10	[10001; 11251)	92	2,22%
11	[11251; 12501)	78	1,88%
12	[12501; 13751)	69	1,67%
13	[13751; 15001)	63	1,52%
	Total	**4.141**	**100,00%**

$$PEQ = -\frac{1}{929}(X - 929)\ si\ 0 \leq x\ \leq 929$$

$$MED = \begin{cases} \frac{1}{572}(x - 352) si\ 352 \leq x \leq 29 \\ 1\ si\ 929 \leq x \leq 1909 \\ -\frac{1}{1434}(x - 3343) si\ 1909 \leq x \leq 3343 \end{cases} \quad (3.11)$$

$$GRAN = \begin{cases} \frac{1}{1434}(x - 1909)\ si\ 1909 \leq x \leq 3343 \\ 1\ si\ x > 3343 \end{cases}$$

4. **Variación de la deuda en el último semestre.**

   Los conjuntos difusos definidos para esta variable son: PEQUEÑA, MODERADA y PEQUEÑA. La construcción de las respectivas funciones de pertenencia se hace de forma similar a las dos primeras variables y los resultados se muestran a continuación.

Tabla 3.6 Variación de la deuda último semestre – Modelo I

Nro.	Intervalo	Frecuencia	Frec. Relativa
1	[0; 0,083)	150	3,62%
2	[0,083; 0,166)	25	0,60%
3	[0,166; 0,249)	39	0,94%
4	[0,249; 0,332)	42	1,01%
5	[0,332; 0,415)	63	1,52%
6	[0,415; 0,498)	64	1,55%
7	[0,498; 0,581)	118	2,85%
8	[0,581; 0,664)	145	3,50%
9	[0,664; 0,747)	209	5,05%
10	[0,747; 0,830)	428	10,34%
11	[0,830; 0,913)	630	15,21%
12	[0,913; 1]	1439	34,75%
13	>1	789	19,05%
	Total	**4.141**	**100,00%**

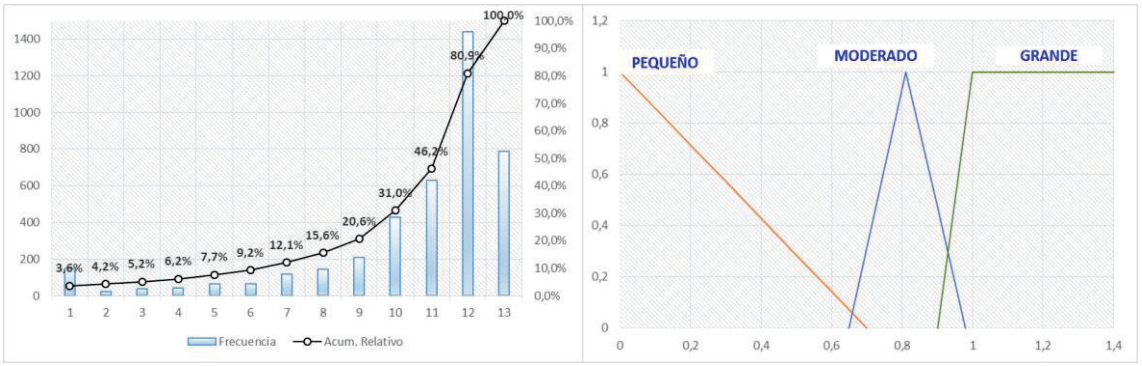

Figura 3.11 Histograma y función de pertenencia de variación de la deuda – Modelo I

$$PEQ = -\frac{1}{0{,}7}(x - 0.7) si\ 0 \leq x \leq 0{,}7$$

$$MOD = \begin{cases} \frac{1}{0{,}16}(x - 0{,}65) si\ 0{,}65 \leq 0{,}81 \\ -\frac{1}{0{,}17}(x - 0{,}98) si\ 0{,}81 \leq x \leq 0{,}98 \end{cases} \quad (3.12)$$

$$GRAN = \begin{cases} \frac{1}{0{,}1}(x - 0{,}9) si\ 0{,}9 \leq x \leq 1 \\ 1\ si\ x > 1 \end{cases}$$

MODELO II.

La construcción de los conjuntos difusos para las variables del modelo II, se ha realizado con una lógica similar a las variables, así que a continuación se presentan: a) la tabla de frecuencias, b) histograma, c) funciones de pertenencia; y, d) ecuaciones asociadas a las funciones de pertenencia.

1. **Perfil de riesgo.**

Tabla 3.7 Perfil de riesgo – Modelo II

Nro.	Intervalo	Frecuencia	Frec. Relativa
1	[549,00; 584,44)	72	25,35%
2	[619,88; 655,32)	48	16,90%
3	[690,76; 726,20)	40	14,08%
4	[655,32; 690,76)	33	11,62%
5	[584,44; 619,88)	62	21,83%
6	[797,08; 832,52)	7	2,46%
7	[726,20; 761,64)	11	3,87%
8	[761,64; 797,08)	9	3,17%
9	[832,52; 868,00)	2	0,70%
	Total	**284**	**100,00%**

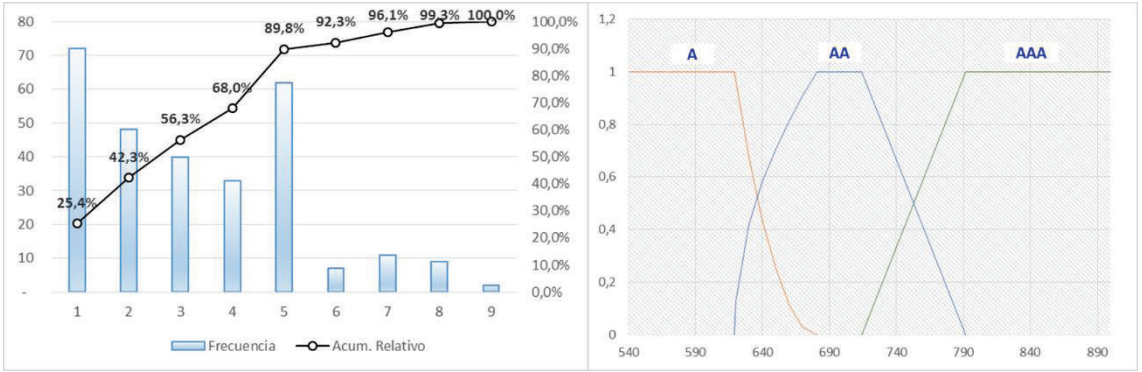

Figura 3.12 Histograma y función de pertenencia de perfil de riesgo – Modelo II

$$A = \begin{cases} 1 & si\ 540 \leq x \leq 619 \\ \left(\dfrac{-(x-681)}{62}\right)^2 & si\ 619 < x \leq 681 \end{cases}$$

$$AA = \begin{cases} \sqrt{\dfrac{1}{62}(x-619)} & si\ 619 \leq x \leq 681 \\ 1 & si\ 681 < x \leq 714 \\ -\dfrac{1}{78}(x-792) & si\ 714 < x \leq 792 \end{cases} \qquad (3.13)$$

$$AAA = \begin{cases} \dfrac{1}{78}(x-714) & si\ 714 \leq x \leq 792 \\ 1 & si\ x > 792 \end{cases}$$

2. Máximo endeudamiento en el último año.

Tabla 3.8 Máximo endeudamiento último año – Modelo II

Nro.	Intervalo	Frecuencia	Frec. Relativa
1	0	93	32,75%
2	[0,001; 1625)	86	30,28%
3	[1625; 3250)	50	17,61%
4	[3250; 4875)	18	6,34%
5	[4875; 6500)	13	4,58%
6	[6500; 8125)	6	2,11%
7	[8125; 9750)	7	2,46%
8	[9750; 11375)	3	1,06%
9	[11375; 13000)	8	2,82%
	Total general	**284**	**100,00%**

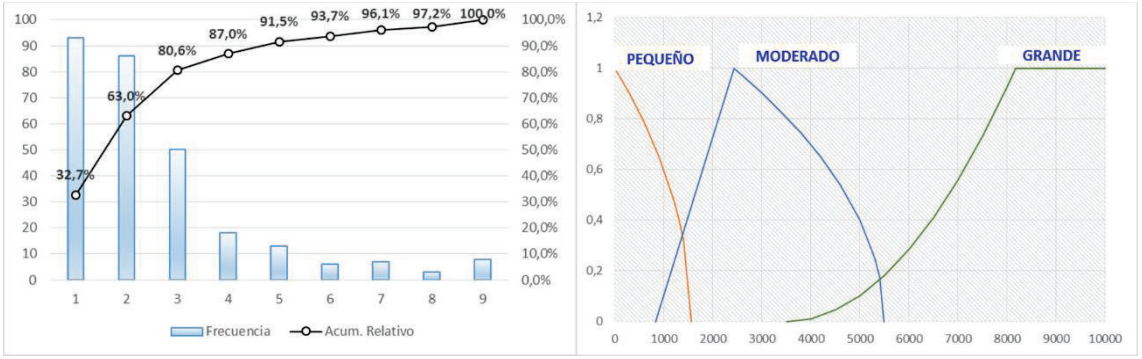

Figura 3.13 Histograma y función de pertenencia de máximo endeudamiento – Modelo II

$$PEQ = \sqrt{-\frac{1}{1560}(x - 1560)} \; si \; 0 \leq x \leq 1560$$

$$MOD = \begin{cases} \frac{1}{1595}(x - 829) & si \; 829 \leq x \leq 2424 \\ \sqrt{-\frac{1}{3068}(x - 5492)} & si \; 2424 < x \leq 5492 \end{cases} \quad (3.14)$$

$$GRAN = \begin{cases} \left(\frac{x - 3500}{4675}\right)^2 & si \; 3500 \leq x \leq 8175 \\ 1 & si \; x > 8175 \end{cases}$$

3. Saldo adeudado en la competencia.

Tabla 3.9 Endeudamiento en la competencia – Modelo II

Nro.	Intervalo	Frecuencia	Frec. Relativa
1	0	101	35,56%
2	[0,001; 1375)	83	29,23%
3	[1375; 2750)	46	16,20%
4	[2750; 4125)	16	5,63%
5	[4125; 5500)	13	4,58%
6	[5500; 6875)	6	2,11%
7	[6875; 8250)	9	3,17%
8	[8250; 9625)	6	2,11%
9	[9625; 11000)	4	1,41%
	Total general	**284**	**100,00%**

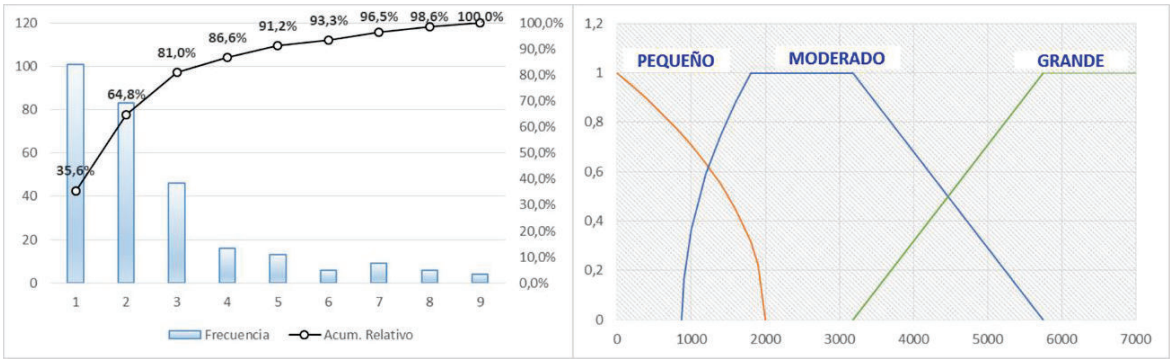

Figura 3.14 Histograma y función de pertenencia de saldo adeudado – Modelo II

$$PEQ = \sqrt{-\frac{1}{2000}(x - 2000)} \; si \; 0 \leq x \leq 2000$$

$$MOD = \begin{cases} \sqrt{\frac{1}{935}(x - 874)} & si \; 874 \leq x \leq 1809 \\ 1 & si \; 1809 < x \leq 3179 \\ -\frac{1}{2571}(x - 5750) & si \; 3179 < x \leq 5750 \end{cases} \quad (3.15)$$

$$GRAN = \begin{cases} \frac{1}{2571}(x - 3179) & si \; 3179 \leq x \leq 5750 \\ 1 & si \; x > 5750 \end{cases}$$

4. Variación del endeudamiento en el último semestre.

Tabla 3.10 Variación del endeudamiento último semestre – Modelo II

Nro.	Intervalo	Frecuencia	Frec. Relativa
1	0	11	3,87%
2	[0,01; 0,14)	2	0,70%
3	[0,14; 0,28)	1	0,35%
4	[0,28; 0,42)	10	3,52%
5	[0,42; 0,56)	8	2,82%
6	[0,56; 0,7)	16	5,63%
7	[0,70; 0,84)	41	14,44%
8	[0,84; 1,00]	85	29,93%
9	>1	110	38,73%
	Total general	**284**	**100,00%**

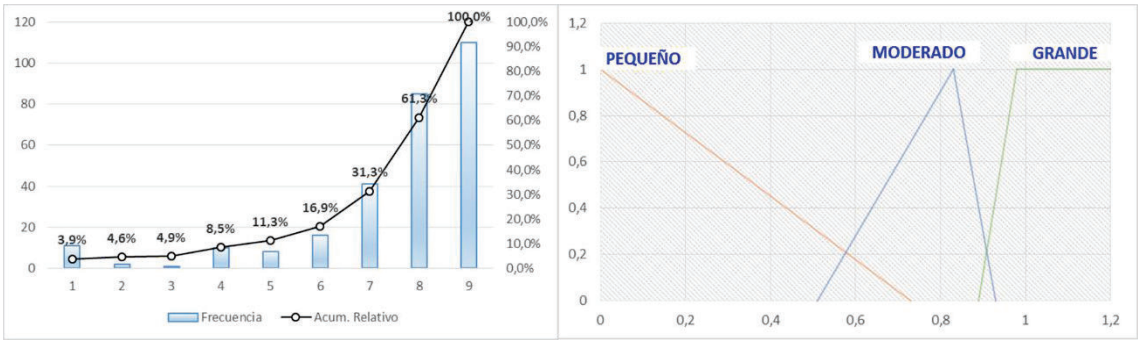

Figura 3.15 Histograma y función de pertenencia de variación de endeudamiento – Modelo II

$$PEQ = -\frac{1}{0{,}73}\ si\ 0 \leq x \leq 0{,}73$$

$$MOD = \begin{cases} \frac{1}{0{,}32}(x - 0{,}51) & si\ 0{,}51 \leq x \leq 0{,}83 \\ -\frac{1}{0{,}1}(x - 0{,}93) & si\ 0{,}83 < x \leq 0{,}93 \end{cases} \quad (3.16)$$

$$GRAN = \begin{cases} \frac{1}{0{,}09}(x - 0{,}89) si\ 0{,}89 \leq x \leq 0{,}98 \\ 1\ si\ x > 0{,}98 \end{cases}$$

VARIABLE DE SALIDA DEL MODELO I Y II.

Evidentemente el propósito de los dos modelos es estimar un cupo de crédito para que sea ofertado como pre-aprobado por la plataforma comercial de ABC. Las funciones de pertenencia se han construido, en primera instancia, con conjuntos difusos de igual amplitud en el eje de las abscisas, pero durante la construcción del modelo se han realizado modificaciones para ajustar la estimación del cupo de crédito a los requerimientos de la institución, tales como el cumplimiento de objetivos de ventas y posicionamiento en el mercado financiero. En ese sentido, los conjuntos difusos que se han construido para el cupo de crédito (del modelo I y II), se presenta en la figura 3.16 y más abajo las ecuaciones asociadas.

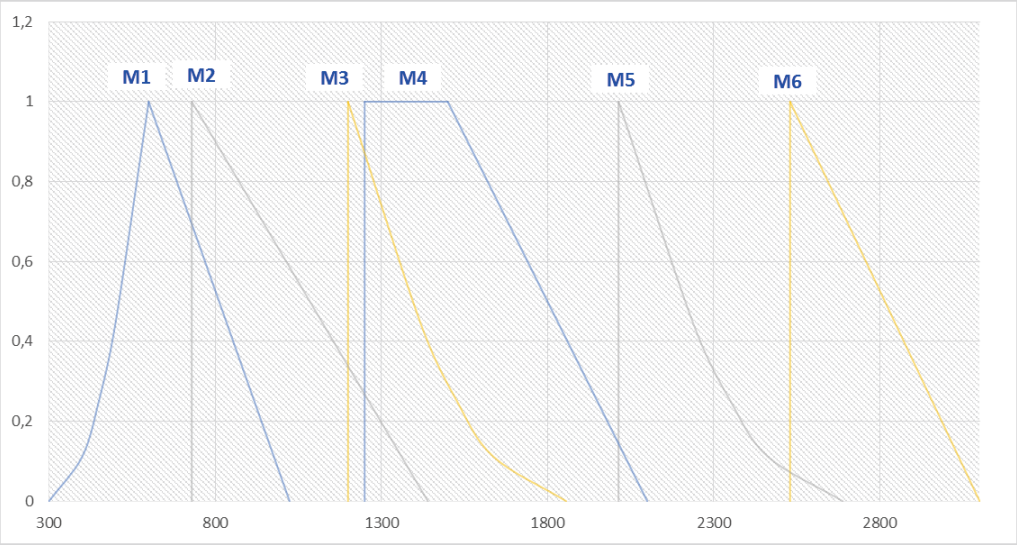

Figura 3.16 Conjuntos difusos del cupo de crédito – Modelo I y II

$$
\begin{aligned}
M1 &= \begin{cases} \left(\dfrac{x-300}{300}\right)^2 & si\ 300 \le x \le 600 \\ -\dfrac{1}{424}(x-1024) & si\ 600 < x \le 1024 \end{cases} \\
M2 &= \begin{cases} 1 & si\ x = 730 \\ -\dfrac{1}{710}(x-1440) & si\ 730 < x \le 1440 \end{cases} \\
M3 &= \begin{cases} 1 & si\ x = 1200 \\ \left(\dfrac{-(x-1856)}{656}\right)^2 & si\ 1200 < x \le 1856 \end{cases} \\
M4 &= \begin{cases} 1 & si\ 1250 \le x \le 1500 \\ -\dfrac{1}{600}(x-2100) & si\ 1500 < x \le 2100 \end{cases} \\
M5 &= \begin{cases} 1 & si\ x = 2014 \\ \left(\dfrac{-(x-2688)}{674}\right)^2 & si\ 2014 < x \le 2688 \end{cases} \\
M6 &= \begin{cases} 1 & si\ x = 2530 \\ -\dfrac{1}{570}(x-3100) & si\ 2530 < x \le 3100 \end{cases}
\end{aligned}
\tag{3.17}
$$

3.3.4 REGLAS TIPO MANDANI.

En la sección 3.1.3 se habló sobre las reglas Mandani como un método para relacionar las variables de entrada y salida de un modelo de inferencia difusa. En este apartado se expone la metodología para la construcción de las reglas tipo Mandani; los pasos a seguir son los mismos para el modelo I y II, lo que varía son los parámetros y se detallan a continuación:

1. A los valores lingüísticos de cada variable se asigna un valor numérico, según se muestra en la tabla 3.11.
2. Se forman las combinaciones posibles entre los valores lingüísticos de todas las variables; es decir, para cada modelo se tienen 81 reglas para la asignación de cupos, por ejemplo, para el modelo I, se tiene la regla que se muestra en la tabla 3.12.

3. Determinar el promedio de los números asociados a los valores lingüísticos de cada regla. En el ejemplo de la tabla 3.12, el promedio de los valores numéricos asociados a la regla es 1,75.

4. Sobre el promedio obtenido con las 81 reglas se calcula los percentiles para obtener una partición de 6 grupos de esas observaciones. En el ejemplo de la tabla 3.12, al valor 1,75 le corresponde un Monto 2, de acuerdo a lo escrito en la tabla de asignación de montos del modelo I (tabla 3.13); es decir, bajo las características de endeudamiento que presenta el potencial cliente de ABC, se sugiere un cupo de crédito para que sea gestionado por la plataforma comercial de la entidad.

Tabla 3.11 Equivalencia entre valores lingüísticos y numéricos

MODELO I		MODELO II	
Perfil de riesgo		**Perfil de riesgo**	
AAA	3	AAA	3
AA	2	AA	2
A	1	A	1
Carga financiera máxima		**Máximo endeudamiento en el último año**	
PEQUEÑO	1	PEQUEÑO	1
MODERADO	2	MODERADO	2
GRANDE	3	GRANDE	3
Saldo adeudado a la competencia		**Saldo adeudado a la competencia**	
PEQUEÑO	3	PEQUEÑO	3
MODERADO	2	MODERADO	2,5
GRANDE	1	GRANDE	1,5
Variación de la deuda en el último semestre		**Variación de la deuda en el último semestre**	
PEQUEÑA	3	PEQUEÑA	3
MODERADA	2	MODERADA	2
GRANDE	1	GRANDE	1

Tabla 3.12 Ejemplo de equivalencia entre valores lingüísticos y numéricos

Variable	Valor lingüístico	Valor numérico
Perfil de riesgo	AAA	3
Carga financiera máxima	Pequeña	1
Saldo adeudado en la competencia	Moderada	2
Variación de la deuda en el último semestre	Grande	1

Tabla 3.13 Percentiles asociados a las reglas tipo Mandani

MODELO I		
Lim. Inf.	**Lim. Sup.**	**Cupo de crédito**
0	1,5	Monto1
1,501	1,75	Monto2
1,751	2	Monto3
2,001	2,25	Monto4
2,251	2,5	Monto5
2,501	3	Monto6

MODELO II		
Lim. Inf.	**Lim. Sup.**	**Cupo de crédito**
0	1,625	Monto1
1,6251	1,875	Monto2
1,8751	2,125	Monto3
2,1251	2,25	Monto4
2,2501	2,5	Monto5
2,5001	3	Monto6

Con esta metodología, cada regla relaciona los valores lingüísticos de las variables de entrada con los valores de la variable de salida. En el anexo II se presentan todas las reglas del modelo I, mientras que las del modelo II se encuentran en el anexo III.

3.3.5 AGREGACIÓN.

En la sección 3.1.4 se realizó una formulación teórica de la etapa de agregación y se ha presentado el problema en el que se conoce la regla tipo Mandani y un antecedente nítido. Para ejemplificar la metodología con los datos obtenidos de los clientes que recibieron su primera operación en ABC, se toma como ejemplo a un individuo de la muestra del modelo I que presenta las siguientes características.

Tabla 3.14 Vector de características de un individuo – Modelo I

Variable	Valor numérico
Perfil de riesgo	770
Carga financiera máxima	$ 180,00
Saldo adeudado en la competencia	$ 958,52
Variación de la deuda en el último semestre	94,57%

Al ubicar los valores en las funciones de pertenencia respectivas, se tiene el siguiente escenario:

a) Por las ecuaciones (3.9), al individuo pertenece al conjunto difuso “PERFIL AA” con grado de membresía 0,53 y al conjunto difuso “PERFIL AAA” con grado de membresía 0,47.
b) Respecto a la variable carga financiera máxima, las ecuaciones (3.10) indican que el individuo pertenece al conjunto difuso “CFM PEQUEÑA” con grado 1; es decir, su pertenencia es absoluta y únicamente a ese conjunto.
c) Por el valor del saldo adeudado en la competencia, las ecuaciones (3.11) indican que el individuo pertenece únicamente al conjunto difuso “SALDO MODERADO” con grado de membresía 1.
d) La variación de la deuda en el último semestre de 94,57% sugiere que el individuo pertenece al conjunto difuso “VARIACIÓN MODERADA” con grado de membresía 0,20 y al conjunto “VARIACIÓN GRANDE” con grado 0,46.

Por lo tanto, las reglas tipo Mandani que intervienen para la evaluación del individuo en cuestión se detallan en tabla 3.15, mostrándose que los montos asignados por el modelo son 1, 2 y 3.

Tabla 3.15 Reglas tipo Mandani para un vector de características

	Variable	Regla 1	Regla 2	Regla 3	Regla 4
Antecedente	Perfil de riesgo	AA	AA	AAA	AAA
	Carga financiera máxima	PEQUEÑA	PEQUEÑA	PEQUEÑA	PEQUEÑA
	Saldo adeudado en la competencia	MODERADO	MODERADO	MODERADO	MODERADO
	Variación de la deuda en el último semestre	MODERADA	GRANDE	MODERADA	GRANDE
Consecuente	Cupo de crédito	MONTO 2	MONTO 1	MONTO 3	MONTO 2
	Nro. Regla	40	43	67	70

El número de regla que se menciona en la tabla 3.15, es el mismo que se encuentra en la tabla del anexo II. Al reemplazar los valores lingüísticos por sus respectivos grados de pertenencia a los conjuntos difusos del antecedente de las reglas tipo Mandani y calcular la T-norma del mínimo definida en la sección 3.1.4, se tienen los resultados que se explican en la tabla 3.16. Gráficamente, el cálculo de la T-norma del

mínimo se muestra en la figura 3.17, donde el trapecio de color plomo es el resultado de aplicar el mínimo entre los valores nítidos que caracterizan la regla 1.

Tabla 3.16 T-norma del minimo para el antecedente de las reglas Mandani

Variable	Regla 1	Regla 2	Regla 3	Regla 4
Perfil de riesgo	0,53	0,53	0,47	0,47
Carga financiera máxima	1,00	1,00	1,00	1,00
Saldo adeudado en la competencia	1,00	1,00	1,00	1,00
Variación de la deuda en el último semestre	0,20	0,46	0,20	0,46
T-norma (mínimo)	**0,20**	**0,46**	**0,20**	**0,46**

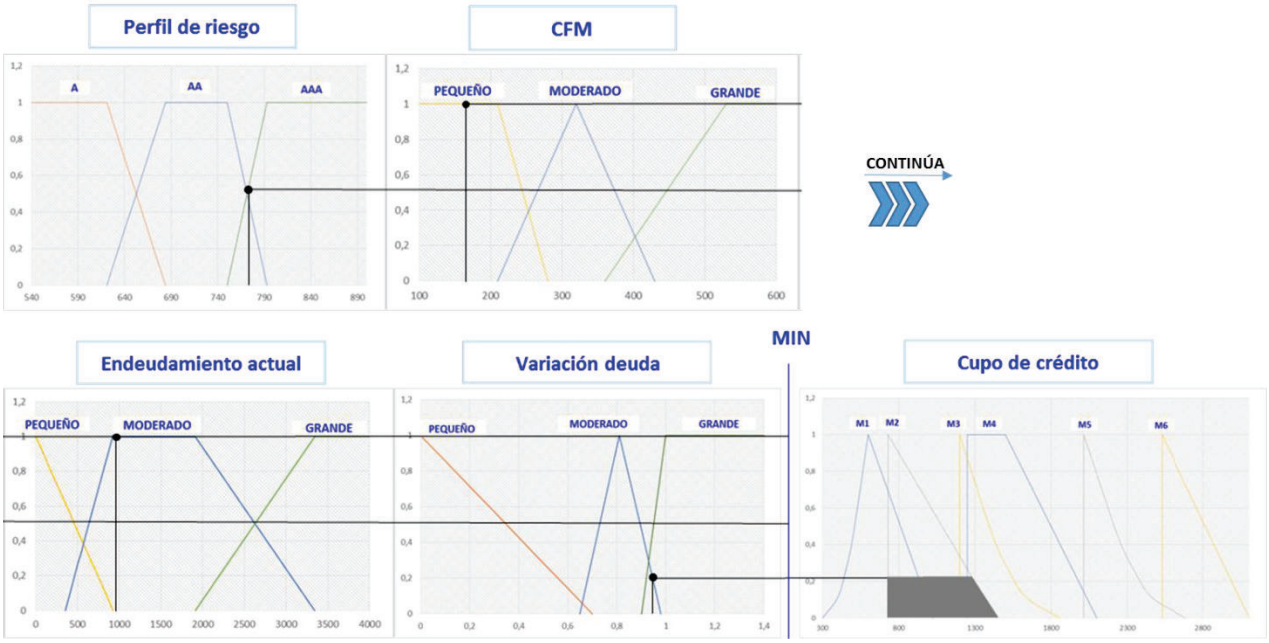

Figura 3.17 T-norma del mínimo sobre las reglas tipo Mandani

Al replicar el resultado de la T-norma sobre las reglas tipo Mandani para esta persona y luego agregarlas en un gráfico se obtiene el resultado de la figura 3.18. De esta manera se concluye la etapa de agregación para la construcción de los modelos I y II, evidentemente que la metodología se extiende para cada uno de los individuos de las dos muestras consideradas.

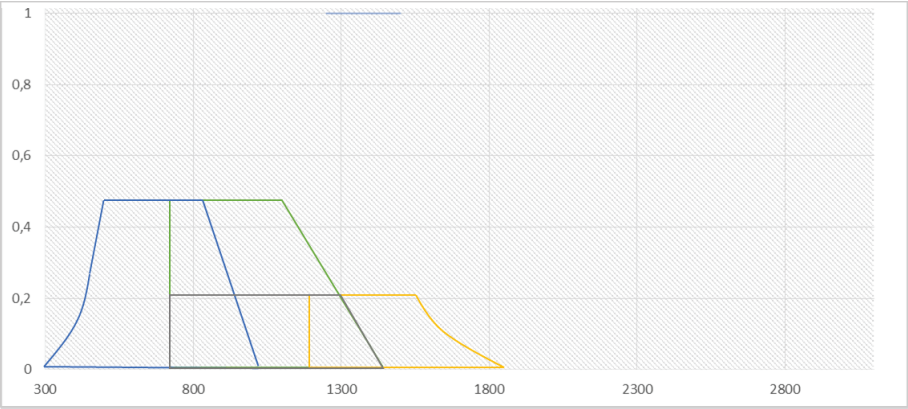

Figura 3.18 Agregación de la T-norma

3.3.6 VALORES CONCRETOS PARA EL CUPO DE CRÉDITO.

En la sección 3.1.5 se ha recalcado que la etapa de concreción consiste en transformar un conjunto o conjuntos difusos en un valor nítido o concreto. Entre los métodos presentados en esa sección, se ha escogido MOM (middle of maximun).

Siguiendo con el ejemplo presentado, de todos los conjuntos agregados se toma el de mayor valor en el eje de las ordenadas, en este caso 0,46. Entre ellos se busca el punto medio en el eje de las abscisas. El valor concreto o cupo de crédito que le corresponde al individuo es 809,17, que es punto medio entre 502,80 y 1115,56.

Este esquema de trabajo se replica sobre toda la muestra seleccionada para los modelos I y II, de esta manera todos disponen de un cupo de crédito sugerido.

3.4 EVALUACIÓN DE RESULTADOS.

Luego de presentar la formulación teórica y práctica para la construcción de un modelo de inferencia difusa, en esta sección se muestran los resultados obtenidos. Los principales puntos que se han considerado, como parte de la evaluación de resultados, son:

a) El cupo que se obtiene con el modelo difuso, se compara con el verdadero monto que fue otorgado al cliente en ABC. Es importante mencionar que los valores se desembolsan luego de la respectiva evaluación crediticia, la misma que se basa en la documentación presentada por el solicitante y la información consultada en el buró de créditos. En la evaluación crediticia se consideran las recomendaciones emitidas por la SBS que ya se ha mencionado anteriormente, en lo que se refiere a la estimación de los cupos de crédito. Por lo tanto, el cupo otorgado por ABC, es una variable que contiene información muy confiable y es por esta razón se la usa en la comparación de las metodologías.
b) Los cupos estimados por el modelo difuso se ofertan a un plazo de 12 meses. Esto se debe a la lógica usada en la relación de las variables de entrada y salida del modelo.
c) Para que los cupos real y estimado sean comparables entre sí, se debe calcular los montos reales a un plazo de 12 meses, pues debido a la lógica de colocación de la entidad, los créditos se otorgan con diferentes plazos.

Estas consideraciones aplican para la evaluación de los resultados obtenidos con los dos modelos.

MODELO I.

La tabla 3.17 muestra las diferencias entre el cupo de crédito real y estimado. Las sub estimaciones no deben preocupar debido a que un cambio de metodologías para el cálculo de cupos de crédito es esperado ese tipo de situaciones; en cambio, en el caso de las sobre estimaciones se debe fijar un umbral para determinar las bondades del modelo propuesto. En este caso, la institución fija como umbral de sobre estimación hasta el intervalo 8 de la tabla 3.17; es decir, la metodología propuesta en el modelo se ajusta de manera similar a los cupos de crédito reales en el 80% de la muestra.

El argumento que sustenta la aceptación de una sobre estimación de los cupos de crédito hasta el umbral fijado de 467 USD es que, ese valor representa una cuota

mensual aproximada de 42 dólares durante 12 meses, que es el plazo fijado en las estimaciones.

Tabla 3.17 Evaluación de resultados – Modelo 1

Nro.	Intervalo	Frecuencia	Frec. Relativa
1	[-2400,00; -2041,54)	28	0.68%
2	[-2041,53; -1683,08)	35	0.85%
3	[-1683,07; -1324,62)	58	1.40%
4	[-1324,61; -966,16)	111	2.68%
5	[-966,15; -607,7)	374	9.03%
6	[-607,69; -249,24)	703	16.98%
7	[-249,23; 109,22)	1,040	25.11%
8	[109,22; 467,68)	880	21.25%
9	[467,68; 826,14)	548	13.23%
10	[826,14; 1184,6)	206	4.97%
11	[1184,60; 1543,06)	84	2.03%
12	[1543,06; 1901,52)	49	1.18%
13	[1901,52; 2260,00)	25	0.60%
	Total general	**4,141**	**100.00%**

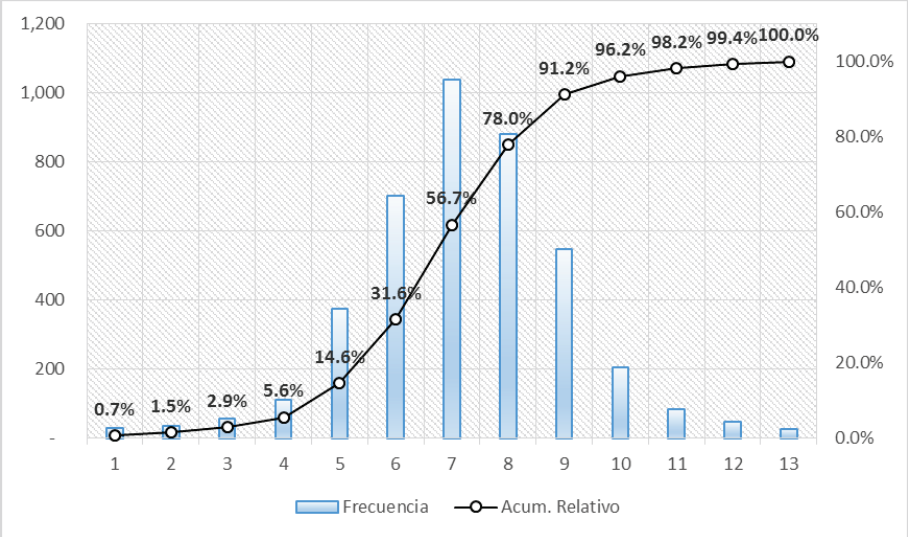

Figura 3.19 Evaluación de resultados – Modelo I

Respecto al 20% restante, para el que se sobre estima el cupo de crédito a ofertar como pre aprobado, el lector debe recordar que cuando el asesor de crédito contacta a una persona que dispone de un cupo pre aprobado y está sobre estimado, tiene la

opción de negociar el plazo de la operación, de tal manera que la cuota mensual sea menor y de esta manera el potencial cliente puede acceder al crédito.

Es importante recordar que en cualquiera de los casos anteriores, previo al desembolso del crédito, la institución tiene la obligación de verificar el origen de los ingresos, el valor real del salario y los gastos mensuales de los clientes nuevos; esta información debe registrarse en el sistema interno del banco y con estos datos es necesario verificar si el cupo de crédito pre aprobado es pertinente, de no serlo, el asesor de crédito puede negociar las condiciones del crédito o negar el concesión del mismo. Las prácticas anteriores de ABC indican que las decisiones más comunes de los clientes nuevos son: a) aceptar el cupo de crédito con un plazo mayor al ofertado por el proceso de pre aprobación; y, b) aceptar un cupo de crédito menor al ofertado y con plazo de 12 meses.

Otro aspecto importante a considerar es que, el volumen total de ofertas comerciales no se ve afectado al introducir una nueva metodología en el proceso *batch* para la estimación de cupos de crédito; este hecho se puede evidenciar en la figura 3.20. Esa situación es trascendental, pues los asesores comerciales trabajan en base a objetivos de ventas mensuales, y cualquier variación en el volumen de ofertas puede afectar drásticamente en el cumplimiento de sus objetivos.

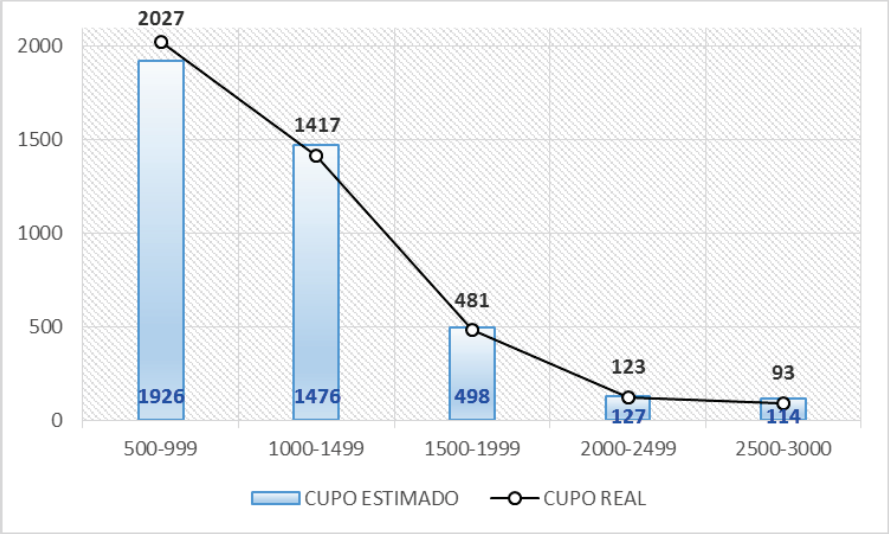

Figura 3.20 Comparación de los cupos de crédito reales y estimados – Modelo I

MODELO II.

Los resultados del modelo II son similares a los obtenidos en el modelo I. En este caso, la sobre estimación representa el 18% de la muestra y sobre este grupo de individuos aplican las mismas consideraciones realizadas con los individuos del modelo I.

Notar que en la figura 3.22, al igual que en el modelo I, se evidencia que el volumen de ofertas comerciales es similar al volumen de colocación real. Por lo tanto el modelo cumple con el objetivo fundamental de inferir un cupo de crédito similar al que reciben los potenciales clientes en su primera operación en ABC, a pesar de que al momento de ejecutar el proceso de pre calificación *batch* no se conoce el salario de los potenciales clientes. Es así que se logra optimizar el uso de los recursos de información de la institución, incluyendo en la estimación de los cupos de crédito para los potenciales clientes que no disponen de la información del salario.

Tabla 3.18 Evaluación de resultados – Modelo II

Nro.	Intervalo	Frecuencia	Frec. Relativa
1	[-2400,00; -1924,44)	3	1.06%
2	[-1924,43; -1448,88)	2	0.70%
3	[-1448,87; -973,32)	9	3.17%
4	[-973,31; -497,76)	28	9.86%
5	[-497,75; -22,2)	75	26.41%
6	[-22,19; 453,36)	114	40.14%
7	[453,36; 928,92)	36	12.68%
8	[928,92; 1404,48)	13	4.58%
9	[1404,48; 1880,00)	4	1.41%
	Total general	**284**	**100.00%**

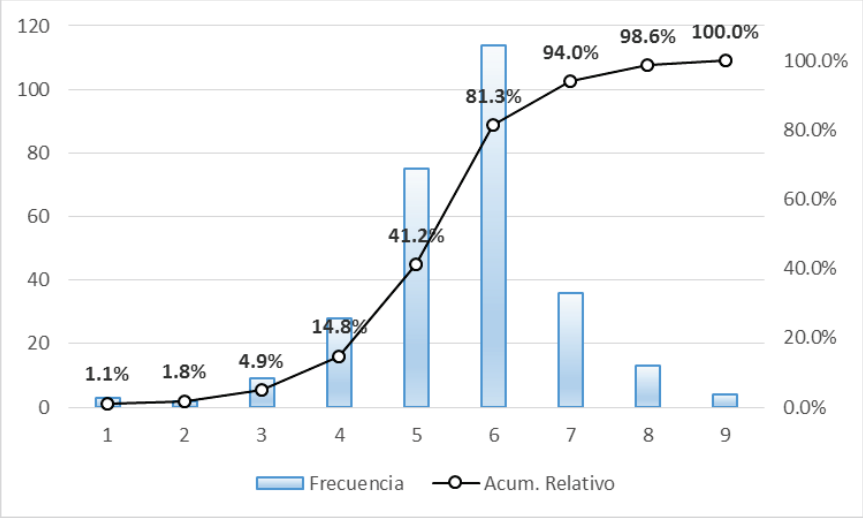

Figura 3.21 Evaluación de resultados – Modelo II

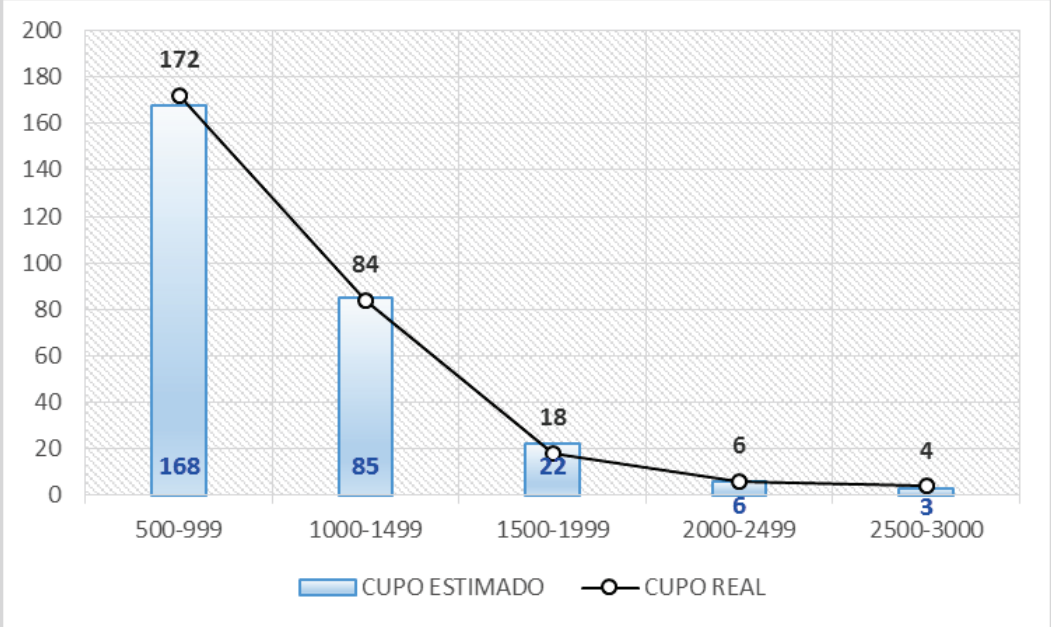

Figura 3.22 Comparación de los cupos de crédito reales y estimados – Modelo II

CAPÍTULO 4
CONCLUSIONES Y RECOMENDACIONES.

4.1 CONCLUSIONES.

Este proyecto se ha centrado en el desarrollo de una metodología para la evaluación crediticia integral de potenciales clientes de la institución, la misma que incluye el análisis del perfil de riesgo y la estimación de un cupo de crédito dentro del proceso de pre-aprobación *batch* de ABC. La metodología se convierte en un aporte a los procesos operativos de la entidad en lo que se refiere al otorgamiento de créditos. En ese sentido, los resultados obtenidos en el presente trabajo permiten exponer las siguientes conclusiones.

1. La normativa vigente emitida por la SBS exige que ABC enfoque su gestión y administración de riesgo de crédito en el desarrollo de metodologías de evaluación crediticia para la segmentación de solicitantes y otorgamiento de créditos, en particular los relativos a las carteras de consumo de las instituciones financieras reguladas. Así mismo, establece como factor importante para el otorgamiento de créditos, la determinación de la capacidad de pago y la estabilidad de la fuente de sus ingresos.

   Las mejores prácticas recomiendan el uso de modelos *scoring* para la evaluación del perfil de riesgo, los mismos que varían dependiendo de la etapa del ciclo de crédito en la que se construyen, pudiendo ser: preventa, análisis, seguimiento y/o cobranzas. Los beneficios que traen consigo son numerosos y se enfocan fundamentalmente en la optimización de la segmentación de las carteras de clientes y potenciales clientes de ABC, esto con el propósito de mejorar el rendimiento de la institución y obtener un mejor posicionamiento en el mercado financiero.

2. La construcción del modelo difuso exige un trabajo multidisciplinario, con la intervención de los integrantes de las diferentes unidades de ABC: Gerencia, Riesgos, Auditoría y Gestión Comercial. Cada uno desempeña funciones específicas tanto en la etapa de construcción de la metodología propuesta como en su futura implementación. Esta interacción permite retroalimentar y enriquecer la metodología propuesta, pues además del análisis matemático y estadístico se incluye el criterio de los expertos.

3. La incorporación de una correcta segmentación de potenciales clientes y estimación de cupos de crédito al proceso *batch* de ABC, permite identificar el segmento de mercado con el que la institución podría tener una oportunidad de negocio, a la vez permite determinar el valor monetario que se va ofertar en calidad de cupo de crédito pre-aprobado. Estas acciones contribuyen al desarrollo del CRM de la institución, el mismo que busca identificar, captar, satisfacer, retener y potenciar relaciones rentables con los mejores clientes de la institución o los potenciales clientes con mejor perfil.

4. La regresión logística binaria es una herramienta estadística confiable dentro de los mecanismos para analizar el perfil de riesgo crediticio de los clientes y potenciales clientes de la institución. Dentro de la etapa de construcción del modelo se logra incorporar criterios estadísticos importantes para garantizar que el modelo obtenido es robusto y estable. Es así que, desde la elección de la variable dependiente se garantiza que la definición escogida en este proyecto es estable, por medio de los indicadores estadísticos: IPS (Índice de estabilidad de la población) y test de Dickey Fuller, los cuales han mostrado resultados favorables. Por otro lado, la selección de las variables que discriminan de mejor manera la población de potenciales clientes buenos y malos ha sido posible gracias a que se ha incluido los criterios: análisis exploratorio, árboles de decisión, estabilidad de la población (IPS), IV (valor de la información) y test Chi-Cuadrado, los mismos que han permitido garantizar la estabilidad en el tiempo de la proporción de individuos buenos y malos de las variable incluida en el modelo final. Así mismo, se incluye el criterio de inflación de la varianza

para tratar el problema de multicolinealidad, y de esta manera se ha obtenido una regresión preliminar, a partir de la cual, por el criterio *bootstrap* se ha determinado un modelo óptimo, el mismo que resulta de haber realizado 1000 simulaciones con 5130 individuos de la muestra de aprendizaje.

Los indicadores de bondad de ajuste: R cuadrado, Hosmer-Lemeshow, estadístico de Wald y prueba de Omnibus demuestran estadísticamente que el modelo ajusta bien a los datos y que los coeficientes son significativos (diferentes de cero).

Por obvias razones la proporción de potenciales clientes buenos y malos es diferente (76% y 24% respectivamente) en la muestra de aprendizaje, eso provoca un desplazamiento sobre la curva logística, viéndose afectada la predicción de la variable dependiente del modelo. El problema se ha solventado con la corrección propuesta por Maddala, afectando únicamente la constante del modelo. El efecto se observa cuando el punto de corte de las curvas de sensibilidad y especificidad, pasa de 0,74 a 0,47 en el eje de las ordenadas.

5. La evaluación de la eficiencia del modelo para discriminar la población de potenciales clientes buenos y malos, se ha realizado por medio de los indicadores KS y Gini, obteniendo resultados satisfactorios luego de aplicar *bootstrap* con 1000 muestras de tamaño 500 (elementos tomados de la muestra de aprendizaje) para estimar los intervalos de confianza. Los indicadores han permitido concluir que la calidad discriminante del modelo se mantiene si se usa una ventana de datos diferente a la considerada en la etapa de construcción.

6. La metodología planteada para la construcción del modelo logístico cumple con los parámetros exigidos por la SBS. En el estricto sentido, para seleccionar la muestra de estudio se ha considerado una ventana de 3 años y a partir de ella se ha realizado el análisis y construcción del modelo, tal como lo menciona la normativa; respecto a los indicadores KS y Gini, se han obtenido valores que superan el mínimo que sugiere la SBS para modelos de originación, 25% y 35% respectivamente. Más aún, los límites de los intervalos de confianza que se han

estimado para los indicadores KS y Gini superan los valores mínimos sugeridos por la SBS.

7. El capital de los créditos de consumo (sin considerar las tarjetas de crédito) es amortizado mensualmente, en la mayoría de los casos, esto permite evidenciar una curva de endeudamiento lineal decreciente, siempre y cuando el individuo pague a tiempo las cuotas pactadas en la tabla de amortización. Este es el principal hecho que motiva la construcción de un modelo difuso, pues la metodología planteada permite establecer reglas que relacionan las variables que describen la forma en que un potencial cliente administra su deuda en la competencia con el cupo de crédito pre-aprobado, por medio de una base de conocimiento que incluye por un lado, el criterio estadístico que se obtiene con el análisis de los datos y por otro, el criterio de los profesionales expertos. Los criterios mencionados hacen de la metodología una herramienta útil para la estimación de cupos de crédito, especialmente en el caso de los individuos para los que no se conoce el salario al momento de ejecutar el proceso *batch*, el mismo que representa el 14% de la base de clientes nuevos de ABC.

8. Para estimar el cupo de crédito para los potenciales clientes se han construido dos modelos: el primero incluye a los individuos que disponían la información de salario para generar una oferta pre-aprobada en el proceso *batch* a partir del cual se otorgó su primera operación en ABC, en cambio el segundo considera a los todos individuos para los que no se conocía esa información. Ambos modelos incluyen estrategias comerciales diferenciadas por perfil de riesgo, orientando las ofertas mayores hacia los individuos con mejor perfil de riesgo crediticio, de esta manera se logra combinar el resultado de la regresión logística con la metodología de los sistemas de inferencia difusa. Las variables que complementan el modelo son aquellas que describen la forma en que una persona administra su deuda en la competencia; para lograr esto se ha contado con el criterio de profesionales expertos. En el desarrollo del proyecto se ha decidido utilizar las reglas tipo Mandani para relacionar las variables de entrada del modelo con el cupo de crédito.

9. Por un lado se logra ajustar la curva de ofertas comerciales en términos del valor monetario total de las ofertas estimadas, y por otro se establece una curva que mide las diferencias entre la metodología sugerida por la SBS y la que propone este proyecto. Este análisis ha permitido concluir que los resultados obtenidos son satisfactorios para ABC, debido a que es muy importante no alterar el cupo promedio de ofertas comerciales ya que los asesores comerciales trabajan en base a objetivos de ventas mensuales, cualquier variación en el volumen de las ofertas puede afectar el cumplimiento de las metas planteadas. Por otro lado, la curva de diferencias muestra que la metodología sobre estima los cupos para el 20% de los individuos que intervienen en el modelo I y el 18% para el modelo II; esto no es una situación alarmante, pues como se dijo anteriormente, la metodología necesariamente debe articularse a los procedimientos actuales de ABC para el otorgamiento de créditos a clientes nuevos que se pre-califican en el proceso *batch*.
10. La construcción de una metodología analítica y estadística se convierte en una herramienta objetiva y estandarizada que ayuda a minimizar el riesgo operativo en los procesos de otorgamiento de crédito de ABC. El modelo propuesto no busca reemplazar el criterio de los asesores de crédito al momento de realizar la evaluación crediticia antes de otorgar un crédito, el propósito es que se convierta en una herramienta de ayuda y que permita optimizar la calidad de servicio al cliente.

4.2 RECOMENDACIONES.

1. Como parte del proceso de administración del riesgo de crédito, los modelos *scoring* que utiliza ABC deben ser monitoreados constantemente y en base a ello actualizar las políticas de aprobación de crédito, y de ser el caso realizar

una actualización de los modelos; esto con la finalidad de definir el nivel de riesgo considerado como aceptable para la institución.

2. En las nuevas propuestas para la construcción de modelos *scoring*, podría evaluarse la posibilidad de aplicar otros métodos de selección de variables adicionales a los que se han presentado en este trabajo, por ejemplo basados en criterios de información (AIC, BIC) y con estrategias de selección hacia adelante o hacia atrás. Así mismo, en la evaluación del perfil de riesgo podría considerarse, como alternativa, los modelos *probit* o lineal generalizado tipo *Poisson* truncado, pues permitirían obtener diferentes niveles de riesgo a ser incluidos en el modelo de inferencia difuso.

3. ABC debe ser cuidadosa al momento de implementar los modelos en el proceso *batch*, los resultados deben ser transparentes para todos los usuarios: técnicos expertos en la programación de procesos *batch*, usuarios funcionales (validan los resultados de la implementación) y usuarios comerciales. Esta transparencia ayudará a que la implementación de una nueva metodología adquiera confiabilidad en proceso de evaluación crediticia.

4. El cambio de metodología debe ser socializado con los usuarios comerciales, no con el detalle técnico y estadístico que se ha empleado en la construcción de los modelos, pero sí de una forma en la que se informe de los futuros cambios que pueden observar al momento de contactar a los potenciales clientes registrados en el proceso *batch*. Estas acciones también ayudan a incrementar el nivel confiabilidad de la nueva metodología.

5. La metodología propuesta forma parte de los procesos operativos de ABC, por ello la entidad debe ser cuidadosa al momento de vincular la nueva metodología a los procesos propios de concesión de créditos, especialmente los que se relacionan al cumplimiento de la normativa vigente de la SBS, como por ejemplo la verificación de los documentos habilitantes para el desembolso del dinero.

6. La base de conocimiento se ha diseñado con las reglas tipo Mandani en base a criterios estadísticos y al aporte de los profesionales expertos. Esta matriz de

decisión puede y debe ser actualizada acorde a los requerimientos de la institución. Una práctica recomendable es la evaluación del primer trimestre de vigencia de la metodología para medir el impacto de los nuevos cambios y en función de ello decidir si ameritan actualizaciones sobre las reglas. Un método adicional que podría considerarse en trabajos futuros, es el TSK como alternativa a las reglas tipo Mandani.

7. En los nuevos trabajos de ABC se podría estudiar la estimación de cupos para las carteras de microempresa y tarjeta de crédito; teniendo en cuenta que la evaluación crediticia debe profundizar en el análisis de los ratios financieros del negocio del solicitante cuando se trata de microempresarios, y considerando que en el caso de la tarjeta de crédito tiene una dinámica de línea de crédito aprobada que el cliente la utiliza de acuerdo a sus requerimientos.

REFERENCIAS BIBLIOGRÁFICAS

[1] S. Medina & G. Paniagua, "*Modelo de inferencia difuso para estudio de crédito*", Universidad Nacional de Colombia, Medellín, Colombia, Dyna, Mar. 2008.

[2] A. Novales, *Econometría*, 2da ed. Madrid, España: McGraw-Hill, 1993.

[3] G. S. Maddala, *Introducción a la Econometría*, 2nd ed. México: Prentice-Hall Hispanoamericana, 1996.

[4] A. M. Soto & S. Medina, "*Desarrollo de un sistema de inferencia difuso para la evaluación de crédito por parte de una empresa prestadora de servicios*", Universidad Nacional de Colombia, Medellín, Colombia, Dyna, Abr. 2004.

[5] C. Morcillo, "*Lógica Difusa. Una introducción práctica*", Técnicas de Softcomputing.

[6] Superintendencia de Bancos y Seguros del Ecuador. *LIBRO I.- NORMAS GENERALES PARA LAS INSTITUCIONES DEL SISTEMA FINANCIERO. TÍTULO IX.- DE LOS ACTIVOS Y DE LOS LÍMITES DE CRÉDITO.*

[7] Superintendencia de Bancos y Seguros del Ecuador. *LIBRO I.- NORMAS GENERALES PARA LAS INSTITUCIONES DEL SISTEMA FINANCIERO. TÍTULO X.- DE LA GESTIÓN Y ADMINISTRACIÓN DE RIESGOS.*

[8] R. Anderson, *The Credit Scoring Toolkit,* Oxford University Press Inc., New York, 2007.

ANEXO I

REQUERIMIENTO DE INFORMACIÓN SOLICITADO A LA UNIDAD DE RIESGOS DE ABC.

Grupo variable	Nro. Variable	Variable	Definición
Informativas	1	Fecha de corte	Es el último día del mes correspondiente a la fecha de desembolso del crédito
	2	Identificación	Cédula de identidad del cliente nuevo de ABC
	3	Numero de operación	Número del crédito otorgado al cliente nuevo
Comportamiento de pago	4	Mora máxima 12 meses	Máximo de los días de mora de la operación identificada en el punto 3, calculada durante los 12 meses posteriores al desembolso
	5	Mora promedio 12 meses	Promedio de los días de mora de la operación identificada en el punto 3, calculada durante los 12 meses posteriores al desembolso
	6	Provisión	Provisión de la operación, de acuerdo a lo que la institución reporta a la SBS.
	7	Monto	Monto financiado en el crédito al cliente nuevo de ABC
Socio demográficas	8	Numero de hijos	Indica el número de cargas familiares que tiene el cliente
	9	Salario	Determina el salario registrado por el cliente en el sistema interno de la institución
	10	Antigüedad de residencia	Indica el número de meses que el cliente reside en su vivienda
	11	Género	Género de la persona
	12	Edad	Indica la edad del cliente al momento del desembolso
	13	Tipo de vivienda	Indica el tipo de vivienda del cliente (Propia, arrendada, etc.)
	14	Estado civil	Indica el estado civil del cliente
	15	Nivel de escolaridad	Indica el grado de escolaridad del cliente
	16	Zona de la vivienda	Indica la categoría de la parroquia en donde vive el cliente (urbano, rural)
	17	Tiene teléfono de domicilio	Indica 1 si el cliente tiene teléfono en su vivienda
	18	Tiene teléfono de trabajo	Indica 1 si el cliente tiene teléfono en el lugar donde trabaja
	19	Actividad específica	Actividad económica específica a la que se dedica el cliente
	20	Línea de negocio	Línea del negocio a la que se dedica el cliente
	21	Actividad económica	Actividad económica del cliente.
	22	Sector económico	Sector económico al que pertenece la actividad que desarrolla
	23	Cargo en la empresa	Indica el cargo que tiene el cliente en la empresa donde trabaja

	24	Ciudad de desembolso del crédito	Ciudad donde se desembolsó el crédito
	25	Parroquia de nacimiento	Parroquia de nacimiento del cliente
	26	Ciudad de nacimiento	Cuidad de nacimiento del cliente
	27	Ciudad de domicilio	Ciudad donde vive el cliente
	28	Parroquia de domicilio	Parroquia donde vive el cliente
	29	Barrio de domicilio	Barrio donde vive el cliente
	30	Antigüedad de relación de dependencia	Número de meses desde el primer aporte al IESS del cliente
	31	Provincia de domicilio	Provincia en la que vive el cliente
	32	Provincia de desembolso del crédito	Provincia en la que se desembolsó el crédito
	33	Provincia de nacimiento	Provincia en la que nació el cliente
	34	Score genérico	Decil del score vigente
Información del sistema financiero (SF)	35	Número de calificaciones "A" (12 último meses)	Número de veces que el cliente ha sido reportado con peor calificación "A" en los 12 últimos meses en el sistema financiero
	36	Número de calificaciones "B" (12 último meses)	Número de veces que el cliente ha sido reportado con peor calificación "B" en los 12 últimos meses en el sistema financiero
	37	Número de calificaciones "B", "C" o "D" (12 último meses)	Número de veces que el cliente ha sido reportado con peor calificación "C", "D" o "e" en los 12 últimos meses en el sistema financiero
	38	Número de calificaciones en blanco	Número de veces que el cliente no tiene calificación en los 12 últimos meses
	39	Peor calificación al corte	Peor calificación del cliente registrado en el buró al corte
	40	Mora ponderada 12 últimos meses	Promedio entre el máximo de días de mora mensuales durante los 12 últimos meses
	41	Días mora corte	Máximo de días de mora al corte
	42	Días mora promedio (36 último meses)	Promedio de los días de mora de las deudas del cliente reportadas por el buró de crédito en los últimos 36 meses
	43	Días mora promedio (24 últimos meses)	Aplica la definición de la variable 42, pero con una temporalidad de 24 meses
	44	Días mora promedio (12 últimos meses)	Aplica la definición de la variable 42, pero con una temporalidad de 12 meses
	45	Veces en mora (12 últimos meses)	Número de veces en que el cliente ha registrado morosidad en los 12 últimos meses
	46	Veces en mora (36 últimos meses)	Número de veces en que el cliente ha registrado morosidad en los 36 últimos meses
	47	Máximo monto histórico	Máximo monto que ha recibido el cliente por operaciones crediticias con la competencia
	48	Máximo monto histórico consumo	Máximo monto que ha recibido el cliente por operaciones crediticias de consumo con la competencia
	49	Máximo monto histórico comercial	Máximo monto que ha recibido el cliente por créditos comerciales con la competencia

	50	Máximo monto histórico vivienda	Máximo monto que ha recibido el cliente por créditos de vivienda con la competencia
	51	Máximo monto histórico microcrédito	Máximo monto que ha recibido el cliente por operaciones de microcrédito con la competencia
	52	Máximo monto histórico olla de oro	Máximo monto que ha recibido el cliente por operaciones de olla de oro con la competencia
	53	Tiene crédito comercial	Indica 1 si el cliente tiene al menos un crédito comercial activo en la competencia
	54	Tiene crédito de consumo	Aplica la definición de la variable 53 para el caso de operaciones de consumo
	55	Tiene microcrédito	Aplica la definición de la variable 53 para el caso de operaciones de microcrédito
	56	Tiene crédito de olla de oro	Aplica la definición de la variable 53 para el caso de operaciones de olla de oro
	57	Tiene crédito de vivienda	Aplica la definición de la variable 53 para el caso de operaciones de vivienda
	58	Tiene deuda	Indica 1 si el cliente tiene al menos un crédito activo en la competencia
	59	Cartera en riesgo (12 últimos meses)	Es la relación definida por: (Mayor valor vencido + Mayor valor no devenga intereses)/Mayor deuda. Cada uno de los valores se calcula dentro de los últimos 12 meses.
	60	Cartera en riesgo (corte)	Es la relación definida por: (Mayor valor vencido + Mayor valor no devenga intereses)/Mayor deuda. Cada uno de los valores se calcula con información del último mes reportado por el buró de créditos.
	61	Cartera en riesgo (24 últimos meses)	Aplica la misma definición de la variable 59, pero con una temporalidad de 24 meses
	62	Cartera en riesgo (36 últimos meses)	Aplica la misma definición de la variable 59, pero con una temporalidad de 36 meses
	63	Cartera por vencer	Saldo por vencer en el último mes reportado por el buró de créditos
	64	Cartera no devenga intereses	Saldo no devenga intereses en el último mes reportado por el buró de créditos
	65	Cartera vencida	Saldo total vencido en el último mes reportado por el buró de créditos
	66	Cartera en demanda judicial	Saldo en demanda judicial en el último mes reportado por el buró de créditos
	67	Cartera castigada	Saldo castigado en el último mes reportado por el buró de créditos
	68	Operaciones vigentes	Número de deudas vigentes en el último mes reportado por el buró de créditos
	69	Operaciones sin mora	Número operaciones sin mora reportadas en el último mes
	70	Operaciones históricas con demanda	Número de operaciones históricas con demanda judicial que registra el cliente en el sistema financiero
	71	Operaciones castigadas históricas	Número de operaciones castigadas históricas que registra el cliente en el sistema financiero
	72	Entidades competencia	Número de entidades que reportan al cliente en el últimos mes

	73	Entidades de la competencia con saldo vencido	Número de entidades que reportan al cliente con saldo vencido en el último mes
	74	Entidades de la competencia con saldo vencido (36 últimos meses)	Número de entidades que reportan al cliente con saldo vencido en los últimos 36 meses
	75	Saldo vencido de 1 a 30 días (12 últimos meses)	Valor del saldo del cliente con días de mora entre 1 y 30 días
	76	Saldo vencido de 31 a 60 días (12 últimos meses)	Valor del saldo del cliente con días de mora entre 31 y 60 días
	77	Máximo saldo histórico vencido	Máximo saldo del cliente que ha sido reportado por el cliente
	78	Máximo valor no devenga intereses histórico	Máximo saldo histórico del cliente una vez que ha caído en mora
	79	Máximo valor histórico en demanda judicial	Máximo saldo histórico en demanda judicial que registra el cliente en el sistema financiero
	80	Veces al día (12 últimos meses)	Número de periodos en los que el cliente ha estado al día en los 12 últimos meses
	81	Valor total por vencer histórico	Suma de todos los saldos por vencer históricos del cliente
	82	Valor histórico no devenga intereses	Suma de todos los saldos históricos que no devengan intereses del cliente
	83	Valor histórico vencido	Suma de todos los saldos vencidos históricos del cliente
	84	Saldo total al corte	Saldo total del cliente en el mes de desembolso de la operación
	85	Endeudamiento promedio (12 últimos meses)	Promedio de todos los sados que registra el cliente en los 12 últimos meses
	86	Endeudamiento promedio (antes de los 12 últimos meses)	Promedio de todos los sados que registra el cliente desde el mes 12 hacia atrás
	87	Endeudamiento promedio al corte	Promedio de todos los sados que registra el cliente en el mes de desembolso de la operación en ABC
	88	Porcentaje de endeudamiento (12 últimos meses)	Resulta de dividir la variable 85 por el sueldo del individuo
	89	Porcentaje de endeudamiento (antes de los 12 últimos meses)	Resulta de dividir la variable 86 por el sueldo del individuo
	90	Antigüedad en el sistema financiero	Número de meses que el cliente ha sido reportado por el buró de créditos
	91	Número de veces reportado en los 12 últimos meses	Número de periodos en los que el cliente ha sido reportado en los 12 últimos meses
	92	Bancarizado en los últimos 36 meses	Indica 1 si el cliente ha sido bancarizado en los 12 últimos meses

ANEXO II

REGLAS TIPO MANDANI PARA EL MODELO I.

Nro	VALORES LINGÜÍSTICOS				VALORES NUMÉRICOS				Promedio	Cupo Crédito
	Perfil Riesgo	CFM	Saldo Corte	Varicación endeud.	Perfil Riesgo	CFM	Saldo Corte	Varicación endeud.		
1	A	PEQ	PEQ	PEQ	1	1	3	3	2	Monto3
2	A	MED	PEQ	PEQ	1	2	3	3	2,25	Monto4
3	A	GRAN	PEQ	PEQ	1	3	3	3	2,5	Monto5
4	A	PEQ	PEQ	MOD	1	1	3	2	1,75	Monto2
5	A	MED	PEQ	MOD	1	2	3	2	2	Monto3
6	A	GRAN	PEQ	MOD	1	3	3	2	2,25	Monto4
7	A	PEQ	PEQ	GRAN	1	1	3	1	1,5	Monto1
8	A	MED	PEQ	GRAN	1	2	3	1	1,75	Monto2
9	A	GRAN	PEQ	GRAN	1	3	3	1	2	Monto3
10	A	PEQ	MOD	PEQ	1	1	2	3	1,75	Monto2
11	A	MED	MOD	PEQ	1	2	2	3	2	Monto3
12	A	GRAN	MOD	PEQ	1	3	2	3	2,25	Monto4
13	A	PEQ	MOD	MOD	1	1	2	2	1,5	Monto1
14	A	MED	MOD	MOD	1	2	2	2	1,75	Monto2
15	A	GRAN	MOD	MOD	1	3	2	2	2	Monto3
16	A	PEQ	MOD	GRAN	1	1	2	1	1,25	Monto1
17	A	MED	MOD	GRAN	1	2	2	1	1,5	Monto1
18	A	GRAN	MOD	GRAN	1	3	2	1	1,75	Monto2
19	A	PEQ	GRAN	PEQ	1	1	1	3	1,5	Monto1
20	A	MED	GRAN	PEQ	1	2	1	3	1,75	Monto2
21	A	GRAN	GRAN	PEQ	1	3	1	3	2	Monto3
22	A	PEQ	GRAN	MOD	1	1	1	2	1,25	Monto1
23	A	MED	GRAN	MOD	1	2	1	2	1,5	Monto1
24	A	GRAN	GRAN	MOD	1	3	1	2	1,75	Monto2

25	A	PEQ	GRAN	GRAN	1	1	1	1	1	Monto1
26	A	MED	GRAN	GRAN	1	2	1	1	1,25	Monto1
27	A	GRAN	GRAN	GRAN	1	3	1	1	1,5	Monto1
28	AA	PEQ	PEQ	PEQ	2	1	3	3	2,25	Monto4
29	AA	MED	PEQ	PEQ	2	2	3	3	2,5	Monto5
30	AA	GRAN	PEQ	PEQ	2	3	3	3	2,75	Monto6
31	AA	PEQ	PEQ	MOD	2	1	3	2	2	Monto3
32	AA	MED	PEQ	MOD	2	2	3	2	2,25	Monto4
33	AA	GRAN	PEQ	MOD	2	3	3	2	2,5	Monto5
34	AA	PEQ	PEQ	GRAN	2	1	3	1	1,75	Monto2
35	AA	MED	PEQ	GRAN	2	2	3	1	2	Monto3
36	AA	GRAN	PEQ	GRAN	2	3	3	1	2,25	Monto4
37	AA	PEQ	MOD	PEQ	2	1	2	3	2	Monto3
38	AA	MED	MOD	PEQ	2	2	2	3	2,25	Monto4
39	AA	GRAN	MOD	PEQ	2	3	2	3	2,5	Monto5
40	AA	PEQ	MOD	MOD	2	1	2	2	1,75	Monto2
41	AA	MED	MOD	MOD	2	2	2	2	2	Monto3
42	AA	GRAN	MOD	MOD	2	3	2	2	2,25	Monto4
43	AA	PEQ	MOD	GRAN	2	1	2	1	1,5	Monto1
44	AA	MED	MOD	GRAN	2	2	2	1	1,75	Monto2
45	AA	GRAN	MOD	GRAN	2	3	2	1	2	Monto3
46	AA	PEQ	GRAN	PEQ	2	1	1	3	1,75	Monto2
47	AA	MED	GRAN	PEQ	2	2	1	3	2	Monto3
48	AA	GRAN	GRAN	PEQ	2	3	1	3	2,25	Monto4
49	AA	PEQ	GRAN	MOD	2	1	1	2	1,5	Monto1
50	AA	MED	GRAN	MOD	2	2	1	2	1,75	Monto2
51	AA	GRAN	GRAN	MOD	2	3	1	2	2	Monto3
52	AA	PEQ	GRAN	GRAN	2	1	1	1	1,25	Monto1
53	AA	MED	GRAN	GRAN	2	2	1	1	1,5	Monto1
54	AA	GRAN	GRAN	GRAN	2	3	1	1	1,75	Monto2
55	AAA	PEQ	PEQ	PEQ	3	1	3	3	2,5	Monto5

56	AAA	MED	PEQ	PEQ	3	2	3	3	2,75	Monto6
57	AAA	GRAN	PEQ	PEQ	3	3	3	3	3	Monto6
58	AAA	PEQ	PEQ	MOD	3	1	3	2	2,25	Monto4
59	AAA	MED	PEQ	MOD	3	2	3	2	2,5	Monto5
60	AAA	GRAN	PEQ	MOD	3	3	3	2	2,75	Monto6
61	AAA	PEQ	PEQ	GRAN	3	1	3	1	2	Monto3
62	AAA	MED	PEQ	GRAN	3	2	3	1	2,25	Monto4
63	AAA	GRAN	PEQ	GRAN	3	3	3	1	2,5	Monto5
64	AAA	PEQ	MOD	PEQ	3	1	2	3	2,25	Monto4
65	AAA	MED	MOD	PEQ	3	2	2	3	2,5	Monto5
66	AAA	GRAN	MOD	PEQ	3	3	2	3	2,75	Monto6
67	AAA	PEQ	MOD	MOD	3	1	2	2	2	Monto3
68	AAA	MED	MOD	MOD	3	2	2	2	2,25	Monto4
69	AAA	GRAN	MOD	MOD	3	3	2	2	2,5	Monto5
70	AAA	PEQ	MOD	GRAN	3	1	2	1	1,75	Monto2
71	AAA	MED	MOD	GRAN	3	2	2	1	2	Monto3
72	AAA	GRAN	MOD	GRAN	3	3	2	1	2,25	Monto4
73	AAA	PEQ	GRAN	PEQ	3	1	1	3	2	Monto3
74	AAA	MED	GRAN	PEQ	3	2	1	3	2,25	Monto4
75	AAA	GRAN	GRAN	PEQ	3	3	1	3	2,5	Monto5
76	AAA	PEQ	GRAN	MOD	3	1	1	2	1,75	Monto2
77	AAA	MED	GRAN	MOD	3	2	1	2	2	Monto3
78	AAA	GRAN	GRAN	MOD	3	3	1	2	2,25	Monto4
79	AAA	PEQ	GRAN	GRAN	3	1	1	1	1,5	Monto1
80	AAA	MED	GRAN	GRAN	3	2	1	1	1,75	Monto2
81	AAA	GRAN	GRAN	GRAN	3	3	1	1	2	Monto3

ANEXO III

REGLAS TIPO MANDANI PARA EL MODELO II.

Nro	VALORES LINGÜÍSTICOS				VALORES NUMÉRICOS				Promedio	CupoCredito
	PerfilRiesgo	SaldoCorte	MaxSaldo12UM	IndEndeu6UM	PerfilRiesgo	SaldoCorte	MaxSaldo12UM	IndEndeu6UM		
1	A	PEQ	PEQ	PEQ	1	3	1	3	2	Monto3
2	A	PEQ	PEQ	MED	1	3	1	2	1,75	Monto2
3	A	PEQ	PEQ	GRAN	1	3	1	1	1,5	Monto1
4	A	PEQ	MOD	PEQ	1	3	2	3	2,25	Monto4
5	A	PEQ	MOD	MED	1	3	2	2	2	Monto3
6	A	PEQ	MOD	GRAN	1	3	2	1	1,75	Monto2
7	A	PEQ	GRAN	PEQ	1	3	3	3	2,5	Monto5
8	A	PEQ	GRAN	MED	1	3	3	2	2,25	Monto4
9	A	PEQ	GRAN	GRAN	1	3	3	1	2	Monto3
10	A	MOD	PEQ	PEQ	1	2,5	1	3	1,875	Monto2
11	A	MOD	PEQ	MED	1	2,5	1	2	1,625	Monto1
12	A	MOD	PEQ	GRAN	1	2,5	1	1	1,375	Monto1
13	A	MOD	MOD	PEQ	1	2,5	2	3	2,125	Monto3
14	A	MOD	MOD	MED	1	2,5	2	2	1,875	Monto2
15	A	MOD	MOD	GRAN	1	2,5	2	1	1,625	Monto1
16	A	MOD	GRAN	PEQ	1	2,5	3	3	2,375	Monto5
17	A	MOD	GRAN	MED	1	2,5	3	2	2,125	Monto3
18	A	MOD	GRAN	GRAN	1	2,5	3	1	1,875	Monto2
19	A	GRAN	PEQ	PEQ	1	1,5	1	3	1,625	Monto1
20	A	GRAN	PEQ	MED	1	1,5	1	2	1,375	Monto1
21	A	GRAN	PEQ	GRAN	1	1,5	1	1	1,125	Monto1
22	A	GRAN	MOD	PEQ	1	1,5	2	3	1,875	Monto2
23	A	GRAN	MOD	MED	1	1,5	2	2	1,625	Monto1

24	A	GRAN	MOD	GRAN	1	1,5	2	1	1,375	Monto1
25	A	GRAN	GRAN	PEQ	1	1,5	3	3	2,125	Monto3
26	A	GRAN	GRAN	MED	1	1,5	3	2	1,875	Monto2
27	A	GRAN	GRAN	GRAN	1	1,5	3	1	1,625	Monto1
28	AA	PEQ	PEQ	PEQ	2	3	1	3	2,25	Monto4
29	AA	PEQ	PEQ	MED	2	3	1	2	2	Monto3
30	AA	PEQ	PEQ	GRAN	2	3	1	1	1,75	Monto2
31	AA	PEQ	MOD	PEQ	2	3	2	3	2,5	Monto5
32	AA	PEQ	MOD	MED	2	3	2	2	2,25	Monto4
33	AA	PEQ	MOD	GRAN	2	3	2	1	2	Monto3
34	AA	PEQ	GRAN	PEQ	2	3	3	3	2,75	Monto6
35	AA	PEQ	GRAN	MED	2	3	3	2	2,5	Monto5
36	AA	PEQ	GRAN	GRAN	2	3	3	1	2,25	Monto4
37	AA	MOD	PEQ	PEQ	2	2,5	1	3	2,125	Monto3
38	AA	MOD	PEQ	MED	2	2,5	1	2	1,875	Monto2
39	AA	MOD	PEQ	GRAN	2	2,5	1	1	1,625	Monto1
40	AA	MOD	MOD	PEQ	2	2,5	2	3	2,375	Monto5
41	AA	MOD	MOD	MED	2	2,5	2	2	2,125	Monto3
42	AA	MOD	MOD	GRAN	2	2,5	2	1	1,875	Monto2
43	AA	MOD	GRAN	PEQ	2	2,5	3	3	2,625	Monto6
44	AA	MOD	GRAN	MED	2	2,5	3	2	2,375	Monto5
45	AA	MOD	GRAN	GRAN	2	2,5	3	1	2,125	Monto3
46	AA	GRAN	PEQ	PEQ	2	1,5	1	3	1,875	Monto2
47	AA	GRAN	PEQ	MED	2	1,5	1	2	1,625	Monto1
48	AA	GRAN	PEQ	GRAN	2	1,5	1	1	1,375	Monto1
49	AA	GRAN	MOD	PEQ	2	1,5	2	3	2,125	Monto3
50	AA	GRAN	MOD	MED	2	1,5	2	2	1,875	Monto2
51	AA	GRAN	MOD	GRAN	2	1,5	2	1	1,625	Monto1
52	AA	GRAN	GRAN	PEQ	2	1,5	3	3	2,375	Monto5
53	AA	GRAN	GRAN	MED	2	1,5	3	2	2,125	Monto3

54	AA	GRAN	GRAN	GRAN	2	1,5	3	1	1,875	Monto2
55	AAA	PEQ	PEQ	PEQ	3	3	1	3	2,5	Monto5
56	AAA	PEQ	PEQ	MED	3	3	1	2	2,25	Monto4
57	AAA	PEQ	PEQ	GRAN	3	3	1	1	2	Monto3
58	AAA	PEQ	MOD	PEQ	3	3	2	3	2,75	Monto6
59	AAA	PEQ	MOD	MED	3	3	2	2	2,5	Monto5
60	AAA	PEQ	MOD	GRAN	3	3	2	1	2,25	Monto4
61	AAA	PEQ	GRAN	PEQ	3	3	3	3	3	Monto6
62	AAA	PEQ	GRAN	MED	3	3	3	2	2,75	Monto6
63	AAA	PEQ	GRAN	GRAN	3	3	3	1	2,5	Monto5
64	AAA	MOD	PEQ	PEQ	3	2,5	1	3	2,375	Monto5
65	AAA	MOD	PEQ	MED	3	2,5	1	2	2,125	Monto3
66	AAA	MOD	PEQ	GRAN	3	2,5	1	1	1,875	Monto2
67	AAA	MOD	MOD	PEQ	3	2,5	2	3	2,625	Monto6
68	AAA	MOD	MOD	MED	3	2,5	2	2	2,375	Monto5
69	AAA	MOD	MOD	GRAN	3	2,5	2	1	2,125	Monto3
70	AAA	MOD	GRAN	PEQ	3	2,5	3	3	2,875	Monto6
71	AAA	MOD	GRAN	MED	3	2,5	3	2	2,625	Monto6
72	AAA	MOD	GRAN	GRAN	3	2,5	3	1	2,375	Monto5
73	AAA	GRAN	PEQ	PEQ	3	1,5	1	3	2,125	Monto3
74	AAA	GRAN	PEQ	MED	3	1,5	1	2	1,875	Monto2
75	AAA	GRAN	PEQ	GRAN	3	1,5	1	1	1,625	Monto1
76	AAA	GRAN	MOD	PEQ	3	1,5	2	3	2,375	Monto5
77	AAA	GRAN	MOD	MED	3	1,5	2	2	2,125	Monto3
78	AAA	GRAN	MOD	GRAN	3	1,5	2	1	1,875	Monto2
79	AAA	GRAN	GRAN	PEQ	3	1,5	3	3	2,625	Monto6
80	AAA	GRAN	GRAN	MED	3	1,5	3	2	2,375	Monto5
81	AAA	GRAN	GRAN	GRAN	3	1,5	3	1	2,125	Monto3