



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

**DISEÑO DE LOS PROCESOS DE EVALUACIÓN DE CLIENTES Y
MANTENCIÓN DEL MODELO DE SCORING Y DE SEGMENTACIÓN,
EN BANDESARROLLO MICROEMPRESAS**

PROYECTO DE GRADO PARA OPTAR AL GRADO DE MAGÍSTER EN
INGENIERÍA DE NEGOCIOS CON TECNOLOGÍAS DE INFORMACIÓN

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

CRISTIAN ANDRÉS JULIO AMDAN

PROFESOR GUÍA
Sr. Óscar Barros Vera

MIEMBROS DE LA COMISIÓN
Sr. Richard Weber Haas
Sr. Ezequiel Muñoz Krsulovic
Sr. Andrés Bernal Espinosa

SANTIAGO DE CHILE
ENERO 2009

*A mis padres, Abelardo y Viviana, pilares
fundamentales de mi vida.*

AGRADECIMIENTOS

A mis padres, por apoyarme incondicionalmente y brindarme una vida llena de oportunidades, en base a mucho esfuerzo y amor. A mis hermanos, Abelardo y Vivian, por su preocupación y compañía en todos estos años. A mis abuelas, Alicia y Elena, por acompañarme a la distancia. A mi abuelo José y mi tío Pablo, por esos maravillosos momentos que vivimos cuando estaban junto a mí.

A María José, por darme la oportunidad de crecer juntos, y por su comprensión y apoyo en todos mis proyectos.

A mis grandes amigos de la Universidad, Germán, Iván y Felipe M., por su ayuda y amistad sincera dentro y fuera de las aulas. A mis amigos de Iquique, Patricio, Ricardo, Felipe E. y Sebastián, por escucharme y compartir conmigo sus vidas, consejos y pensamientos.

A todos mis compañeros y profesores del DII de la Universidad de Chile, por contribuir en mi desarrollo personal y profesional. En especial, al profesor Óscar Barros, por confiar en mis capacidades y darme la oportunidad de crecer profesionalmente, y a Jonathan Scott, por su tiempo y ayuda desinteresada. A la Sra. Ana María Valenzuela, por su excelente disposición y simpatía.

A la Gerencia de Bandedesarrollo Microempresas, por recibirme durante un año y medio para realizar mi proyecto de grado. A todo el personal de la empresa, que me acogió de forma muy cálida desde el primer día. En particular, a Juan Carlos Dintrans, por brindarme su amistad y confiar en mis capacidades.

A todos los que ingratamente olvido nombrar, pero que contribuyeron de alguna u otra forma al logro de esta gran meta.

TABLA DE CONTENIDO

1. INTRODUCCIÓN	13
2. ANTECEDENTES DE LA INDUSTRIA Y LA EMPRESA	15
2.1. La Industria de las Microfinanzas en Chile.....	15
2.2 La Empresa	18
2.2.1 Descripción General.....	18
2.2.2 Estrategia de Bandesarrollo Microempresas.....	21
2.2.3. Modelo de Negocio de Bandesarrollo Microempresas.....	22
3. ANÁLISIS DEL MERCADO	30
3.1 Clientes	30
3.2 Tamaño y Tendencias del Mercado	32
3.2.1 Tamaño y Evolución del Mercado Potencial.....	32
3.2.2 Formalidad Tributaria de las Microempresas.....	34
3.2.3 Ventas y Operaciones Bancarias de las Microempresas	36
3.2.4 Segmentación de Clientes por Rubro	37
3.3 Competencia y Participación de Mercado	38
4. MARCO TEÓRICO CONCEPTUAL.....	40
4.1 Regresión Logística.....	41
4.1.1 Selección de Variables.....	42
4.1.2 Ajuste del Modelo.....	45
4.1.3 Bondad de Ajuste.....	46
4.1.4 R^2 de Nagelkerke	48
4.1.5 Validación Cruzada y Punto de Corte.....	51
4.1.6 Stepwise Logistic Regression	55
4.2 Mantención del Modelo.....	57
4.2.1 Estabilidad Poblacional.....	57
4.2.2 Relevancia de las Variables en los Modelos	59
4.2.3 Índice de Concordancia (C-Index)	64
5. ANÁLISIS DEL MEDIO INTERNO Y EXTERNO	68
5.1 Análisis de 5 Fuerzas de Porter	68
5.2 Análisis FODA.....	70
6. EL PROYECTO	72
6.1 Planteamiento Estratégico	72
6.1.1 El Problema.....	73

6.1.2. Posicionamiento Estratégico.....	75
6.2 Descripción de la Oportunidad	75
6.3 Definición del Proyecto	76
6.3.1 Objetivos	77
6.3.2 Producto.....	78
6.3.3 Alcance.....	79
6.4 Propuesta de Valor.....	80
6.5 Factores Críticos de Éxito	81
6.6 Justificación Económica	83
6.6.1 Medición de Beneficios	83
6.6.2 Medición de Costos.....	83
6.6.3 Construcción del Flujo de Caja.....	85
6.6.4 Análisis de Sensibilidad	91
7. DISEÑO DEL PROCESO DE EVALUACIÓN Y ANÁLISIS DE CLIENTES.....	92
7.1 Modelamiento BPMN del Rediseño.....	92
7.1.1 Macroprocesos involucrados	92
7.1.2 Administración Relación con el Cliente.....	94
7.1.3 Venta y Atención al Cliente.....	95
7.1.4 Marketing y Análisis de Clientes.....	97
7.1.5 Análisis Comportamiento de Ventas y Clientes.....	98
7.1.6 Diagramas de Pista.....	100
7.2 Lógicas de Negocio.....	140
7.2.1 Evaluación del Cliente.....	141
7.2.2 Mantenimiento Modelo de Comportamiento de Clientes.....	147
7.3 Tecnología Habilitante	166
7.3.1 Introducción.....	166
7.3.2 Tecnología ASP.....	167
7.4 Arquitectura de Apoyos Computacionales	169
7.5 Casos de Uso.....	171
7.5.1 Ingresar Datos del Cliente.....	172
7.5.2 Evaluación del Cliente.....	173
7.5.3 Decidir Crédito	174
7.5.4 Monitorear Desempeño de Ventas	176
7.6 Diagramas de Secuencia	177
7.6.1 Caso de Uso: Ingresar Datos del Cliente.....	178
7.6.2 Caso de Uso: Llenar Evaluación del Cliente	179
7.6.3 Caso de Uso: Modificar Evaluación del Cliente.....	180

7.6.4 Caso de Uso: Monitorear Desempeño de Clasificación	181
7.6.5 Caso de Uso: Modificar Parámetros de Lógica de Clasificación.....	182
7.7 Diagramas de Secuencia de Sistema	183
7.7.1 Llenar Evaluación del Cliente	183
7.7.2 Modificar Evaluación del Cliente.....	184
7.7.3 Monitorear Desempeño de Clasificación	185
7.7.4 Modificar Parámetros de Lógica de Clasificación.....	185
7.8 Diagrama de Secuencia Físicos.....	186
7.8.1 Llenar Evaluación del Cliente	186
7.8.2 Modificar Evaluación del Cliente.....	187
7.8.3 Monitorear Desempeño de Clasificación	188
7.8.4 Modificar Parámetros de Lógica de Clasificación.....	188
7.9 Diagrama de Clases.....	189
7.10 Modelo de Datos Físico.....	191
8. IMPLEMENTACIÓN ORGANIZACIONAL.....	193
8.1 Contexto Organizacional	193
8.2 Desafíos para la Gestión del Cambio.....	195
8.3 Estrategia para Gestión de Cambio.....	196
8.3.1 Sentido de Urgencia.....	196
8.3.2 Primeros Pasos.....	198
8.3.3 Gestión del Poder	199
8.3.4 Definición de Coalición Conductora.....	202
8.3.5 Gestión de Narrativas	203
8.3.6 Estrategia Comunicacional	205
8.3.7 Evaluación y Cierre del Proceso de Cambio	212
9. PRUEBA DE CONCEPTO	214
9.1 Contexto.....	214
9.2 Recolección y Transformación de los Datos.....	216
9.3 Mantención del Modelo de Comportamiento de Clientes	218
9.3.1 Análisis del Modelo Actual.....	219
9.3.2 Seleccionar Criterios de Re-calibración.....	228
9.3.3 Comparar Desempeño de los Modelos	229
9.3.4 Solución Alternativa	238
10. GENERALIZACIÓN DE LA EXPERIENCIA.....	251
10.1 Elección y Definición del <i>Framework</i>	252
10.2 Construcción del <i>Framework</i>	254

11. CONCLUSIONES	262
11.1 Del Proceso de Evaluación de Clientes.....	262
11.2 Del Proceso de Mantenimiento del Modelo de <i>Scoring</i>	263
11.3 De la Ingeniería de Negocios	266
12. BIBLIOGRAFÍA	268
13. ANEXOS: SALIDAS DE SPSS	271
13.1 Análisis del Modelo Actual	271
13.1.1 <i>Análisis Univariado</i>	271
13.1.2 <i>Curva ROC y C- Index</i>	277
13.2 Re-calibración con Selección Total de Variables	278
13.2.1 <i>Análisis Univariado</i>	278
13.2.2 <i>Desempeño de Puntos de Corte del Modelo Nuevo</i>	283
13.2.3 <i>Estabilidad del Modelo Nuevo</i>	290
13.3 Solución Alternativa	300
13.3.1 <i>Análisis Univariado</i>	300
13.3.2 <i>Análisis Multivariado</i>	306
13.3.3 <i>Desempeño de Puntos de Corte del Modelo Alternativo</i>	307
13.3.4 <i>Estabilidad del Modelo Alternativo</i>	314

ÍNDICE DE ILUSTRACIONES Y CUADROS

<i>Ilustración 1: Estructura del Banco del Desarrollo.....</i>	<i>19</i>
<i>Ilustración 2: Gráfico de Evolución del N° de Microempresarios Bancarizados.....</i>	<i>33</i>
<i>Ilustración 3: Gráfico de Formalización Tributaria de Microempresas Bancarizadas.....</i>	<i>34</i>
<i>Ilustración 4: Gráfico de Empresas Formales con Deuda Comercial Bancaria.....</i>	<i>35</i>
<i>Ilustración 5: Gráfico de distribución de Microempresas por Ventas Anuales.....</i>	<i>36</i>
<i>Ilustración 6: Gráfico de Tamaño de las Operaciones de Microempresas.....</i>	<i>37</i>
<i>Ilustración 7: Gráfico de Segmentación por Rubros.....</i>	<i>38</i>
<i>Ilustración 8: Matriz de Confusión Genérica.....</i>	<i>54</i>
<i>Ilustración 9: Matriz de Frecuencias Observadas.....</i>	<i>63</i>
<i>Ilustración 10: Matriz de Frecuencias Esperadas.....</i>	<i>63</i>
<i>Ilustración 11: Escala para interpretar valores del índice de concordancia.....</i>	<i>67</i>
<i>Ilustración 12: Gráfico de sensibilidad del VAN al aumento de MOB/Cliente.....</i>	<i>91</i>
<i>Ilustración 13: Macroprocesos Involucrados.....</i>	<i>92</i>
<i>Ilustración 14: Proceso Administración Relación con el Cliente.....</i>	<i>94</i>
<i>Ilustración 15: Proceso Venta y Atención al Cliente.....</i>	<i>95</i>
<i>Ilustración 16: Proceso Marketing y Análisis de Clientes.....</i>	<i>97</i>
<i>Ilustración 17: Proceso Análisis Comportamiento de Ventas y Clientes.....</i>	<i>98</i>
<i>Ilustración 18: Proceso Evaluación del Cliente.....</i>	<i>101</i>
<i>Ilustración 19: Proceso Decidir Crédito.....</i>	<i>103</i>
<i>Ilustración 20: Proceso Desempeño de Ventas.....</i>	<i>104</i>
<i>Ilustración 21: Proceso Mantener Modelo Comportamiento de Clientes.....</i>	<i>106</i>
<i>Ilustración 22: Árbol Comercial.....</i>	<i>142</i>
<i>Ilustración 23: Árbol de Pesca Artesanal.....</i>	<i>143</i>
<i>Ilustración 24: Árbol Silvoagropecuario.....</i>	<i>144</i>
<i>Ilustración 25: Arquitectura de Apoyo Computacional para Evaluación del Cliente.....</i>	<i>170</i>
<i>Ilustración 26: Caso de Uso de Ingresar Datos Cliente.....</i>	<i>172</i>
<i>Ilustración 27: Casos de Uso de Evaluación del Cliente.....</i>	<i>173</i>
<i>Ilustración 28: Caso de Uso de Decidir Crédito.....</i>	<i>175</i>
<i>Ilustración 29: Caso de Uso de Monitorear Desempeño de Ventas.....</i>	<i>176</i>
<i>Ilustración 30: Diagrama de Secuencia de Ingresar Datos del Cliente.....</i>	<i>178</i>
<i>Ilustración 31: Diagrama de Secuencia de Llenar Evaluación del Cliente.....</i>	<i>179</i>
<i>Ilustración 32: Diagrama de Secuencia de Modificar Evaluación del Cliente.....</i>	<i>180</i>
<i>Ilustración 33: Diagrama de Secuencia de Monitorear Desempeño de Clasificación.....</i>	<i>181</i>
<i>Ilustración 34: Diagrama de Secuencia de Modificar Parámetros de Lógica de Clasificación.....</i>	<i>182</i>
<i>Ilustración 35: DSS de Llenar Evaluación del Cliente.....</i>	<i>183</i>
<i>Ilustración 36: DSS de Modificar Evaluación del Cliente.....</i>	<i>184</i>
<i>Ilustración 37: DSS de Monitorear Desempeño de Clasificación.....</i>	<i>185</i>
<i>Ilustración 38: DSS de Modificar Parámetros Lógica de Clasificación.....</i>	<i>185</i>
<i>Ilustración 39: DSF de Llenar Evaluación del Cliente.....</i>	<i>186</i>

<i>Ilustración 40: DSF de Modificar Evaluación del Cliente.....</i>	<i>187</i>
<i>Ilustración 41: DSF de Monitorear Desempeño de Clasificación.....</i>	<i>188</i>
<i>Ilustración 42: DSF de Modificar Parámetros Lógica de Clasificación.....</i>	<i>188</i>
<i>Ilustración 43: Diagrama de Clases.....</i>	<i>189</i>
<i>Ilustración 44: Modelo de Datos Físico.....</i>	<i>192</i>
<i>Ilustración 45: Ficha de Evaluación Agrícola (Word).....</i>	<i>197</i>
<i>Ilustración 46: Mapa de Poder.....</i>	<i>201</i>
<i>Ilustración 47: Prototipo de Evaluación Silvoagropecuaria de PPA (1° Parte).....</i>	<i>208</i>
<i>Ilustración 48: Prototipo de Evaluación Silvoagropecuaria de PPA (2° Parte).....</i>	<i>209</i>
<i>Ilustración 49: Prototipo de Evaluación Silvoagropecuaria de PPA (Detalle de Costos y Deudas).....</i>	<i>210</i>
<i>Ilustración 50: Prototipo de Evaluación Silvoagropecuaria de PPA (3° Parte).....</i>	<i>211</i>
<i>Ilustración 51: Gráfico de Distribución de Predicción de Clientes (Modelo Actual)....</i>	<i>223</i>
<i>Ilustración 52: Gráfico de Utilidad Relativa al Punto de Corte Actual (Modelo Actual).....</i>	<i>227</i>
<i>Ilustración 53: Gráfico de Distribución de Predicción de Clientes (Modelo Nuevo)...</i>	<i>234</i>
<i>Ilustración 54: Gráfico de Distribución de Predicción de Clientes (Modelo Stepwise).</i>	<i>248</i>
<i>Ilustración 55: Diagrama de Clase de Control del Framework.....</i>	<i>255</i>
<i>Ilustración 56: Diagrama de Clase Entity del Framework.....</i>	<i>260</i>
<i>Cuadro 1: Líneas de Productos ofrecidas a Microempresas.....</i>	<i>16</i>
<i>Cuadro 2: Número de Intermediarios con orientación Microfinanciera.....</i>	<i>16</i>
<i>Cuadro 3: Tamaño de empresa, basada en criterio de Ventas.....</i>	<i>30</i>
<i>Cuadro 4: Variables cualitativas que caracterizan las microempresas.....</i>	<i>31</i>
<i>Cuadro 5: Plan de Ventas, Trienio 2008-2010.....</i>	<i>85</i>
<i>Cuadro 6: Proporción de Variables del Margen Financiero, Trienio 2005-2007.....</i>	<i>85</i>
<i>Cuadro 7: Proyección del MOB Promedio Anual por Cliente, Trienio 2008-2010.....</i>	<i>86</i>
<i>Cuadro 8: Proyección del MOB Promedio Anual por Cliente, Escenario Conservador.....</i>	<i>87</i>
<i>Cuadro 9: Proyección del MOB Promedio Anual por Cliente, Escenario Pesimista...</i>	<i>87</i>
<i>Cuadro 10: Flujo de Caja con Aumento del MOB por Cliente de 3%.....</i>	<i>89</i>
<i>Cuadro 11: Flujo de Caja con Aumento del MOB por Cliente de 5%.....</i>	<i>89</i>
<i>Cuadro 12: Flujo de Caja con Aumento del MOB por Cliente de 10%.....</i>	<i>90</i>
<i>Cuadro 13: Miembros de Coalición Conductora.....</i>	<i>203</i>
<i>Cuadro 14: Promedio y Rango permitido de cada Variable (Modelo Actual).....</i>	<i>220</i>
<i>Cuadro 15: Análisis Univariado de Variables del Modelo Actual (con predicción binarizada).....</i>	<i>222</i>
<i>Cuadro 16: Matriz de Confusión e Indicadores (Modelo Actual).....</i>	<i>224</i>
<i>Cuadro 17: Desempeño de distintos Puntos de Corte (Modelo Actual).....</i>	<i>226</i>
<i>Cuadro 18: Análisis Univariado de Variables del Modelo Actual (con resultados reales).....</i>	<i>229</i>
<i>Cuadro 19: Porcentaje de casos para entrenar y probar el Modelo Nuevo.....</i>	<i>230</i>
<i>Cuadro 20: Coeficientes Beta del Modelo Nuevo.....</i>	<i>231</i>
<i>Cuadro 21: Ajuste del Modelo Nuevo.....</i>	<i>232</i>

<i>Cuadro 22: R^2 de Nagelkerke del Modelo Nuevo.....</i>	<i>232</i>
<i>Cuadro 23: Desempeño de distintos Puntos de Corte (Modelo Nuevo).....</i>	<i>233</i>
<i>Cuadro 24: Matrices de Confusión de Modelos Actual y Nuevo.....</i>	<i>234</i>
<i>Cuadro 25: Índices de Concordancia de Modelos Actual y Nuevo.....</i>	<i>234</i>
<i>Cuadro 26: Predicción del Modelo Nuevo en Set de Entrenamiento y Prueba.....</i>	<i>235</i>
<i>Cuadro 27: Efectividad de Predicción del Modelo Nuevo, en 10 Particiones distintas</i> <i>.....</i>	<i>236</i>
<i>Cuadro 28: Cálculo de Coeficientes del Modelo Nuevo, en 10 Particiones distintas..</i>	<i>237</i>
<i>Cuadro 29: Análisis Univariado de Variables Candidatas (con resultados real de pago)</i> <i>.....</i>	<i>240</i>
<i>Cuadro 30: Ajuste del Modelo Stepwise.....</i>	<i>241</i>
<i>Cuadro 31: R^2 de Nagelkerke del Modelo Stepwise.....</i>	<i>241</i>
<i>Cuadro 32: Variables ingresadas al Modelo por el Método Stepwise.....</i>	<i>242</i>
<i>Cuadro 33: Matriz de Correlación de Variables seleccionadas por Método Stepwise</i> <i>.....</i>	<i>244</i>
<i>Cuadro 34: Porcentaje de casos para entrenar y probar el Modelo Stepwise.....</i>	<i>245</i>
<i>Cuadro 35: Coeficientes Beta del Modelo Stepwise.....</i>	<i>245</i>
<i>Cuadro 36: Desempeño de distintos Puntos de Corte (Modelo Stepwise).....</i>	<i>246</i>
<i>Cuadro 37: Matriz de Confusión de Modelo Stepwise.....</i>	<i>247</i>
<i>Cuadro 38: Predicción del Modelo Stepwise en Set de Entrenamiento y Prueba.....</i>	<i>249</i>
<i>Cuadro 39: Efectividad de Predicción del Modelo Stepwise, en 10 Particiones distintas</i> <i>.....</i>	<i>249</i>
<i>Cuadro 40: Cálculo de Coeficientes del Modelo Stepwise, en 10 Particiones distintas</i> <i>.....</i>	<i>250</i>

RESUMEN

En industrias altamente competitivas, como la industria bancaria, la estrategia de relación estrecha con el cliente, que incluye la oferta de productos o servicios a la medida y la prestación de una atención de excelencia, ha demostrado ser la única estrategia que genera a las empresas rentabilidades sobre el promedio y de largo plazo. En particular, en un mercado que sólo se encuentra bancarizado en un 40%, como las microempresas, la importancia de captar clientes, conocerlos, y generar vínculos fuertes antes que la competencia, es vital para todo actor involucrado.

Este proyecto propone un rediseño de procesos para la atención, evaluación y control segmentado de clientes en Banderrollo Microempresas, que permita ofrecerles el producto que mejor resuelva sus necesidades de crédito e identificar los segmentos de mayor/menor rentabilidad y riesgo. Para lograr lo anterior, el proyecto integra a las evaluaciones de clientes un modelo de *credit scoring*, para el cual propone un proceso de actualización en el tiempo, con el fin de incorporar cambios en los perfiles de riesgo de los clientes.

El rediseño de procesos se basa en la metodología de diseño a partir de patrones de procesos, y detalla desde la arquitectura de procesos hasta el diseño del apoyo computacional y lógicas de negocio requeridos para

implementarlo. Además, se realizó una prueba de concepto para el proceso de mantención del *scoring*, donde se corroboró la desactualización del modelo y se incrementó su acierto en clientes no pagadores en un 7%, a igual acierto en clientes pagadores. Finalmente, la experiencia de esta prueba de concepto es generalizada en la construcción de un *Framework*, desarrollado para cualquier empresa que utilice modelos predictivos y desee actualizarlos en el tiempo.

En la evaluación económica del proyecto, se determinó que basta un aumento del 2% del Margen Operacional Bruto (MOB) por cliente, para que el proyecto se justifique económicamente. El VAN aumentará en la medida que el incremento en el MOB sea mayor y/o se incorporen más clientes a la cartera. La evaluación del proyecto excluye el uso y mantención del modelo de *scoring*, por no contar con suficiente información al momento de desarrollar este trabajo.

1. INTRODUCCIÓN

En el presente trabajo, se propone un rediseño de procesos para la evaluación de clientes en Banderarrollo Microempresas, que incluye el proceso de mantención de un modelo de *credit scoring*, para el cual se realizó una prueba de concepto.

Primero, como forma de contextualizar el proyecto, se describirán la industria microfinanciera en Chile, caracterizada por la alta concentración de la oferta, y el mercado, cuyos clientes tienen características muy distintas a los que atiende al resto de la Banca. Para analizar cuán robusta es la posición de la empresa en el mercado, se realizará un análisis FODA y de las 5 fuerzas de Porter.

A continuación, se expondrán los lineamientos estratégicos que motivan el proyecto, señalando por qué es una oportunidad la mejor selección, evaluación y control de los microempresarios. La propuesta de valor se concreta en un producto, cuyo objetivo principal es la estandarización del criterio de evaluación de clientes y la actualización continua de los modelos de segmentación y *scoring*. También se incluye la evaluación económica de su implementación.

Utilizando la metodología propuesta en la Ingeniería de Negocios, luego se realiza el modelamiento BPMN del rediseño de procesos, detallando los apoyos computacionales necesarios para una adecuada implementación del producto, mediante la notación UML, y las lógicas de negocio asociadas a cada proceso. El último nivel de detalle es el diagrama de clases de la aplicación y el modelo físico de datos.

Para la implementación organizacional se diseña un plan de gestión del cambio, con el objetivo de lograr que el proyecto forme parte del funcionamiento natural de la empresa, integrando a todas las personas que se participen en él.

Finalmente, se realiza una prueba de concepto del proceso de mantención del *scoring*, con el fin de mostrar empíricamente la validez de la propuesta que se realiza a la empresa. La experiencia de la realización de la prueba deriva en la construcción de un *Framework*, que permite extender el problema de actualización de un modelo de riesgo crediticio, a cualquier modelo predictivo basado en regresión logística o árboles de decisión.

2. ANTECEDENTES DE LA INDUSTRIA Y LA EMPRESA

2.1. La Industria de las Microfinanzas en Chile

Durante la época de los ochenta, la actividad microfinanciera se desarrollaba a muy baja escala y era canalizada principalmente por organismos no gubernamentales con apoyo de la cooperación internacional.

A partir de la década de los noventa, el escenario cambió profundamente. Entre los aspectos más sobresalientes, se destaca la activa participación del Estado en la implementación de programas de fomento, la incorporación de intermediarios financieros a la oferta de microcréditos, el aumento de la escalas de operación y, por cierto, la diversificación de las líneas de productos.

Actualmente, se observa que las líneas de productos, inicialmente orientadas al crédito, se han ampliado, incluyendo productos de ahorro, administración de efectivo, medios de pago, seguros y otros servicios destinados a las necesidades familiares de los microempresarios.

Por otra parte, las redes de sucursales se han expandido tanto numérica como geográficamente.

En el cuadro 1, se observa las líneas de productos ofrecidas a los microempresarios, con sus productos asociados:

Cuadro 1: Líneas de Productos ofrecidas a Microempresarios

Líneas de Productos	Productos
Ahorro	Cuentas de Ahorro
	Fondos Mutuos
	Depósitos a Plazo
Crédito	Líneas
	Cuotas
Administración de Efectivo	Cuenta corriente
	Tarjeta de crédito
	Cuenta vista
Seguros	Cobertura familiar
	Coberturas de Crédito

Fuente: Morales y Yáñez (2007)

En el cuadro 2, se identifica las instituciones especializadas fiscalizadas por la SBIF, que presentan orientación al segmento de empresas de menor tamaño. Si bien existe otorgamiento de servicios financieros a microempresarios fuera de estas instituciones, en ellas tiene un carácter no especializado.

Cuadro 2: Número de Intermediarios con orientación Microfinanciera

	N° de Instituciones	N° de Sucursales
Divisiones Bancarias especializadas	3	306
CAC	4	97
Total	7	403

Fuente: Morales y Yáñez (2007)

Las divisiones bancarias especializadas son BancoEstado Microempresas, Bandedesarrollo Microempresas y Santander BANEFE. Las

Cooperativas de Ahorro y Crédito (CAC) son Oriencoop, Capual, Detacoop y Coocretal.

Las 3 divisiones bancarias especificadas anteriormente, atienden un 88,5% de los clientes bancarizados a nivel nacional, lo que revela una alta concentración de la oferta¹.

Sin embargo, una competencia potencial puede estar dada por actores del mercado enfocados a los créditos de consumo, que podrían direccionar sus productos a los microempresarios. Tal es el caso de Banco Falabella y Banco Ripley.

¹ Morales y Yáñez (2007).

2.2 La Empresa

2.2.1 Descripción General

2.2.1.1 El Banco del Desarrollo

El Banco del Desarrollo es una empresa privada orientada a apoyar a sectores de la población que tienen menos oportunidad para acceder al sistema financiero tradicional. Su énfasis se centra en las medianas, pequeñas y microempresas, y personas de menores ingresos.

Desde su fecha de fundación, en 1983, ha apoyado a más de 420.000 clientes. Una amplia red de 74 sucursales y 50 Centros especializados en la atención a Microempresas a lo largo de todo Chile, le ha permitido lograr una mayor proximidad con la gente, un mejor conocimiento de sus necesidades locales y un claro compromiso con la regionalización del país.

Su objetivo principal es apoyar la consolidación de múltiples iniciativas empresariales; mejorar la calidad de vida de los chilenos, fundamentalmente en materia de vivienda, salud y educación y contribuir a hacer de Chile un lugar donde todos puedan vivir mejor.

El Grupo Banco del Desarrollo está conformado por filiales y empresas relacionadas, que brindan servicios financieros especializados e integrales a sus clientes. Éstas entregan apoyo en aspectos tales como vivienda y financiamiento para microempresas, a través de profesionales de alto nivel y con la más amplia gama de instrumentos de inversión para la consolidación de proyectos de desarrollo. En la ilustración 1, se muestra un esquema de las filiales que componen el banco:

Ilustración 1: Estructura del Banco del Desarrollo



Fuente: www.intranet.bdd.cl

2.2.1.2 Bandesarrollo Microempresas

Bandesarrollo Microempresas es líder privado en el apoyo a microempresarios formales e informales del país, para el logro de sus metas productivas o de servicio. Microempresas prioriza sus esfuerzos en la atención

de trabajadores independientes, comerciantes, quiosqueros, feriantes, prestadores de servicios y trabajadores por cuenta propia.

Su misión es ser un instrumento de desarrollo de las microempresas y de los sectores emprendedores de menores ingresos de Chile, apoyándolos en su progreso y bienestar económico y social.

Sus valores son Excelencia, Innovación, Integridad y Trabajo en Equipo.

El objetivo de la empresa es atender y bancarizar masivamente a los microempresarios y a los sectores de menores ingresos a lo largo del país, de manera tal de ofrecerles apoyo en el desarrollo de sus negocios e iniciativas y mejorar su calidad de vida.

Un Poco de Historia

Luego un plan piloto de otorgamiento de créditos, en junio de 1992, el BDD inauguró la primera sucursal bancaria en el país dedicada exclusivamente al otorgamiento de microcréditos. En 1993 se expandió en la Región Metropolitana, y entre 1994 y 1995, se amplió a otras zonas del país, llegando a 47 centros de atención especializados desde Arica a Punta Arenas. A partir del año 1998, pasó de ser una banca centrada en el otorgamiento de créditos, a

una banca de multiproductos, y desde el 2004, ofrece productos y servicios microfinancieros.

Actualmente, además de atender microempresarios dedicados a actividades productivas, comercio y servicios, Bandesarrollo Microempresas atiende a pequeños productores agrícolas, pescadores artesanales, transportistas, personas dependientes y a mujeres jefes de hogar, a través del Programa de Desarrollo Solidario (PDS).

Este crecimiento ha sido posible gracias a las alianzas con FOSIS, SERCOTEC, INDAP, CORFO, SENCE, Municipalidades, Organismos de Capacitación, Asociaciones de Microempresarios y Sindicatos de Trabajadores Independientes.

2.2.2 Estrategia de Bandesarrollo Microempresas

La estrategia adoptada por Bandesarrollo Microempresas, para hacer frente al entorno competitivo, se define en tres partes:

1. Fortalecer el servicio al cliente, a través de una agilización del proceso crediticio y del desarrollo de estrategias comerciales innovadoras.
2. Diversificación de servicios, pasando de un enfoque de microempresas a microfinanzas, acompañando e identificando, de esta manera, las necesidades de los microempresarios. Un claro ejemplo son las tarjetas de crédito VISA y las cuentas corrientes con línea de crédito.
3. Incursión en otros nichos de mercado, de niveles más modestos, con el Programa de Desarrollo Solidario (PDS).

2.2.3. Modelo de Negocio de Bandedesarrollo Microempresas

2.2.3.1 Oportunidad de la Empresa

Actualmente, la microempresa se enfrenta a un mercado global y altamente competitivo. Sin embargo, aún existen nichos donde se pueden encontrar oportunidades de negocio. Dichas posibilidades dependen de su capacidad de producir y colocar en el mercado, a precios competitivos, bienes y

servicios en cantidades reducidas, con tecnología intermedia, con mano de obra no especializada y, por consiguiente, con escaso capital.

Aunque pareciera difícil competir en bienes y servicios que tienen economías de escala, las microempresas aún tienen a su favor la flexibilidad, dada por un menor costo de cambio respecto de las grandes empresas, y la relación personal con el cliente.

La escasez de recursos propios con que cuentan los microempresarios, hace muy difícil su desarrollo, sin la existencia de recursos suficientes de crédito que los apoyen. Adicionalmente, la existencia de rubros diversos, como el silvoagropecuario, requieren de productos financieros que se adecúen a su realidad productiva y que, por ende, tengan una estructura de pago acorde a su producción.

En este contexto nace el modelo de negocio de Banderrollo Microempresas, ofrecer a los microempresarios los productos más convenientes del mercado, que satisfagan sus necesidades y requerimientos específicos, basándose en una estrategia de diferenciación en calidad de servicio, a través de:

- Atención personalizada y eficiente al cliente.
- Asesoría integral y atención en terreno de ejecutivos especializados.
- Productos y servicios flexibles.
- Alianza con instituciones gubernamentales y no gubernamentales, para créditos a todos los segmentos atendidos.
- La mejor cuota del mercado.

2.2.3.2 Productos

Acorde su estrategia, Banderarrollo Microempresas ofrece una amplia gama de productos de ahorro y crédito, que se pueden agrupar en los siguientes ámbitos:

1. Créditos Tradicionales: específicamente diseñados para el microempresario y su entorno familiar. Existen diversos subproductos segmentados por sector (agrícola, pesca artesanal, transporte y comercial), aunque destacan los siguientes grupos generales: créditos de fomento productivo en sus distintas formas (capital de trabajo, inversión, cuenta corriente con línea de crédito), créditos de capacitación (de empleados de microempresas), y crédito familiar.

2. Crédito Hipotecario y Tarjeta de Crédito: ofrecidos a los microempresarios más formalizados.

3. Programa de Desarrollo Solidario (PDS): destinado a mujeres jefes de hogar con niveles de ingreso muy por debajo del microempresario tradicional, como estrategia de crecimiento *down scaling*. Este crédito no se concentra únicamente en actividades productivas, pues sigue una metodología solidaria.

4. Ahorro: Los microempresarios pueden acceder a cualquier producto de ahorro. Los más comunes son cuenta corriente, cuenta vista, chequera electrónica, depósitos a plazo y libretas de ahorro convencionales y para la vivienda.

2.2.3.3 Modelo de Atención a Clientes CAR

La estrategia de diferenciación en calidad de servicio, se ha conceptualizado en un Modelo de Atención a Clientes (CAR), que se basa en tres conceptos principales: *Calidez, Asesoría y Rapidez*. A continuación, se detallarán y explicarán los componentes del CAR.

a. Calidez

“[Se trata de crear un] Banco con vocación local y regional, caracterizado por la calidad y calidez en el servicio que está llamado a proporcionar. Un buen servicio comienza con una buena acogida”

La Calidez considera 2 componentes principales:

a1. Amabilidad: Saludar al Cliente. Tratar al Cliente de Usted, sin importar sexo ni edad. Usar un lenguaje adecuado, formal y no técnico.

a2. Empatía: Ponerse en el lugar del cliente. Dejar hablar al cliente sin interrumpirlo.

b) Asesoría

“El Banco debe otorgar a sus asociados asesorías, orientaciones y apoyo en las materias que le son propias. Es necesario descubrir formas y áreas de interés, en las cuales el Banco pueda ir creando una acción de fomento que lo haga distinguirse de la competencia (...) establecer algún tipo de servicio distinto del servicio financiero tradicional, que haga que el cliente sienta más necesario este banco, en comparación con otro”

La Asesoría se compone de:

b1. Orientación: Informar sobre los productos y servicios que podrían solucionarle las necesidades al cliente.

b2. Asertividad: Dar la respuesta inmediatamente, tanto “Sí” como “No”.

b3. Solución: Dirigir la asesoría a la mejor solución para el cliente.

c) Rapidez

“El principal atractivo que un Banco de Asociados puede exhibir es la calidad de servicio que ofrece. En este aspecto debe encontrarse la principal diferencia a los ojos del cliente y asociados. (...) el Banco debe ser ágil y eficiente. La rapidez en las respuestas, sean positivas o negativas, y en el procesamiento de los servicios, son de importancia crucial”

La rapidez, refiere a:

c1. Mínima Espera: Ser concreto en las explicaciones. Realizar transacciones y atenciones cortas en cajas y mesón de atención.

c2. Cumplimiento: Cumplir con los plazos entregados a los clientes.

2.2.2.4 Descripción de las Operaciones Actuales

Cuando la calidad del servicio es parte integral de la estrategia de diferenciación de la empresa, éste debe contar con operaciones eficientes, que se encuentren alineadas con sus fundamentos estratégicos.

El proceso actual de evaluación de clientes, funciona de la siguiente forma: Cuando un cliente llega a solicitar un producto financiero, el ejecutivo verifica que tenga los antecedentes necesarios para acceder al crédito, dependiendo del producto que está solicitando, por ejemplo, pago de los últimos 12 arriendos de su casa. Cuando se cuenta con toda la información, el ejecutivo analiza si el cliente cumple con los requisitos mínimos para postular a un crédito, como tener un nivel de endeudamiento razonable.

Si el cliente cumple con dichos requisitos, el ejecutivo decide qué evaluación utilizará para evaluar al cliente, basándose en la normativa, y su experiencia y criterio.

Dependiendo del perfil del cliente y tipo de crédito solicitado, la evaluación puede incluir una visita a terreno para corroborar si los datos proporcionados por el cliente concuerdan con la realidad de su negocio. Con

todo esto reunido, se envía la solicitud de Producto a la instancia de aprobación.

En caso que la respuesta a la solicitud sea positiva, se entrega el producto; en caso negativo, se avisa al cliente y se cambian las condiciones del crédito, como tasa o monto a otorgar.

3. ANÁLISIS DEL MERCADO

3.1 Clientes

Los clientes de Bandedesarrollo Microempresas, son los microempresarios. En el país, no hay consenso absoluto sobre la definición de microempresa, pero a nivel de los Organismos de Fomento, el criterio de mayor difusión para determinar el tamaño de una empresa, es el de las ventas anuales. En el cuadro 3, se muestra la definición del tamaño de las empresas, basada en sus ventas anuales:

Cuadro 3: Tamaño de empresa, basada en criterio de ventas anuales

Tamaño	Ventas Anuales: V [UF]
Grandes Empresas	$V > 100.000$
Mediana Empresa	$25.000 < V < 100.000$
Pequeña Empresa	$2.400 < V < 25.000$
Microempresa	$V < 2.400$

Fuente: Morales y Yáñez (2007)

En este contexto, las microempresas son empresas cuyas ventas anuales no exceden de UF2400.

Si bien las definiciones de carácter cuantitativo, proveen de criterios objetivos que facilitan la medición y la focalización de los recursos de fomento, son las aproximaciones cualitativas las que más ayudan a entender los factores

que determinan el desempeño de las microempresas. El cuadro 4, especifica y describe las variables cualitativas para definir las microempresas:

Cuadro 4: Variables cualitativas que caracterizan las microempresas

Variable	Descripción
Situación Socioeconómica	Bajo nivel socioeconómico de las personas vinculadas a la actividad, la cual tiende a ser la principal fuente de ingreso familiar.
Formalización	Bajo nivel de formalización de las actividades. Muchas de estas empresas carecen de iniciación de actividades, patentes municipales, permisos específicos y no tributan.
Tecnología	Tecnologías anticuadas o tradicionales, o bien difundidas masivamente.
Tamaño	Escaso personal, pocos activos y reducido nivel de operación.
Mercado	Productos y servicios orientados fundamentalmente a mercados locales o domésticos.
Propiedad y gestión	Estrecha vinculación entre la propiedad, la gestión del negocio, el trabajo y la familia. Los procesos de decisión son extremadamente centralizados, existe concentración en funciones organizacionales y una reducida segregación entre la función financiera del negocio y la "economía familiar". Lo anterior sumado a la informalidad de las actividades, redundan en la escasa generación y uso de información financiera del negocio.

Fuente: Morales y Yáñez (2007)

En este nuevo contexto, las microempresas son empresas (formales o informales) que desarrollan una actividad productiva, de servicios o comercial, caracterizada tanto por su reducida escala de operación como por los otros elementos expuestos en el cuadro 4 (Morales y Yáñez, 2007).

Perfil del Microempresario

- Género: Son más mujeres que hombres (53% - 47%).
- Edad: Son adultos. El 16% es menor de 40 años, el 61% tiene entre 41 y 60 años, y el 21% restante es mayor de 60 años.

- Educación: Sólo el 5% tiene educación universitaria completa o incompleta, el 14% técnica, el 49% Media y el 31% sólo tiene Educación Básica.
- Nivel de Experiencia: Medio. Sólo el 41% tiene más de 10 años en el rubro. Un 25% tiene entre 5 y 10 años, y el 32% restante menos de 5 años.
- Son dueños únicos: Un 97% de los microempresarios son únicos dueños. El 3% tiene socios².

3.2 Tamaño y Tendencias del Mercado

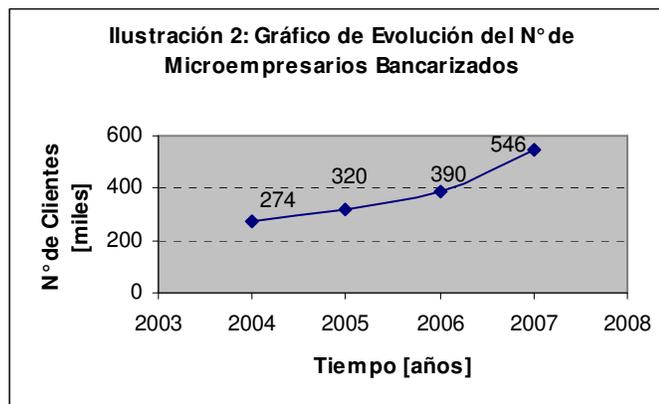
3.2.1 Tamaño y Evolución del Mercado Potencial

El mercado potencial de microempresas se estima en 1.339.000, de los cuales 579.000 (42%) son formales y 760.000 (58%) son informales (Morales y Yáñez, 2007). De todas ellas, sólo se encuentran bancarizadas 534.167, que corresponden al 40%.

En los últimos años, la cartera de microempresas asociada a las divisiones bancarias especializadas, se ha expandido considerablemente. En

² Extraído de: *Valoración de Atributos e Impacto de Microempresas* (2004).

efecto, si se compara la cartera de clientes activos a la fecha, con los existentes a principios de la década (115 mil deudores a diciembre de 2001), la tasa de crecimiento anual cifra supera ampliamente la tasa de expansión de la base de clientes comerciales (10% promedio en los últimos dos años) y de consumo de la banca (8% promedio en los últimos 5 años). En la ilustración 2, se muestra el gráfico de evolución de la cartera desde el año 2004:

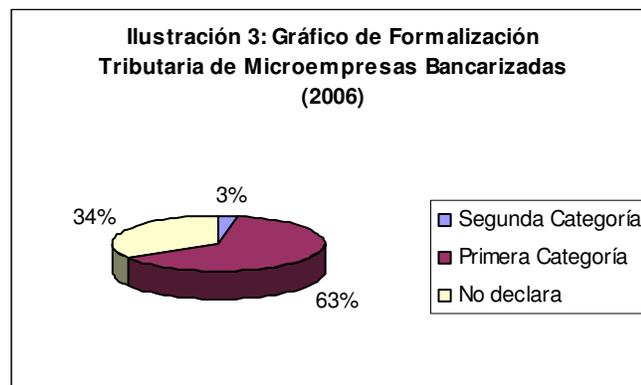


Fuente: Elaboración propia, en base a: *Gestión Comercial con Impacto Social* (2007), *BancoEstado Microempresas premió a los más emprendedores* (2007) e *Informe Estado de las Microfinanzas en Chile* (2006)

Como se aprecia en la ilustración 2, la bancarización de microempresarios anterior ha experimentado un crecimiento exponencial en los últimos años, lo que se condice con la estrategia de captación de clientes de las instituciones financieras más importantes.

3.2.2 Formalidad Tributaria de las Microempresas

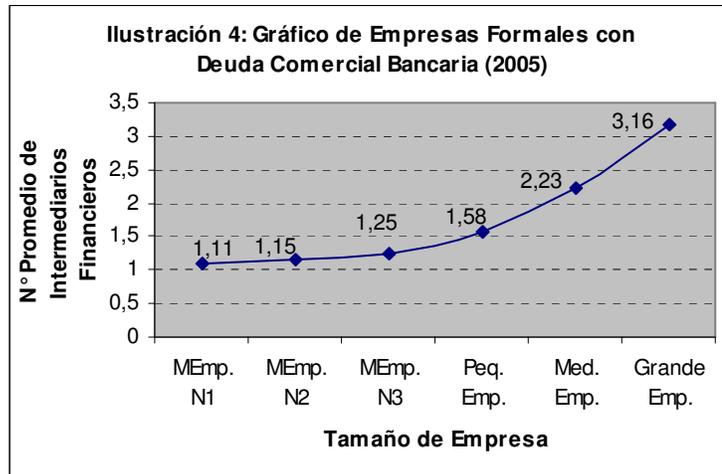
La participación de las microempresas informales dentro de la cartera de los intermediarios microfinancieros especializados, es materialmente relevante. Las cifras presentadas indican que de cada 100 microempresas con cobertura crediticia, 34, son informales. No obstante lo anterior, dicha cobertura es limitada, si se considera que el 56% de las microempresas del país son informales. En la ilustración 2, se muestra la distribución de las microempresas bancarizadas, de acuerdo a su formalidad tributaria:



Fuente: Morales y Yáñez (2007)

Otro aspecto a destacar, es que las microempresas formales con cobertura crediticia tienden a relacionarse con sólo un intermediario financiero. En efecto, cuando se considera la cartera comercial segregada por tamaño de empresa, se observa que el número de instituciones donde las empresas

mantienen deudas, crece con el tamaño de las empresas. Tal situación se aprecia en el gráfico de la ilustración 4:



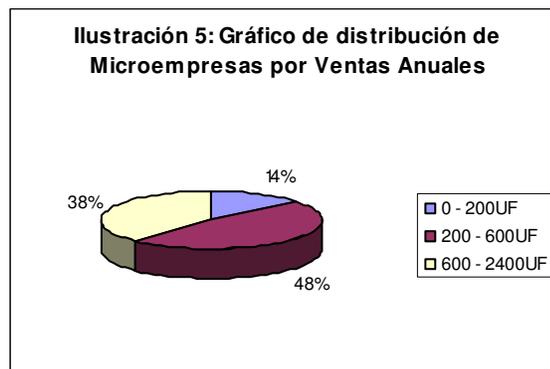
El segmento de microempresas ha sido dividido en tres estratos, según ventas anuales declaradas expresadas en UF: N1 (0-200); N2 (200-600); y N3 (600-2400).

Fuente: Morales y Yáñez (2007)

La ilustración 4, ratifica la importancia de la calidad de servicio, mencionada en el modelo de negocio. Si un cliente decide cambiarse de Banco, es muy probable que no vuelva. Por ello, todos los esfuerzos deben estar enfocados a retener y fidelizar a los clientes, con operaciones eficientes y productos que atiendan sus necesidades particulares.

3.2.3 Ventas y Operaciones Bancarias de las Microempresas

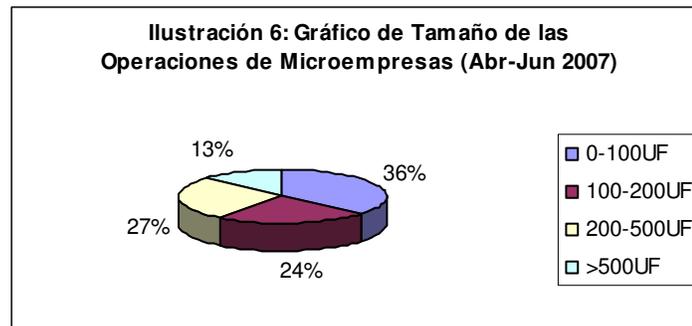
La distribución de las microempresas atendidas por la Banca, de acuerdo a sus ventas anuales, es la que se señala en la ilustración 5:



Fuente: Morales y Yáñez (2007)

Los datos anteriores, muestran que el nivel de ventas de las microempresas es superior a UF200 anuales en el 86% de los casos. Es más, es superior a UF600 en el 38% de los casos.

Por otro lado, el tamaño de las operaciones de las microempresas, es decir, del monto de los créditos, se distribuye según se muestra el gráfico de la ilustración 6:

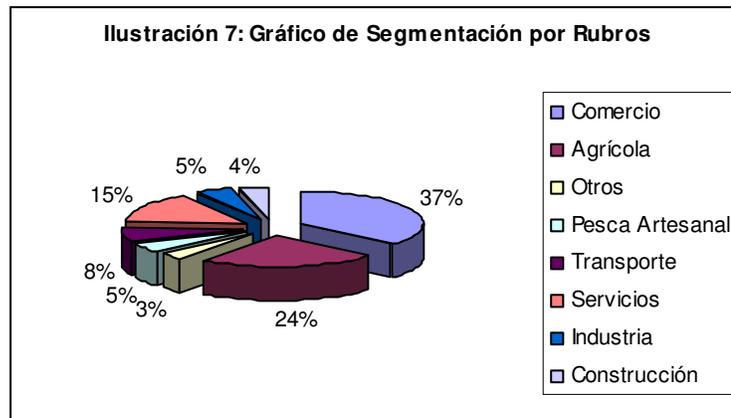


Fuente: Morales y Yáñez (2007)

Por lo tanto, hay una gran oportunidad de atender clientes más grandes y que solicitan créditos por montos mayores. No obstante lo anterior, el segmento con operaciones menores a UF100, constituye el 36% del total de las microempresas, por lo que no es despreciable.

3.2.4 Segmentación de Clientes por Rubro

En la ilustración 7, se muestra la segmentación de las microempresas por rubro:



Fuente: Elaboración propia, en base a: *La Microempresa en Chile* (2005) y *Registro Pesquero Artesanal (RPA)* (2007)

En el gráfico anterior, se observa que el 24% de los clientes pertenecen al sector agrícola y el 37% al rubro comercial. Mención especial merece el rubro pesquero, alcanza los 59.578 pescadores artesanales registrados. Estos rubros son agrupados por Bandedesarrollo Microempresas en 3 segmentos: Agrícola, Pesca Artesanal y Comercial, que incluye todos los rubros restantes. Algunos, como el Banco del Estado, separan de este último segmento el rubro de Transporte, y lo atienden como un segmento aparte.

3.3 Competencia y Participación de Mercado

Como ya se ha mencionado, sólo 3 instituciones financieras concentran el 88,5% del mercado bancarizado, distribuyéndose de la siguiente manera:

- a. BancoEstado: 284.000 Clientes (52%)³
- b. Banderarrollo Microempresas: 119.450 Clientes (21,9%)⁴
- c. Santander BANEFE: 80.000 Clientes (14,7%)⁵

Pero aún queda un 60% del mercado potencial por bancarizar. De ahí que estos últimos años, la competencia se ha centrado en la captación de clientes, que se ha traducido en un crecimiento exponencial de la participación de mercado de estas entidades. Sin duda, la más exitosa será la capaz de captar, retener y rentabilizar más de los mejores clientes.

³ *BancoEstado Microempresas premió a los más emprendedores (2007).*

⁴ *Resultados de Banderarrollo Microempresas (2007) [no publicado].*

⁵ *Memoria de Sostenibilidad (2007).*

4. MARCO TEÓRICO CONCEPTUAL

En un mercado financiero competitivo, se hace necesario contar con herramientas avanzadas, que permitan aumentar la rentabilidad y controlar el riesgo del negocio, mediante una selección adecuada clientes.

La aplicación Minería de Datos, permite a las instituciones financieras realizar un tratamiento a los datos de sus clientes, para transformarlos en información sobre su comportamiento crediticio, y así predecir su probabilidad de no pago, antes de otorgar el crédito. Pero la construcción del modelo también les permite conocer mejor a los clientes, en tanto identifica qué variables determinan (y en qué medida) el cumplimiento del compromiso de pago.

El objetivo de desarrollar modelos matemáticos para la gestión de clientes no es reemplazar la decisión de otorgamiento del crédito, sino más bien apoyarla, cuantificando el riesgo que se asume al incorporar ese cliente a la cartera. Al respecto, lo que se busca es acotar el riesgo de operar en el mercado, y por ningún motivo eliminarlo. Sin riesgo no hay negocio, pero un riesgo desmesurado puede acabar con él.

En este contexto, cabe preguntarse cuál es la vigencia de un modelo predictivo. Pese a que existe vasta literatura sobre cómo éstos deben ser contruidos, muy pocos autores se refieren la importancia de su mantención en el tiempo. El uso de modelos desactualizados puede tener fuertes repercusiones en la rentabilidad de una empresa, por cuanto identificará riesgos inexistentes y obviará riesgos relevantes en los clientes buenos y malos, respectivamente.

El Marco Teórico que se presenta a continuación, incluye los conceptos necesarios para la comprensión y aplicación del proceso de mantención del modelo de *scoring*, propuesto en este trabajo. Particularmente, se detalla el proceso para el método de Regresión Logística, utilizado actualmente en la empresa.

4.1 Regresión Logística

Los métodos regresivos se han convertido en un componente fundamental de cualquier análisis de datos, cuya finalidad sea describir la relación entre una variable de objetivo y una o más variables explicativas. Históricamente, la regresión logística se ha convertido, en diversas áreas, en el método estándar para predecir una variable discreta.

Antes de comenzar el estudio de la regresión logística, es importante entender que el objetivo de un análisis basado en este método, es el mismo que el de cualquier modelo construido con técnicas estadísticas: encontrar un modelo parsimonioso, con sentido y alto poder para predecir de una variable de salida, utilizando un set de variables independientes (Hosmer & Lemeshow, 1989).

4.1.1 Selección de Variables

A continuación, se esbozará la metodología para la selección de variables de la regresión logística. Si bien es cierto, esta metodología se aplica sólo cuando el modelo es desarrollado por primera vez, se muestra para una mejor comprensión de su origen.

Primero, se define un conjunto de variables a estudiar, esto es, variables que potencialmente podrían formar parte del modelo predictivo. La definición de dichas variables se hace en función de las características del problema, del conocimiento de los ejecutivos del negocio, y la disponibilidad de información en las bases de datos.

Una vez que las variables se encuentran definidas, se debe analizar si la calidad de los datos de cada una de ellas, les permitirán aportar información al modelo. Es decir, se debe realizar un análisis de calidad de datos, específicamente en dos dimensiones: los datos faltantes (NULL) y la concentración en un valor. En el análisis de los datos nulos se detectan las variables que, por tener gran cantidad de campos sin información, no son aptas para ser incorporadas al modelo. Por otro lado, en el análisis de concentración, se detectan aquellas variables que, por tener una alta concentración en un valor específico, no aportarán capacidad discriminante al modelo. Por ejemplo, si la nacionalidad está concentrada en un 99% en “chilena”, carece de sentido incorporarla al modelo predictivo. Como resultado de esta etapa, se ha reducido el primer conjunto de variables seleccionadas, a aquéllas que poseen datos de buena calidad.

La siguiente etapa es el análisis univariado. En ella, se mide el grado de relación que tiene cada una de las variables independientes (de la etapa anterior) con la variable objetivo. Para ello, se realizan pruebas estadísticas, que dependen de la naturaleza de la variable: si es continua u ordinal, se utiliza la prueba de Kolmogorov-Smirnov (K-S), y si es una variable nominal, se utiliza la prueba Chi-cuadrado de Pearson. Al finalizar esta etapa, se cuenta con las variables independientes que tienen mayor grado de asociación con la variable

objetivo, y son candidatas a ser incorporadas en el modelo multivariado. Esta etapa también sirve como filtro cuando se cuenta con un gran número de variables independientes.

Finalmente, se construye un modelo multivariado, escogiendo el subconjunto de variables que mejor predigan la variable dependiente. Esta etapa tiene varias sub-etapas, que van desde ajustar el modelo hasta evaluar su capacidad predictiva⁶.

El desarrollo previo del modelo, debidamente documentado, tiene gran valor para el proceso de re-calibración, por cuanto deja registrado el conocimiento sobre: qué variables fueron utilizadas, cómo fueron categorizadas, cuáles son más difíciles de obtener, qué variables fueron eliminadas, cuáles requieren un tratamiento especial, etc. Toda esta información, será muy importante para el analista, a la hora de definir el set de datos a ingresar al proceso de re-calibración.

⁶ Extraído de: *Clasificación de Créditos mediante Técnicas avanzadas de Análisis de Datos para Banco Desarrollo* (2007).

4.1.2 Ajuste del Modelo

Si se supone una muestra con 'n' observaciones independientes del vector (\vec{x}_j, y_j) , donde y_j es la variable dependiente y \vec{x}_j es el vector de variables independientes para la persona j-ésima. Ajustar la regresión logística, consiste en encontrar el vector $\vec{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$, asociado al vector $\vec{x} = (x_1, x_2, \dots, x_p)$, que maximice la *probabilidad* de obtener el resultado (y) de todos los individuos que se observan en la muestra. Por lo tanto, la relación funcional de la regresión logística es como en la expresión [1]:

$$\pi(\vec{x}) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} \quad [1]$$

El método para estimar el vector $\vec{\beta}$ es el de máxima verosimilitud. Para aplicar el método de máxima verosimilitud, se debe construir la llamada función de verosimilitud, que expresa la probabilidad de ocurrencia del fenómeno en función de parámetros desconocidos. Su ecuación es la de la expresión [2]:

$$l(\vec{\beta}) = \prod_{i=1}^n \pi(x_i)^{y_i} * [1 - \pi(x_i)]^{1-y_i} \quad [2]$$

Los estimadores de máxima verosimilitud son aquéllos que maximizan esa función, y corresponden a los β_i . Matemáticamente, es más fácil trabajar con el logaritmo natural de la función de verosimilitud (especificada en [2]), para calcular el vector $\vec{\beta}$. Su forma se muestra en la expresión [3].

$$L(\vec{\beta}) = \ln[l(\vec{\beta})] = \sum_{i=1}^n y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)] \quad [3]$$

Diferenciando $L(\vec{\beta})$ con respecto a los $(p+1)$ coeficientes, se obtienen los parámetros que mejor se ajustan a la muestra (Hosmer & Lemeshow, 1989).

4.1.3 Bondad de Ajuste

Después de estimar los coeficientes, se debe verificar la significancia de las variables ingresadas al modelo. Es decir, se busca responder la siguiente pregunta. ¿Explica mejor el modelo la variable de salida con las nuevas variables, que sin ellas?

Para responder la pregunta anterior, se debe comparar los valores observados de la variable dependiente con y sin las variables en cuestión. En

regresión logística, esta comparación se basa en la función de verosimilitud definida en el punto 3.1.2. (Hosmer & Lemeshow, 1989).

La medida de bondad de ajuste propuesta en esta tesis, es el *likelihood ratio test*, donde se construye el estadístico G, definido en la expresión [4]:

$$G = -2 * \ln \frac{(\text{likelihood con todas las variables})}{(\text{likelihood sólo con la constante})} \quad [4]$$

El valor de G puede ser interpretado como la “diferencia de ajuste” entre el modelo que contiene sólo en intercepto (constante) y el modelo con todas las variables independientes. Bajo la hipótesis nula que $\beta_1 = \beta_2 = \dots = \beta_p = 0$, este estadístico sigue una distribución χ^2 con ‘p’ grados de libertad. Si G es estadísticamente significativo (a un nivel $p < 0.05$), entonces se rechaza H_0 y se concluye que las variables independientes permiten realizar mejores predicciones, que usar sólo la constante (Menard, 2002). Es importante recordar que “el rechazo de H_0 , (...) permite concluir que al menos una, y quizás todos los ‘p’ coeficientes son distintos de cero” (Hosmer & Lemeshow, 1989).

4.1.4 R^2 de Nagelkerke

En regresión lineal, la medida de ajuste R^2 es muy popular. Hay una tendencia natural a buscar un número, que se encuentre entre 0 y 1, y que indique, a medida que crece, que el modelo se “ajusta mejor”. Estas medidas parecieran proveer, de manera informal, información muy útil para los investigadores (Magee, 1990).

De allí, que surge la “necesidad” de buscar una medida, análoga al R^2 de OLS, para la regresión logística. Muchos R^2 han sido contruidos para este modelo, en afán de encontrar una medida comparable al “porcentaje de la varianza explicada”, de OLS. Dado que no han sido contruidos en base a la varianza explicada, han sido denominados “pseudo- R^2 ”. Magee (1990), señala que éstos sólo son utilizados, porque a los investigadores les gusta tener medidas parecidas entre modelos regresivos, y no por razones estadísticas formales.

De hecho, Garson (2008) previene que los pseudo- R^2 , que se mostrarán posteriormente, no son medidas de bondad de ajuste, sino más bien, de fortaleza de asociación entre la variable dependiente y las variables independientes (...) Por ejemplo, para muestras pequeñas, puede ocurrir que el

R^2 tenga un valor alto, a pesar que el *likelihood ratio test* (G) haya rechazado que el modelo se ajusta a los datos.

Una consideración importante, es que los " R^2 no pueden ser comparados entre sets de datos distintos, y son válidos sólo para comparar distintos modelos que predigan la misma variable, en el mismo set de datos. En otras palabras, el estadístico pseudo- R^2 sin contexto carece de significado"⁷.

A continuación, se describen los R^2 utilizados por la mayoría de los programas estadísticos, como SAS y SPSS:

1. R^2 de Cox y Snell: Se construye de la siguiente manera:

$$R_C^2 = 1 - \left(\frac{l_0}{l_m} \right)^{2/n} \quad [5]$$

Donde: l_0 es el *likelihood* del modelo, que contiene sólo la constante.

l_m es el *likelihood* del modelo, con todas las variables escogidas.

⁷ FAQ: What are pseudos R-squared? [s.a.].

La división de los *likelihoods* refleja la mejoría del modelo final respecto del modelo con sólo la constante. Cabe destacar que su valor es siempre menor a 1, por lo que es difícil interpretar cuándo es “suficientemente bueno”.

2. R^2 de Nagelkerke: Ajusta el R^2 de Cox y Snell, para que alcance un valor igual a 1 cuando el modelo sea perfecto. Por lo tanto, el R^2 de Nagelkerke tendrá un valor generalmente mayor. Se define como:

$$R_N^2 = \frac{1 - \left(\frac{l_0}{l_m}\right)^{2/n}}{1 - l_0^{2/n}} \quad [6]$$

El R^2 de Nagelkerke, “es el coeficiente más requerido de todos los pseudo- R^2 existentes” (Garson, 2008).

3. R^2 de McFadden: Es una medida no idéntica, pero análoga a la varianza explicada de la regresión lineal, en tanto $\ln(l_0)$ se asemeja a la suma total de cuadrados, y $\ln(l_m)$ a la suma de cuadrados de la regresión. La expresión [7], muestra este estadístico:

$$R_L^2 = 1 - \frac{\ln(l_0)}{\ln(l_m)} \quad [7]$$

Es decir, el R^2 de McFadden indica cuánto reduce la variación total, medida en $\ln(l_0)$, la inclusión de nuevas variables en el modelo (Menard, 2002).

La elección de un R^2 es difícil, y no hay consenso sobre qué medida es mejor, ya que no existen fundamentos estadísticos sólidos que justifiquen la elección de un R^2 sobre otro (Magee, 1990). Tomando eso en consideración, se propone la utilización del R^2 de Nagelkerke, tanto porque mejora propuestas de R^2 anteriores (por ejemplo, el de Cox y Snell), como por su amplia utilización y preferencia de los investigadores. Sin embargo, se recomienda usarlo de forma complementaria al *likelihood ratio test*, para contar con medidas tanto de ajuste a los datos, como de fortaleza de asociación, para decidir la calidad del modelo.

4.1.5 Validación Cruzada y Punto de Corte

Hasta el momento, se han mencionado sólo medidas de ajuste a los datos y de fortaleza de asociación. La eficiencia de predicción es una medida diferente, que establece cuán bien el modelo predice la variable dependiente para observaciones futuras, por lo que es utilizada frecuentemente para la selección de modelos.

La validación cruzada consiste en dividir el set de datos del que se dispone, de forma aleatoria⁸, en dos subconjuntos disjuntos: el set de entrenamiento, que se utilizará para calibrar el modelo (calculando el vector $\vec{\beta}$, estadísticos G y R^2 , etc.), y el set de prueba, con el que se determinará cuán eficiente es el modelo para predecir nuevos clientes.

Para la calibración del modelo, se debe tener en cuenta que, si existe una diferencia significativa en el porcentaje de clientes buenos y el porcentaje de clientes malos, se debe ponderar los buenos por un factor menor a 1, que viene dado por la razón entre buenos y malos. De esta manera, se balancea la base de entrenamiento, para “evitar que el modelo se entrene para predecir un alto porcentaje de los clientes buenos (alta efectividad), pero una pequeña proporción de los clientes malos (pésima discriminación)”⁹. Por otro lado, es importante señalar, que eficiencia del modelo en el set de prueba, dependerá del punto de corte escogido

El punto de corte es un número entre 0 y 1, a partir del cual se “binariza” la predicción del modelo, es decir, se determina que un cliente es bueno o malo.

⁸ Se debe cuidar que la fracción de clientes buenos (y malos) observados, sea similar en ambos sets de datos.

⁹ *Clasificación de Créditos mediante Técnicas avanzadas de Análisis de Datos para Banco Desarrollo* (2007).

El criterio de elección de este punto depende de la estrategia de la empresa y de la naturaleza del fenómeno a predecir, por ejemplo: maximizar el acierto global, lograr utilidad máxima, eliminar un porcentaje de clientes malos, etc.

Menard (2002) muestra que en la regresión logística, generalmente, para la eficiencia de predicción, los estadísticos G y R^2 producirán resultados consistentes. Es totalmente posible, sin embargo, que un modelo se ajuste bien, pero haga un trabajo pobre prediciendo la variable dependiente. Lo anterior ocurre cuando el modelo se ha sobreajustado a la muestra de entrenamiento. Que un modelo esté sobreajustado, quiere decir que predice significativamente mejor los datos de entrenamiento que los datos de prueba, por lo que su capacidad de predecir correctamente nuevos casos será baja.

Una forma de verificar que el modelo no se encuentra sobreajustado, es revisando que el error (o acierto) de predicción es “similar” en el set de entrenamiento y de prueba. Para ello, es necesario construir las matrices de confusión en ambos sets. La matriz de confusión, es una matriz de 2x2, que contiene en sus filas los valores predichos de la variable dependiente (dado un punto de corte), y en sus columnas el valor observado.

La ilustración 8 muestra una Matriz de Confusión genérica.

Ilustración 8: Matriz de Confusión Genérica

		<i>Observado</i>	
		$Y_i = 1$	$Y_i = 0$
<i>Predicho</i>	$\hat{Y}_i = 1$	A	C
	$\hat{Y}_i = 0$	B	D

En la matriz se tiene que: A es el número de verdaderos positivos (malos acertados) predichos por el modelo, B el de falsos negativos (buenos errados), C el de falsos positivos (malos errados), y D el número de verdaderos negativos (buenos acertados). De aquí se deduce, por ejemplo, que la expresión $\frac{A+D}{(A+B+C+D)}$, es el porcentaje de acierto global del modelo, y que $\frac{A}{(A+B)}$, es el porcentaje de acierto de malos, del total de malos observados.

Luego, si $\frac{A}{(A+B)_{entrenamiento}} \approx \frac{A}{(A+B)_{prueba}}$, el modelo no se estará sobreajustando a los datos, en la predicción de clientes malos. Lo anterior debe cumplirse también para cada una de las demás componentes $(\frac{B}{(A+B)}, \frac{C}{(C+D)}$ y $\frac{D}{(C+D)})$. Nótese que el sobreajuste no depende del punto de corte que se elija.

Ahora bien, para determinar el mejor punto de corte a utilizar, se deberá escoger aquella matriz de confusión que maximice, en el set de prueba, el

criterio de elección del punto de corte. Por ejemplo, en el caso que el criterio sea la maximización del acierto global del modelo, se escogerá la matriz que alcance la mayor expresión para $\frac{A+D}{(A+B+C+D)}$, de entre todos los puntos de corte candidatos.

4.1.6 Stepwise Logistic Regression

Anteriormente, la descripción del modelo se ha enfocado en la selección de variables candidatas, estimación de coeficientes, bondad de ajuste y efectividad de predicción, y es posible que se haya creado la percepción que existe sólo un modelo posible. Aunque hay muchas situaciones en que esto es así, es más típico que haya muchas variables independientes que podrían ser incluidas en el modelo.

Una técnica para abordar la selección de variables, es el uso del método *stepwise*, en el cual las variables son seleccionadas e incluidas (o excluidas) del modelo, sólo en base a criterios estadísticos. Hay dos versiones principales: selección *forward* y selección *backward*.

La selección *forward* comienza sólo con un constante, agregando en cada paso, la variable que maximiza el estadístico G del modelo. Por otro lado, la selección *backward* comienza con todas las variables candidatas, eliminando en cada paso, la variable que realiza el aporte menos importante. El criterio de detención de ambos métodos, es cuando el paso siguiente produce (para *backward*) o no (para *forward*) un cambio significativo en el modelo, medido en la variación de G.

Un aspecto crucial para juzgar si el cambio en el *likelihood ratio* del modelo, es efectivamente significativo, es la elección del nivel de significancia α . Constanza y Afifi (citados en Hosmer & Lemeshow, 1989), consideran que el criterio de $\alpha=0.05$ puede ser restrictivo, y recomiendan utilizar valores entre 0.15 y 0.20. Obviamente, la decisión dependerá, entre otras cosas, de la cantidad de variables de las que se disponga y de las variables que se quiera conservar en el modelo.

Los procedimientos de selección *stepwise*, han sido criticados porque pueden dar resultados sin sentido y elegir variables irrelevantes y ruidosas. Pero el problema no es el hecho que el algoritmo pueda elegir tales modelos, sino que el analista falle en inspeccionar cuidadosamente el modelo resultante y suponga que corresponde al mejor modelo final posible. Es sólo cuando el

analista comprende las fortalezas, y especialmente las limitaciones de estos métodos, para que sean una herramienta efectiva en la construcción de modelos. Es el analista, no el computador, el último responsable de la revisión y evaluación del modelo (Hosmer & Lemeshow, 1989).

4.2 Mantención del Modelo

En este ítem, se explicarán los conceptos directamente relacionados con el proceso “Mantención del Modelo de Comportamiento de Clientes”, sin perjuicio que los conceptos mostrados anteriormente, surjan de alguna u otra manera en dicho proceso. Lógicamente, se asumirá que la regresión logística ya se encuentra construida.

4.2.1 Estabilidad Poblacional

Para estudiar si han ocurrido cambios importantes en la población, a lo largo del tiempo, se debe realizar una prueba que dependa de los parámetros obtenidos en el modelo. En particular, se debe verificar que los Beta se

encuentren dentro de un intervalo de confianza, que permita inferir si la población se ha mantenido estable en el tiempo.

En primer lugar, es necesario recordar que cada coeficiente Beta corresponde a “un valor”¹⁰ que pondera un atributo de la población (no la muestra). Pero dado que este valor es calculado en base a una muestra, tiene un error asociado, que incluye elementos propios de la información que no se está observando. Entonces, como los estimadores betas admiten cierta variabilidad, es posible exista otra población, no muy distinta a la actual, tal que éstos aún la “representen”. Dicha variabilidad, se captura en los intervalos de confianza para el coeficiente Beta, que se muestran en la expresión [8]:

$$\beta_{min} \leq \beta \leq \beta_{max} \quad [8]$$

Donde: $\beta_{min} = (\beta - 2\sigma_\beta)$ y $\beta_{max} = (\beta + 2\sigma_\beta)$.

σ_β es la desviación estándar muestral de los coeficientes.

De esta forma, se puede considerar que, si la media de las variables acompañadas por los coeficientes Beta varía, es equivalente a que variara el valor del coeficiente para la población inicial. Esta variación, permite generar un

¹⁰ Ese “valor único” sólo podría ser calculado, si se contara con toda la población para calibrar el modelo.

intervalo de confianza para las medias de las variables, en base a la simulación de movimientos en los parámetros del modelo. Así, si se tiene una población distinta, el cambio en la media de la variable $\bar{X}' = C\bar{X}$, puede asociarse a un cambio en el coeficiente Beta, dado por $\beta\bar{X}' = \beta(C\bar{X}) = (C\beta)\bar{X} = \beta'\bar{X}$. Por lo tanto, si β' pertenece al intervalo dado por las variaciones máximas permitidas por el coeficiente Beta inicial, se tendrá que la población no ha cambiado significativamente. Luego, se cumplirá lo que se establece en [9]¹¹:

$$\text{Si } \beta_{min} \leq \beta' \leq \beta_{max} \Rightarrow \frac{\beta_{min}}{\beta} \leq C \leq \frac{\beta_{max}}{\beta} \quad [9]$$

4.2.2 Relevancia de las Variables en los Modelos

Para estudiar la importancia de una variable para el modelo, es posible utilizar los mismos instrumentos estadísticos del análisis univariado: el Test K-S y el Test Chi-Cuadrado de Pearson. Sin embargo, en este caso se debe estudiar si la diferencia de la distribución de cada variable explicativa es significativamente distinta, para clientes buenos y malos, de acuerdo a las

¹¹ Basado en: *Clasificación de Créditos mediante Técnicas avanzadas de Análisis de Datos para Banco Desarrollo* (2007).

predicciones del modelo. Es decir, se utiliza como variable dependiente la predicción del modelo “binarizada” de acuerdo al criterio de corte escogido, y no al comportamiento observado, como en el análisis univariado. De esta forma, se obtiene una medida de cuán útil es la variable para el modelo y no de su poder de predicción¹².

Ambas pruebas estadísticas propuestas, tienen la particularidad de ser no paramétricas, es decir, no hacen supuestos sobre el valor de los parámetros o las distribuciones de las variables. Esto adquiere especial relevancia en los casos en que la distribución de las variables es desconocida.

4.2.2.1 Test de Kolmogorov Smirnov (K-S)

Esta prueba se realiza para las variables continuas y ordinales, y se utiliza para probar si las distribuciones poblacionales de dos variables son iguales o distintas. En este trabajo, se comparará la misma variable en dos poblaciones, la de clientes buenos y la de malos. Para realizar el contraste, el test se basa en las discrepancias entre las frecuencias relativas acumuladas de

¹² *Clasificación de Créditos mediante Técnicas avanzadas de Análisis de Datos para Banco Desarrollo* (2007).

las dos muestras. Lo anterior, propicia que la prueba pueda advertir diferencias no tan sólo entre los promedios, sino también debidas a la dispersión, simetría u oblicuidad de los datos.

A la vez, esta prueba admite que los tamaños de las muestras sean distintos.

Las Hipótesis son:

H_0 : Las distribuciones de ambas poblaciones son iguales.

H_1 : Las distribuciones de ambas poblaciones son distintas.

El estadístico de contraste es D , y corresponde a la máxima diferencia entre las proporciones escalonadas calculadas para cada valor. El parámetro viene dado de la siguiente forma:

$$D = \max[S_1(x) - S_2(x)] \quad [10]$$

En la expresión anterior, S_1 es la proporción de valores de la primera muestra, que son iguales o menores que x , S_2 es la proporción de valores de la segunda muestra, que son iguales o menores que x (Álvarez, 1994).

El valor de este estadístico sigue una distribución normal tipificada, por lo que si su valor Z^* es mayor que $Z_{1-\alpha}$, es posible rechazar que las poblaciones son iguales, y concluir que esa variable le es útil al modelo para distinguir clientes buenos y malos.

4.2.2.2 Test Chi-Cuadrado de Pearson

Este test se realiza para variables categóricas, y se utiliza para averiguar si éstas se encuentran están asociadas. Para el caso en estudio, si se suponen dos variables categóricas A y B, la prueba permite determinar si la variable predictiva (sea ésta B, con r categorías), influye o no en el pago del crédito (A, con 2 categorías). Para realizar el contraste, el test dispone los datos en dos tablas de frecuencias, y para cada intervalo, calcula la diferencia entre la frecuencia observada y la esperada.

Las Hipótesis son:

H_0 : Las variables A y B son independientes (no existe asociación).

H_1 : Las variables A y B son dependientes (asociación).

Primero, se construye la matriz de frecuencias observadas, como se muestra en la ilustración 9:

Ilustración 9: Matriz de Frecuencias Observadas

	A_0	A_1	$\sum n_{i.}$
B_0	n_{00}	n_{01}	$n_{0.}$
B_1	n_{10}	n_{11}	$n_{1.}$
B_2	n_{20}	n_{21}	$n_{2.}$
\vdots	\vdots	\vdots	\vdots
B_r	n_{r0}	n_{r1}	$n_{r.}$
$\sum n_{.j}$	$n_{.0}$	$n_{.1}$	n

Bajo la H_0 , de no asociación, se debe cumplir que: $P(A_i \cap B_j) = P(A_i) * P(B_j)$, $\forall i, j$.

Entonces, la tabla de frecuencias esperadas será como muestra la ilustración 10:

Ilustración 10: Matriz de Frecuencias Esperadas

	A_0	A_1	$\sum n_{i.}$
B_0	e_{00}	e_{01}	$n_{0.}$
B_1	e_{10}	e_{11}	$n_{1.}$
B_2	e_{20}	e_{21}	$n_{2.}$
\vdots	\vdots	\vdots	\vdots
B_r	e_{r0}	e_{r1}	$n_{r.}$
$\sum n_{.j}$	$n_{.0}$	$n_{.1}$	n

Donde:
$$e_{ij} = \frac{n_{i.} * n_{.j}}{n}$$

Finalmente, se calcula el estadístico χ^2 , como la suma de las “distancias” entre la frecuencia esperada y observada, como lo especifica la expresión [11]:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^2 \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad [11]$$

Este estadístico sigue una distribución chi-cuadrado, con $(r-1)$ y (1) grados de libertad, es decir, tiene la forma $\chi^2_{(r-1)(1)}$. Para un nivel de significancia α , se tendrá que si $\chi^2_{(r-1)(1)} > \chi^2_{(1-\alpha)}$, entonces se rechaza H_0 , y se concluye que existe asociación significativa entre las variables. En caso contrario, no será posible concluir que existe dicha asociación entre las variables¹³.

4.2.3 Índice de Concordancia (C-Index)

Hacer un *ranking* de regresiones logísticas, basándose únicamente en sus matrices de confusión, en una muestra determinada, tiene los siguientes problemas:

¹³ Ítem basado en: Cadarso-Suárez, C. [n.d.].

1. La elección del punto de corte 'c' es arbitraria. La efectividad del modelo cambiará si se elige un punto diferente.
2. Al estimar la variable dicotómica \hat{Y}_i (del cliente 'i'), mediante la discretización de la variable continua $\hat{\pi}_i$, hay una pérdida de información.
3. Las componentes de la matriz de confusión, son determinadas por la muestra, y no dependen del modelo. Como resultado, estas cantidades no serán medidas absolutas de calidad del modelo.

Existe la posibilidad de utilizar todos los puntos de corte de forma simultánea, a través de la curva ROC. Bamber y, Hanley y McNeil (citados en Harrell, 2001), han mostrado que “el *C-Index* es idéntico al área bajo la curva ROC”, por lo que el análisis se centrará en la deducción e interpretación únicamente del índice.

Para deducir el *C-Index*, se supondrá un set de datos con n_1 registros positivos y n_0 negativos. Luego, se debe construir una matriz, que contenga todos los pares posibles de ceros y unos ($n_1 * n_0$). Cada componente de la matriz, contendrá la variable U_i , definida en la expresión [12]:

$$U_i = \begin{cases} 1 & \text{si } \hat{\pi}_{1i} > \hat{\pi}_{0i} \\ 0 & \text{en otro caso} \end{cases} \quad [12]$$

En la variable anterior, se tiene que $\hat{\pi}_{1i}$ y $\hat{\pi}_{0i}$ son las probabilidades estimadas (por la regresión logística) que cada cliente del par i -ésimo, sea malo y bueno, respectivamente. Así, $U_i=1$ si el modelo asigna una mayor probabilidad de no pago al cliente malo que al cliente bueno. Cuando esto ocurre, se dice que las observaciones son “concordantes”.

A partir de lo anterior, se puede calcular el índice de concordancia del modelo, que se define en la expresión [13]:

$$C - Index = \frac{\sum_{i=1}^{n_1 * n_0} U_i}{n_1 * n_0} \quad [13]$$

Si el índice (o equivalentemente, el área ROC) es igual a 0.5, entonces el modelo no lo está haciendo mejor que el azar. Por otro lado, mientras más se acerque a 1.0, significa que su poder discriminante aumenta. Se ha propuesto una escala arbitraria, para interpretar los valores del índice de concordancia, que se muestra en la ilustración 11¹⁴:

¹⁴ Ítem basado en: Weiss, J. (2006)

Ilustración 11: Escala para interpretar valores del índice de concordancia

$$C - Index = \begin{cases} 0.5 & \text{No hay Discriminación} \\ 0.6 - 0.7 & \text{Pobre} \\ 0.7 - 0.8 & \text{Aceptable} \\ 0.8 - 0.9 & \text{Excelente (Bueno)} \\ > 0.9 & \text{Sobresaliente} \end{cases}$$

5. ANÁLISIS DEL MEDIO INTERNO Y EXTERNO

Para un mejor entendimiento de la dinámica de la industria y de la posición de la empresa frente a ella, se realizará primero un análisis de las 5 Fuerzas de Porter y luego un análisis FODA.

5.1 Análisis de 5 Fuerzas de Porter

1. Amenaza de entrada de Nuevos Competidores: Media.

De acuerdo a lo mencionado en el punto 3.2, se tiene que aún hay gran cantidad de microempresarios por bancarizar, lo que hace al segmento atractivo para la entrada de nuevos competidores. Tal podría ser el caso de Banco Falabella y Banco Ripley, dedicados a los créditos de consumo. Sin embargo, los desincentivos son altos, por el “peso” de las marcas predominantes en el mercado, “fidelidad”¹⁵ de los microempresarios ya bancarizados y alta concentración de la oferta.

¹⁵ Esta “fidelidad”, refiere a la relación casi exclusiva que tienen los microempresarios formales con una institución financiera, tal como se mencionó en el punto 3.2.2 de este informe. Que esto así ocurra, dependerá seguramente del servicio que éstas le presten a los clientes.

2. Rivalidad entre competidores: Alta.

Gran competencia por captación de clientes, con productos atractivos para cada segmento, y de retención de clientes, basada en estrategia de calidad de servicio.

3. Poder de Negociación de los Proveedores: Baja.

El costo del dinero, es similar para todas las instituciones bancarias con riesgo comparable.

4. Poder de Negociación de los Compradores: Alto.

El cliente puede cambiar fácilmente de proveedor. Por ello, la estrategia de retención de clientes, toma especial relevancia para la institución financiera proveedora del servicio.

5. Amenaza de Productos Sustitutos: Baja.

Créditos de consumo ofrecidos por empresas de Retail. Sin embargo, éstos no han sido diseñados especialmente para los microempresarios.

5.2 Análisis FODA

1. Fortalezas:

- Experiencia del mercado en el segmento de microempresas.
- Relación de confianza y trato personalizado al cliente.
- Alto compromiso del personal con la empresa, que además se encuentra sensibilizado con el enfoque de rentabilización de clientes

2. Debilidades:

- Operaciones poco eficientes: Baja automatización, criterios no estandarizados de evaluación, bajo control de desempeño de la cartera.
- Baja calidad de servicio a los clientes, consecuencia del punto anterior.
- Falta de seguimiento de los clientes: Sólo depende de los ejecutivos, por lo que se pierden muchas de oportunidades de negocio.

3. Oportunidades:

- Gran parte del mercado potencial se encuentra sin bancarizar.
- Traspaso de INDAP de 100.000 clientes al sistema financiero, con buena clasificación de riesgo.

- Relación del cliente microempresario con una única institución bancaria. Oportunidad de captación y retención de clientes buenos.

4. Amenazas:

- Altos costos de energía y agua, que afecta principalmente al sector agropecuario.
- Variabilidad del precio del dólar, que afecta los productos de exportación y los servicios a exportadores que realizan los microempresarios. Lo anterior, viene acompañado de un escenario económico internacional inestable, debido a la crisis sub-prime de Estados Unidos, desde mediados del 2007, y recesión de los países europeos, el año 2008.
- Variabilidad del precio del petróleo, que afecta principalmente al rubro del transporte, y áreas productivas que requieren de este insumo.
- Aumento de riesgo de los microempresarios, consecuencia de las razones anteriores.

6. EL PROYECTO

6.1 Planteamiento Estratégico

A principios del 2007, Banderrollo Microempresas apuntó a aumentar su participación de mercado, estableciendo como meta duplicar el número de clientes para fines de año. La captación masiva de clientes, era la tendencia natural de un mercado potencial que sólo alcanzaba el 40% de bancarización. Para llevar a cabo el plan, se incentivó los canales de productos masivos, se generó créditos automáticos, se creó nuevos productos (crédito fácil) y atendió nuevos segmentos (pesca artesanal). Con el fin de agilizar las atenciones, se utilizaron principalmente instrumentos de evaluación simples, que escasamente reflejaban la capacidad de pago del cliente.

Todo lo anterior, se tradujo en un aumento del riesgo de la cartera, que no fue reflejado en la tasa de los créditos. Por ello, la rentabilidad fue menor a la esperada y las provisiones y castigos aumentaron ostensiblemente, cerrando el año con pérdidas.

La fusión entre el Banco del Desarrollo y el Scotiabank, hacia fines del 2007, trajo como consecuencia directa un cambio en la estrategia comercial,

que podría resumirse en el lema “rentabilidad con criterios de masividad”¹⁶. La rentabilidad, se lograría mediante una mejor selección de los clientes y adecuada gestión de la cartera actual, identificando y reteniendo a los mejores clientes. La masividad, reconoce la necesidad de tener una alta participación en un mercado creciente.

Para la selección de clientes, se encontraba en fase final de desarrollo un modelo de *scoring*. Sin embargo, los instrumentos evaluadores seguían reflejando de manera muy simple la situación financiera de cada cliente, por lo que complementaban pobremente la sugerencia del modelo para el otorgamiento del crédito. Adicionalmente, no existía un proceso definido para controlar el desempeño tanto de las ventas como del modelo, que permitiera analizar si la rentabilidad y el riesgo de la cartera eran concordantes con lo definido por la empresa.

6.1.1 El Problema

El problema de evaluación de clientes tiene tres aristas:

¹⁶ Anteriormente el lema era “Masividad con criterios de Rentabilidad”, invirtiéndose las prioridades.

1. Diseño de las Evaluaciones: Lo primero, es que no todos los instrumentos de evaluación se encuentran automatizados. Varios de éstos están hechos en Word o Excel, con el consiguiente flujo de archivos por correo electrónico, dificultad de levantamiento de información y cálculos manuales que esto conlleva.

Lo segundo, es que la información contenida en las evaluaciones de algunos segmentos, obvian aspectos relevantes y propios de cada uno de ellos, o bien los describen muy precariamente.

2. Ingreso a las Evaluaciones: Dado que no todas las evaluaciones se encuentran automatizadas, cada ejecutivo decide cuál de ellas utilizará para evaluar al cliente. Generalmente, optan por utilizar la evaluación que más les “parezca”, que no necesariamente es la que más se adecúa al perfil del cliente.
3. Ausencia de Control: No existen procesos definidos para controlar el desempeño de los créditos otorgados, en relación a la rentabilidad y riesgo de los distintos segmentos, ni la efectividad del modelo de *scoring*, para la discriminación de clientes.

6.1.2. Posicionamiento Estratégico

Las microempresas, son un segmento en el que pocas instituciones financieras tienen experiencia en el país. Justamente, lo que diferencia al Banco del Desarrollo de la competencia, es su conocimiento de este segmento y la creación de lazos de confianza con los clientes. Éstos se sienten identificados con el Banco, porque entiende su realidad y demuestra un compromiso social asociado a su actividad. Como ya se ha explicado, esto es una ventaja competitiva que el Banco ya posee, pero que debe ser complementada con excelencia operacional, que la refuerce y consolide al Banco como una empresa que apoya a los microempresarios con un servicio de calidad.

6.2 Descripción de la Oportunidad

Para describir la oportunidad del proyecto, se expondrá la siguiente cita, respecto del desafío de las instituciones financieras que atienden microempresas para lograr la eficiencia económica:

“(...) el trabajo de los intermediarios financieros con orientación al segmento de empresas de menor tamaño, consiste en implementar tecnologías

crediticias específicas que permitan “seleccionar” a los clientes viables y enfrentar las particularidades del segmento. En otros términos, implementar procesos de crédito especializados que permitan entre otras cosas, la evaluación del perfil de riesgo de los microempresarios, de su capacidad de pago y el establecimiento de condiciones de crédito consistentes con la evaluación anterior” (Morales y Yáñez, 2007).

La oportunidad es clara: cómo reflejar la situación financiera del cliente, de manera de determinar lo más certeramente posible su excedente y, finalmente, su capacidad de pago del crédito. Pero el desafío debe ir más allá: cómo identificar a los clientes riesgosos antes del otorgamiento el crédito, mediante el uso de herramientas avanzadas, y cómo retener a los clientes buenos, a través de un servicio de calidad.

6.3 Definición del Proyecto

Para definir el proyecto, se abordarán 3 ámbitos:

1. Objetivos, derivados de la oportunidad expuesta en el punto 5.2.
2. Producto del rediseño.
3. Alcance.

6.3.1 Objetivos

Aumentar la rentabilidad de la empresa y controlar el riesgo de los distintos segmentos de la cartera de clientes, mediante el cumplimiento de los objetivos que se exponen a continuación.

6.3.1.1 Objetivo General

Rediseñar el proceso de evaluación de clientes, de forma que permita a la empresa estandarizar el criterio de evaluación y controlar tanto el desempeño de los segmentos que atiende, como el funcionamiento del *scoring*.

6.3.1.2 Objetivos Específicos

- Evaluar los clientes customizadamente.
- Medir el desempeño de créditos por segmento.
- Centralizar el criterio de clasificación de clientes, para alinear la estrategia comercial con la ejecución operacional.
- Mantener actualizado el modelo de *scoring*, para identificar adecuadamente los cambiantes perfiles de riesgo de los clientes.

6.3.2 Producto

Para cumplir los objetivos descritos anteriormente, se incluirán los siguientes elementos en el proceso de evaluación de clientes, que será modelado y explicado posteriormente:

1. Rediseño de las evaluaciones automatizadas y manuales. Respecto a las primeras, deben caracterizar adecuadamente a cada segmento atendido. Por otro lado, el rediseño y automatización de las fichas “manuales” es estrictamente necesario para la clasificación del punto siguiente.
2. Clasificación automática de clientes, mediante la aplicación centralizada de la lógica de negocio, que redirija al ejecutivo a la evaluación que corresponda al perfil del cliente.
3. Mantención de la lógica de clasificación, a través del seguimiento y evaluación constante de las variables de clasificación de los clientes. Para ello, se establecerán indicadores de rentabilidad y riesgo, que den cuenta del desempeño los segmentos atendidos.
4. Integración de evaluaciones con modelo de *scoring*, que apoyará a la instancia de aprobación, en la decisión de otorgamiento del crédito.

5. Medición de la efectividad del *scoring*, estableciendo criterios para la re-calibración del modelo, en caso que sea necesario.

6.3.3 Alcance

El alcance del proyecto propuesto, incluye todos los aspectos relacionados con la evaluación de los clientes:

- La estandarización del criterio de clasificación de clientes.
- Rediseño y automatización de evaluaciones para cada perfil de clientes.
- Control de desempeño de los segmentos atendidos.
- Control de efectividad y re-calibración del *scoring*.

Adicionalmente, el Proyecto tendrá un rol integrador, por cuanto:

- Las evaluaciones actuales y rediseñadas, se encontrarán en una plataforma.
- Integra el modelo de *scoring* a las evaluaciones.

- Propone una mirada holística del proceso de evaluación, desde la clasificación y cuantificación del riesgo del cliente, hasta la medición de desempeño de cada segmento.

6.4 Propuesta de Valor

“La estrategia competitiva consiste en ser diferente. Significa elegir deliberadamente un conjunto diferente de actividades para brindar una mezcla única de valor” (Porter, 1997).

El proyecto busca reforzar la estrategia competitiva de la empresa, que define una propuesta de valor basada en una atención customizada y eficiente al cliente. Dicho esto, la propuesta de valor del proyecto se centra en dos aspectos principales:

1. Reforzar la estrategia de calidad de servicio al cliente (CAR):
 - 1.1 Mejor comprensión de la realidad del cliente: Evaluaciones customizadas a cada segmento, lo que permitirá “asesorarlos” mejor.

1.2 Mayor excelencia operacional: Evaluaciones automatizadas e información en línea de los clientes, aumentando la “rapidez” de atención y ciclo de venta.

2. Alinear las operaciones con la estrategia comercial del Banco:

2.1 Criterio estandarizado de evaluación de clientes.

2.2 Control de desempeño de los distintos segmentos atendidos.

2.3 Control del riesgo de la cartera, mediante el uso de un *scoring* debidamente actualizado.

6.5 Factores Críticos de Éxito

Se identifican los siguientes factores, que son cruciales para la implementación del proyecto sea exitosa:

1. Diseño de las evaluaciones: Las evaluaciones rediseñadas, deben reflejar lo mejor posible la realidad comercial (productiva) y financiera del cliente, para determinar adecuadamente las condiciones de crédito.

2. Ingreso de las variables de clasificación: Los ejecutivos pueden tener incentivos a inducir la lógica a alguna evaluación en particular. Para disminuir este riesgo, se utilizarán campos distractores.
3. Incorporación de nuevas evaluaciones: Al pasar el tiempo, nuevos segmentos serán atendidos y otros serán dejados de atender. Es de suma importancia que las evaluaciones que se creen (eliminen) para esos segmentos, sean automatizadas e incorporadas a la lógica de clasificación. Si esto no es así, se volverá a una situación similar a la actual, de automatización parcial y criterio de clasificación no completamente centralizado.
4. Mantenimiento del *scoring*: Si el modelo de riesgo se encuentra desactualizado, al punto de afectar su capacidad discriminante, éste comenzará a recomendar la exclusión de clientes poco riesgosos y aceptación de clientes muy riesgosos, afectando la rentabilidad de la empresa.
5. Campaña comunicacional: Dar a conocer oportunamente el proyecto, para que las personas estén alineadas y comprometidas con sus objetivos y funcionamiento.

6.6 Justificación Económica

6.6.1 Medición de Beneficios

El beneficio del proyecto será medido en el aumento del Margen Operacional Bruto (MOB) promedio por cliente. Se define el MOB como:

$$\text{MOB} = \text{Margen Financiero} - \text{Comisiones} + \text{Otros Ingresos} - \text{Provisiones} - \text{Castigos}$$

Notar que el riesgo, medido en provisiones, está incluido en el MOB. El beneficio total, será el aumento del MOB promedio por cliente multiplicado por el número de clientes.

6.6.2 Medición de Costos

Para determinar los costos, se tomó como punto de comparación el costo promedio de desarrollo informáticos similares, implementados con anterioridad

en el Banco. Adicionalmente, dada la naturaleza del proyecto y sus requerimientos, se consideró un plazo de 6 meses¹⁷ para su desarrollo.

En cuanto al hardware, se necesitará usar un servidor web, para la capa de presentación de las nuevas evaluaciones, y un servidor de aplicaciones, que almacene la lógica de negocio y envíe la información del desempeño de las ventas y efectividad del *scoring*. Los costos de hardware no serán incluidos en la evaluación del proyecto, pues dado que ya existe hardware apropiado para su desarrollo, el costo de su uso es marginal.

Por lo tanto, los costos mensuales son:

- Ingeniero MBE: \$1.500.000.
- Equipo de desarrollo informático: \$2.000.000 (1 ingeniero informático y 1 programador ASP).

Adicionalmente, se considerará un costo de mantención de las evaluaciones, lógica de clasificación y modelo de *scoring*, en los años 2 y 3, equivalentes a dos meses del equipo de desarrollo.

¹⁷ Este periodo es una estimación, y debe ser sometida a un estudio riguroso de requerimientos y plazos, al igual que los costos expuestos.

6.6.3 Construcción del Flujo de Caja

Para la construcción del flujo de caja, primero se debe proyectar el margen que desea obtener la empresa durante el próximo trienio. El plan de ventas, estima que el margen financiero de los activos y pasivos, muestra en el cuadro 5:

Cuadro 5: Plan de Ventas, Trienio 2008-2010

Cifras en MM\$	2.008	2.009	2.010
Margen de Activos	11.835	14.740	17.490
Margen de Pasivos	360	280	320
Margen Financiero	12.195	15.020	17.810

Pero para el cálculo del MOB también se requiere tener una estimación de las provisiones, castigos, recuperaciones y comisiones. Esta información será calculada en base a los datos del trienio anterior, obteniendo el porcentaje que representa cada variable del margen financiero. Por ejemplo, si en los últimos tres años las provisiones fueron un 22% del margen financiero (en promedio), se considerará que en el periodo de evaluación esa tendencia se mantendrá. El resumen de estos porcentajes se muestra en el cuadro 6:

Cuadro 6: Proporción de Variables del Margen Financiero, Trienio 2005-2007

Datos Históricos	2005	2006	2007	Promedio
Comisiones Netas/Margen Fin	26,7%	12,9%	26,3%	21,9%
(Provisiones y Castigos)/Marg.F.	19,5%	30,6%	37,3%	29,1%
Recuperaciones/Marg. Financ	53,4%	21,0%	16,8%	30,4%

Adicionalmente, para el cálculo del MOB promedio por cliente, se necesita una proyección del número de clientes. Dado que sólo existe una meta de clientes para el año 2008 (45.000 clientes más), se supondrá que en los siguientes años, el crecimiento será constante e igual a 40.000.

Con lo expuesto anteriormente, se tiene que la proyección del MOB promedio anual por cliente es como se adjunta en el cuadro 7:

Cuadro 7: Proyección del MOB Promedio Anual por Cliente, Trienio 2008-2010

Cifras en MM\$	2.008	2.009	2.010
Margen de Activos	11.835	14.740	17.490
Margen de Pasivos	360	280	320
Margen Financiero	12.195	15.020	17.810
Comisiones Netas (22%)	-2682,9	-3304,4	-3918,2
Resultado Operacional Bruto	9.512	11.716	13.892
Provisiones y Castigos (29%)	-3.537	-4.356	-5.165
Recuperaciones (30%)	1.061	1.307	1.549
MOB Anual	7.037	8.667	10.276
N° Clientes	165.000	205.000	245.000
MOB Anual/Cliente [\$]	\$ 42.646	\$ 42.276	\$ 41.944

Para efectos de calcular un MOB esperado por cliente, se pensará en tres escenarios posibles: optimista, conservador y pesimista. En cada uno de ellos, se obtendrá un porcentaje (100%, 75% y 50% correspondientemente) de las proyecciones del margen financiero y número de clientes, que se encuentran de la tabla anterior.

En base a juicio experto, se le ha otorgado una probabilidad de ocurrencia a cada escenario de 20% para optimista, 60% para conservador y 20% para pesimista.

Dado que el escenario optimista es equivalente al cuadro 7, éste no será presentado. Los otros escenarios se muestran en los cuadros 8 y 9:

Cuadro 8: Proyección del MOB Promedio Anual por Cliente, Escenario Conservador (75%)

Cifras en MM\$	2.008	2.009	2.010
Margen de Activos	8.876	11.055	13.118
Margen de Pasivos	270	210	240
Margen Financiero	9.146	11.265	13.358
Comisiones Netas (22%)	-2.012	-2.478	-2.939
Resultado Operacional Bruto	7.134	8.787	10.419
Provisiones y Castigos (29%)	-2.652	-3.267	-3.874
Recuperaciones (30%)	796	980	1.162
MOB	5.277	6.500	7.707
Número Clientes	153.750	183.750	213.750
MOB/Cliente [\$]	\$ 34.324	\$ 35.374	\$ 36.057

Cuadro 9: Proyección del MOB Promedio Anual por Cliente, Escenario Pesimista (75%)

Cifras en MM\$	2.008	2.009	2.010
Margen de Activos	5.918	7.370	8.745
Margen de Pasivos	180	140	160
Margen Financiero	6.098	7.510	8.905
Comisiones Netas (22%)	-1.341	-1.652	-1.959
Resultado Operacional Bruto	4.756	5.858	6.946
Provisiones y Castigos (29%)	-1.768	-2.178	-2.582
Recuperaciones (30%)	530	653	775
MOB	3.518	4.333	5.138
Número Clientes	142.500	162.500	182.500
MOB/Cliente [\$]	\$ 24.690	\$ 26.666	\$ 28.154

Para el cálculo del ingreso esperado del proyecto, se deben tener las siguientes consideraciones:

- El Proyecto sólo considera la evaluación uno a uno de clientes, es decir, no incluye los productos masivos, cuentas de ahorro, etc. Por ello, se debe obtener sólo el beneficio del MOB incremental, multiplicado por el 50% del número total de clientes.
- Las evaluaciones rediseñadas, son las del rubro de pesca artesanal, silvoagropecuario, más las empresas emergentes. Estos clientes representan un 20% del total, por lo que también se deberá obtener esa fracción para calcular el beneficio del proyecto. No obstante lo anterior, el proyecto debería tener un impacto positivo sobre las evaluaciones de los otros clientes, pues estandarizará los criterios de evaluación y permitirá controlar la rentabilidad y riesgo de los distintos segmentos.
- Para el año 2008, sólo se considerará el ingreso del segundo semestre (la mitad del ingreso anual esperado), lo que reduce el horizonte de evaluación a 2 años y medio. Lo anterior, es porque a partir de mitad de año se espera que el proyecto genere algún impacto.

Dicho lo anterior, para calcular el Flujo de Caja, se supondrán 3 escenarios de aumento de MOB por cliente (3%, 5% y 10%). Luego, se podrá

estimar los ingresos esperados, dado un aumento en el MOB anual promedio por cliente, producto de la implementación del proyecto. Con esto, se busca ilustrar cuál es el impacto que tendrá en el VAN el incremento porcentual del MOB, para determinar una curva de sensibilidad del valor del proyecto.

Finalmente, se debe mencionar que la tasa de descuento a ocupar será la tasa libre de riesgo $R_f = 6\%$, dado que el riesgo ya se encuentra incluido en el MOB, a través de las provisiones, según ya se explicitó anteriormente. Los 3 escenarios se muestran en los cuadros 10, 11 y 12:

Cuadro 10: Flujo de Caja con Aumento del MOB por Cliente de 3%

Concepto/Año	2008	2009	2010
MOB incremental	7.916.079	19.499.715	23.121.833
Costos Desarrollo	-12.000.000	-4.000.000	-4.000.000
Costo Ingeniero MBE	-9.000.000	0	0
Utilidad Antes de Impuesto	-13.083.921	15.499.715	19.121.833
Impuesto	0	-2.634.952	-3.250.712
Ahorro Impuesto	2.224.267	0	0
Utilidad Después de Impuesto	-10.859.654	12.864.763	15.871.121
Flujo de Caja Operacional	-10.859.654	12.864.763	15.871.121
VAN (Rf = 6%)	\$ 14.530.336		

Cuadro 11: Flujo de Caja con Aumento del MOB por Cliente de 5%

Concepto/Año	2008	2009	2010
MOB incremental	13.193.466	32.499.525	38.536.388
Costos Desarrollo	-12.000.000	-4.000.000	-4.000.000
Costo Ingeniero MBE	-9.000.000	0	0
Utilidad Antes de Impuesto	-7.806.534	28.499.525	34.536.388
Impuesto	0	-4.844.919	-5.871.186
Ahorro Impuesto	1.327.111	0	0
Utilidad Después de Impuesto	-6.479.424	23.654.606	28.665.202
Flujo de Caja Operacional	-6.479.424	23.654.606	28.665.202
VAN (Rf = 6%)	\$ 39.007.707		

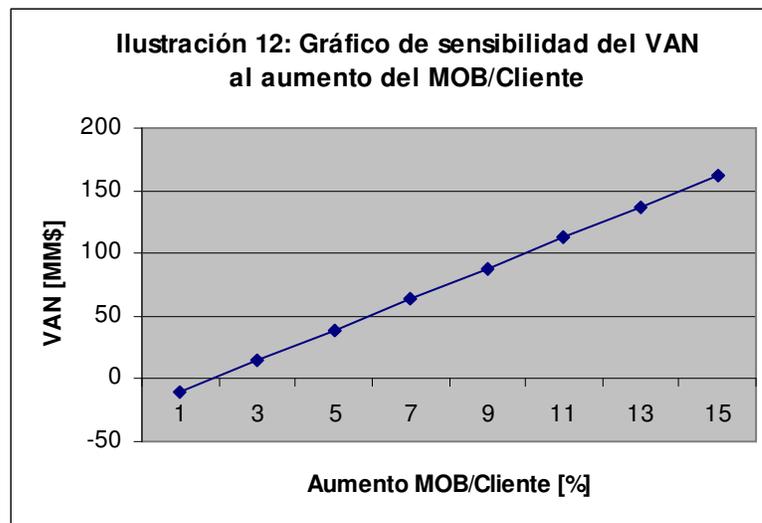
Cuadro 12: Flujo de Caja con Aumento del MOB por Cliente de 10%

Concepto/Año	2008	2009	2010
MOB incremental	26.386.931	64.999.050	77.072.775
Costos Desarrollo	-12.000.000	-4.000.000	-4.000.000
Costo Ingeniero MBE	-9.000.000	0	0
Utilidad Antes de Impuesto	5.386.931	60.999.050	73.072.775
Impuesto	-915.778	-10.369.839	-12.422.372
Utilidad Después de Impuesto	4.471.153	50.629.212	60.650.403
Flujo de Caja Operacional	4.471.153	50.629.212	60.650.403
VAN (Rf = 6%)	\$ 100.201.135		

Es decir, si se logra incrementar el MOB promedio por cliente en un 10%, el VAN del proyecto es de **\$100.201.135**.

6.6.4 Análisis de Sensibilidad

El análisis de sensibilidad busca mostrar la variación del VAN ante el incremento porcentual del MOB promedio por cliente. Se obtiene el gráfico de la ilustración 12:



La recta resultante para distintos incrementos del MOB, da cuenta que basta que un incremento del 2% del margen bruto por cliente, para que el proyecto se justifique económicamente.

Cabe mencionar que, por cada unidad porcentual que se incremente el MOB promedio por cliente, el VAN aumentará en 12 millones.

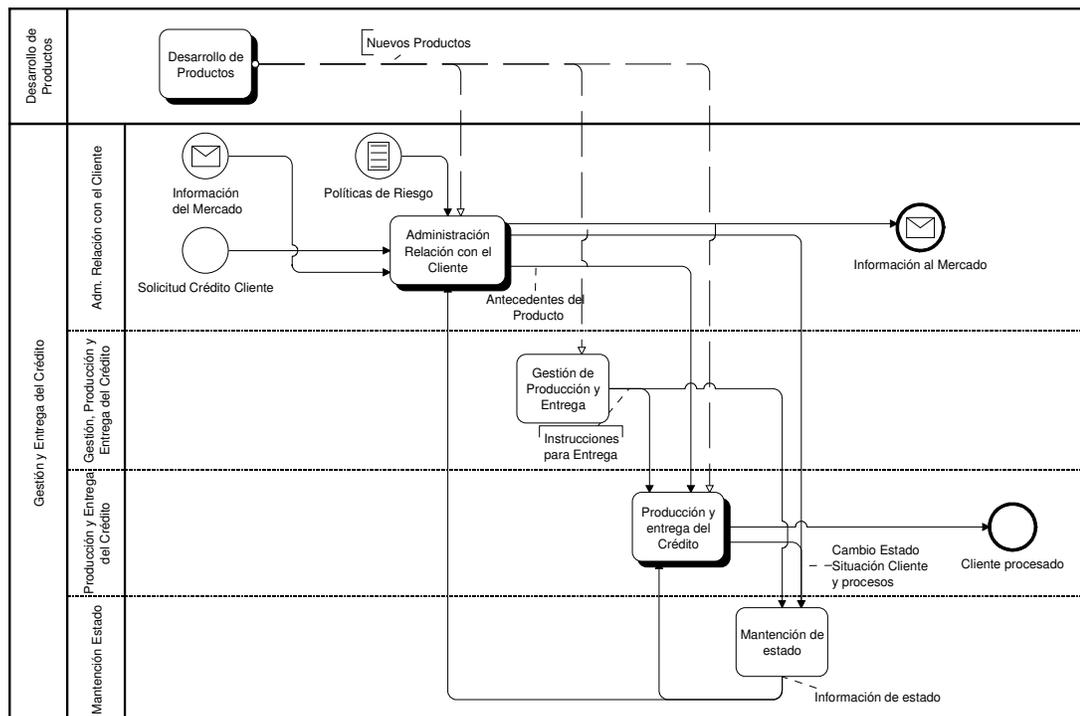
7. DISEÑO DEL PROCESO DE EVALUACIÓN Y ANÁLISIS DE CLIENTES

El diseño de los procesos de evaluación y análisis de clientes, es una particularización de los patrones de procesos de negocios (Barros, 2000) y ha sido modelado con el software iGrafx®, utilizando la notación estándar de procesos de negocio BPMN.

7.1 Modelamiento BPMN del Rediseño

7.1.1 Macroprocesos involucrados

Ilustración 13: Macroprocesos involucrados



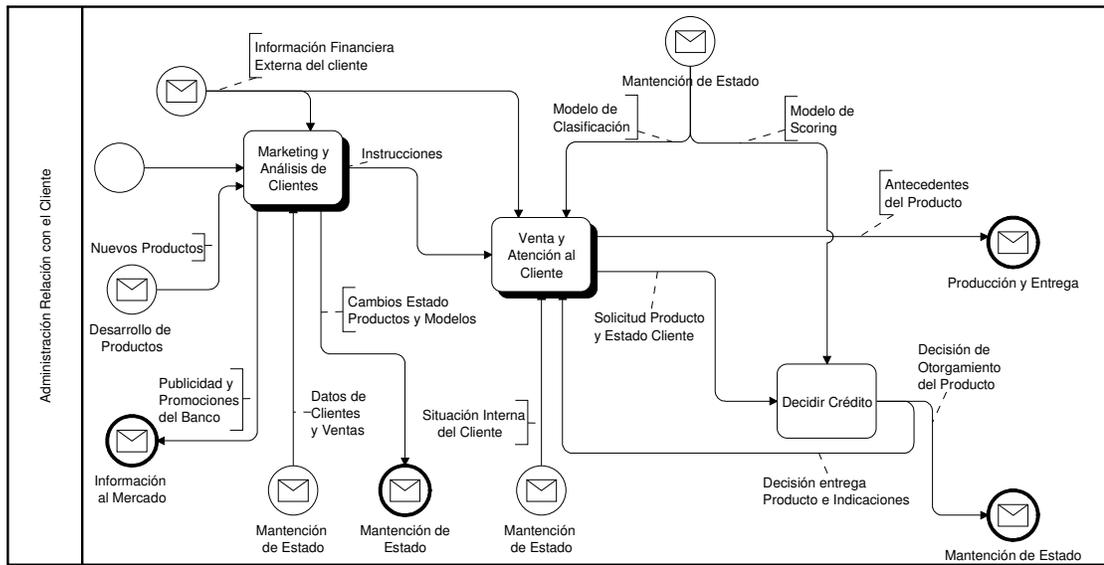
Este proceso se desarrolla en el macroproceso “Gestión y Entrega del Crédito”, que a su vez contiene los procesos “Administración Relación con el Cliente”, “Gestión de producción y entrega del Crédito”, “Producción y entrega del Crédito” y “Mantención de Estado”.

En el primero de los cuatro procesos mencionados, es donde tomará lugar el proyecto de esta tesis. Este proceso necesita de entrada la normativa de crédito, que determinará qué clientes serán aceptados, e información del mercado, principalmente relativa a la situación financiera del cliente. La salida es información al mercado del cliente, cuya situación financiera varía si el crédito fue otorgado, y los antecedentes del producto, para que “Producción y entrega del Producto” haga efectiva la venta.

El otro macroproceso involucrado es “Desarrollo de Productos”, que provee de nuevos productos a la oferta que se hace al cliente.

7.1.2 Administración Relación con el Cliente

Ilustración 14: Proceso Administración Relación con el Cliente



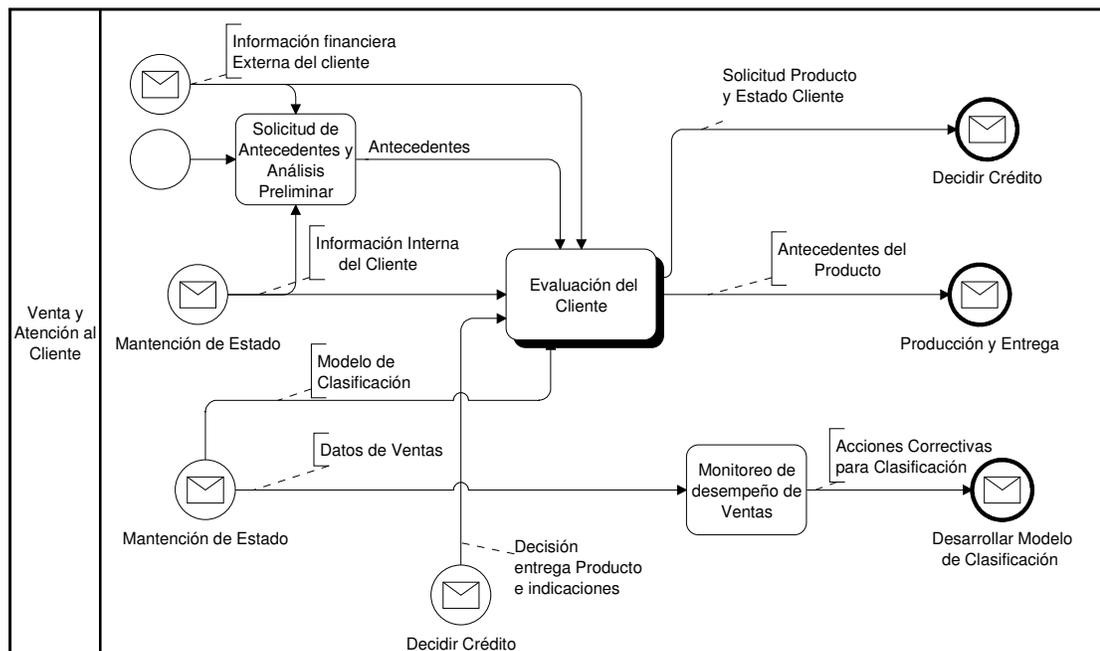
Dentro de este proceso se identifica “Marketing y Análisis de Clientes”, “Venta y Atención al Cliente” y “Decidir Crédito”. En este último proceso es donde se encuentra la instancia de aprobación del crédito, que basará su decisión en la información contenida en la evaluación y *scoring* del cliente.

La filial de Microempresas no posee un departamento de Marketing, y cualquier iniciativa de promoción o publicidad, debe ser presentada a la Gerencia Comercial. Por otro lado, el análisis de los clientes es un proceso que no existe y será definido, especialmente para la mantención del modelo de *scoring*.

El proceso de Venta recibe las instrucciones de Marketing (Gerencia Comercial), y la situación financiera externa e información interna del cliente. Aquí se clasificará al cliente y evaluará de acuerdo al segmento que pertenece, para luego enviar la solicitud del producto a la instancia de aprobación, adjuntando el estado del cliente.

7.1.3 Venta y Atención al Cliente

Ilustración 15: Proceso Venta y Atención al Cliente



Como lo muestra la ilustración 15, dentro de “Venta y Atención al Cliente”, lo primero que se realiza es la “Solicitud de Antecedentes y Análisis

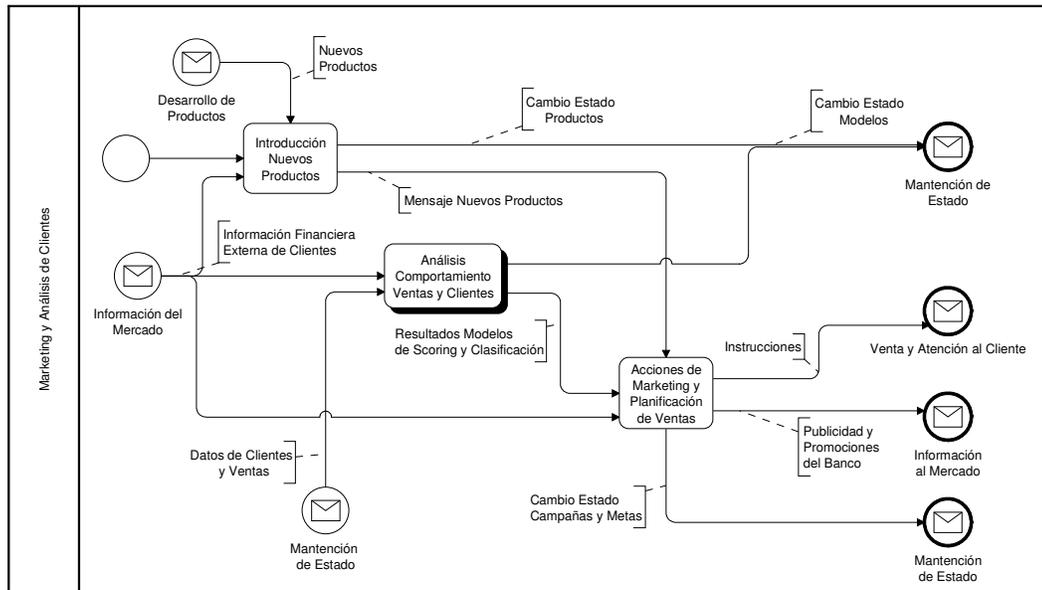
Preliminar”, que corresponde a la petición, obtención y análisis de los documentos requeridos para el producto que se está solicitando. Éstos pueden ser proporcionados por el cliente u obtenidos externamente, como el informe de DICOM. Aquí, se eliminará del proceso a aquellos clientes que, dado su perfil, no podrán acceder a un crédito. Luego, una vez que el ejecutivo ha decidido que el cliente es sujeto de crédito, deberá ingresar sus datos personales al sistema, o actualizarlos en caso que sea antiguo.

A continuación, se procede con la “Evaluación del Cliente”, que recibe los antecedentes del proceso anterior, y utiliza el modelo de clasificación, para clasificar centralizadamente al cliente que está siendo atendido. Una vez que el cliente ha sido segmentado y derivado a la evaluación de su perfil, se procede a llenar la solicitud correspondiente, a través de la cual se determinará el producto y monto a proponer a la instancia de aprobación.

Pero la evaluación individual del cliente no es suficiente, por lo que se incluye el proceso “Monitoreo de Ventas”, que es donde se analiza el desempeño de los créditos de los distintos segmentos. En caso que se requiera una acción correctiva del modelo de clasificación, la necesidad es enviada a “Desarrollo de Modelo de Comportamiento de Ventas”, que es donde se rediseña la lógica que este modelo utiliza.

7.1.4 Marketing y Análisis de Clientes

Ilustración 16: Proceso Marketing y Análisis de Clientes



Como se especifica en la ilustración 16, el primer proceso es “Introducción de Nuevos Productos”, que recibe los nuevos productos del macroproceso “Desarrollo de Nuevos Productos”, y la información del mercado. Los detalles de los productos desarrollados son enviados a “Acciones de Marketing y Planificación de Ventas”.

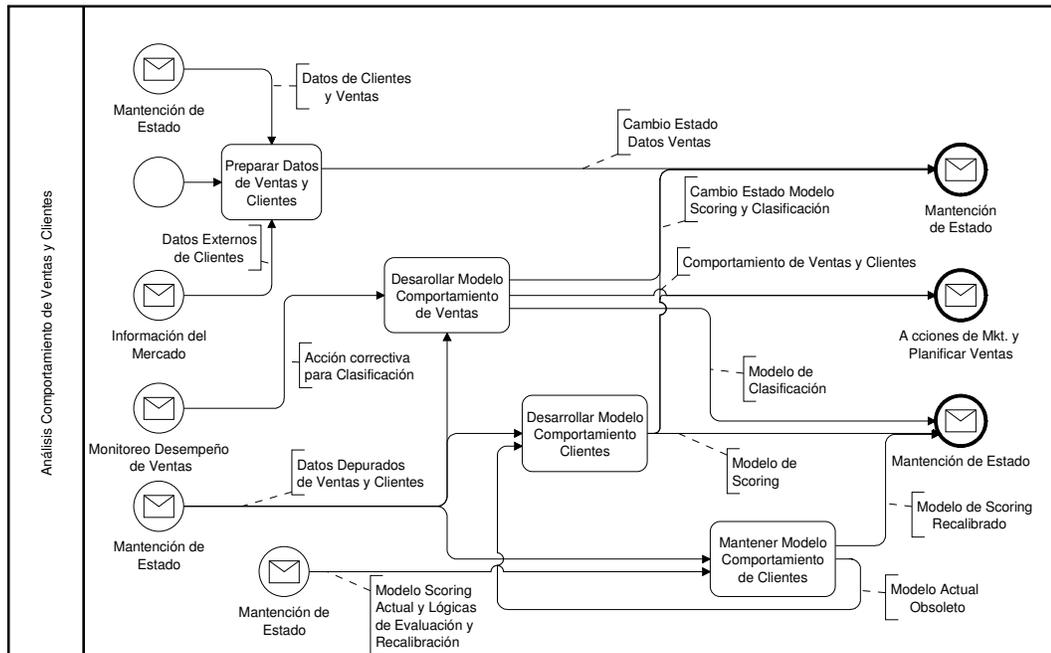
El segundo proceso, corresponde a “Análisis Comportamiento de Ventas y Clientes”, y es donde se realizará la parte del proyecto que incluye el desarrollo y actualización del modelo de clasificación, y la mantención del

scoring. Como ya se ha mencionado, el desarrollo del *scoring* fue externalizado, por lo que sólo se deja propuesto en la estructura de procesos.

Del último proceso, “Acciones de Marketing y Planificación de Ventas”, salen las instrucciones que van a “Venta y Atención al Cliente. Por otro lado, al mercado se envían las promociones y publicidad. Este proceso se alimentará de información generada consecuencia del proyecto, pero no forma parte de él.

7.1.5 Análisis Comportamiento de Ventas y Clientes

Ilustración 17: Proceso Análisis Comportamiento de Ventas y Clientes



El proceso “Preparar Datos de Ventas y Clientes”, tiene como fin consolidar la información desde múltiples fuentes de datos (internas y externas) en una base de datos, que permita el desarrollo y mantenimiento de los modelos de comportamiento de ventas y clientes. Dicha información, es almacenada en Mantención de Estado, para su posterior uso.

En lo siguiente, el proceso “Desarrollo del Modelo de Comportamiento de Ventas” recibe las acciones correctivas propuestas desde “Monitoreo de Desempeño de Ventas”, para analizar si cambiar parámetros de la clasificación de clientes, o bien realizar alguna modificación de las variables de la lógica de clasificación. Además, este proceso recibe los datos depurados desde mantención de estado, para realizar dichos cambios al modelo de clasificación, en caso que sea necesario. Sus salidas son la información sobre el comportamiento de las ventas, a “Acciones de Marketing y Planificación de Ventas”, y la actualización del modelo mismo, a “Mantención de Estado”, sólo si corresponde.

El proceso “Desarrollo del Modelo Comportamiento de Clientes” recibiría los datos depurados de los clientes, y tendría como salida el nuevo modelo creado a “Mantención de Estado”. Nótese que este proceso también tendría

como entrada un mensaje del proceso “Mantener Modelo de Clientes”, cuando es necesario cambiarlo por obsolescencia.

Por su parte, el proceso “Mantener Modelo de Comportamiento de Clientes” tiene como entrada los datos depurados de los clientes, que deberán ser definidos por el analista, para determinar si el modelo actual requiere recalibración. De ser así, también se requerirá que el analista ingrese un set de datos, que puede o no ser igual al anterior. Si el modelo resulta estar obsoleto, entonces se enviará un mensaje a “Desarrollo de Modelo de Comportamiento de Clientes”.

7.1.6 Diagramas de Pista

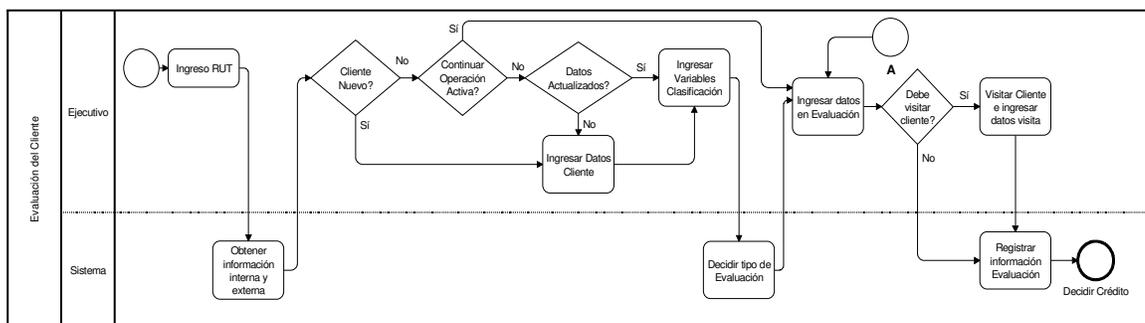
En estos diagramas se describirá todas actividades del último nivel de cada proceso incluido en el proyecto. Como los seudocódigos, asociados a las operaciones que deba realizar el sistema, serán detallados posteriormente en el informe, sólo serán referenciados en la actividad correspondiente. Éstos servirán de base para una eventual automatización futura.

Sin perjuicio de lo anterior, dado que el proceso de mantención del modelo de comportamiento de clientes requerirá conexión con un paquete estadístico (solicitud de funciones, intercambio de información, etc.), sólo se citará la funcionalidad requerida y la respuesta recibida, pues su automatización dependerá exclusivamente del paquete utilizado.

Cuando la actividad esté designada a un actor distinto al sistema, se especificará los análisis que éste debe realizar, cuyos conceptos asociados se encuentran en el Marco Teórico.

7.1.6.1 Evaluación del Cliente

Ilustración 18: Proceso Evaluación del Cliente



Como lo detalla la ilustración 18, lo primero que se realiza al evaluar un cliente es ingresar su RUT. Con éste, el sistema procederá a obtener su

información externa (DICOM, SBIF, etc.) e interna (para clientes antiguos, etc.). Si el cliente es nuevo, entonces el ejecutivo debe registrar sus datos personales.

En caso que el cliente sea antiguo, el ejecutivo tiene la opción de modificar una solicitud de crédito que se encuentre pendiente, ya sea porque aún no la ha completado, o porque sus condiciones deben ser cambiadas. Esta decisión corresponde a “continuar operación activa”. Si se elige esta opción, se pasa directamente a ingresar (o modificar) los datos de la evaluación anterior. Aquí no se realiza la clasificación del cliente, es porque ésta ya fue hecha anteriormente.

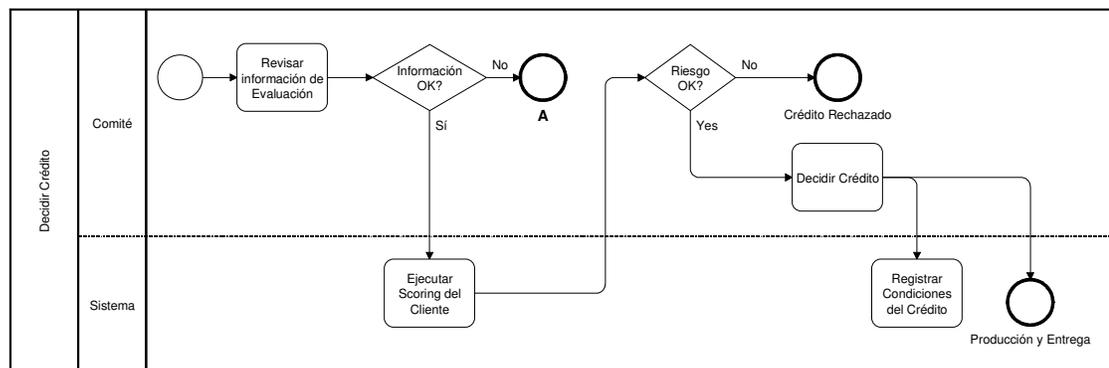
En cualquier otro caso, el ejecutivo deberá ingresar los datos del cliente que permitirán al sistema aplicar la lógica de clasificación. El *Seudocódigo N°1*, detalla el modelo de clasificación.

Finalmente, el ejecutivo deberá ingresar los datos que la evaluación requiera, que puede incluir los datos de una visita posterior al cliente, con el fin de corroborar que la situación de la microempresa coincide con la entregada por éste. La entrada de ‘A’ ocurre cuando la instancia de aprobación detectó información faltante o mal ingresada en la solicitud, la cual deberá ser corregida

por el ejecutivo. Toda la información de la evaluación es registrada en el sistema y enviada la actividad “Decidir Crédito”.

7.1.6.2 Decidir Crédito

Ilustración 19: Proceso Decidir Crédito

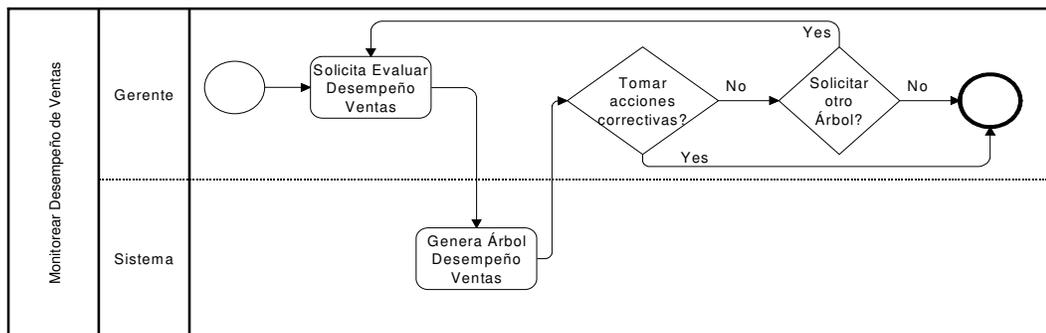


Primero, el Comité revisa que toda la información esté ingresada correctamente y los documentos requeridos del cliente estén adjuntos. En caso de existir un dato erróneo o faltante, éste deberá ser corregido por el ejecutivo. Si toda la información está correcta, se ejecutará y desplegará el *scoring* del cliente, que determinará su nivel de riesgo. Si éste no es adecuado, considerando tanto el resultado del *scoring* como los datos de la evaluación, el Comité deberá revisar las condiciones del crédito, evaluar alternativas y decidir si otorgarlo o no. Por otro lado, si el nivel de riesgo es adecuado al producto

que se está solicitando, se otorga el crédito. Luego, la información y condiciones del crédito son registradas en el sistema.

7.1.6.3 Monitorear Desempeño de Ventas

Ilustración 20: Proceso Desempeño de Ventas



El Gerente solicita el desempeño de las ventas a los segmentos, para lo cual debe ingresar el rubro (comercial, agrícola o pesca artesanal) y el periodo en que se desea obtener los indicadores. Estos indicadores, que serán calculados por el sistema, son: número total de clientes, % de clientes buenos, % de clientes malos, rentabilidad y riesgo, y serán desplegados para cada nodo del árbol, para visualizar cómo se distribuyen sus valores en la ruta de clasificación. El *Seudocódigo N°2*, corresponde al cálculo de los indicadores mencionados.

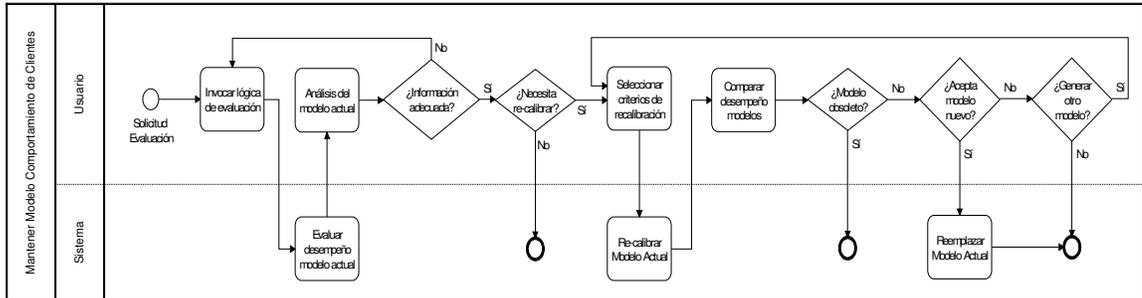
Con dichos indicadores, el Gerente decidirá si proponer acciones correctivas para el modelo de clasificación de clientes, a ser consideradas en el proceso de desarrollo del modelo. En caso que no decida enviar propuestas de acciones correctivas del modelo, puede generar otro árbol de desempeño.

Es importante mencionar que aquí no se corrige directamente el modelo de clasificación, porque éste es generado en base a juicio experto de todos los Gerentes y la decisión debe ser tomada en conjunto.

7.1.6.4 Mantener Modelo de Comportamiento de Clientes

Este Diagrama será expuesto más extensamente, porque requiere de análisis y cálculos más complejos, por parte del analista y del sistema, respectivamente. Es importante recalcar que la ejecución de este proceso requiere de una persona con conocimientos previos de Inteligencia de Negocios, dado que su correcta ejecución dependerá, en gran medida, del criterio utilizado en cada actividad. El Diagrama de Pistas se detalla en la ilustración 21:

Ilustración 21: Proceso Mantener Modelo Comportamiento de Clientes



7.1.6.4.1 Invocar Lógica de Evaluación

En la actividad “Invocar Lógica de Evaluación”, el analista deberá elegir el set de datos sobre el que desea probar el modelo actual de Scoring. Se sugiere hacer dicha selección, especificando un periodo de tiempo que contenga una cantidad de datos “suficiente” para el análisis del modelo. De forma adicional, se recomienda saber a priori cuál es la proporción de buenos y malos del set, ya que será un aspecto importante a considerar durante todo el proceso de re-calibración.

Se debe tener presente que los registros incluidos deben pertenecer al mismo universo de análisis definido para el proyecto de *scoring*.

7.1.6.4.2 Evaluar Desempeño del Modelo Actual

Esta actividad será descrita por etapas, para especificar claramente los objetivos de cada una de ellas, e indicadores y variables de desempeño a considerar. Las etapas son las siguientes:

1. Estabilidad Poblacional

- *Objetivo:* Determinar si la población evaluada con el modelo ha cambiado significativamente, basándose en las variables que éste utiliza para discriminar entre clientes buenos y malos.
- *Descripción:* Tal como se ha argumentado en el Marco Teórico, es posible establecer un intervalo de confianza para las medias poblacionales de cada una de las variables, a partir de una muestra (con la que se calibró el modelo). Este intervalo es obtenido gracias al establecimiento de una variación máxima para los valores de los coeficientes Beta.

Si se definen las variables:

\bar{x}_{it} = Media de la Variable i, de la muestra de la población del periodo t.

$\overline{x_{it-1}^{sup}}$ = Límite Superior permitido para la Media de la Variable i, obtenida de la muestra de la población del periodo anterior (t-1).

$\overline{x_{it-1}^{inf}}$ = Límite Inferior permitido para la Media de la Variable i, obtenida de la muestra de la población del periodo anterior (t-1).

La salida será una tabla que contenga los promedios $\overline{x_{it}}$, $\overline{x_{it-1}}$, las cotas $\overline{x_{it-1}^{inf}}$ y $\overline{x_{it-1}^{sup}}$, y una columna que especifica el estado de la variable (si dicha media se encuentra o no dentro del intervalo).

Con el fin de complementar el análisis, esta etapa termina con la confección de un gráfico por variable, que refleje la evolución temporal (mensual) de su media. El cálculo mensual de la media de cada variable, debe realizarse con los datos de los clientes que tengan crédito vigente ese mes, y que se encuentren dentro del Universo de estudio.

- *Lógicas de Negocio*: El *Seudocódigo N°3*, detalla el cálculo de las cotas inferiores y superiores, y si la media poblacional ha variado significativamente. El *Seudocódigo N°4* muestra cómo se debe construir el gráfico de la evolución promedio de cada variable.

2. Análisis Univariado

- *Objetivo:* Analizar si cada una de las variables le siguen siendo útiles al modelo para discriminar entre buenos y malos clientes.
- *Descripción:* Para ver la importancia de cada variable para el modelo, se deben realizar las pruebas estadísticas de Kolmogorov-Smirnov (K-S), para las variables continuas y ordinales, y de Chi-cuadrado de Pearson, para las categóricas. Como se ha explicado en el marco teórico, éstas tienen por objetivo determinar si existe alguna diferencia significativa entre los clientes buenos y malos, en la distribución de la variable en estudio. Ahora bien, como se necesita ver la relevancia que tienen las variables *para el modelo*, se debe utilizar como variable dependiente la predicción del modelo (“binarizada” de acuerdo al criterio de corte escogido).
- *Lógicas de Negocio:* El *Seudocódigo N°5*, muestra cómo realizar este análisis univariado.

En las etapas anteriores¹⁸ aún no se ha cuantificado la capacidad de predicción del modelo. Esto será realizado en las siguientes etapas.

¹⁸ Etapas basadas en: *Clasificación de Créditos mediante Técnicas avanzadas de Análisis de Datos para Banco Desarrollo* (2007).

3. Coherencia del Modelo

- *Objetivo:* Realizar una primera aproximación a cuánto concuerdan las predicciones del modelo con la realidad.
- *Descripción:* Para analizar la coherencia del modelo, primero se realiza un gráfico, que muestre la distribución de probabilidades que éste entrega, para las solicitudes de crédito que resultaron buenas y malas. Se espera que el modelo prediga probabilidades de pérdida más altas, para los clientes que efectivamente resultaron no pagadores, y más bajas para los pagadores¹⁹.

Segundo, y como una forma de resumir el gráfico anterior en un indicador, se construye el índice de concordancia. Este índice refleja cuánto mejor que el azar discrimina el modelo, independiente del punto de corte. Es una medida de coherencia, en el sentido que “compara si las probabilidades que asigna a los clientes buenos, son efectivamente menores que las que asigna a los clientes malos” (Weiss, 2006). Debido a lo anterior, también se dice que el “*C-Index* refleja el poder discriminante de la ecuación logística” (Garson, 2008).

¹⁹ *Clasificación de Créditos mediante Técnicas avanzadas de Análisis de Datos para Banco Desarrollo (2007).*

- *Lógicas de Negocio:* El *Seudocódigo N°6*, muestra cómo se debe generar del gráfico, y cómo calcular el índice de concordancia, para analizar la coherencia del modelo.

4. Capacidad Predictiva del Modelo

- *Objetivo:* Cuantificar la capacidad predictiva del modelo, dado el punto de corte escogido. Determinar si ésta ha cambiado, respecto de cuando el modelo fue calibrado.
- *Descripción:* Para cuantificar la capacidad predictiva del modelo, se debe comparar cuánto coincide su pronóstico con los resultados reales de los créditos, mediante la construcción de una matriz de confusión.
- *Lógicas de Negocio:* En el *Seudocódigo N°7*, se especifica la construcción de la matriz de confusión.

5. Alternativas de Punto de Corte

- *Objetivo:* Mostrar desempeños alternativos del modelo, variando el punto de corte.

- *Descripción:* Para el análisis de desempeños alternativos del modelo actual, se moverá el punto de corte (en intervalos de 0,05), y se calculará la variación de Utilidad asociada a la mayor o menor aceptación de malos y buenos clientes.

Los resultados deberán mostrarse en una tabla que incluya, para los distintos puntos de corte, la **variación** respecto del punto de corte actual, medida en: clientes buenos aceptados y malos rechazados, % de acierto global, % de efectividad, % de cobertura, utilidad (beneficio - costo) y utilidad relativa.

El acierto global, corresponde al porcentaje ponderado de clientes que el modelo predice correctamente. Cuando la proporción de clientes es distinta, se debe analizar su resultado cuidadosamente, ya que puede guiar a decisiones que dependen de la composición de la cartera (% de buenos y malos) más que de la calidad del modelo.

La efectividad, es el porcentaje de clientes que el modelo predice correctamente, sin importar la composición de la cartera. Es decir, pondera el acierto a los buenos y malos por un 50%. En ese sentido, no tiene el problema del acierto global, pero puede subestimar la capacidad predictiva real del modelo, ya que es poco común que la proporción de buenos y malos sea igual (o viceversa). Por eso, estos dos indicadores son complementarios y no sustitutos.

La cobertura, es el porcentaje de clientes que no son rechazados por el modelo. Se recomendó para analizar cómo variará el número de clientes atendidos con el cambio de corte, lo que repercutirá directamente en la participación de mercado de la empresa.

Finalmente, se generará un gráfico, que muestre la utilidad relativa para los distintos puntos de corte, permitiendo visualizar el punto que proporcionaría la utilidad máxima.

- *Lógicas de Negocio*: El *Seudocódigo N°8* detalla el cálculo del desempeño relativo al punto de corte actual.

7.1.6.4.3 Análisis del Modelo Actual

En esta actividad, se deben analizar los resultados obtenidos en las 5 etapas de la actividad anterior, para determinar si el modelo actual debe ser recalibrado. A continuación, se mostrará **una guía de base** para el análisis, que **debe** ser sometida al criterio y conocimientos del analista, para tomar la decisión de re-calibración. Para exponer más claramente los aspectos relevantes que deben ser considerados en el análisis, éstos serán expuestos nuevamente por etapas.

1. Análisis de Estabilidad Poblacional

- *Salida 1:* Tabla con las columnas $\overline{x_{it}}$, $\overline{x_{it-1}}$, cotas $\overline{x_{it-1}^{inf}}$ y $\overline{x_{it-1}^{sup}}$, y estado de $\overline{x_{it}}$, para todas las variables del modelo.

- *Análisis:* Expresado matemáticamente, se tiene que la media de **toda** variable de la muestra poblacional actual, ($\overline{x_{it}}$) debe cumplir la desigualdad siguiente:

$$\overline{x_{it-1}^{inf}} \leq \overline{x_{it}} \leq \overline{x_{it-1}^{sup}} \quad [14]$$

Es decir, la media de cada variable en la población actual, debe encontrarse dentro del intervalo generado con la muestra de la población, cuando se calibró el modelo que está siendo utilizado.

Luego, se tiene que:

- a) Si $\overline{x_{it}}$ se encuentra dentro del intervalo y cerca de la media del periodo anterior ($\overline{x_{it-1}}$), entonces el valor de la variable no ha cambiado significativamente desde que se calibró el modelo. Por lo tanto, la población se ha mantenido **estable** en esa variable.
- b) Si $\overline{x_{it}}$ se encuentra dentro del intervalo, pero más bien cerca de los límites, entonces esa variable deberá ser supervisada en análisis futuros.

- c) Si $\overline{x_{it}}$ se encuentra fuera del intervalo, significa que la población actual ha cambiado significativamente en esa variable. Por lo tanto, es probable que: se deba re-calibrar el Beta (la población se desplazó en el valor de la variable, pero ésta aún discrimina) o la variable ya no discrimine.

Sin prejuicio de lo anterior, se deberá verificar que cada una de las variables aún contribuya al modelo a discriminar entre clientes buenos y malos.

- *Salida 2:* Gráficos de Valor del Promedio de cada Variable en el tiempo.
- *Análisis:* Se debe estudiar la evolución del valor del promedio de cada variable, de forma de tener una perspectiva dinámica de su comportamiento, en términos de tendencias, aleatoriedad, etc.

2. Análisis Univariado

- *Salida 1:* Tabla que contiene, para cada variable, el valor del estadístico y p-valor.
- *Análisis:* Antes de comenzar el análisis, es importante recordar que el estadístico Z, del test K-S, sigue una distribución normal tipificada, y que el estadístico χ^2 de Pearson, sigue una distribución

chi-cuadrado con $k-1$ grados de libertad, donde 'k' es el número de categorías de la variable.

Luego, si para el estadístico de cada variable (denotada por el subíndice 'i'), se tiene que $\chi_{i,k-1}^2 > \chi_{k-1,1-\alpha}^2$, o bien que $Z_i > Z_{1-\alpha}$, entonces dicha variable le ayuda al modelo a discriminar los clientes, con un nivel de confianza de $(1 - \alpha)$. Se sugiere utilizar un $\alpha=0.05$, para tener mayor seguridad de la diferencia entre buenos y malos en esa variable. Bajo esa sugerencia, si el p-valor de la variable es menor a 0.05, se acepta que ésta contribuye al modelo a discriminar entre buenos y malos clientes.

Antes de rechazar el aporte de una variable al modelo (sobre todo cuando el p-valor está muy cerca del valor 0.05), es importante recordar que este análisis se hizo con la variable dependiente "binarizada", según el **punto de corte**. Por ello, en este caso, se recomienda hacer más pruebas para esa variable, variando levemente el punto de corte o aumentando el valor de α .

3. Análisis de Coherencia del Modelo

- *Salida 1*: Gráfico de predicción de probabilidades para clientes buenos y malos.

- *Análisis:* Un modelo “coherente”, debe predecir probabilidades menores para los clientes buenos, y mayores para los malos, por lo que sus distribuciones porcentuales (curvas) deberían estar más desplazadas hacia la izquierda (más cercana a cero) y derecha (más cercana a uno), correspondientemente. Si las curvas se confunden en demasía, sin distinguirse una separación clara entre buenos y malos, el modelo podría no estar discriminando los clientes satisfactoriamente. Para ver esto más claramente, se debe observar el valor del *C-Index*.

- *Salida 2:* Valor del Índice de Concordancia actual y de calibración.

- *Análisis:* El valor del índice varía entre 0.5 y el 1.0. Valores muy cercanos a 0.5, indican predicciones equivalentes al azar, por lo cual el modelo estaría siendo incoherente (poco discriminante), y podría ser reemplazado por el lanzamiento de una moneda. Por el contrario, “valores muy cercanos a 1.0, significa que el modelo discrimina casi perfectamente” (Harrell, 2001).

Utilizando el C-Index del modelo, cuando fue calibrado, se podrá comparar la variación de su capacidad discriminante, en el periodo seleccionado. Sea ésta superior o inferior, se deberá proceder con el análisis de su capacidad predictiva, en la etapa siguiente.

4. Análisis de Capacidad Predictiva del Modelo

- *Salidas:* Matrices de Confusión del modelo, para la muestra actual y de calibración, que incluyan el porcentaje de acierto y desacierto de predicción, en cada componente.
- *Análisis:* Con las matrices de confusión de ambas muestras, se podrá determinar si el desempeño actual del modelo (con los nuevos clientes), ha cambiado desde su calibración. Dado que las muestras para cada matriz son distintas, sólo tiene sentido hacer la comparación de forma porcentual.

Si se están aceptando más clientes malos, es posible que la política de punto de corte esté siendo poco rígida, y su valor deba ser disminuido. Por otro lado, si el porcentaje de buenos aceptados ha disminuido sustancialmente, es probable que deba optarse por un punto de corte mayor. Lo anterior es válido siempre y cuando el modelo sea coherente, ya que algunas variaciones pueden deberse a aumento de incoherencia del modelo.

5. Análisis de Puntos de Corte Alternativos

- *Salida 1:* Tabla con las columnas punto de corte, número de clientes buenos aceptados y malos rechazados, % de acierto global, % de efectividad, % de cobertura, utilidad y utilidad relativa.
- *Análisis:* Se debe tener presente que la utilidad relativa, es respecto a la política de corte actual. Por lo tanto, si la diferencia de utilidad es ' $U > 0$ ', se tiene que, si se hubiera utilizado ese punto de corte **en vez** del actual, se hubiera obtenido una ganancia ' U ' adicional.

Pero la elección del punto de corte no necesariamente debe basarse en resultados financieros. Los indicadores de acierto global, efectividad y cobertura, constituyen información adicional importante para decidir el mejor punto de corte *para la empresa*. Es decir, el criterio de elección deberá depender de la *estrategia* que ésta tenga. De esta forma, si la empresa desea aumentar su participación de mercado drásticamente, es probable que le sea preferible cambiar a un corte con menor utilidad relativa, pero que incremente su cobertura.

Aunque estos indicadores están pensados para complementar la decisión del mejor punto de corte, también pueden ser convenientes

de utilizar si no fuese posible atribuir un valor a la aceptación de clientes buenos y malos. En tal caso, se recomienda utilizar el mismo valor para el beneficio y el costo.

- *Salida 2:* Gráfico de Variación de Utilidad v/s Punto de Corte
- *Análisis:* Permite observar gráficamente los valores de la utilidad relativa, para efectos de escoger el punto de corte. Si el valor más alto (*peak* de la curva) es mayor que cero, es un indicio que ese punto de corte puede ser mejor.

El análisis de las salidas de las 5 etapas, en conjunto, permitirá al analista:

- Determinar si la información generada es adecuada, en el sentido que permita tomar una decisión de re-calibración (o no re-calibración) del modelo. Es decir, si el set de datos fue idóneo, en términos: temporales (¿El periodo fue el adecuado?), número de registros incluidos (¿Fueron suficientes registros?), representatividad de la muestra (¿Representa a la población?), confiabilidad y coherencia de la información mostrada (¿Tienen sentido los resultados?), etc.

En caso que la información sea inadecuada, deberá volver a la actividad “Invocar Lógica de Evaluación”, para reingresar otro set de datos. Si la información es adecuada, se debe tomar la decisión siguiente.

- Determinar si el modelo necesita ser re-calibrado. De ser así, el analista deberá pasar a la actividad “Seleccionar criterios de re-calibración”. De lo contrario, se mantiene el modelo actual, y el proceso concluye.

7.1.6.4.4 Seleccionar Criterios de Re-calibración

La re-calibración del modelo, refiere a cualquier cambio que se le realice a configuración actual, es decir, en términos de puntos de corte, supresión o adición de variables, nuevo cálculo de los coeficientes beta, etc.

Antes de seleccionar los criterios de re-calibración, el analista debe decidir qué tipo de re-calibración desea hacer. Se identifican 3 tipos: de Punto de Corte, con Selección Total de Variables y sin Selección de Variables. Cada

una de ellas, requerirá de distintas entradas (criterios), dado que se realizarán procedimientos diferentes.

1. Cambio del Punto de Corte

- *Descripción:* Es la re-calibración más simple, y sólo consiste cambiar el punto de corte que utiliza actualmente el modelo, por uno nuevo. Los elementos para tomar esta decisión, le fueron entregados al analista en la actividad “Análisis del Modelo Actual”.
- *Entrada:* Punto de Corte nuevo.

2. Selección Total de Variables

- *Descripción:* Consiste en seleccionar exactamente las variables, con las cuales se desea examinar si el modelo discriminaría mejor que con la estructura actual. Lo anterior, incluye el caso de utilizar las mismas variables que las actuales, para re-calcular los parámetros beta asociados.

Este tipo de re-calibración, se justifica en conocimiento, intuición, o nueva información que tenga el analista, y que considere que podría predecir mejor el comportamiento de los clientes.

- *Entradas:* Nivel de significancia α para el análisis univariado, número de iteraciones 'n' para ver estabilidad del modelo, beneficio de aceptar un cliente bueno y costo de aceptar uno malo.

3. Selección Estadística de Variables

- *Descripción:* Consiste en buscar, sólo en base a criterios estadísticos, relaciones interesantes entre variables, cuya inclusión le contribuya al modelo a predecir mejor el riesgo de los clientes. Sin embargo, se debe tener en cuenta que “pueden encontrarse relaciones que provengan únicamente del ruido existente en la muestra y sean difíciles de replicar en otras” (Menard, 2002). Por esta razón, “se debe someter a juicio experto, la coherencia de los resultados encontrados” (Hosmer & Lemeshow, 1989). De esta forma, este método tiene como objetivo apoyar al analista en la búsqueda del mejor modelo, y en ningún caso sustituirlo.

- *Entradas:* Niveles de significancia: α para análisis univariado, α' de entrada y α'' de salida para el método *stepwise*, tipo de *stepwise* que se desea hacer (*forward* o *backward*), número de iteraciones 'n' para analizar estabilidad del modelo, beneficio de aceptar un cliente bueno y costo de aceptar uno malo.

7.1.6.4.5 Re-calibrar Modelo Actual

Una vez que se ha seleccionado el tipo de re-calibración que se realizará, y se hayan ingresado las entradas requeridas, se procede a re-calibrar el modelo. A continuación, se describirá el procedimiento de cada una de ellas, referenciando las lógicas de negocio asociadas a cada etapa.

1. Re-calibración del Punto de Corte

Se reemplaza el punto de corte anterior, por el corte ingresado por el analista. En adelante, el modelo utilizará el nuevo punto de corte para discriminar clientes buenos y malos.

2. Re-calibración del Modelo, con Selección Total de Variables

Es posible dividir el procedimiento en las siguientes etapas:

- a) *Dividir el set de Datos en sets de Entrenamiento y Prueba*

Como se menciona en el Marco Teórico, se deberá particionar los datos, aleatoriamente, en 2 subconjuntos²⁰: el set de entrenamiento, que contendrá el 80% de los casos, y el set del prueba, con el 20% de los casos restantes. Para corroborar que estas particiones son adecuadas, se debe verificar que cada una tenga igual proporción de buenos y malos.

b) *Análisis Univariado*

Aunque “generalmente este análisis se utiliza para escoger variables candidatas al modelo” (Hosmer & Lemeshow, 1989), aquí se realiza de forma complementaria, a todas las variables escogidas por el analista. El seudocódigo para este análisis, es análogo al Seudocódigo N°5, y sólo difiere en que se utiliza como variable dependiente, el resultado real de las solicitudes (en vez de la predicción del modelo “binarizada”).

²⁰ *Clasificación de Créditos mediante Técnicas avanzadas de Análisis de Datos para Banco Desarrollo* (2007).

c) *Entrenamiento del Modelo Nuevo*

Consiste en balancear los registros del set de prueba adecuadamente, para luego proceder a calibrar los betas de las variables (con su estadístico, p-valor y desviación asociados) y calcular el R^2 de Nagelkerke y el estadístico G del modelo. El “balanceo”, consiste en ponderar los registros que tienen con la mayor frecuencia (malos o buenos), por un factor menor a 1.

El *Seudocódigo N^o 9*, especifica el balanceo de datos y entrenamiento del modelo.

d) *Comparar Modelo Actual y Nuevo*

Una vez que el modelo nuevo ha sido entrenado, se debe proceder a compararlo con el actual. Para ello, se predice la probabilidad de no pago para cada uno de los registros del set de prueba, utilizando ambos modelos. Dadas sus predicciones, primero se busca el punto de corte óptimo para

el modelo nuevo (en base a utilidad máxima)²¹. Luego, se construye la matriz de confusión y se calcula el índice de concordancia para los dos modelos. Adicionalmente, se genera una tabla con indicadores para los distintos puntos de corte del modelo nuevo, que le permita al analista corroborar la elección automática de dicho punto.

El *Seudocódigo N°10*, muestra la comparación entre ambos modelos, e incluye el cálculo de todos los elementos expuestos en esta etapa.

e) Evaluar Sobreajuste y Verificar Estabilidad del Modelo Nuevo

Una vez que el modelo nuevo ha sido calibrado, y se ha determinado el mejor²² punto de corte, se debe corroborar que no se encuentre sobreajustado y que sea estable. Para lo primero, es necesario verificar que la efectividad del modelo en el set de entrenamiento y prueba sea similar. Para ratificar

²¹ Esta parte de la etapa, será análoga a la presentada en el Seudocódigo N°8, y consistirá en mover el punto de corte del modelo nuevo, en intervalos de 0.05, para encontrar el que maximice su utilidad. Esto incluye, además, el cálculo de sus indicadores respectivos.

²² Por construcción, el “mejor” punto de corte, **depende** de los valores que se haya asignado al beneficio de aceptar un cliente bueno y costo de aceptar uno malo.

lo segundo, se debe probar que, si se hubiera tomado distintas muestras para entrenar y probar el modelo, el valor de sus parámetros beta y eficiencia de predicción, habrían sido “similares”.

Para determinar si el modelo es estable, se dividirá el set de datos (completo) ‘n’ veces²³ en muestras de entrenamiento y prueba, cuidando que la proporción de buenos sea la misma en ambas muestras. En cada iteración, se entrenará un modelo y se medirá su capacidad predictiva. Finalmente, se obtendrá la desviación estándar de cada parámetro beta y de la efectividad en buenos y malos, utilizando los resultados de todos los modelos calibrados. En caso que éstas sean pequeñas, se podrá concluir que el modelo es estable, en términos de sus variables y de su capacidad predictiva²⁴.

El *Seudocódigo N°11*, detalla el cálculo de la efectividad del modelo en ambos sets, para evaluar si está sobreajustado, y el *Seudocódigo N°12*, muestra los cálculos para verificar su estabilidad.

²³ Para cada iteración se utilizan todos los datos disponibles, es decir, se repondrán y volverán a utilizar todos los registros.

²⁴ *Clasificación de Créditos mediante Técnicas avanzadas de Análisis de Datos para Banco Desarrollo* (2007).

3. Re-calibración del Modelo, con Selección Estadística de Variables

Este procedimiento también será dividido en etapas, pero dada la similitud que algunas de ellas tendrán con las etapas de la re-calibración anterior, sólo serán citadas y no se describirán en detalle.

a) *Dividir el set de Datos en sets de Entrenamiento y Prueba*

Igual a la re-calibración anterior.

b) *Análisis Univariado*

“Este análisis se utiliza para escoger las variables candidatas a entrar el modelo” (Hosmer & Lemeshow, 1989). Esta etapa puede ser considerada como un filtro, para evitar que el método *stepwise* reciba variables que sólo generen ruido y no tengan relación con la variable objetivo. Sólo las que resulten significativas continuarán en el proceso.

c) *Análisis Multivariado*

Ahora que ya se tienen las variables que, por sí solas, más podrían contribuir a predecir el comportamiento de pago de los clientes, corresponde encontrar cuál es la mejor combinación de variables que, en conjunto, predigan más asertivamente dicho comportamiento.

Esta elección será hecha mediante el algoritmo *stepwise regression*, considerado útil para el contexto de identificar el set de variables que más contribuyan al modelo a predecir un fenómeno (Menard, 2002). Sin embargo, “los métodos *stepwise* no necesariamente identifican los mejores modelos, (...) pues corren el peligro de sobreajustarse al ruido propio de la muestra de entrenamiento” (Garson, 2008).

Como se señala en el marco teórico, el criterio de $\alpha=0.05$ puede ser restrictivo, por lo que se recomienda utilizar valores entre 0.15 y 0.20. Evidentemente, cuán estricto sea el criterio de significancia, depende de la cantidad de variables que se desea incluir en el procedimiento.

Dado que los paquetes estadísticos poseen métodos de *stepwise (forward y backward)*, no se adjuntará el seudocódigo para llevarlo a cabo.

Para finalizar esta etapa, se debe calcular la matriz de correlación para todas las variables que entraron al método, con el fin de generar una visión más completa sobre la calidad, utilidad y relación de las variables que fueron incluidas y excluidas del modelo.

d) Entrenamiento del Modelo Propuesto

Igual a la re-calibración anterior, pero utilizando el conjunto de variables seleccionadas en el análisis multivariado.

e) Comparar Modelo Actual y Propuesto

Igual a la re-calibración anterior, se encuentra el punto de corte óptimo para el modelo nuevo (en base al criterio de utilidad máxima), se calcula el índice de concordancia y la matriz de confusión de ambos modelos, y se genera una tabla

con indicadores de desempeño para los distintos puntos de corte del modelo nuevo.

f) *Evaluar Sobreajuste y Verificar Estabilidad del Modelo Nuevo*

Al igual que a re-calibración con selección total de variables. Para evitar el riesgo de sobreajuste, mencionado en el análisis multivariado, se compara la efectividad del modelo propuesto en el set de entrenamiento y prueba. Asimismo, se hacen los cálculos necesarios para inferir estabilidad del modelo.

7.1.6.4.6 Comparar Desempeño de los Modelos

Esta actividad sólo se realizará para las re-calibraciones con selección total y estadística de variables. Aquí, se debe decidir si el modelo actual será sustituido por el modelo nuevo. Por lo tanto, al igual que en la decisión de re-calibración del modelo actual, se mostrará una **guía de base** para tomar dicha decisión, pero que **debe** ser sometida al criterio y conocimientos del analista. Para exponer claramente los aspectos a considerar, éstos serán abarcados por etapas:

a) *Análisis Univariado*

- *Salida:* Tabla que contiene, para cada variable, el valor del estadístico y p valor.
- *Análisis:* Es muy similar al análisis univariado del modelo actual. Es decir, si el estadístico de cada variable 'i', se tiene que $\chi_{i,k-1}^2 > \chi_{k-1,1-\alpha}^2$, o bien, $Z_i > Z_{1-\alpha}$, entonces dicha variable le ayuda al modelo a discriminar los clientes, con un nivel de confianza de $(1 - \alpha)$.

Cuando la selección de variables candidatas sea hecha por el sistema, se debe tener en cuenta que valores pequeños de α (por ejemplo 0.05,) puede excluir variables relevantes para el modelo. Mickey y Greenland (citados en Hosmer & Lemeshow, 1989), recomiendan valores más altos (hasta 0.25) para este análisis, en caso que pocas o ninguna variable sean significativas a un nivel de 0.05. Por esta razón es importante "revisar críticamente cada variable añadida al modelo, antes de decidir el modelo final" (Hosmer & Lemeshow, 1989).

b) *Análisis Multivariado (sólo para selección estadística de variables)*

- *Salida 1:* Tabla que contiene las variables que se incluyeron en el modelo (con los Betas, estadísticos y p-valores asociados), y otra tabla que contiene las que se excluyeron (con sus p-valores correspondientes).

- *Descripción:* La elección de un conjunto de variables, en base a criterios estadísticos, tiene como ventaja el descubrimiento de nuevas variables que predigan el comportamiento de pago. Aunque el modelo propuesto no mejore el desempeño del modelo actual, es posible encontrar relaciones novedosas entre variables, que podrían ser consideradas en una re-calibración con selección total de variables, que se realice posteriormente.

No se debe olvidar que la elección del nivel de significancia, determina cuántas variables entrarán, eventualmente, al modelo, por lo que elecciones muy estrictas podrían dejar fuera variables significativas para la predicción del fenómeno en estudio. Por ello, es importante observar la tabla de las

variables que fueron excluidas del modelo, y analizar la posibilidad de su inclusión.

- *Salida 2:* Matriz de correlación de todas las variables ingresadas al método *stepwise*.

- *Descripción:* A través de la matriz de correlación, es posible determinar si las variables escogidas por el método *stepwise*, realizan un aporte de calidad y carente de ruido al modelo. Mientras más alta sea la correlación entre variables, mayor es el ruido que generan. Por lo tanto, lo “deseable” es que las correlaciones sean menores a 0,3.

c) *Entrenamiento del Modelo Nuevo*

- *Salida 1:* Tabla que contiene en sus columnas, el valor de los Betas de cada variable, con su correspondiente desviación estándar, estadístico y p-valor asociado.

- *Descripción:* Esta salida busca mostrar la importancia de las variables, al actuar en conjunto, y que se refleja en el p-valor de los Betas. “Cada coeficiente beta, representa el cambio en el logit, cuando la variable independiente cambia en una

unidad” (Hosmer & Lemeshow, 1989). En base a eso, es posible establecer a qué variables éste es más sensible.

- *Salida 2:* R^2 de Nagelkerke y *Likelihood Ratio Test G* (con su nivel de significancia), del modelo actual y nuevo.

- *Descripción:* El R^2 de Nagelkerke, puede ser interpretado como cuánto reduce la variación del modelo (con sólo una constante), la inclusión de nuevas variables. Aunque la comparación del R^2 de dos modelos no es directa, si el R^2 del modelo nuevo es drásticamente menor que el actual, significa que existe una variación en los datos, que el modelo nuevo no puede reducir en la muestra nueva, y que el modelo actual si podía reducir en la muestra en la que se calibró.

Si ‘G’ es significativo, quiere decir que el modelo se ajusta bien a los datos. Es importante considerar que, “si la muestra es lo suficientemente grande, incluso pequeñas asociaciones pueden resultar estadísticamente significativas, por lo que G no permite, por sí solo, concluir la fuerza predictiva del modelo” (Harrell, 2001). Adicionalmente, el modelo puede estar sobreajustado. Por ello, se deberá evaluar la capacidad predictiva del modelo y el sobreajuste, en las etapas posteriores.

d) *Comparar Modelo Actual y Propuesto*

- *Salida 1:* Índice de Concordancia de ambos modelos.
- *Descripción:* El modelo que tenga el índice con valor más cercano a uno, significa que discrimina mejor los clientes, en el set de prueba.

- *Salida 2:* Matrices de confusión de ambos modelos.
- *Descripción:* En las matrices de confusión, se podrá comparar la habilidad predictiva de ambos modelos, medida en el porcentaje de efectividad de cada uno. La diferencia de las utilidades, permitirá determinar cuál habría sido la ganancia de haber usado el modelo nuevo en vez del actual, por lo que constituye un criterio importante de decisión para su eventual reemplazo.

- *Salida 3:* Tabla con indicadores de acierto global, efectividad, cobertura y utilidad (no relativa) para los distintos puntos de corte del modelo nuevo.
- *Descripción:* Esta salida es tan importante como las anteriores, pues proporciona una visión general de la capacidad predictiva del nuevo modelo, al mostrar diversos

indicadores para los distintos puntos de corte. Si bien es cierto que se propone la elección del corte que genere la utilidad máxima, también se ha mencionado que es posible que este criterio no se condiga con la estrategia de la empresa, por lo que esta tabla proporcionará al analista, la información necesaria para corroborar el punto escogido o modificarlo.

g) Evaluar Sobreajuste y Verificar Estabilidad del Modelo Nuevo

- *Salida 1:* Efectividad del modelo nuevo, en el set de entrenamiento y de prueba.

- *Descripción:* Las efectividades en ambos sets deben ser similares, o ligeramente mayores o menores. Si hay diferencias muy grandes (por ejemplo, superior al 10%), existe sobreajuste del modelo, y su capacidad de predicción será pobre.

- *Salida 2:* Tabla que posee en sus columnas, los valores del parámetro de cada variable y efectividad del modelo, y en sus filas, las 'n' iteraciones más el promedio y desviación de los valores en las columnas (betas y efectividad).

- *Descripción:* La baja desviación estándar de la efectividad, permite concluir que el modelo es estable, en términos de su capacidad predictiva. Si es muy alta, es difícil asegurar que la efectividad será similar en el futuro, por lo que la sustitución del modelo no es directa.

La baja desviación de cada variable, permite confirmar su relevancia en la discriminación de clientes. Se fuese muy alta, su aporte no necesariamente será útil en el futuro, por lo que deberá revisarse minuciosamente su inclusión en el modelo.

En caso que el modelo actual no esté discriminando los clientes de acuerdo a las expectativas de la empresa, y que no sea posible mejorar su desempeño satisfactoriamente mediante alguna re-calibración, se dirá que el modelo está obsoleto. En ese caso, se debe requerir asesoría experta, que identifique las razones de obsolescencia y desarrolle, si es posible, un modelo de predicción distinto a regresión logística. La salida de este proceso será, entonces, la necesidad de generar otro modelo de comportamiento de clientes.

7.1.6.4.7 Reemplazar Modelo Actual

En caso que el analista decida que el modelo nuevo no es mejor que el actual, puede generar otro modelo, bajo distintos criterios. En caso que decida que el modelo nuevo sí debe reemplazar al actual, se debe almacenar en el sistema:

- Coeficientes beta, con desviación estándar y significancia.
- R^2 de Nagelkerke
- Likelihood Ratio Test G
- Índice de concordancia
- Punto de corte
- Efectividad de predicción y matriz de confusión

El proceso termina con el modelo de regresión logística re-calibrado.

7.2 Lógicas de Negocio

A continuación, se expondrán los pseudocódigos que contienen la lógica de negocio elaborada para los procesos de “Evaluación del Cliente” y “Mantenimiento del Modelo de Comportamiento de Clientes”.

7.2.1 Evaluación del Cliente

7.2.1.1 Seudocódigo N°1: Modelo de Clasificación

Este modelo fue desarrollado en conjunto con la Gerencia de Banderrollo Microempresas, y se basa en juicio experto para segmentar a los clientes. Pese a que la segmentación podría haber sido hecha con técnicas de Minería de Datos, la opción de hacerla cualitativamente fue tomada por la Gerencia misma. Elseudocódigo de la segmentación, puede ser resumido en un árbol de decisión.

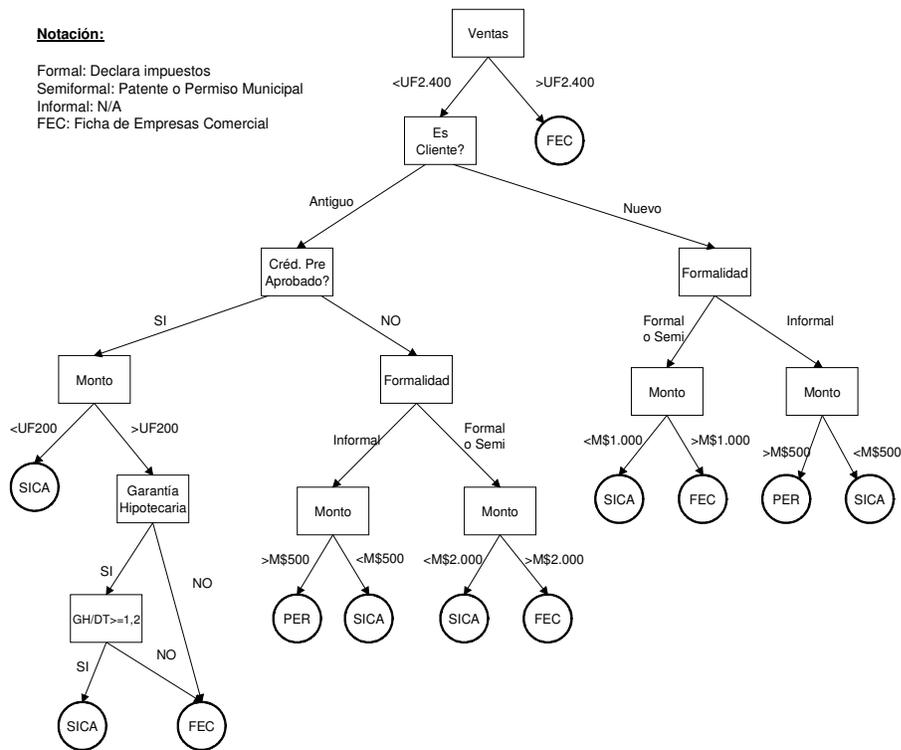
El primer criterio de división del árbol es el rubro del cliente. Por simplicidad, se presentarán 3 árboles, uno correspondiente a cada rubro: comercial, silvoagropecuario y pesca artesanal. Los nodos finales contienen las evaluaciones a las que se redirigirá al ejecutivo. Éstas son: SICA, Ficha de Empresas y PER.

1. Árbol Comercial

Éste es el árbol más general, pues considera todos los rubros de los clientes, salvo el de pesca artesanal y silvoagropecuario.

Tal como se puede observar las variables de segmentación son: nivel de ventas anuales (UF), cliente antiguo o nuevo, grado de formalidad, si posee un crédito pre-aprobado, si posee garantía hipotecaria (más su valor) y monto a solicitar (en M\$). Además, se estableció la razón $GH/DT \geq 1,2$, donde GH es garantía hipotecaria y DT es la deuda total del cliente (interna + externa), incluyendo la nueva. Este árbol se muestra en la ilustración 22.

Ilustración 22: Árbol Comercial



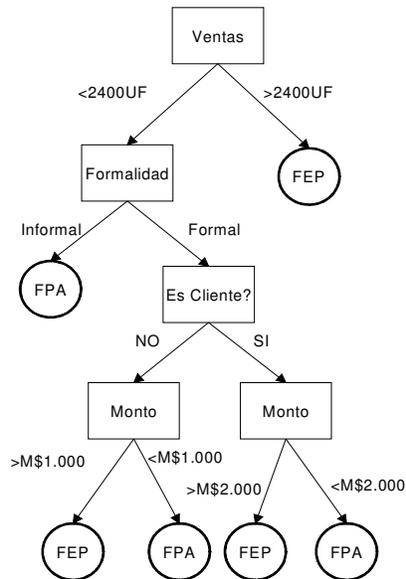
2. *Árbol de Pesca Artesanal*

En este árbol, se clasifica por nivel de ventas anuales, formalidad, si es cliente antiguo o nuevo y el monto a solicitar. Como se aprecia en la ilustración 23, si bien es cierto que este árbol es mucho más pequeño que los de los otros rubros, ello se justifica en que la cantidad de clientes pesqueros es mucho menor.

Ilustración 23: Árbol de Pesca Artesanal

Notación

FORMAL: Iniciación Actividades
SEMIFORMAL: Concesión Marítima
INFORMAL: N/A
FPA: Ficha de Pesca Artesanal
FEP: Ficha de Empresas Pesquera



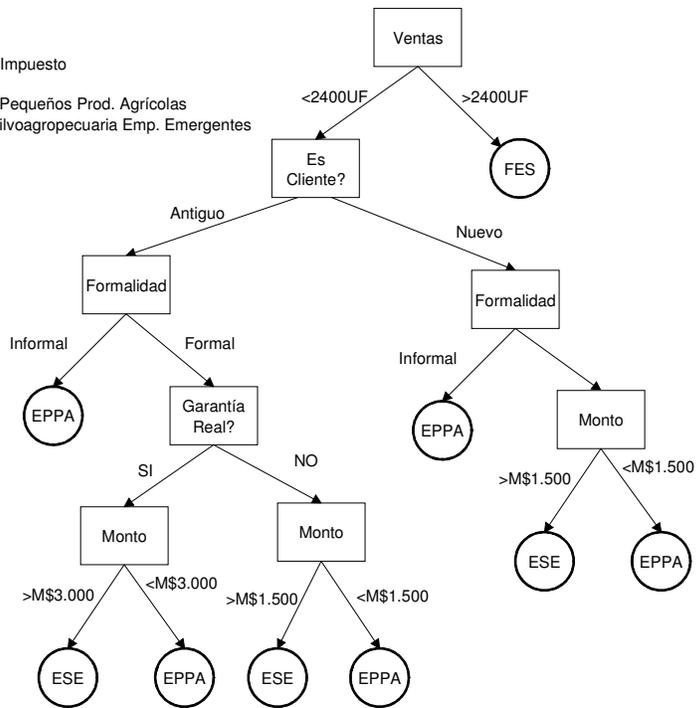
3. Árbol Silvoagropecuario

No se incluyen variables de clasificación, adicionales a las mencionadas anteriormente. La lógica de este árbol, se detalla en la ilustración 24:

Ilustración 24: Árbol Silvoagropecuario

Notación

FORMAL: Declara Impuesto
INFORMAL: N/A
EPPA: Evaluación Pequeños Prod. Agrícolas
ESE: Evaluación Silvoagropecuaria Emp. Emergentes



7.2.1.2 Seudocódigo N°2: Medición de Desempeño de los Segmentos

Esteseudocódigo recorre el árbol nodo por nodo, calculando en cada uno de ellos los indicadores de desempeño mencionados en el proceso “Monitoreo de Desempeño de Venta”.

- Método para recorrer el Árbol

```
Recorrer todos los nodos del árbol (tipo_arbol){
  Niveles_Arbol=Consultar niveles para el árbol (tipo_arbol);
  Mientras Niveles_Arbol>0
    For(j=0; j en número de nodos por nivel, j++){
      Calcular indicadores nodos j del Nivel_Arbol i;
    }
    Niveles_Arbol - - ;
  Fin Mientras
}
```

- Calcular Indicadores en Nodo j (Cualquier Nodo distinto de la Raiz)

```
Seleccionar todos los clientes que pasaron por ese nodo (a través de la
Lógica);
Pasar Clientes desde la tabla a Lista_Clientes_Nodoj;
Mientras haya Clientes en Lista_Clientes_Nodoj
  Si (Cliente.Comp_Pago <= 30 días de mora promedio)
    Clientes_Buenos_Nodoj++;
  Fin Si
  Si (Cliente.Comp_Pago >= 30 días de mora promedio)
    Clientes_Malos_Nodoj++;
  Fin Si
  Rentabilidad_Nodoj= Rentabilidad_Nodoj + Cliente.Rentabilidad;
  Clientes_Nodoj++;
Fin Mientras
```

- Calcular Indicadores en Primer Nodo (Raíz del Árbol del Rubro)

Pasar todos los clientes desde la tabla a Lista_Clientes (del rubro que corresponda)

Mientras haya Clientes en Lista_Clientes

Si (Cliente.Comp_Pago <= 30 días de mora promedio)

 Clientes_Buenos_Primer_Nodo++;

Fin Si

Si (Cliente.Comp_Pago >= 30 días de mora promedio)

 Clientes_Malos_Primer_Nodo++;

Fin Si

Rentab._Primer_Nodo= Rentab._Primer_Nodo + Cliente.Rentab.;

Clientes_Totales++;

Fin Mientras

7.2.2 *Mantenimiento Modelo de Comportamiento de Clientes*

Los seudocódigos asociados a la mantención del modelo de *scoring*, son las siguientes:

7.2.2.1 *Seudocódigo N°3: Variación de Media Poblacional*

La primera parte, corresponde al cálculo de los límites inferior y superior permitidos para la media de cada variable. La segunda parte, utiliza estos límites para determinar si la media actual de cada variable, se encuentra en ese rango.

- Calcular límites superior e inferior de la media de cada variable

Obtener Promedio ($\overline{x_{it-1}}$), Beta (β_{t-1}) y Desv. Estándar de Beta ($\sigma_{\beta_{t-1}}$), para todas las Variables del Modelo (de cuando el modelo fue calibrado);
Obtener Set de Datos ingresado por el usuario;

//Se declaran las variables

double $\beta_{max} = 0$; //Valor máximo que puede tomar β_{t-1}

double $\beta_{min} = 0$; //Valor mínimo que puede tomar β_{t-1}

double deltaMax = 0; //Aumento porcentual máximo de β_{t-1}

double deltaMin = 0; //Aumento porcentual mínimo de β_{t-1}

double $\overline{x_{it-1}^{inf}} = 0$; //Valor del límite inferior de la variable $\overline{x_{it-1}}$

double $\overline{x_{it-1}^{sup}} = 0$; //Valor del límite superior de la variable $\overline{x_{it-1}}$

//Se calculan los límites del promedio de cada variable

For (i en {variables del modelo}; i++) {

$\beta_{max} = \beta_{t-1} + 2 * \sigma_{\beta_{t-1}}$;

```


$$\beta_{min} = \beta_{t-1} - 2 * \sigma_{\beta_{t-1}};$$


$$\text{deltaMax (\%)} = (\beta_{max} / \beta_{t-1}) * 100;$$


$$\text{deltaMin (\%)} = (1 - \beta_{min} / \beta_{t-1}) * 100;$$


$$\overline{x_{it-1}^{sup}} = \overline{x_{it-1}} * \text{deltaMax};$$


$$\overline{x_{it-1}^{inf}} = \overline{x_{it-1}} * \text{deltaMin};$$

Almacenar valores de  $L_{sup}$  y  $L_{inf}$ ;
}
/*Estos límites pueden estar previamente calculados, en cuyo caso el
cálculo también se realizaría de la forma recién expuesta.* /

```

- Calcular Media de las Variables de la muestra actual ($\overline{x_{it}}$) y pertenencia al intervalo

```

Obtener Clientes del Set de Datos ingresado por el usuario;
For (i en {variables del modelo}; i++) {
  Si (variable es continua)
    Calcular el promedio de la variable en el Set de Datos;
  Fin Si
  Si (variable discreta)
    Calcular promedio de variable en el Set de Datos;
  Fin Si
  Si ( $\overline{x_{it-1}^{inf}} \leq \overline{x_{it}} \leq \overline{x_{it-1}^{sup}}$ )
    Almacenar estado de variable es OK;
  Fin Si
  Si ( $\overline{x_{it-1}^{inf}} > \overline{x_{it}}$  ó  $\overline{x_{it}} > \overline{x_{it-1}^{sup}}$ )
    Almacenar estado de variable FUERA DE RANGO;
  Fin Si
}

```

7.2.2.2 Seudocódigo N°4: Gráficos de evolución promedio de cada variable

- Construcción de gráficos de evolución del promedio de cada variable

```
For (i en {últimos 12 meses}; i++) {  
  Obtener clientes (y sus variables) con créditos vigentes del mes;  
  Para cada una de las variables del modelo {  
    Obtener el valor de la variable de todos los clientes;  
    Calcular y guardar el promedio de dichos valores;  
  }  
}  
Graficar Valores Promedio de la Variable (eje Y) v/s Meses (eje X)
```

7.2.2.3 Seudocódigo N°5: Análisis univariado

- Análisis univariado de las variables del modelo

```
Obtener el set de datos ingresado por usuario;  
Calcular Vector Y = vector de valores de la variable dependiente  
"binarizada";  
For (i en {variables del modelo}; i++) {  
  Vector X = vector de valores de la variable independiente;  
  Si (variable es continua u ordinal)  
    Hacer Test K-S, con Vectores X e Y;  
    Almacenar Valor Estadístico Z y su Significancia;  
  Fin Si  
  Si (variable es categórica)  
    Hacer Test Chi Cuadrado de Pearson con Vectores X e Y;  
    Almacenar Valor Estadístico  $\chi^2$  y su Significancia;  
  Fin Si  
}
```

7.2.2.4 Seudocódigo N°6: Coherencia del Modelo

Aquí, se detalla tanto la construcción del gráfico de probabilidades predichas por el modelo, para determinar si discrimina entre clientes buenos y malos, como el índice de concordancia, que resume el gráfico anterior en un número.

- Construcción de gráfico de predicción de probabilidades

```
Obtener desde el set de datos ingresado por el usuario;  
Calcular Vector R = Resultados reales de pago de los clientes;  
Calcular Vector P = Valores de la probabilidad predicha por el modelo;
```

```
/*Para generar la distribución de buenos y malos predichos, de crearán 20  
sub-intervalos de tamaño 0,05, dentro del intervalo [0,1], y se contará la  
frecuencia de individuos en cada sub-intervalo respecto del total.*/  
*/
```

```
//Se declaran las variables que se necesitarán  
int IntervaloB[] = new int[20]; //contador de buenos en cada sub-intervalo  
int IntervaloM[] = new int[20]; //contador de malos de cada sub-intervalo  
int NumeroBuenos = 0; //contador del total de buenos en el intervalo [0,1]  
int NumeroMalos = 0; //contador del total de malos en el intervalo [0,1]
```

```
/*Se recorre el vector con resultados reales de pago, y se ve en qué sub-  
intervalo se encuentra la probabilidad predicha para cada cliente, para  
obtener la frecuencia de clientes por subintervalo.*/  
*/
```

```
For (int i =1; i en Vector R, i++) {  
  Si (R[i] == 1) //si el cliente fue malo  
    For (int j=0; j<20; j++) { //Recorre sub-intervalos de Malos  
      Si (P[i] pertenece al sub-intervalo j)  
        IntervaloM[j]++;  
        //suma 1 al contador del subintervalo  
      }  
    }  
  Fin Si  
}
```

```

        NúmeroMalos++; //cuenta el total de malos
    Fin Si
    Si (Y[i] == 0) //si el cliente fue bueno
        For (int j=0; j<20; j++) { //Recorre sub-intervalos de Buenos
            Si (P[i] pertenece al sub-intervalo j)
                IntervaloB[j]++;
                //suma 1 al contador del subintervalo
            Fin Si
        }
        NúmeroBuenos++; //cuenta el total de buenos
    Fin Si
}

```

/*Ahora se tiene el número de clientes buenos y malos por cada sub-intervalo. Pero se busca la proporción de clientes por cada sub-intervalo.*/

```

ProporcionB[]= new int[20] //porcentaje de buenos por sub-intervalo
ProporcionM[]=new int[20] //porcentaje de malos por sub-intervalo

```

```

For (int i=1; i<=20; i++) {
    ProporcionB[i] = IntervaloB[i]/NumeroBuenos;
    ProporcionM[i] = IntervaloM[i]/NumeroMalos;
}

```

/*Finalmente, teniendo en cuenta que en el sub-intervalo j hay (ProporcionX[j-1])% de clientes, es posible graficar las distribuciones de los clientes buenos y malos.*/

Graficar % casos v/s sub-intervalos de probabilidad del modelo, utilizando las proporciones ProporcionB[i] y ProporcionM[i] en el eje Y, y los sub-intervalos de probabilidad en el eje X.

- Cálculo del Índice de Concordancia²⁵

```

Obtener del set de Datos ingresado por el usuario:
Calcular Vector R = Resultados reales de pago de los clientes.
Calcular Vector P = valores de la probabilidad predicha por el modelo;
int contador=0; //contador de "probabilidades coherentes"
int Nbuenos=0; //contador del número de buenos
int Nmalos=0; //contador del número de malos
double indice; //almacena el C-Index

For (int i en Vector R; i++) {
  Si (R[i] == 0) //si el cliente es bueno
    Nbuenos++;
    For (int j en Vector R; j ≠ i; j++) {
      Si (R[j]==1) //se compara sólo con los malos
        Si (P[i]<P[j]) //si la prob. del malo es mayor
          contador++; //es coherente.
        Fin Si
      //si la prob. del malo es menor, no suma.
    Fin Si
    }
  Fin Si
  Si (R[i] == 1) //si el cliente es malo
    Nmalos++;
  Fin Si
}

//Finalmente, se calcula el índice de concordancia.
indice = contador/(Nmalos+Nbuenos);

```

²⁵ Seudocódigo basado en Weiss, J. (2006).

7.2.2.5 Seudocódigo N°7: Construcción de Matriz de Confusión

- Construcción de Matriz de Confusión

```
Obtener del set de Datos ingresado por el usuario;
Calcular Vector Y = valores de la variable dependiente "binarizada";
Obtener Vector R = Resultados reales de pago de los clientes.
int TP = 0; //contador de verdaderos positivos;
int FP = 0 //contador de falsos positivos;
int TN = 0 //contador de verdaderos negativos;
int FN = 0 //contador de falsos negativos;
int MC[][] = new int [2][2] //Matriz de Confusión del modelo;
double TPR = 0 //Ratio (%) de verdaderos positivos;
double FNR = 0 //Ratio (%) de falsos negativos;
double TNR = 0 //Ratio (%) de verdaderos negativos;
double FPR = 0 //Ratio (%) de falsos positivos;

//se recorren y comparan los vectores para construir la matriz
For (int i en Vector Y; i++) {
    Si (Y[i] == 0) //si se predijo que el cliente era bueno
        Si (R[i]==0) //y efectivamente fue bueno
            TP++;
        Fin Si
    Sino //si el cliente fue malo
        FP++;
    Fin Sino
    Fin Si
    Si (Y[i]==1) //Si se predijo que el cliente era malo
        Si (R[i]==1) //y efectivamente era malo;
            TN++;
        Fin Si
        Sino //si el cliente fue bueno
            FN++;
        Fin Sino
    Fin Si
}

//Se llenan las componentes de la Matriz de Confusión
MC [0][0] = TP;
```

```
MC [0][1] = FP;  
MC [1][0] = FN;  
MC [1][1] = TN;
```

```
//se calculan los porcentajes de acierto y desacierto del modelo  
TPR = TP*100/(TP + FP);  
TNR = TN*100/(TN + FN);  
FNR = FN*100/(FN + TN);  
FPR = FP*100/(TP + FP);
```

7.2.2.6 Seudocódigo N°8: Cálculo de Desempeño de Cortes Alternativos

- Cálculo de desempeños alternativos relativos al punto de corte actual:

```
Obtener del set de datos ingresado por el usuario;  
Calcular Vector P = valores de la probabilidad predicha por el modelo;  
Obtener Vector R = Resultados reales de pago de los clientes.  
Calcular Vector  $Y_c$  = valores de la var. dep. binarizada con corte actual ('c')
```

```
//Se declaran las variables que se necesitarán  
int B[] = new int[20] //contador de buenos aceptados para c/ pto. de corte i  
int BR[] = new int[20] //contador de buenos rechazados para cada corte i;  
int M[] = new int[20] //contador de malos aceptados para cada pto. de corte i  
int MR[] = new int[20] //contador de malos rechazados para cada corte i;  
int bc = 0; //contador de buenos aceptados según punto de corte actual ('c');  
int mc = 0; //contador de malos aceptados según punto de corte actual ('c');  
int brc=0; //contador de buenos rechazados según pto. de corte actual ('c')  
int mrc=0; //contador de malos rechazados según pto. de corte actual ('c');  
int NB = 0; //contador del total de buenos reales (atendidos);  
int NM = 0; //contador del total de malos reales (atendidos);  
Inicializar B[], BR[], M[] y MR[] en cero;
```

```
/*Se recorre el Vector  $Y_c$  y se compara con el resultado real en R. Además,  
se obtiene el número total real de buenos y malos.*/
```

```
For (int j =1; j en Vector R, j++) {  
    Si (R[j] == 1) //si el cliente fue malo
```

```

        NM++;
        Si (Yc[j]==1) //si el modelo actual lo rechaza
            mrc++;
        Fin Si
        Si(Yc[j]==0) //si el modelo con corte actual lo acepta
            mc++;
        Fin Si
    Fin Si
    Si (R[j] == 0) //si el cliente fue bueno
        NB++;
        Si (Yc[j]==1) //si el modelo actual lo rechaza
            brc++;
        Fin Si
        Si(Yc[j]==0) //si el modelo con corte actual lo acepta
            bc++;
        Fin Si
    Fin Si
}

//Ahora se repite lo anterior para el vector de cada punto de corte 'i' (Yi).
For (i =0,05; i≠c; i+0,05) {
    Calcular Vector Yi = valores de la var. dep. "binarizada" con el corte 'i';

    //Se recorre el Vector Yi con, y se compara con el resultado real en R
    //Para lo siguiente, notar que (20*i-1) sitúa a la variable 'i' entre 0 y 19.
    For (int j =1; j en Vector R, j++) {
        Si (R[j] == 1) //si el cliente fue, en realidad, malo
            Si (Yi[j]==1) //si el modelo con corte en 'i' lo rechaza.
                MR[(20*i)-1]++;
            Fin Si
            Si(Yi[j]==0) //si el modelo con corte en 'i' lo acepta.
                M[(20*i)-1]++; //suma malos
            Fin Si
        Fin Si
        Si (R[j] == 0) //si el cliente fue, en realidad, bueno
            Si (Yi[j]==1) //si el modelo con corte en 'i' lo rechaza.
                BR[(20*i)-1]++;
            Fin Si
            Si(Yi[j]==0) //si el modelo con corte en 'i' lo acepta
                B[(20*i)-1]++; //suma buenos
    }
}

```

```

                Fin Si
            Fin Si
        }
    }

    //Se calcula el % de acierto global del modelo actual y de cada corte
    double acierto[] = new double[20]; //almacena los aciertos con cada corte;
    double a_c = Porcentaje de acierto global del modelo actual
    a_c=( b_c+ m_r_c)*100/(NB+NM) //acierto global actual
    For (int k=0; k<20; k++) {
        acierto[k] = (B[k]+MR[k])*100/(NB+NM) //acierto global de corte 'k'
    }

    //se calcula la efectividad de cada corte del modelo nuevo
    double efe[] = new double[20]; //almacena la efectividad con cada corte;
    For (int k=0; k<20; k++) {
        efe[k] = [(B[k]/NB)*100+(MR[k]/NM)*100]/2 //% de efectividad de corte 'k'
    }

    //se calcula la cobertura que se tendría con cada corte en el modelo nuevo
    double cobertura[] = new double[20]; //almacena la cobertura de cada corte;
    For (int k=0; k<20; k++) {
        cobertura[k] = (B[k]+M[k])*100/(NB+NM) //% de cobertura de corte 'k'
    }

    /*Se calcula la variación de utilidad de cada corte respecto al modelo
    actual, la diferencia de buenos y malos aceptados*/
    //Se obtienen los parámetros: costo de aceptar un malo y aceptar un bueno
    Obtener costoMalo (C_m) y beneficioBueno (B_b)
    int deltaB[] = new double[20]; //variación relativa de buenos
    int deltaM[] = new double[20]; //variación relativa de malos
    double U[] = new double[20]; //Utilidad de cada punto de corte
    double deltaU[] = new double[20]; //variación relativa de utilidad
    For (int j=0; j<19; j++) {
        deltaB[j]= (B[j] - b_c);
        deltaM[j]= (M[j] - m_c);
        U[j]=B[j]*B_b - M[j]*C_m;
        deltaU[j]=deltaB[j]*B_b - deltaM[j]*C_m;
    }

```

//Se hace la tabla comparativa entre los cortes alternativos y el actual
 Generar tabla con las siguientes columnas: punto de corte, número de buenos y malos aceptados (B[] y M[]), % de acierto global, % de efectividad, % de cobertura, Utilidad (U[]) y variación relativa de utilidad (deltaU[])

//Se grafica la variación de utilidad
 Graficar variación relativa de utilidad (deltaU[]) en eje Y, v/s pto. corte en X

7.2.2.7 Seudocódigo N°9: Entrenamiento del Modelo Nuevo

- Entrenar el Modelo Nuevo

```
//Primero se balancea el set de datos de Entrenamiento
Obtener Vector R = Resultados reales de pago de los clientes del set de
entrenamiento;
int NBtrain=0; //contador de buenos
int NMtrain=0; //contador de malos
For (int i=1; i en Vector R, i++) {
  Si (R[i]==1)
    NMtrain++;
  Fin Si
  Si (R[i]==0)
    NBtrain++;
  Fin Si
}
Si (NBtrain>NMtrain)
  Ponderar Buenos del set por (NMtrain/(NMtrain+NBtrain))
Fin Si
Si (NBtrain<NMtrain)
  Ponderar Buenos del set por (NBtrain/(NMtrain+NBtrain))
Fin Si

//Se declaran las variables que se almacenarán luego del entrenamiento
double Rcuadrado=0; //R2 de Nagelkerke
```

```

double G=0; //likelihood ratio
double[]Betastrain = 0; //coeficientes Betas
double[]DesvBetas=0; //Desviación de los Betas
double[]EstadisticoBetas=0; //Valor de los estadísticos de los Betas
double[]PvalorBetas=0; //Nivel de significancia de los Betas
/*El largo de los arreglos anteriores, dependerá del número de variables,
pero se deben inicializar en cero.*/

//Se entrena el modelo
Entrenar Modelo con set de datos de entrenamiento;
Almacenar los resultados correspondientes en las variables;

```

7.2.2.8 Seudocódigo N°10: Comparación del Modelo Actual y Nuevo

- Comparar Modelo Actual y Nuevo

```

Obtener el set de datos de prueba (con variables nuevas)
Obtener  $R_N^2$  y G del modelo actual;
Obtener Vector R = Resultados reales de pago de los clientes del set de
prueba;
Obtener costoMalo ( $C_m$ ) y beneficioBueno ( $B_b$ );
Obtener Vector  $Y_A$  = valores de la var. dep. “binarizada” del modelo actual;

//se calcula la predicción de cada modelo para cada registro del set de
Prueba
Calcular Vector  $P_N$  = valores de la prob. predicha por el modelo nuevo;
Calcular Vector  $P_A$  = valores de la prob. predicha por el modelo actual;

//Se declaran los contadores a utilizar en distintos cortes del modelo nuevo
int  $B_N[]$  = new int[20] //n° de buenos aceptados en cada pto. de corte i;
int  $BR_N[]$  = new int[20] //n° de buenos rechazados en cada corte i;
int  $M_N[]$  = new int[20] //n° de malos aceptados para en pto. de corte i
int  $MR_N[]$  = new int[20] //n° de malos rechazados en cada corte i;
//Se declaran contadores para el modelo actual
int  $b_A = 0$ ; //contador de buenos aceptados;
int  $m_A = 0$ ; //contador de malos aceptados;

```

```
//Se declaran los contadores para los resultados reales de pago de clientes
int NBR = 0; //contador del total de buenos reales (atendidos);
int NMR = 0; //contador del total de malos reales (atendidos);
```

//Determinación del mejor punto de corte para el modelo nuevo

/*Primero, se obtiene el número total real de buenos y malos, recorriendo el Vector R. Como una forma de aprovechar la iteración, se obtiene el total predicho del modelo actual, para la posterior construcción de su matriz de confusión*/

```
For (int j =1; j en Vector R, j++) {
```

```
    Si (R[j] == 1) //si el cliente fue malo
```

```
        NMR ++;
```

```
        mA++;
```

```
    Fin Si
```

```
    Si (R[j] == 0) //si el cliente fue bueno
```

```
        NBR ++;
```

```
        bA++;
```

```
    Fin Si
```

/*Los clientes atendidos (reales) han sido aceptados por el modelo, por lo que el número de buenos y malos, coinciden para ambos casos*/

```
}
```

/*Ahora se calcula el desempeño de la predicción del modelo nuevo, **para cada** punto de corte 'i' (se almacenará en el vector P_{Ni})*/

```
For (i =0,05; i<=1; i+0,05) {
```

```
    Calcular Vector PNi = valores de predicción del modelo nuevo con corte i;
```

/*Se recorre el Vector P_{Ni} con, y se compara con el resultado real en R

```
For (int j =1; j en Vector R, j++) {
```

```
    Si (R[j] == 1) //si el cliente fue, en realidad, malo
```

```
        Si (PNi[j]==1) //si el modelo nuevo con corte en i rechaza.
```

```
            MRN[(20*i)-1]++;
```

```
        Fin Si
```

```
        Si(Yi[j]==0) //si el modelo nuevo con corte en 'i' lo acepta.
```

```
            MN[(20*i)-1]++; //suma malos
```

```
        Fin Si
```

```
    Fin Si
```

```
    Si (R[j] == 0) //si el cliente fue, en realidad, bueno
```

```
        Si (Yi[j]==1) //si el modelo con corte en 'i' lo rechaza.
```

```

        BRN[(20*i)-1]++;
    Fin Si
    Si(Yi[j]==0) //si el modelo con corte en 'i' lo acepta
        BN[(20*i)-1]++; //suma buenos
    Fin Si
Fin Si
}
}

```

```

//Ahora, se escogerá el corte que genera mayor utilidad en el modelo nuevo
double UtilidadN[] = new double[20]; //Utilidad por usar punto de corte en 'i'
For (int j=0; j<19; j++) { //se calcula la utilidad de cada punto de corte
    UtilidadN[j] = BN[j]*B0 - M[j]*Cm - BRN[j]*Cop;
}
//se escoge el corte  $\bar{c}$ , que genera la mayor utilidad
double  $\bar{u}$  = 0; //Guarda la utilidad máxima
double  $\bar{c}$  = 0; //Guarda el punto de corte donde se genera  $\bar{u}$ 
For (int j=0; j<19; j++) {
    Si (UtilidadN[j] >  $\bar{u}$ )
         $\bar{c}$  = (j+1)/20;
         $\bar{u}$  = Utilidad[j];
    Fin Si
}

```

//Cálculo indicadores del nuevo modelo, para cada punto de corte

//% de acierto global

```

double aciertoN[] = new double[20]; //almacena los aciertos con cada corte;
For (int k=0; k<20; k++) {
    aciertoN[k] = (BN[k]+MRN[k])*100/(NBN+NMN) //acierto global de corte 'k'
}

```

//% de efectividad

```

double efeN[] = new double[20]; //almacena la efectividad con cada corte;
For (int k=0; k<20; k++) {
    efeN[k] = [(BN[k]/NBN)*100+(MRN[k]/NMN)*100]/2 //% de efectividad de
corte 'k'
}

```

//% de cobertura

```

double coberturaN[] = new double[20]; //almacena la cobertura de cada
corte;
For (int k=0; k<20; k++) {
    coberturaN[k] = (B[k]+M[k])*100/(NB+NM) //% de cobertura de corte 'k'
}
//Generación de Matrices de Confusión
//Se declaran las Matrices para el modelo actual (MCA) y nuevo (MCN)
int MCA[][] = new int[2][2];
int MCN[][] = new int[2][2];
//Se llenan los componentes de las matrices de confusión actual
MCA [0][0] = bA;
MCA [1][1] = mA;
//En la matriz nueva, se usa la posición en contadores, donde utilidad es  $\bar{u}$ 
MCN [0][0] = BN[20* $\bar{c}$ -1];
MCN [0][1] = BRN[20* $\bar{c}$ -1];
MCN [1][0] = MN[20* $\bar{c}$ -1];
MCN [1][1] = MRN[20* $\bar{c}$ -1];

//Cálculo de Índices de Concordancia
//Se declaran las variables que se necesitarán
int contadorA=0; // contador de “probabilidades coherentes” modelo actual
int contadorN=0; //contador de “probabilidades coherentes” modelo nuevo
double indiceA; //almacena el C-Index del modelo actual
double indiceN; //almacena el C-Index del modelo nuevo

//Se recorre el vector de resultados reales
For (int i en Vector R; i++) {
    Si (R[i] == 0) //si el cliente es bueno
        For (int j en Vector R; j ≠ i; j++) {
            Si (R[j]==1) //se compara sólo con los malos
                Si (PA[i]<PA[j]) //si prob. del malo es mayor
                    contadorA++; //es coherente.
                Fin Si
            Si (PN[i]<PN[j]) //si prob. del malo es mayor
                    contadorN++; //es coherente.
            Fin Si
            //si las prob. de malos es menor, no suma.
        }
    Fin Si
}
Fin Si

```

```

    //Si el cliente es malo, no se hace nada (para no sumar dos veces)
}
//Finalmente, se calculan los índices de concordancia.
indiceA = contadorA/(NBR+NMR);
indiceN = contadorN/(NBR+NMR);

//Cálculo de Utilidad del Modelo Actual
double UtilidadA = 0;
UtilidadA = bA*Bb - mA*Cm;

```

7.2.2.9 Seudocódigo N°11: Efectividad del Modelo en Set de Entrenamiento y Prueba

- Evaluar Sobreajuste del Modelo Nuevo

```

Obtener el set de datos de entrenamiento (con variables del modelo)
Obtener el punto de corte  $\bar{c}$  (que generó utilidad máxima en el set de
//prueba)
Obtener Vector RT = Resultados reales de pago de los clientes del set
//entrenamiento
Calcular Vector YTC = valores “binarizados” del modelo en set de
//entrenamiento con corte  $\bar{c}$ 
Obtener MCN [0][0] y MCN [1][1] //buenos aceptados y malos rechazados
//en el set de Prueba, en Seudocódigo N°10.

//contadores para probar la efectividad del modelo nuevo en el set
//entrenamiento
int bTC = 0; //n° de buenos aceptados por modelo con corte  $\bar{c}$ ;
int mrTC = 0; //n° de malos rechazados por modelo, con corte  $\bar{c}$ ;
int NBT = 0; //n° total de buenos reales en set Entrenamiento (atendidos);
int NMT = 0; //n° total de malos reales en set de Entrenamiento (atendidos);

/*Se recorre el Vector YTC y se compara con el resultado real en RT.
Además, se obtiene el número total real de buenos y malos.*
For (int j= 1; j en Vector RT, j++) {

```

```

Si (RT[j] == 1) //si el cliente fue malo
    NMT++;
    Si (YTC[j]==1) //si el modelo lo rechaza
        mrTC++; //acierta
    Fin Si
Fin Si
Si (RT[j] == 0) //si el cliente fue bueno
    NBT++;
    Si (YTC[j]==0) //si el modelo lo acepta
        bTC++; //acierta
    Fin Si
Fin Si
}

```

//Se calcula la efectividad del modelo en el Entrenamiento y Prueba

double E_{TRAIN} = % de efectividad del modelo nuevo en el set de Entrenamiento

double E_{TEST} = % de efectividad del modelo nuevo en el set de Prueba

E_{TRAIN} = (b_{TC}+ mr_{TC})*100/(NB_T + NM_T);

E_{TEST} = (MC_N [0][0]+ MC_N [1][1])*100/(NB_T + NM_T);

7.2.2.10 Seudocódigo N°12: Estabilidad del Modelo Nuevo

- Verificar Estabilidad del Modelo Nuevo

Obtener el set de datos completo (que incluya las variables del modelo)

Obtener n = número de iteraciones a realizar

Obtener \bar{c} ;

double[][] Betas = new double [n][número de variables modelo];

double[] efectividades = new double [n];

//se realizan las 'n' iteraciones

For (int i=1; i<=n; i++) {

Dividir aleatoriamente el set de datos en un set Train (80%) y Test (20%)

Entrenar el modelo con set de Entrenamiento //Análogo a Seudocódigo N°9

```

//el Seudocódigo N°10 obtiene los Betas y los guarda en Betastrain[]
//Luego, se almacena los valores de los Beta de esta iteración
For (int j=0; j< número de variables; j++) {
  Betas[i][j]=Betastrain[j];
}

```

//Ahora se calculará el acierto del modelo de esta iteración en el set de prueba

Obtener Vector R_E = Resultados reales de pago de los clientes en prueba
 Calcular Vector Y_E = Valores de la var. dep. “binarizada” en set de prueba
 //Para el cálculo del Vector anterior, se deberá utilizar el corte \bar{c}
 int $B_E = 0$; //contador de buenos aceptados por el modelo
 int $M_E = 0$; //contador de malos rechazados por el modelo
 int Total $B_E = 0$; //contador del total de buenos del set de Prueba
 int Total $M_E = 0$; //contador del total de malos del set de Prueba

*/*Se recorre el Vector Y, y se compara con el resultado real en R. Además, se obtiene el número total real de buenos y malos.*/*

```

For (int k =1; k en Vector R, k++) {
  Si ( $R_E [k] == 1$ ) //si el cliente fue malo
    Total $M_{E++}$ ;
    Si ( $Y_E [k]==1$ ) //si el modelo lo rechaza
       $M_{E++}$ ; //se acierta
    Fin Si
    //Sólo importa, en este caso, los malos que se rechazan
  Fin Si
  Si ( $R_E [k] == 0$ ) //si el cliente fue bueno
    Total $B_{E++}$ ;
    Si( $Y_{c[j]}==0$ ) //si el modelo lo acepta
       $B_{E++}$ ;
    Fin Si
    //Sólo importa, en este caso, los buenos que se aciertan
  Fin Si
}

```

//Se calcula el % de efectividad global del modelo en esta iteración
 double E_E = Porcentaje de efectividad global del modelo
 $E_E = (B_E + M_E) * 100 / (TotalB_E + TotalM_E)$;
//Se almacena la efectividad del modelo en esta iteración
 efectividades[i] = E_E ;

```
} //fin de las iteraciones
```

//Se calculan los promedios y desviaciones de variables y efectividad

```
Calcular  $\bar{X}_\beta[i]$  = Promedio de cada columna 'i' de la matriz Betas[][]
```

```
Calcular  $\sigma_\beta[i]$  = desv. estándar de cada columna 'i' de la matriz Betas[][];
```

```
Calcular  $\bar{X}_E$  = Promedio de los valores del vector efectividades[];
```

```
Calcular  $\sigma_E$  = desviación estándar del vector efectividades[];
```

```
double MatrizEstabilidad = new double[n+2][n° variables +1];
```

```
//Se llena la matriz de estabilidad, que será la salida de esta etapa
```

```
For (int j=0; j< n° de variables+1; j++) {
```

```
    Si (j<= n° de variables) //Se llenan los datos de Betas de variables
```

```
        For (int k=0; k< n+2; k++) {
```

```
            Si (k< n) //se copian los valores de Betas de la variable
```

```
                MatrizEstabilidad[j][k] = Betas[j][k];
```

```
            Fin Si
```

```
            Si (k= n) //se copia el promedio de Betas de la variable
```

```
                MatrizEstabilidad[j][k] =  $\bar{X}_\beta[j]$ ;
```

```
            Fin Si
```

```
            Si (k= n+1) //copia la desviación de Betas de la variable
```

```
                MatrizEstabilidad[j][k] =  $\sigma_\beta[i]$ ;
```

```
            Fin Si
```

```
        }
```

```
    Fin Si
```

```
    Si (j= n° de variables+1) //Se llena la efectividad de cada iteración
```

```
        For (int k=0; k< n+2; k++) {
```

```
            Si (k< n) //se copia la efectividad de la iteración
```

```
                MatrizEstabilidad[j][k] = efectividad[k];
```

```
            Fin Si
```

```
            Si (k= n) //se copia el promedio de las efectividades
```

```
                MatrizEstabilidad[j][k] =  $\bar{X}_E$ ;
```

```
            Fin Si
```

```
            Si (k= n+1) //copia la desviación de las efectividades
```

```
                MatrizEstabilidad[j][k] =  $\sigma_E$ ;
```

```
            Fin Si
```

```
        }
```

```
    Fin Si
```

```
}
```

7.3 Tecnología Habilitante

7.3.1 Introducción

En un mundo globalizado, la necesidad de construir sistemas seguros y eficientes en el procesamiento de información es indispensable, sobretodo en empresas de servicios, que manejan millonarias transacciones de dinero y cuya estrategia se basa en la eficacia operacional.

Hasta fines de los años 90, la arquitectura cliente servidor (C/S) era ampliamente utilizada en la construcción de sistemas de información. Una de sus características principales, era su arquitectura de 2 capas: la capa de presentación y lógica de negocio, por el lado del cliente, y la capa de datos, por el lado del servidor. Lo anterior genera tráfico excesivo, por el constante flujo de información sin procesar a los usuarios, y dificultad de actualización, por cuanto cada modificación a la lógica de negocio debe ser realizada en cada uno de los hosts.

Por lo tanto, el desafío fue cómo estructurar la tecnología disponible, de forma que la interacción entre las distintas componentes sea rápida y transparente.

En este contexto, se diseñó la arquitectura de 3 capas, que separa la capa lógica y la sitúa en un servidor, que es el que realiza el procesamiento de la información y la envía a los clientes. Con ello, la capa de presentación sólo cumple la función de recibir los requerimientos del usuario, traspasarlos a la capa de negocio y desplegar lo solicitado. Finalmente, la capa de datos es la que recupera y almacena la información, de acuerdo a los requerimientos de la capa lógica. Es importante señalar que las capas sólo se comunican con la capa contigua, es decir, la capa de presentación no se relaciona directamente con la capa de datos.

En esta arquitectura, la tecnología Web adquiere especial relevancia, por cuanto permite hacer requerimientos a servidores ubicados en cualquier punto del mundo, acceder a información actualizada y operar en línea. Lo anterior es particularmente importante para empresas dispersas geográficamente, como el Banco del Desarrollo.

7.3.2 Tecnología ASP

La tecnología propuesta debe integrarse a cabalidad con la existente, por lo que dar continuidad tanto a su arquitectura como su lenguaje orientado a

objetos, es la decisión más obvia. Dicha tecnología es ASP (Active Server Pages), la cual posee una arquitectura de 3 capas, se basa en el lenguaje Visual Basic Script, y que se integra a una Base de Datos SQL.

La tecnología ASP ha sido desarrollada por Microsoft, y diseñada para la creación de aplicaciones en internet. Por lo tanto, es ideal para la utilización de servicios Web, necesarios para el acceso remoto a la información.

ASP utiliza de páginas dinámicas, que son páginas HTML generadas a partir de lenguajes de programación (*scripts*) y que se ejecutan en el servidor web (*server side*). Dichos *scripts* generan el código HTML que se envía al usuario desde el servidor. Este código puede ser modificado en función de la petición realizada por el usuario, por lo que el contenido de la página de respuesta dependerá del resultado que genere. Las ventajas de utilizar páginas dinámicas, es que el cliente no puede ver los *scripts*, son independientes del navegador web, y que aumentan la rapidez de visualización del usuario, ya que el procesamiento lo hace el servidor.

Por otro lado, ASP proporciona una serie de objetos integrados, a los que se puede tener acceso sin necesidad de instanciarlos, y que están disponibles en una librería. Estos objetos son componentes ActiveX Server, que están

especialmente diseñados para intercambiar información entre los usuarios y el servidor. Los objetos son módulos incorporados al lenguaje, que permiten el desarrollo de tareas específicas. A partir de una llamada, el objeto realizará la tarea requerida, por lo que se ahorra la programación de operaciones habituales. Otros objetos pueden ser incorporados a ASP, como si se tratase de componentes accesorios. Obviamente, estos objetos deben estar integrados al servidor para su utilización.

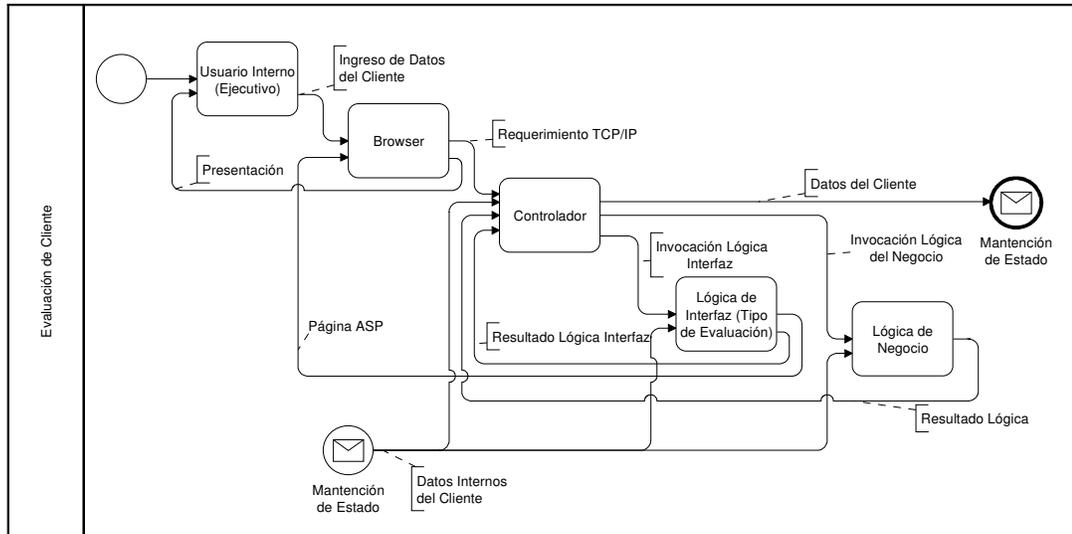
En definitiva, ASP es una tecnología que permite una interacción rápida el usuario y la información, ideal para sistemas en constante actualización y que funcionan con servicios web. Es de gran escalabilidad, de fácil mantención y seguro, pues encapsula el procesamiento de información en el lado del servidor.

7.4 Arquitectura de Apoyos Computacionales

La arquitectura del apoyo computacional será realizada mediante la aplicación del patrón de apoyo computacional basado en internet. Este patrón será aplicado a la “Evaluación del Cliente”, que es el proceso para el cual se detallará el diseño computacional, y se detalla en la ilustración 25. No obstante

lo anterior, esta arquitectura podría ser aplicada a cualquier actividad que requiera apoyo computacional.

Ilustración 25: Arquitectura de Apoyo Computacional para Evaluación del Cliente



En la ilustración anterior, se aprecia la interacción del usuario con el sistema para evaluar al cliente, donde particularmente, se explicita el ingreso de variables de segmentación y el resultado de la lógica (página de la evaluación).

En la página de clasificación, se desplegará una página para el ingreso de las características del cliente, que serán capturadas por el controlador, validadas, y sometidas a la lógica de negocio (clasificación), de donde resultará la evaluación a utilizar. Con dicho resultado, el controlador procederá a invocar

la página de la solicitud de crédito que corresponda, la que será desplegada y posteriormente completada por el usuario

7.5 Casos de Uso

Perfiles de Usuarios

1. Ejecutivo de Cuenta: El ejecutivo utilizará a la aplicación tanto para ingresar (actualizar) los datos del cliente, como para evaluarlo (o modificar una evaluación anterior).
2. Instancia de Aprobación (Comité): Existen tres instancias de aprobación, dependiendo de las características del crédito solicitado: el Jefe de Sucursal, el Supervisor Comercial y el Agente del Banco. Todos cumplen la función de aprobar o rechazar las solicitudes que les correspondan, registrando claramente las condiciones de aprobación o rechazo.
3. Administrador (Gerencia): El administrador puede ingresar a un módulo para levantar indicadores de desempeño de los segmentos de clientes, y eventualmente, modificar los parámetros de dicha lógica.

7.5.1 Ingresar Datos del Cliente

Modela el apoyo computacional para el diagrama de actividades correspondiente a “Ingresar Datos del Cliente”, y cuyo único caso de uso se muestra en la ilustración 26:

Ilustración 26: Caso de Uso de Ingresar Datos Cliente

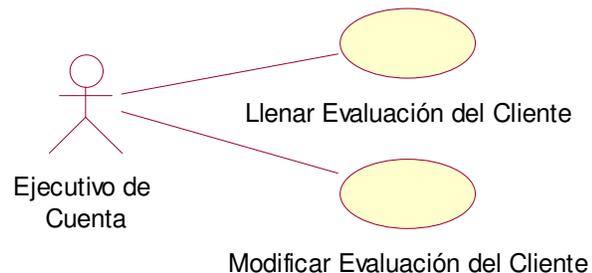


- *Ingresar Datos del Cliente*
 - Propósito: Permite al ejecutivo ingresar (actualizar) los datos de un cliente nuevo (antiguo).
 - Resumen: Si el cliente es nuevo, el ejecutivo de cuenta deberá registrar sus datos. Si el cliente es antiguo, el ejecutivo podrá actualizar sus datos, si éstos han cambiado respecto del último registro. Esto se realiza antes de evaluar al cliente.

7.5.2 Evaluación del Cliente

Modela el apoyo computacional para el diagrama de actividades correspondiente a “Evaluación del Cliente”, y cuyos casos de uso se muestran a en la ilustración 27:

Ilustración 27: Casos de Uso de Evaluación del Cliente



- *Llenar Evaluación del Cliente*
 - Propósito: Permite al ejecutivo evaluar al cliente de forma segmentada, sea éste nuevo o antiguo.
 - Resumen: El ejecutivo ingresa a llenar la solicitud de crédito, mediante el proceso de clasificación. Aquí, el sistema lo redirige a la solicitud de crédito que corresponda al perfil del cliente que será evaluado. Si la evaluación está completa, ésta puede ser enviada al Comité para su aprobación. De lo contrario, su estado será “activa”, y podrá ser modificada posteriormente.

- Precondición: Los datos del cliente deben haber sido ingresados en el sistema.

- *Modificar Evaluación del Cliente*

- Propósito: Permite al ejecutivo cambiar las condiciones de una evaluación anterior, o terminar de llenarla.

- Resumen: El ejecutivo reingresa a llenar una solicitud incompleta, que posee datos faltantes para ser enviada al Comité, o bien a corregir las condiciones de una solicitud anterior que fue rechazada del Comité, como el monto del crédito, tasa o plazo. Una vez completa, es enviada al Comité. En caso de seguir incompleta, su estado permanece como “activo”.

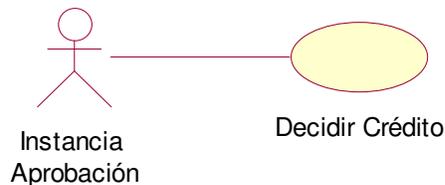
- Precondición: Los datos del cliente deben haber sido ingresados en el sistema. La solicitud debe estar “activa”, es decir, el cliente ya debe haber sido clasificado previamente por el sistema.

7.5.3 Decidir Crédito

Modela el apoyo computacional para el diagrama de actividades correspondiente a “Decidir Crédito”, que se muestra en la ilustración 28 sólo con

fines aclaratorios, porque ya se encuentra implementada y sólo debe ser integrada al proyecto. Por ello, no se detallarán sus diagramas de secuencia.

Ilustración 28: Caso de Uso de Decidir Crédito

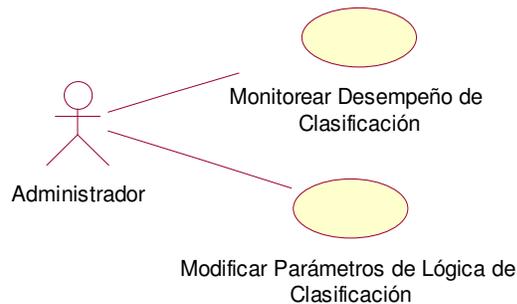


- *Decidir Crédito*
 - Propósito: Permite a la instancia de aprobación (Comité) decidir el otorgamiento del crédito, o si sus condiciones deben ser modificadas por el ejecutivo.
 - Resumen: La instancia de aprobación ingresará a revisar las condiciones del crédito propuesta por el ejecutivo. Si las condiciones son acordes al riesgo del cliente, entonces la solicitud será aprobada. En caso de rechazo, el ejecutivo deberá chequear las condiciones sugeridas con el cliente y modificar la propuesta de crédito a ser enviada a la instancia de aprobación. En ambos casos anteriores, se podrá hacer observaciones o poner condiciones a la propuesta del ejecutivo.
 - Precondición: Que la solicitud de crédito haya sido llenada y enviada previamente por el ejecutivo.

7.5.4 Monitorear Desempeño de Ventas

Modela el apoyo computacional para el diagrama de actividades correspondiente a “Monitorear Desempeño de Ventas”, y cuyo único caso de uso se muestra en la ilustración 29:

Ilustración 29: Casos de Uso de Monitorear Desempeño de Ventas



- *Monitorear Desempeño de Clasificación*
 - Propósito: Permite al administrador ver el desempeño de los distintos segmentos de clientes y evaluaciones, medido en indicadores de riesgo y rentabilidad, mensual y anualmente.
 - Resumen: El administrador ingresará el rubro del árbol y el periodo en el cual desea analizar el desempeño de las ventas. El Sistema le desplegará el árbol solicitado, con el desempeño tanto de las evaluaciones (nodos finales) como de la lógica de clasificación (ramas

intermedias). A través de los indicadores generados, se podrá decidir si proponer cambios al modelo utilizado actualmente.

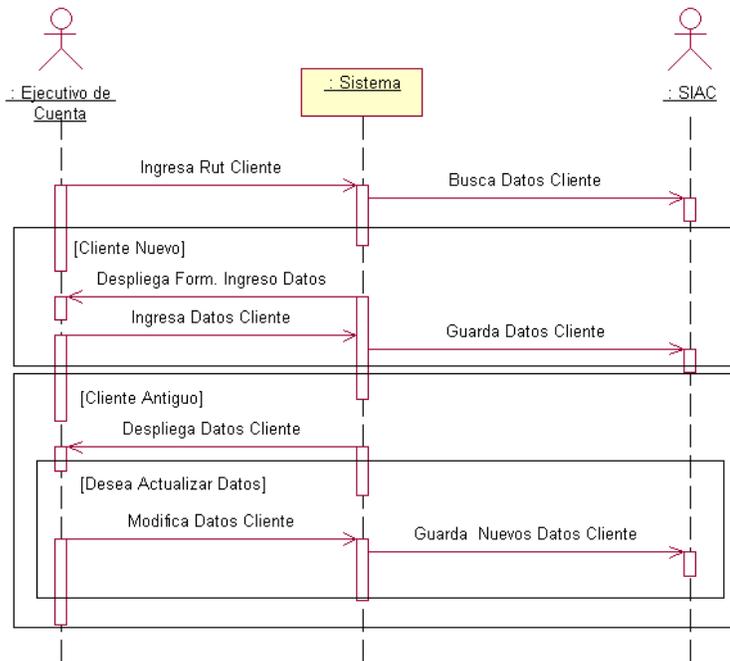
- *Modificar Parámetros de Lógica de Clasificación*
 - Propósito: Cambiar la lógica de clasificación del modelo.
 - Resumen: El administrador deberá escoger el rubro del árbol cuya lógica desea modificar, y posteriormente ingresar el valor del nuevo parámetro en la lógica. Esta modificación será realizada sólo cuando no sean necesarios cambios “estructurales” en la lógica, es decir, que requiera agregación o disminución de variables.

7.6 Diagramas de Secuencia

Para simplificar los diagramas de secuencia posteriores, se asumirá que todos los actores involucrados ya se encuentran dentro del sistema, por lo que éste ya los identifica de acuerdo a uno de los perfiles descritos previamente.

7.6.1 Caso de Uso: Ingresar Datos del Cliente

Ilustración 30: Diagrama de Secuencia de Ingresar Datos del Cliente

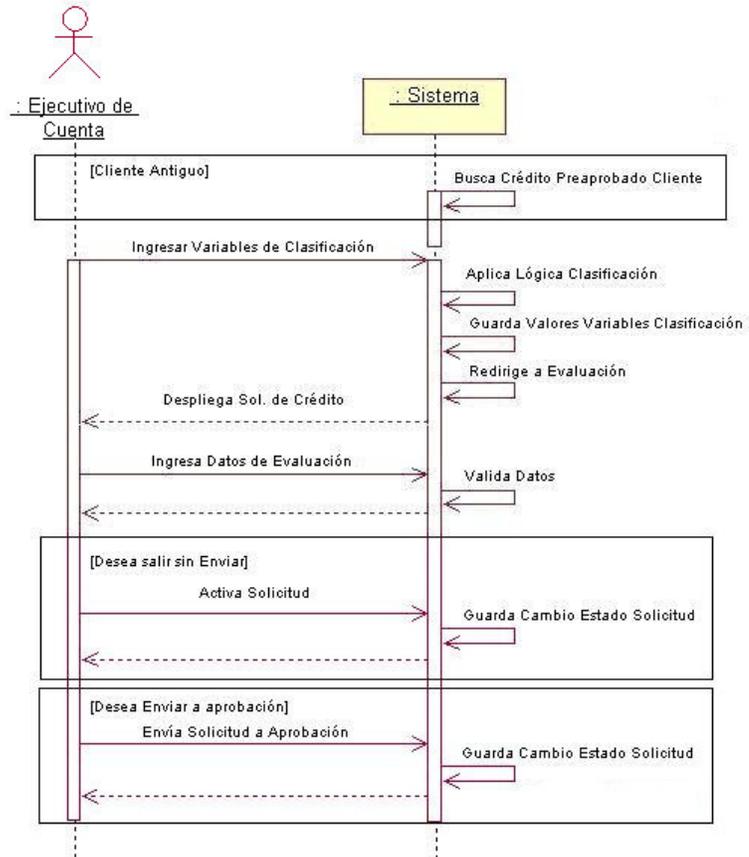


El diagrama de secuencia asociado a este caso de uso, se detalla en la ilustración 30. Aunque éste ya se encuentra implementado, se muestra sólo de manera complementaria. El primer paso es ingresar el RUT del cliente, con el que el sistema buscará sus datos en el SIAC. Si el cliente es nuevo, se deberá ingresar sus, antes de continuar con su evaluación.

Si el cliente es antiguo, el ejecutivo podrá actualizar sus datos. Si elige esta opción, se podrá modificar los datos y guardarlos en las bases del Banco.

7.6.2 Caso de Uso: Llenar Evaluación del Cliente

Ilustración 31: Diagrama de Secuencia de Llenar Evaluación del Cliente

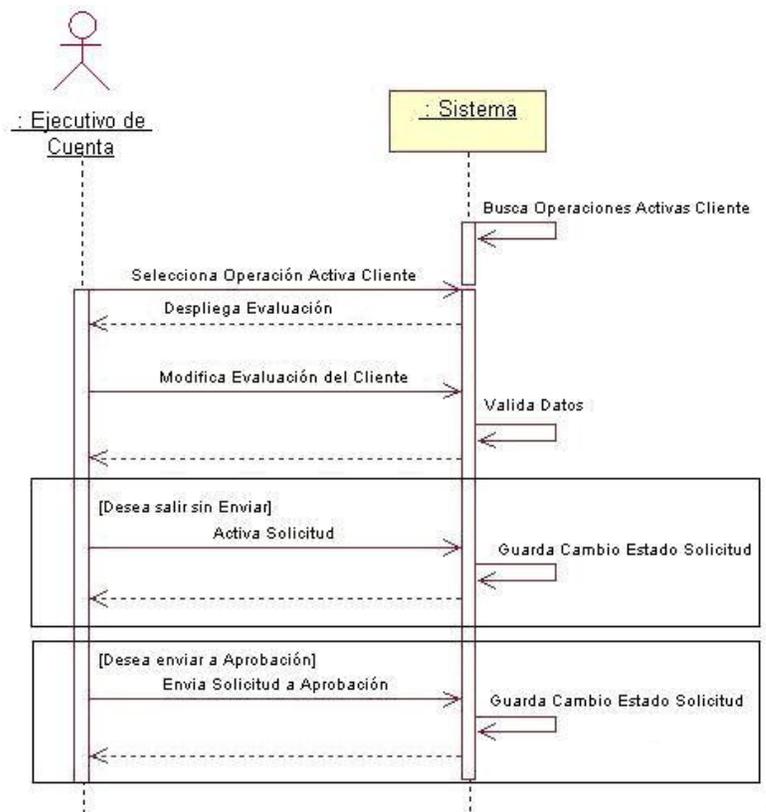


La ilustración 31, muestra el diagrama de secuencia de este caso de uso. La primera acción es realizada por el sistema, que obtendrá, en caso que el cliente sea antiguo, si éste posee algún crédito pre-aprobado. En seguida, se deberán ingresar las variables de clasificación, que serán validadas, almacenadas y sometidas a la lógica de clasificación, con cuyo resultado se redirigirá a la evaluación correspondiente. Luego, el ejecutivo deberá ingresar

los datos que ésta requiera, que serán validados antes de su almacenamiento. Finalmente, el ejecutivo podrá grabar la solicitud del cliente y salir de la aplicación, en cuyo caso la solicitud quedará “activa”, o bien, enviarla a aprobación, quedando la solicitud en estado “pendiente”.

7.6.3 Caso de Uso: Modificar Evaluación del Cliente

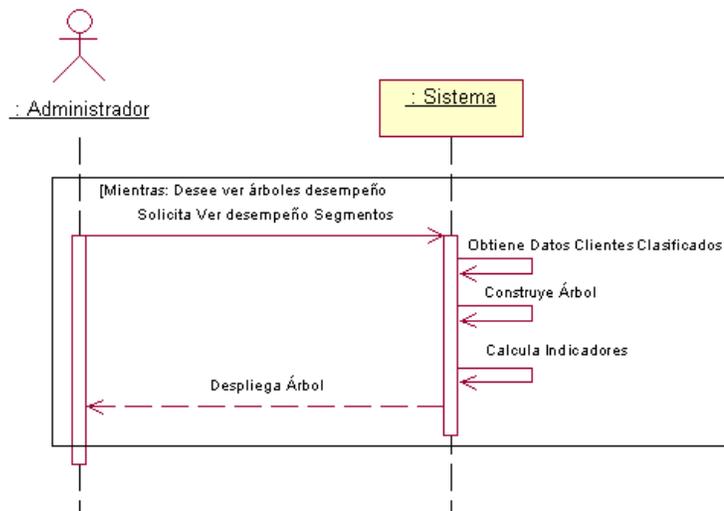
Ilustración 32: Diagrama de Secuencia de Modificar Evaluación del Cliente



Como muestra el diagrama de secuencia en la ilustración 32, primero el sistema buscará las operaciones del cliente que se encuentren activas y las desplegará al ejecutivo. Aquí, éste debe seleccionar a qué operación desea ingresar, para que el sistema lo redirija sin tener que clasificarlo nuevamente. Una vez en la evaluación, se despliegan todos los datos que ya han sido ingresados anteriormente. Los pasos siguientes ya han sido explicados en el diagrama anterior. Se debe tener presente que sólo se llega a este escenario cuando la solicitud aún no ha sido enviada a aprobación, o bien, cuando ésta ha sido rechazada y sus condiciones deben ser modificadas.

7.6.4 Caso de Uso: Monitorear Desempeño de Clasificación

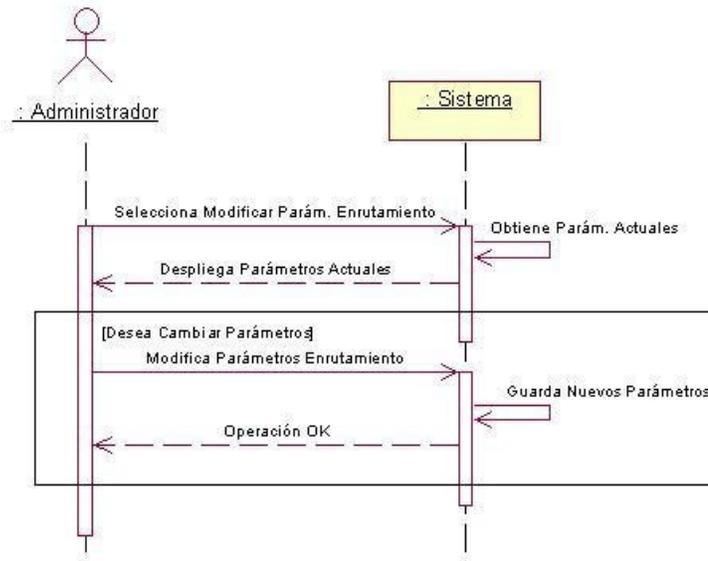
Ilustración 33: Diagrama de Secuencia de Monitorear Desempeño de Clasificación



La ilustración 33, detalla el diagrama de secuencia del caso de uso monitorear desempeño de clasificación. El administrador podrá evaluar el desempeño de los créditos de los distintos segmentos, que ya han tenido algún vencimiento. De esta forma, se podrá visualizar el comportamiento de pago de los clientes de los distintos rubros, en todos los periodos de tiempo que desee.

7.6.5 Caso de Uso: Modificar Parámetros de Lógica de Clasificación

Ilustración 34: Diagrama de Secuencia de Modificar Parámetros de Lógica de Clasificación



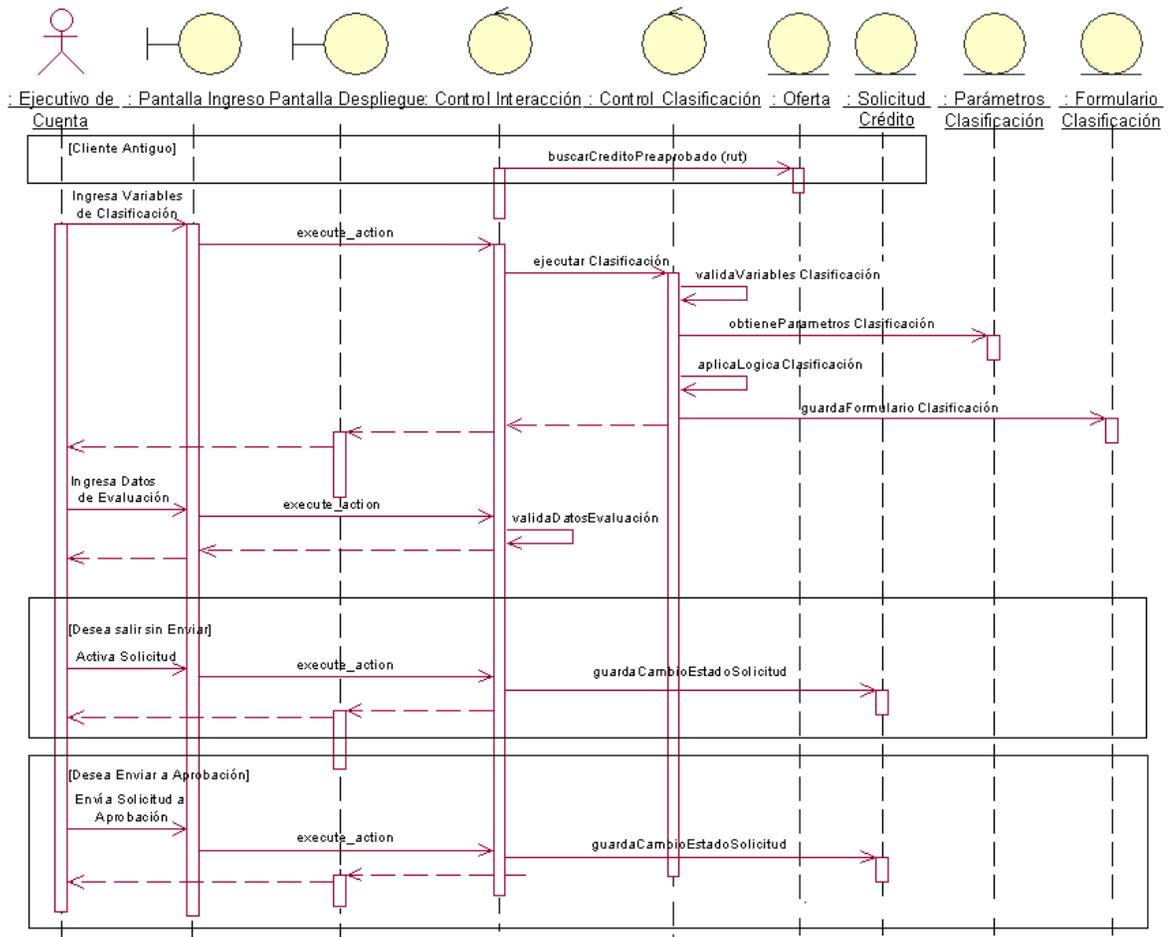
Para modificar los parámetros de la lógica de clasificación, al administrador se le desplegarán los parámetros actuales. En caso que se modifique alguno, su rango será validado y su nuevo valor almacenado.

7.7 Diagramas de Secuencia de Sistema

Sólo se explicitará los diagramas de secuencia de sistema, de aquellos casos de uso que no se encuentran implementados, es decir, se omitirá el DSS de “Ingresar Datos del Cliente”.

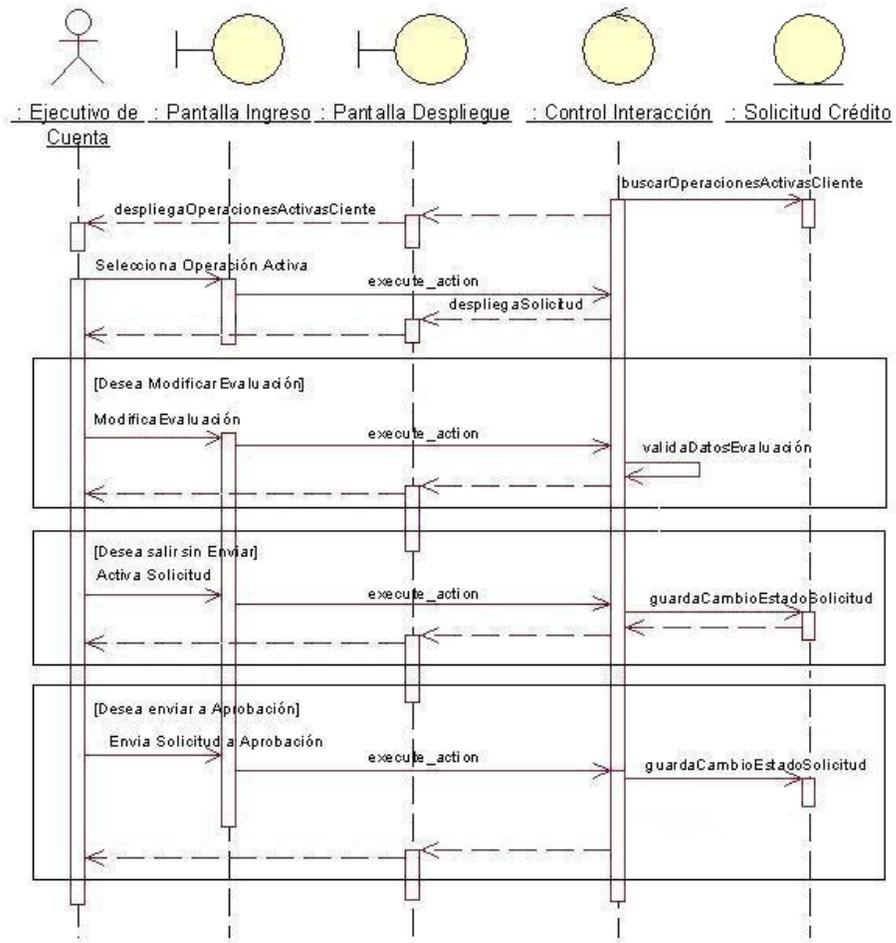
7.7.1 Llenar Evaluación del Cliente

Ilustración 35: DSS de Llenar Evaluación del Cliente



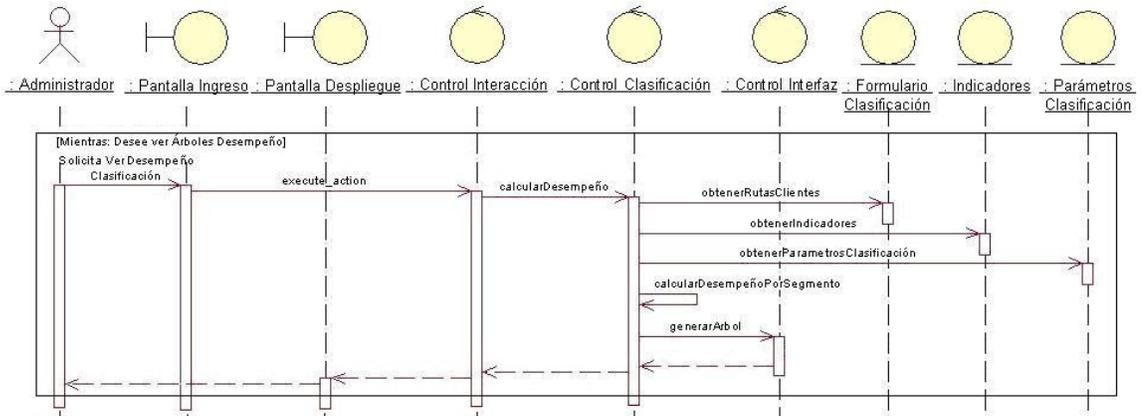
7.7.2 Modificar Evaluación del Cliente

Ilustración 36: DSS de Modificar Evaluación del Cliente



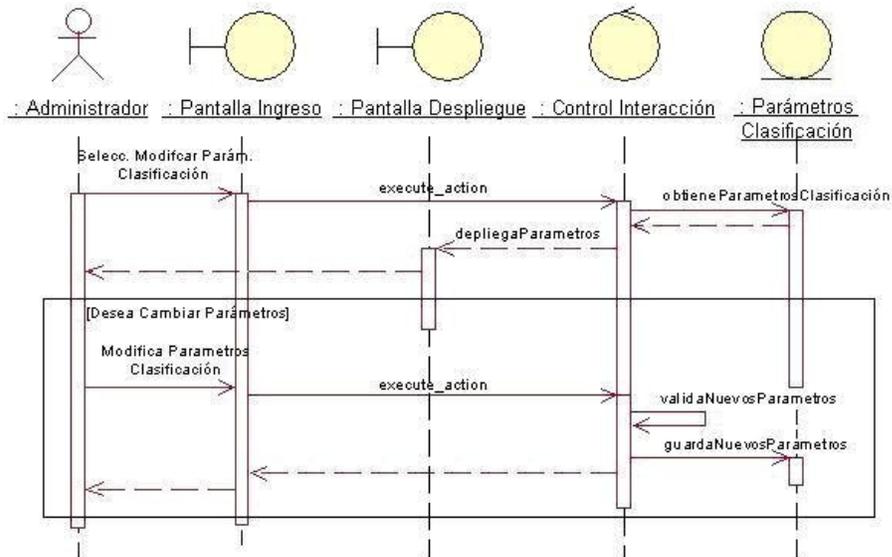
7.7.3 Monitorear Desempeño de Clasificación

Ilustración 37: DSS de Monitorear Desempeño de Clasificación



7.7.4 Modificar Parámetros de Lógica de Clasificación

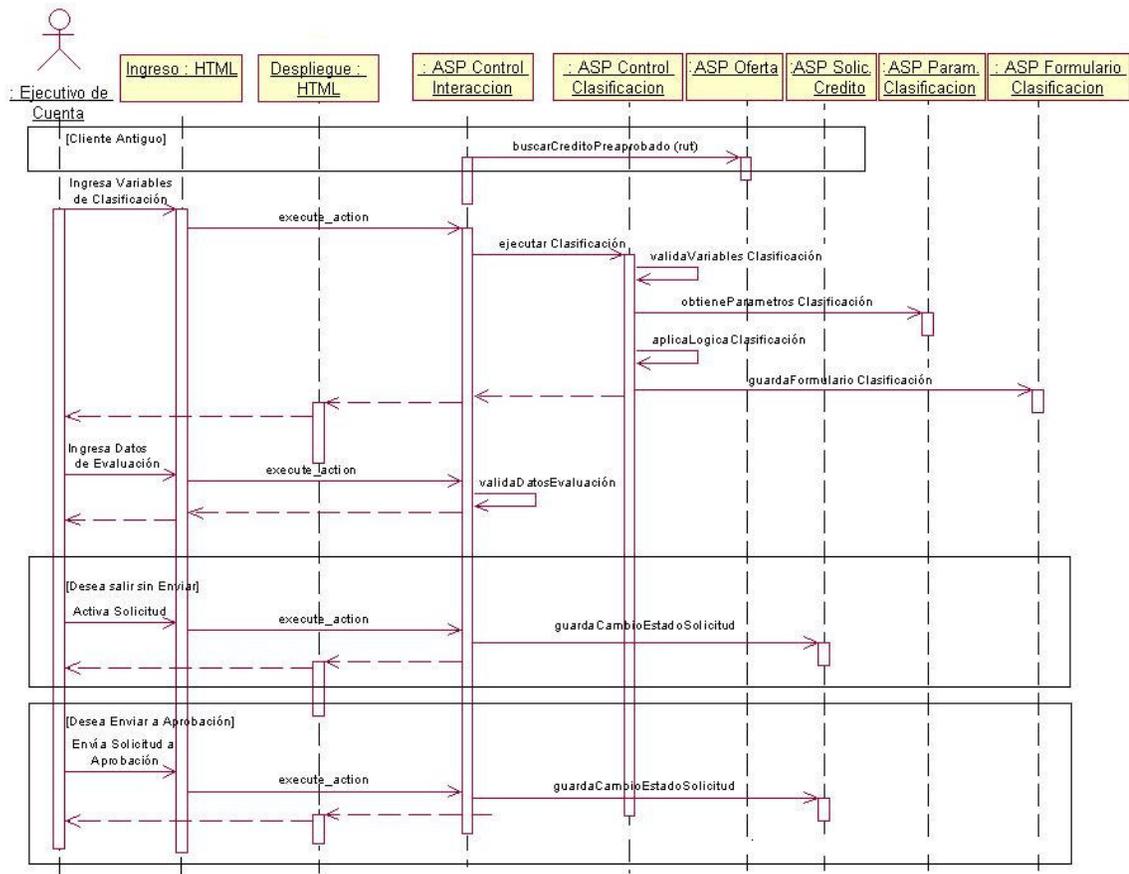
Ilustración 38: DSS de Modificar Parámetros de Lógica de Clasificación



7.8 Diagrama de Secuencia Físicos

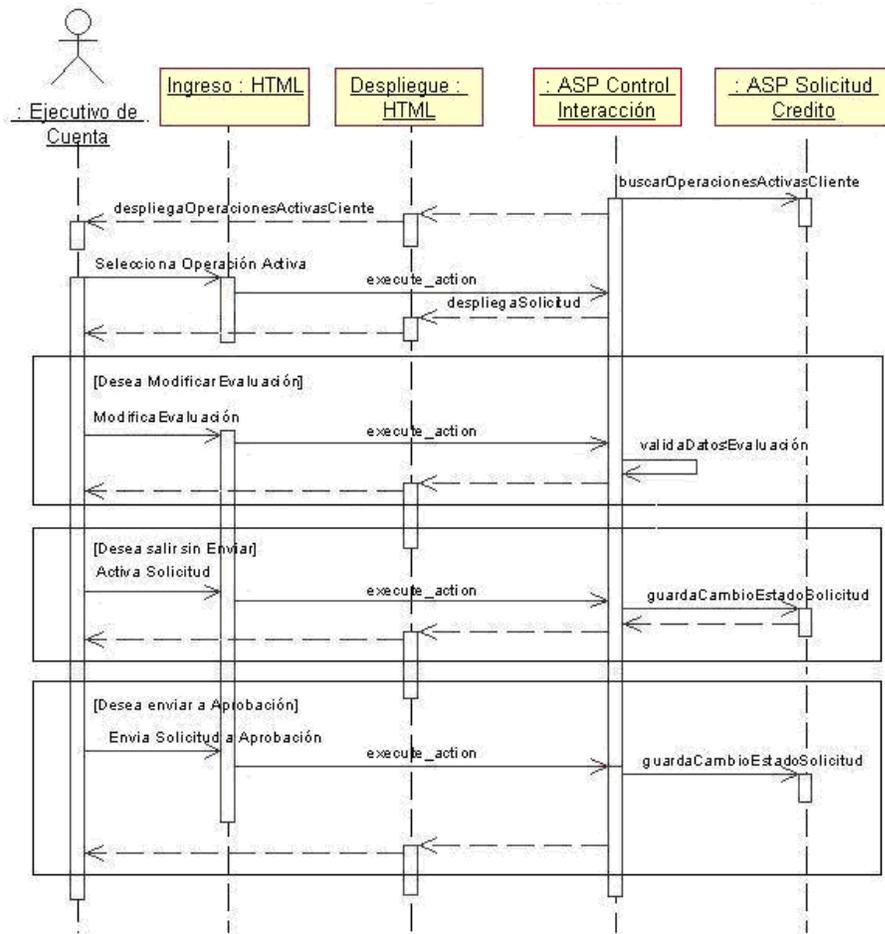
7.8.1 Llenar Evaluación del Cliente

Ilustración 39: DSF de Llenar Evaluación del Cliente



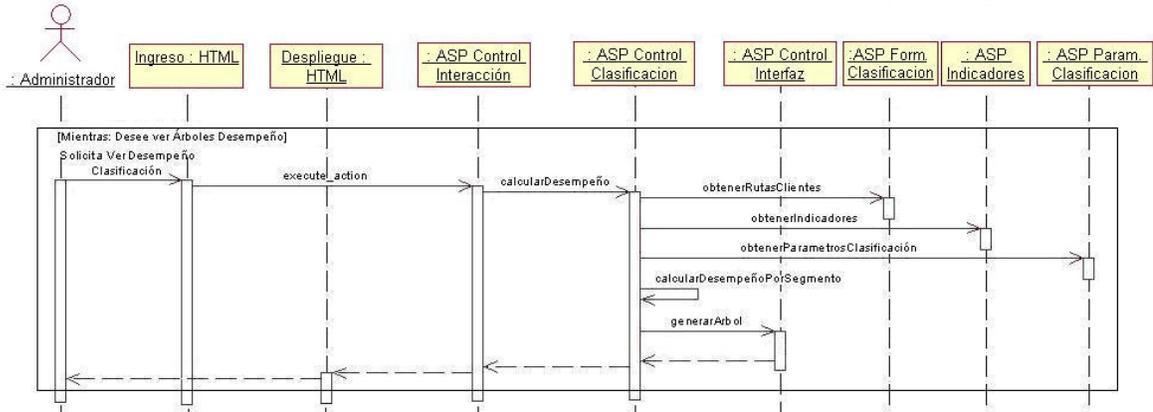
7.8.2 Modificar Evaluación del Cliente

Ilustración 40: DSF de Modificar Evaluación del Cliente



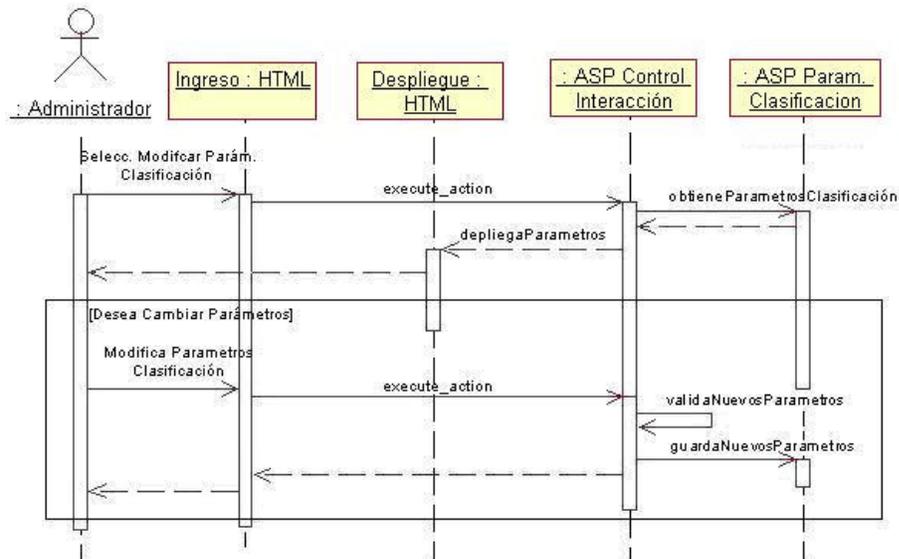
7.8.3 Monitorear Desempeño de Clasificación

Ilustración 41: DSF de Monitorear Desempeño de Clasificación



7.8.4 Modificar Parámetros de Lógica de Clasificación

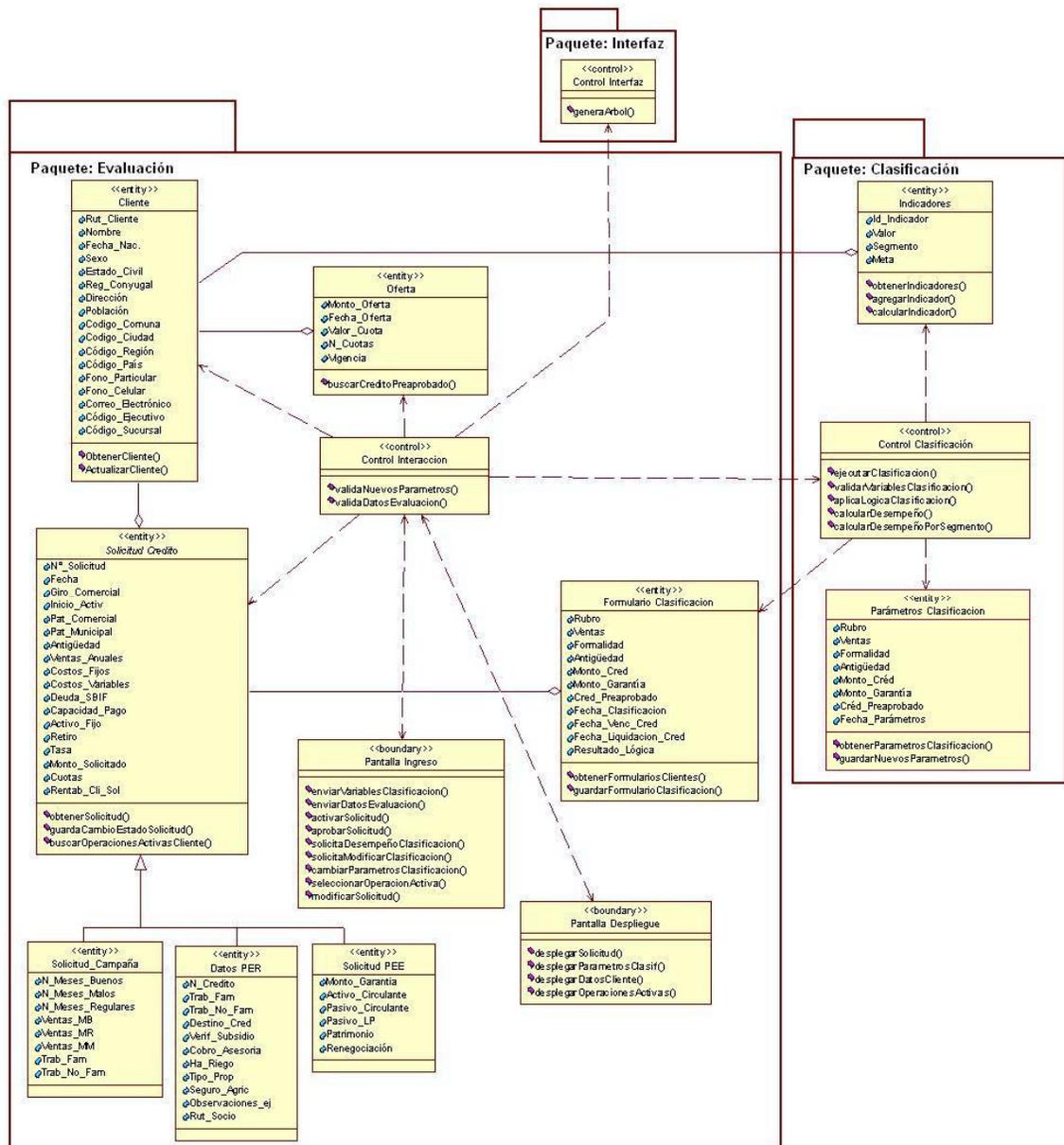
Ilustración 42: DSF de Modificar Parámetros de Lógica de Clasificación



7.9 Diagrama de Clases

Los paquetes del diagrama de clases, se muestran en la ilustración 43:

Ilustración 43: Diagrama de Clases



Las clases fueron agrupadas en 3 paquetes, de acuerdo a las funcionalidades que éstas soportan:

1. ***Paquete de Evaluación:*** Este paquete permite al ejecutivo evaluar al cliente, e incluye todas las relaciones necesarias para que esto ocurra. Así, es posible encontrar el “Control de Interacción”, que orquesta las demás clases para interactuar con el usuario (a través de las clases *boundary*), y que sólo ejerce algunas validaciones básicas sobre los datos que son ingresados. Adicionalmente, se encuentran clases *entity*, como Cliente, Solicitud de Crédito (abstracta), Oferta y Formulario de Clasificación, que son instanciadas directamente por el controlador, y que por sí mismas muestran la estructura de clases utilizada en el proceso de evaluación.

Este paquete se relaciona con el paquete de clasificación, exclusivamente para realizar la segmentación del cliente evaluado. Por último, requiere al paquete de interfaz, para generar la vista de los distintos árboles de desempeño.

2. ***Paquete de Clasificación:*** Este paquete interviene en la clasificación del cliente y en el control del desempeño de los segmentos. Para ello, cuenta con el “Control Clasificación”, que accede directamente a los

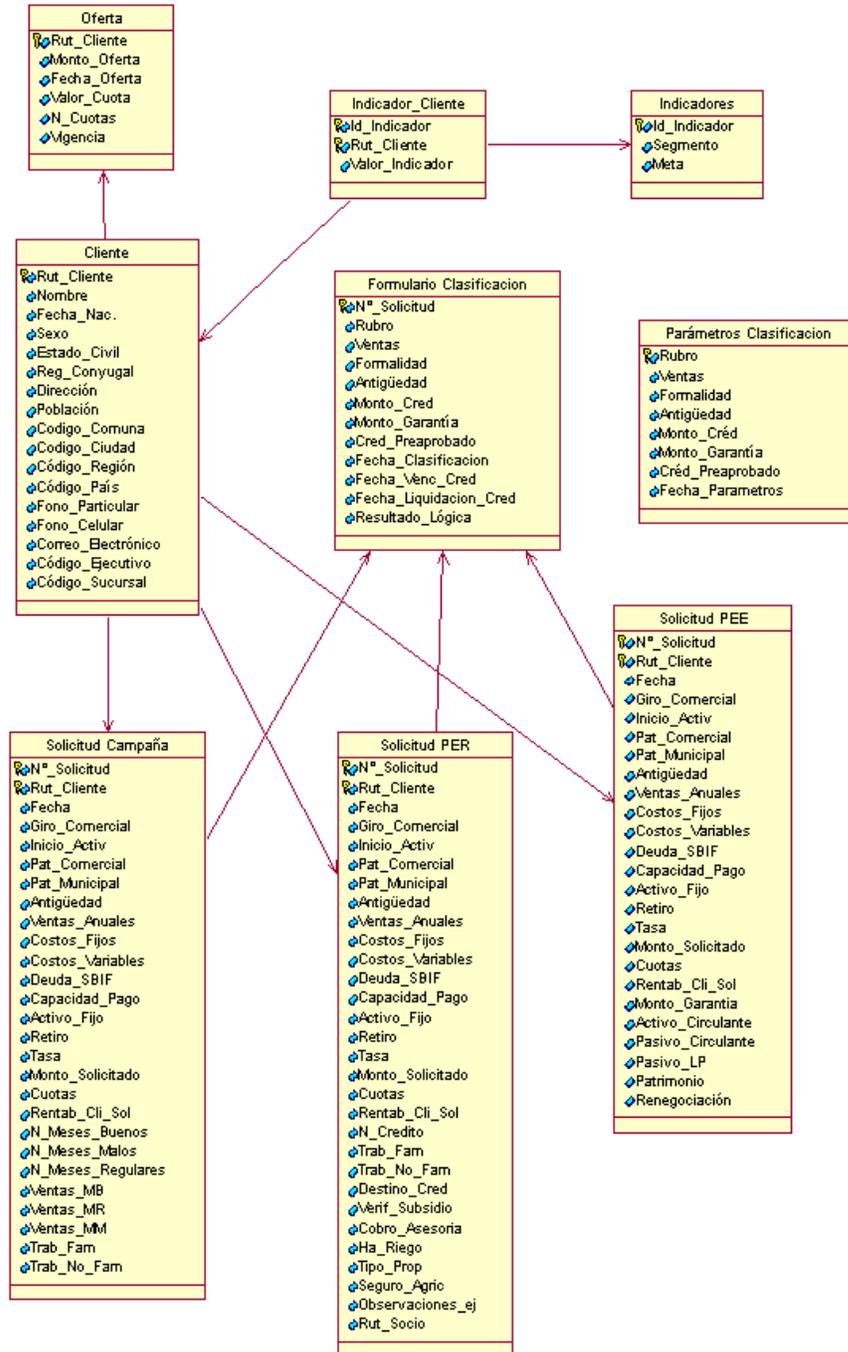
indicadores que necesitará para calcular el desempeño por segmento, y a los parámetros que utilizará para aplicar la lógica de clasificación. A esta información accede directamente en las clases que corresponden. Finalmente, este paquete se relaciona con el paquete de evaluación, para obtener los formularios de clasificación de los clientes, que le permitirán, en definitiva, calcular el desempeño de forma agregada (por segmento).

3. **Paquete de Interfaz:** Este paquete permite la generación de la interfaz de los distintos árboles de clasificación, y que ha sido separado de los otros paquetes, por la complejidad que esta lógica representa. Como ya se ha mencionado, sólo es utilizado cuando es requerido por el paquete de evaluación.

7.10 Modelo de Datos Físico

La estructura de datos existente en la empresa, más la requerida para el proceso de Evaluación de Clientes, dan origen al Modelo de Datos Físico que se muestra en la ilustración 44:

Ilustración 44: Modelo de Datos Físico



8. IMPLEMENTACIÓN ORGANIZACIONAL

8.1 Contexto Organizacional

La fusión entre Scotiabank y el Banco del Desarrollo, ha requerido abocar gran parte de la energía organizacional de ambas instituciones a concretar el proceso exitosamente. La complejidad de la situación es alta, sobre todo si se considera que las personas de la empresa absorbida se deben alinear con una nueva estrategia, deben adquirir prácticas laborales distintas, insertarse en una nueva estructura organizacional, e incluso aprender a utilizar otros sistemas. Justamente, los departamentos de informática de las dos empresas, se encuentran trabajando a tiempo completo en la unificación de sistemas, teniendo como consecuencia la detención de todos los proyectos que requiriesen de esa área, hasta que la fusión se complete.

Lo anterior no queda ajeno al proyecto que se expone en esta tesis, por lo que la implementación organizacional asociada quedará propuesta. Es decir, se describirá desde el punto de vista de la gestión del cambio, lo que se ha realizado hasta ahora y los desafíos enfrentados, para finalmente, proponer una estrategia para lograr un cambio exitoso.

La estrategia para cambiar el proceso de evaluación de clientes consistió, inicialmente, en identificar el problema y diseñar una solución a ser propuesta, para crear el sentido de urgencia. Como la Gestión del Cambio debe incluir cada una de las personas involucradas con el proyecto, lo primero fue identificarlas y establecer su vinculación con éste. De acuerdo a su poder, roles y relaciones, se definió quiénes serían los más idóneos para liderar el proyecto, y se estableció narrativas, para que cada uno entendiera sus ventajas y se comprometiera con su implementación. Finalmente, se propone continuar con una estrategia comunicacional cercana, para dar sentido a la propuesta, recogiendo las sugerencias de los actores que forman parte de él.

Lamentablemente, en un ambiente de incertidumbre como el que generan las fusiones, es difícil asegurar que los aspectos a considerar para la gestión del cambio serán los mismos, por lo que es recomendable esperar a que el proceso termine, para decidir si todavía tiene sentido lo propuesto antes que la fusión se iniciara.

8.2 Desafíos para la Gestión del Cambio

Los desafíos para la Gestión del Cambio, se plantean de acuerdo a cada una de las personas involucradas con el proyecto, desde los que deciden su financiamiento, hasta los relacionados con su operación. Éstos se exponen a continuación:

1. Seducir a la Gerencia de empresa, comprometiendo su apoyo para el financiamiento y desarrollo del proyecto.
2. Seducir a los mandos medios (jefes de producto, jefes zonales y de oficina), para contribuir con el diseño y comunicación del proyecto.
3. Seducir a los ejecutivos de cuenta, que son los que estarán relacionados operativamente con la evaluación de clientes. Lograr que sus prácticas estén alineadas con el proyecto y, en consecuencia, con la estrategia de la empresa.
4. Establecer una coalición conductora creíble y con autoridad suficiente para dar sentido al proyecto, y así lograr los puntos anteriores.

8.3 Estrategia para Gestión de Cambio

8.3.1 Sentido de Urgencia

La primera parte de la estrategia dice relación con la creación del sentido de urgencia en la dirección, para lograr su apoyo y obtener financiamiento para el proyecto. Para hacer surgir la necesidad de cambio, se reflejó, desde una perspectiva externa, el impacto que tiene la atención no estandarizada de clientes y su evaluación no segmentada. La idea, para solucionar lo anterior, sería centralizar y automatizar el criterio de atención a clientes, lo que traería como consecuencia alinear la estrategia con la ejecución operacional.

Además, se mostró que las evaluaciones de algunos segmentos escasamente reflejaban su realidad productiva y financiera, por lo que imperaba la necesidad de rediseñarlas en conjunto con especialistas de cada rubro, esto es, incluyendo a los jefes de producto. Es decir, se mostró la urgencia de rediseñar y darle un sentido a los instrumentos de evaluación, exponiendo sus falencias actuales. Para mostrar lo anterior, se adjunta la ilustración 45, que corresponde a la “ficha de evaluación agrícola”, utilizada en ese momento:

directos de la producción. Además, como está hecha en Word, las operaciones matemáticas deben ser hechas “a mano”. Es decir, es más bien una “encuesta”, donde los ejecutivos sólo escriben los datos proporcionados por el cliente, que no permite hacer mayores inferencias a partir de la información solicitada. Esto es especialmente relevante, por cuanto en los ejecutivos no estaba la idea que el objetivo de las “fichas” debe ser evaluar al cliente, que conlleva comprender su realidad productiva y financiera y, a partir de ahí, ofrecerle el producto más adecuado.

8.3.2 Primeros Pasos

Para comenzar el proceso de cambio, se decidió realizar prototipos de cada una de las evaluaciones que serían rediseñadas en Excel. Las razones para utilizar esta herramienta fueron principalmente su bajo costo, rapidez de producción y flexibilidad para hacer modificaciones. De esta forma, sería posible probar rápidamente las evaluaciones en algunas sucursales, hacer las modificaciones que fuesen necesarias, para luego utilizarlas a nivel nacional. Las nuevas evaluaciones deberían estar mejor estructuradas y tener el mayor grado de automatización posible.

El punto de partida sería la ficha anterior de cada rubro. A partir de ella, se elegirían los elementos que se desea conservar, y se analizaría los nuevos aspectos que se deberían incluir, dependiendo del rubro que se esté evaluando. Con toda esta información, sería necesario darle una nueva estructura a la evaluación, cuya lógica de ingreso de datos fuese ordenada y fácil de entender.

Las evaluaciones de prototipo permitirían, además, familiarizar a los ejecutivos con los nuevos instrumentos de evaluación, transmitirles su nuevo sentido, enseñarles a utilizarlos, responder sus dudas y hacerlos parte del proceso de cambio.

8.3.3 Gestión del Poder

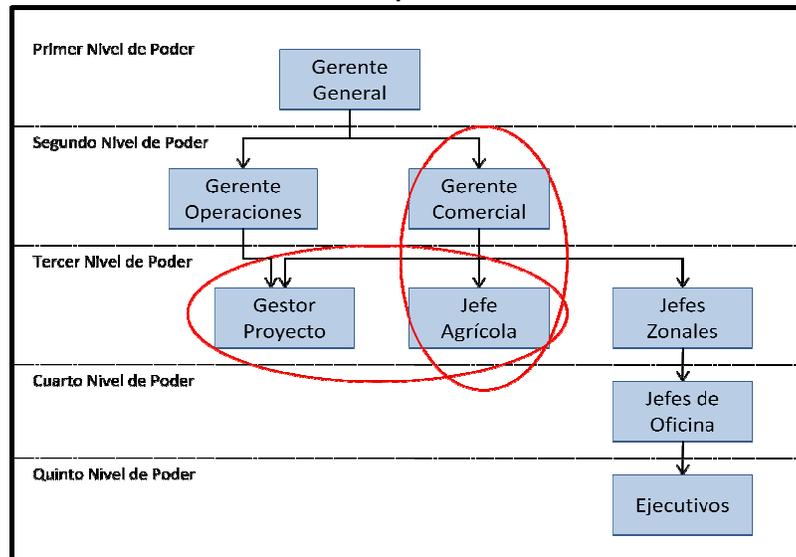
La Gestión del Poder es necesaria para que el proceso de cambio ocurra. Por esto, es importante identificar quiénes tienen poder en la organización, cuánto poder tienen y cuál es su posición frente al proyecto. Además, se debe analizar cuál es la relación interpersonal de los actores involucrados y el tipo de poder que tienen, aspectos que serán determinantes para definir la coalición conductora.

En el mapa de poder, que se encuentra en la ilustración 46, se observa que en el primer nivel se encuentra el Gerente General, Sr. Álvaro Retamales, quien ya se encuentra de acuerdo con el proyecto. Lo mismo ocurre para el Gerente de Operaciones, Sr. Mariano Mazur, y el Gerente Comercial, Sr. Andrés Bernal, que están en el segundo nivel del mapa. Éstos últimos son los que supervisarán los avances del proyecto y lo apoyarán directamente. Tal como se puede apreciar, se cuenta con el poder jerárquico para la realización del proyecto.

El poder experiencial y cognitivo de cada segmento de clientes, se encuentra en los jefes de producto, quienes tienen expertiz suficiente para mostrar a los ejecutivos cuáles aspectos deben considerar para evaluar los rubros y por qué. Adicionalmente, el poder cognitivo del aspecto técnico de la evaluación, es decir, de la herramienta Excel, se atribuye al gestor del proyecto.

Finalmente, se identifica con poder jerárquico a los jefes de oficina y jefes zonales, que ejercen el poder desde mucho más cerca que los Gerentes, por lo que contar con su apoyo es trascendental para lograr compromisos de cada oficina y zona.

Ilustración 46: Mapa de Poder



El contar con el apoyo gerencial, que representa el nivel del poder jerárquico más alto en la empresa, no necesariamente garantiza el protagonismo de los gerentes en el proyecto. Debido a lo anterior, se hace propicio diseñar una estrategia que los vincule periódicamente con el proyecto, de forma que su apoyo trascienda a los niveles jerárquicos menores y se mantengan como parte visible de su desarrollo. De hecho, la buena relación del Gerente Comercial con los mandos medios y la última línea, lo sitúan como la figura más idónea para liderar el proyecto.

El jefe de producto agrícola, Sr. Juan Carlos Dintrans, coincide con la idea de comenzar a “evaluar” a los clientes, transformándose en la primera persona del tercer nivel interesada en el proyecto. Su interés en las

evaluaciones, lo convierte en un factor clave para impulsar el proyecto, y la confianza que le tiene el Gerente Comercial, en un canal de comunicación importante entre el gestor y el segundo nivel de poder.

Por lo tanto, el apoyo del Jefe Agrícola, permitiría al gestor del proyecto mantener una comunicación fluida con el Gerente Comercial, el que, a su vez, le brinda el respaldo y confianza necesaria para tratar con los mandos medios restantes (jefes zonales y de oficina).

8.3.4 Definición de Coalición Conductora

De acuerdo a la distribución del poder, entre los distintos actores involucrados directa o indirectamente con el proyecto, se decide cuál debería ser la coalición conductora, para lograr un proceso de cambio exitoso.

El líder que surge de forma natural es el Gerente Comercial, pues otorga el poder jerárquico necesario para movilizar a la organización en el sentido de la innovación. Adicionalmente, el Gerente de Operaciones debe respaldar los cambios generados, por cuanto él ha sido, históricamente, el encargado del proceso de evaluación de clientes.

Por otro lado, los jefes de producto apoyarán tanto el diseño de las evaluaciones, como el aprendizaje organizacional de su segmento (de los jefes zonales, jefes de oficina, ejecutivos, etc.), a partir de su poder experiencial y cognitivo. Para que todo esto sea posible, el gestor del proyecto debe realizar un diseño adecuado de las evaluaciones, brindar soporte técnico constante y apoyar a los jefes de productos en su función ya mencionada. El cuadro 13 muestra los miembros de la coalición conductora:

Cuadro 13: Miembros de Coalición Conductora

NOMBRE	CARGO
Andrés Bernal	Gerente Comercial
Mariano Mazur	Gerente de Operaciones
Juan Carlos Dintrans	Jefe Producto Agrícola
Jimmy Rodriguez	Jefe Producto Pesca
Cristian Julio	Gestor del Proyecto

8.3.5 Gestión de Narrativas

La generación de narrativas, para los distintos actores de la organización, puede hacer la diferencia entre lograr implementar exitosamente un proyecto o no. Éstas deben depender del perfil y cargo de la persona que escucha, y del rol que se espera que cumpla en el proyecto. Por ello, se generaron narrativas distintas para los actores involucrados directa e indirectamente con el proyecto:

1. **Gerente General:** “Los ejecutivos no evaluarán de acuerdo a su criterio, sino de acuerdo a los lineamientos estratégicos de la empresa. Conoceremos mejor el comportamiento de nuestros clientes, identificando constantemente sus perfiles de riesgo y rentabilidad”.

2. **Gerente Comercial:** “Podrá controlar la rentabilidad y riesgo de los distintos segmentos de clientes. Con ello podrá tomar decisiones de mejora de evaluaciones, exclusión de segmentos no rentables, creación de nuevos productos, etc., logrando aumentar la rentabilidad de los clientes”.

3. **Gerente de Operaciones:** “Las evaluaciones serán automáticas, por lo que se podrá accederlas en línea, y el proceso de evaluación será más rápido y con menos errores”.

4. **Jefes de Producto:** “Evaluar segmentadamente a los clientes del rubro, instruyendo a los ejecutivos sobre sus características específicas, permitirá tomar mejores decisiones de otorgamiento de crédito, aumentando la rentabilidad de su producto”.

5. **Jefes Zonales y Jefes de Oficina:** “El nuevo proceso de evaluación, generará mejores colocaciones, disminuyendo el riesgo y aumentando la rentabilidad de su zona y oficina”.

6. **Ejecutivos de Cuenta:** “La nueva forma de evaluar es nada distinto a lo que han hecho siempre, sólo se han agregado aspectos que les ayudarán a entender mejor el negocio del cliente. Además, en el futuro será automatizada, por lo que podrán evaluar más rápido y atender mayor número de clientes. Con esto, tendrán una cartera de clientes más grande y de mejor calidad, lo que se verá reflejado en su sueldo”.

8.3.6 Estrategia Comunicacional

En esta empresa, se acostumbra “comunicar” todo por mail. De acuerdo a eso, es posible afirmar que las personas se encuentran bien informadas, pero la comunicación es deficiente. Este canal de comunicación, hace muy difícil otorgar un sentido a lo que se transmite, y deja mucho espacio para la libre interpretación.

Por esta razón, se propone una estrategia comunicacional basada en gestiones personales o telefónicas, que dé a la contraparte la instancia de expresar su opinión y aclarar dudas. En otras palabras, se plantea una estrategia que fomente una interacción bilateral, que incorpore la experiencia y opinión de todas las partes relacionadas con el proyecto.

Luego, el sentido del proyecto fue transmitido a los jefes zonales y jefes de oficina, para que éstos, a su vez, se lo transmitieran a los ejecutivos a su cargo. Para lograrlo, se fijaron reuniones telefónicas con cada jefe de oficina, en las que se simuló una evaluación de un cliente al azar utilizando los instrumentos rediseñados. De esta forma, se generaron los espacios para mostrar las consideraciones requeridas en el proceso y analizar sus sugerencias en conjunto, permitiéndoles ser parte del proyecto.

Por otro lado, para acompañar a los ejecutivos en el cambio a las nuevas evaluaciones, se les confeccionó un manual de usuario, que describiese claramente el objetivo del instrumento, segmento que evalúa y procedimiento para completarlo de forma adecuada. Para que la comunicación sea más efectiva, el gestor del proyecto y los jefes de producto, se encargaron de responder telefónicamente las dudas y sugerencias de cada ejecutivo. De esta forma, también sería posible gestionar directamente los estados de ánimo de

los ejecutivos y reforzar el sentido del cambio, aumentando su compromiso con el proyecto.

Como una forma de ilustrar mejor el foco de la estrategia comunicacional, se mostrará el diseño final de una evaluación, que ya se ocupa oficialmente a nivel nacional. Esta versión ha sido lograda gracias a la colaboración de todos los actores involucrados por el proyecto, desde los ejecutivos hasta los gerentes. En la ilustración 47, se mostrará la “Evaluación Silvoagropecuaria de Pequeños Productores”, señalando los factores de cambio que se debió enfrentar con los ejecutivos.

La nueva estructura es el primer cambio importante en la evaluación, ya que separa lógicamente los distintos aspectos que deben tomarse en cuenta al evaluar al microempresario. Por ejemplo, distingue los antecedentes del predio de los personales, ya que el cliente no necesariamente vive en el predio donde produce, ni es su dueño.

Otro cambio importante que debieron enfrentar los ejecutivos, es que la evaluación detalla qué partes del espacio físico de producción es utilizado para los distintos rubros del cliente, especificando cuándo y cuánto recibe por cada uno de ellos. Este aprendizaje no es trivial, si se considera que el rendimiento por hectárea, la frecuencia anual de producción y los meses de cosecha,

dependen del producto explotado. Mientras mejor se completen los datos de esta tabla, más exacta será la estimación de los ingresos reales del cliente.

Ilustración 47: Prototipo de Evaluación Silvoagropecuaria de PPA (1° Parte)



Versión: 01/Sep

EVALUACIÓN SILVOAGROPECUARIA DE PEQUEÑOS PRODUCTORES

Tipo de Proyecto:

Rut Cliente:

Fecha Solic.:

Sucursal:

Nombre Ejecutivo:

Visita Evaluación:

1.- ANTECEDENTES DEL CLIENTE

Nombre:

Dirección:

Comuna:

Teléfono 1:

Teléfono 2:

Estado Civil:

N° Integrantes Grupo Familiar:

Contacto en la Ciudad:

(Nombre, Parentesco, Direcc., Teléf.)

Inicio Actividades:

Patente Comercial:

Permiso Municipal:

Bancarización BDD:

Usuario INDAP:

Clasificación INDAP:

2.- ANTECEDENTES DEL PREDIO

Dirección:

Localidad:

Comuna:

Código:

Cobertura Comuna:

Giro Sil:

N° Trab. Familiares:

N° Trab. NO Familiares:

Tipo de Propiedad:

Rubro Cultivo Principal:

Clase Cultivo Principal:

Código del Rubro Ppal.:

Rubro Cultivo Secundario:

Clase Cultivo Secundario:

Código del Rubro Sec.:

N° Hectáreas físicas:

Clase de Suelo:

Factor de Corrección:

N° Hect. Riego Básico:

3.- ANTECEDENTES PRODUCTIVOS Y VENTAS ANUALES

Rubro	Hectáreas [Hás]	Rend. Anual por [Hás]	Unidad	Rend. Anual Total	Meses de Cosecha (desde - hasta)	Precio Venta (\$)	SubTotal (\$)
				0			0
				0			0
				0			0
				0			0
				0			0
TOTAL:							0

4.- ANTECEDENTES DEL CRÉDITO ANTERIOR PAGADO EN EL BDD

Num. Ope	Monto Otorgado	Fecha de Curse	Plazo	N° Renovaciones	Comp. Pago

5.- ANTECEDENTES DEL CRÉDITO AGRÍCOLA A RENOVAR O RENEGOCIAR

Situación del Crédito:

La evaluación continúa con el levantamiento del estado de situación y la estimación de la capacidad de pago del cliente. Las ventas se derivan de la parte anterior, y los costos y deudas, se detallan en un anexo, al que se accede a través de los links azules. Esta segunda parte, se muestra en la ilustración 48.

Ilustración 48: Prototipo de Evaluación Silvoagropecuaria de PPA (2° Parte)
6.- ESTADO DE SITUACIÓN

ACTIVOS FIJOS	Valor Comercial [\$]	Respaldo Adj	OBSERVACIONES
Bienes Raíces (casa)			
Predio Agrícola Propio			
Predio Agr. De Tercero			
Vehículos			
Maquinaria			
Otros Activos Fijos			
TOTAL:	0		

ENDEUDAMIENTO	Monto [\$]
Deuda Corto Plazo (con Int.)	0
Deuda Largo Plazo (con Int.)	0
Deuda Total (con Int.)	0

[Ir a Detalle de Deuda:](#)

7.- ESTIMACIÓN DE FLUJOS Y MOB DEL CLIENTE

Concepto	Anual [\$]	Mensual [\$]
Ventas Totales	0	0
Retiros Familiares	0	
Costos Fijos	0	0
Costos Variables	0	0
Otros Ing. Respaldados	0	
Excedente Bruto	0	0
Exced. Ponderado (70%)	0	0
Cuotas de Corto Plazo	0	0
Monto Máx. de Crédito	0	0

MOB	Monto [\$]
Flujo Mensual	
Acum. Año Actual	
Acum. Ult. 12 Meses	

[Ir a Detalle de Costo:](#)

8.- CARTA DECLARACIÓN Y MANDATO: SI

9.- DECLARACIÓN JURADA COMPRA DEUDA AGRIC.:

El anexo recién mencionado, significó otro desafío importante para la comunicación del proyecto, principalmente por el detalle de costos fijos y variables, puesto que los ejecutivos desconocían los costos involucrados en la

producción agrícola. Como éstos dependen del rubro del cliente, se les especificó los más comunes, para ayudarlos en la recolección de información. El detalle de costos y deudas, es como se especifica en la ilustración 49.

Ilustración 49: Prototipo de Evaluación Silvoagropecuaria de PPA (Detalle de Costos y Deudas)

DETALLE DE COSTOS FIJOS Y VARIABLES						
COSTOS FIJOS (Anuales)		COSTOS VARIABLES (Anuales)				
Unidad de Costo	Monto [\$]	Unidad de Costo	Monto [\$]			
Acciones de agua		Arriendo Maquinaria, Vehíc. y Equipos				
Agua, Gas, Electricidad		Combustible				
Arriendo del Predio		Preparación de suelos				
Pago Contador		Semillas				
Pago Contribuciones		Fertilizantes				
Fletes		Pesticidas				
Mantención Maquinarias y Equipos		Otros Insumos				
Mediería		Personal NO Permanente				
Personal permanente		Otros Costos Variables (detallar)				
Teléfono y/o Movil						
Otros costos Fijos (detallar)						
Total Costos Fijos:	0	Total Costos Variables:	0			

DETALLE DE ENDEUDAMIENTO						
Plazo (Venc. Menor a 1 año)	Institución Acreedora	Monto Cuota [\$]	Tipo de Cuota	Nº Cuotas por Pagar	Total Deuda [\$]	Estado Deuda
Deuda Financiera BDD	Banco del Desarrollo				0	
Deuda Fin. Otras Inst.					0	
Deuda Comercial					0	
Deuda INDAP	INDAP				0	
Deudas Largo Plazo (Venc. Mayor a 1 año)						
Deuda Hipotecaria					0	
Deuda Financiera LP					0	
Deuda INDAP LP	INDAP				0	
Otra Deuda LP					0	

El último cambio importante en la nueva evaluación, dice relación con el vencimiento recomendado para la primera cuota del crédito. En ese campo, se

recomienda automáticamente un intervalo²⁶ de dos meses, para que el ejecutivo fije el pago de dicha cuota. Esto también requirió de una interacción constante con los ejecutivos, por cuanto estaban acostumbrados a fijar el pago exactamente a un año del otorgamiento del crédito. Lo anterior aumentaba el riesgo de las colocaciones, ya que el vencimiento de la cuota podría estar fijada para meses muy posteriores a los que el cliente recibía los flujos.

Gracias a este cambio, la propuesta de crédito del ejecutivo sería más asertiva y el riesgo de no pago disminuiría. La última parte de la evaluación, donde el ejecutivo establece las condiciones del crédito con el cliente, se muestra en la ilustración 50-

Ilustración 50: Prototipo de Evaluación Silvoagropecuaria de PPA (3° Parte)
10.- PROPUESTA DE CRÉDITO DEL EJECUTIVO

Producto:		Línea de Financiam.:	
Monto Propuesto:		Destino del Crédito:	
Tasa Interés:		Número Crédito BDD:	
Vencim Recomendado:		Verificación Subsidio:	
Fecha 1er. Vencim.:		N° Subsidio:	
Plazo Créd. (días):		Seguro Agrícola:	
Horiz. de Pago (años):			
Código Convenio:			
Detalle de Inversión, Aval y Observaciones			

²⁶ El cálculo del intervalo, se hizo utilizando la “duración” de los flujos estimados de los rubros del cliente. La “duración”, refiere al concepto financiero que equivale al centro de masa de los flujos. A partir del mes resultante, se otorgan dos meses de holgura para que el cliente cancele el crédito, dando origen al intervalo mencionado.

Aunque existe un ítem de compromiso de pago, de aprobación del crédito y un anexo para las renovaciones, éstos no serán expuestos, por ser considerados no relevantes para ilustrar la estrategia de cambio.

Finalmente, como una forma de reforzar los elementos abarcados en la estrategia de cambio, el jefe de producto agrícola solicitó copia de todas las evaluaciones que fueran enviadas a aprobación, y revisó diariamente cada una de ellas. En caso que hubiera errores, se comunicó directamente con el ejecutivo, tomando un rol activo en la comunicación del sentido de cada una de las partes de esta evaluación.

8.3.7 Evaluación y Cierre del Proceso de Cambio

Para la evaluación del proceso de cambio, se dejará propuesto cuáles son los aspectos de deben ser considerados antes de dar el proceso por cerrado. En pos de lograr que el proceso de evaluación de clientes que se esté utilizando, efectivamente esté en la línea de lo que propone este proyecto, debe ocurrir que:

1. Se haya creado la conciencia organizacional, sobre todo a nivel de ejecutivos, de la importancia de evaluar los clientes para mejorar la calidad de la cartera. Esto es, comprender cabalmente la naturaleza y funcionamiento del negocio del cliente, para proponer las condiciones del crédito.
2. El Gerente Comercial, o la persona designada, revise periódicamente los indicadores de desempeño de cada uno de los segmentos atendidos.
3. Se tomen medidas para el mejoramiento continuo de las evaluaciones y de la cartera de clientes, a partir del análisis del punto anterior. Esto implica, entre otras cosas, rediseño constante de las evaluaciones y de la lógica de clasificación de clientes.

En otras palabras, el proceso de cambio recién podrá darse por cerrado, cuando se demuestre que el nuevo proceso de evaluación, control y análisis segmentado de clientes, mejora el desempeño del negocio y entrega una perspectiva más amplia de su funcionamiento. Más aún, cuando los aspectos antes mencionados, sean parte del funcionamiento natural de la empresa.

9. PRUEBA DE CONCEPTO

Según ya se ha comentado previamente, como la implementación de proyectos que requieran desarrollo informático, se encuentra detenida por la fusión, no es posible realizar un prototipo del proceso de evaluación de clientes, que sea más sofisticado que lo desarrollado hasta ahora. Por esta razón, se realizará una prueba de concepto para el proceso de mantención del modelo de comportamiento de clientes (*scoring*). A través de ésta, se busca probar cuantitativamente que el procedimiento propuesto en este trabajo, efectivamente puede ser utilizado para decidir si un modelo predictivo, basado en regresión logística, debe ser re-calibrado o no.

El paquete estadístico que se utilizará para la prueba de concepto, será SPSS®.

9.1 Contexto

En el tiempo en que el modelo de *scoring* fue desarrollado, se debió tomar la decisión de utilizar un único modelo para todos los clientes (global), o bien, un modelo para los clientes nuevos y otro para los antiguos. La primera

opción era la más simple, pues se ocuparían las mismas variables para todos los clientes, y no se requeriría la incorporación de variables de comportamiento, como sería necesario en la segunda opción (para los clientes antiguos). Por otro lado, tener dos modelos (para clientes nuevos y antiguos), tenía la ventaja de predecir de forma más efectiva el riesgo de los clientes, gracias a la utilización de variables distintas y propias de cada segmento. Justamente, por este motivo se decidió utilizar dos modelos en vez de uno.

Luego de escoger aplicar un modelo distinto para nuevos y antiguos, se debió decidir qué punto de corte utilizar. Como el Banco cuenta con una instancia de aprobación de crédito, se decidió utilizar dos puntos de corte, de forma de generar tres zonas de riesgo en cada modelo: alto, medio y bajo. Así, por ejemplo, suponiendo que los cortes escogidos fueron a y b ($a < b$), si la probabilidad de no pago $p < a$, entonces el cliente sería de riesgo bajo y el crédito podría ser aprobado. Por otro lado, si $a < p < b$, el cliente estaría en zona de riesgo medio, por lo que la decisión de otorgamiento del producto es de exclusiva responsabilidad de la instancia de aprobación. Finalmente, si $p > b$, el crédito debe ser rechazado, sin excepción. En este contexto y bajo esta modalidad, el *scoring* comenzó a ser utilizado en Marzo del año 2008.

Lamentablemente, para efectos de probar el proceso de mantención, la obtención de algunas variables de comportamiento tiene un costo de tiempo y recursos muy altos. Por esta razón, se optó por probarlo sólo con el modelo global. Si bien es cierto, éste no es el que se utiliza actualmente en el Banco, será de igual utilidad para establecer argumentos cuantitativos suficientes de validez del proceso en cuestión.

9.2 Recolección y Transformación de los Datos

El requerimiento de datos, que serán utilizados en esta prueba de concepto, se hizo de acuerdo a los mismos criterios que cuando el modelo fue calibrado. Es importante mencionar, que el *scoring* fue calibrado con cerca de 7500 registros de créditos, con una proporción de malos del 20%, que fueron otorgados a partir el año 2004 y que no vencieron después del 31 de Octubre del 2006. Esto dejaría espacio, entonces, para utilizar los créditos desde fines del año 2005, hasta fines del 2007, para probar el proceso de mantención.

En términos generales, se solicitaron los datos de los créditos no renegociados, con determinados códigos de producto, cuyo vencimiento no

fuese posterior al 31 de Agosto del año 2008, y que contuvieran las variables utilizadas en el modelo global. Estas variables son:

- *CasaProp*: Vale 1 si el cliente tiene casa propia, y 0 sino.
- *Edad*: Edad del cliente, en años.
- *Cump_Imp*: Variable categórica, que indica la situación de pago de impuestos del cliente. Si no declara impuestos, toma el valor 1. Si declara y paga menos el 90% de ellos (inclusive), vale 0, y si declara y paga más del 90% de los impuestos, vale 2.
- *NCred1*: Variable que indica el número de crédito que ha tenido el cliente con el Banco del Desarrollo, incluyendo el que se está solicitando.
- *Vehic*: Vale 1 si el cliente tiene vehículo, y 0 sino.
- *Reg_M*: Vale 1 si el cliente se encuentra en la Región Metropolitana, y 0 sino.
- *SerBas*: Vale 1 si el cliente paga sus servicios básicos al día, y 0 sino.
- *SiNuevo*: Vale 1 si el cliente es Nuevo en el Banco del Desarrollo, y 0 sino.
- *Ln (Deuda Comercial + Deuda Consumo)*: aplica logaritmo a la suma de las deudas del cliente, al momento del solicitar el crédito.

Adicionalmente, se solicitaron otras variables, que serían ingresadas como “candidatas” en el proceso de re-calibración, en caso que fuese necesario.

La experiencia previa del desarrollo del modelo, permitió tener líneas claras sobre cómo la variable objetivo debía ser definida, y cómo las variables independientes debían ser transformadas. De forma complementaria, se eliminaron todos los registros que tuvieran datos faltantes o fuera de rango, de forma de contar con una base “limpia” para probar el proceso.

En consecuencia, al final de esta etapa, se contaba con un set con 2816 créditos, con una proporción de malos cercana al 30%, para comenzar la mantención del *scoring*.

9.3 Mantención del Modelo de Comportamiento de Clientes

Dado que el *scoring* empezó a aplicarse en el año 2008, sólo será posible analizar su efectividad utilizando un punto de corte, ya que no se cuenta con el resultado de los créditos que fueron otorgados apoyándose en la predicción del modelo. Es decir, es imposible saber, para los créditos

catalogados de “riesgo medio” en el set de datos obtenido, cuál habría sido la decisión de la instancia de aprobación. Debido a lo anterior, se analizará el modelo actual utilizando el mejor punto de corte de la calibración, y se comparará con el desempeño del mejor corte de los modelos que sean propuestos.

Finalmente, todas las salidas de SPSS® que no sean mostradas explícitamente en el desarrollo del proceso, podrán ser encontradas en Anexos.

9.3.1 Análisis del Modelo Actual

El primer paso correspondió a seleccionar el *sub-set* de datos para analizar la efectividad del modelo actual. Situando al analista en Septiembre del año 2008, se ha supuesto que éste desea realizar un análisis de los créditos entregados en los últimos 15 meses. Es decir, se utilizarán los créditos solicitados después del 1° de Mayo del 2006. La cantidad de registros que cumplen esta condición, son 1327.

1. Estabilidad Poblacional

El análisis de estabilidad poblacional de la muestra, se deriva de la información contenida en el cuadro 14:

Cuadro 14: Promedio y Rango permitido de cada Variable (Modelo Actual)

Variable	Promedio (t)	Promedio (t-1)	Cota Inferior	Cota Superior	Estado
Casaprop	0,381	0,415	0,253	0,576	OK
Edad	44,181	43,604	30,498	56,710	OK
Cump_Imp (1)	0,472	0,529	0,00	1,00	OK
Cum_Imp (2)	0,452	0,403	0,251	0,556	OK
N_Cred	1,546	1,475	0,800	2,150	OK
Vehic	0,250	0,239	0,138	0,341	OK
Reg_M	0,187	0,301	0,092	0,510	OK
SerBas	0,393	0,398	0,154	0,642	OK
SiNuevo	0,518	0,207	0,090	0,324	FUERA DE RANGO
LN(Deuda Com+Cons)	4,141	28,488	5,343	41,879	FUERA DE RANGO

(1) = No declara Impuestos; (2) = Declara, y cumple >90%

En la primera columna, se encuentra el promedio de cada variable en el set de datos incluido. En las columnas siguientes, se encuentra su promedio anterior (de cuando el modelo fue calibrado), y sus cotas inferiores y superiores, respectivamente. La última columna, muestra si el promedio actual está dentro o fuera del rango permitido.

Como se puede observar, la mayoría de las variables han variado muy poco desde la calibración del modelo, salvo las variables SiNuevo y ln(deuda comercial + consumo).

Cuando el *scoring* se calibró, la variable SiNuevo tenía un valor de 0.207, lo que significa que la proporción de clientes nuevos era de un 21%. El hecho que su nuevo valor sea 0.518, indica que la proporción de clientes nuevos ha aumentado en más de un 30%. Esto no es algo inesperado, dada la estrategia de masividad de Banderarrollo Microempresas en los últimos dos años, que ha traído como consecuencia la integración de gran cantidad de clientes nuevos a la cartera. Por lo tanto, dado que la población ha cambiado en esa variable, es probable que su coeficiente deba ser re-calibrado, lo que no significa necesariamente que ésta ya no sea útil para el modelo. Esta decisión deberá ser tomada en base a las salidas de las siguientes etapas.

Por otro lado, el promedio de la variable que incluye la deuda ha disminuido de 24.48 a 4.14 millones. Pese a que este valor no se encuentra tan lejos de la cota inferior, es posible explicar su disminución por el gran aumento de la proporción de clientes nuevos a la cartera. Éstos se caracterizan por tener una deuda mucho menor que los clientes antiguos, por lo que se puede inferir que dicha disminución responde más bien a la nueva composición de la cartera, que a una nueva tendencia de endeudamiento de los clientes.

2. Análisis Univariado

El siguiente paso es analizar si las variables aún contribuyen al modelo a realizar sus predicciones. El análisis univariado, se muestra a continuación:

Cuadro 15: Análisis Univariado de Variables del Modelo Actual (con predicción binarizada)

Variable	Tipo de Test	Valor	Sig.
Casaprop	Chi Pearson	66,246	0,000
Edad	K-S	4,414	0,000
Cump_Imp	Chi Pearson	349,932	0,000
N_Cred1	K-S	8,782	0,000
Vehic	Chi Pearson	31,024	0,000
Reg_M	Chi Pearson	5,782	0,016
SerBas	Chi Pearson	31,091	0,000
SiNuevo	Chi Pearson	282,809	0,000
LN(Deuda Com+Cons)	K-S	4,809	0,000

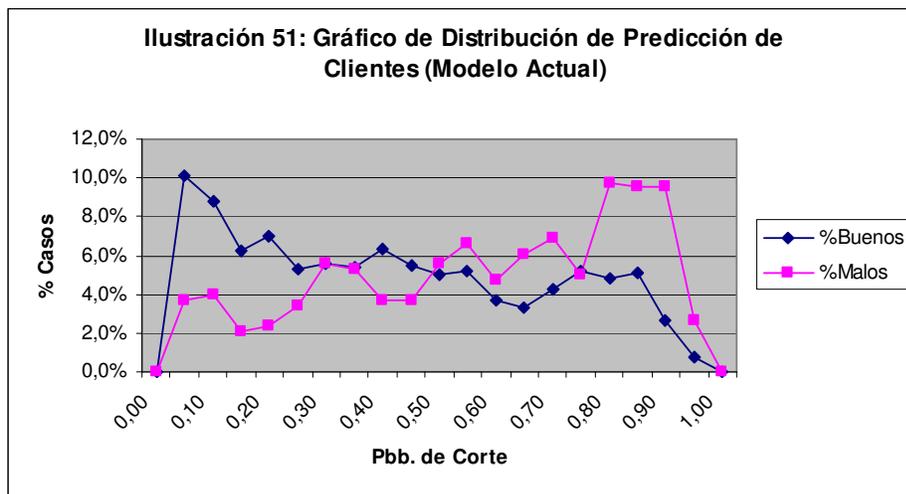
Claramente, se observa que cada variable por sí sola, le es importante al modelo para predecir la variable objetivo, con más de un 98% de confianza. Luego, el análisis univariado indica que todas las variables siguen siendo relevantes.

3. Coherencia del Modelo

Para continuar analizando el modelo actual, es necesario determinar si el modelo efectivamente predice probabilidades menores para los clientes buenos,

y mayores para los malos, en el set de datos ingresado. Para esto, se utilizó el corte 0.7, como se mencionó anteriormente.

El gráfico que refleja la coherencia del modelo, se muestra en la ilustración 51:



La distribución de predicción de clientes no es tan clara. Aunque ciertamente se está prediciendo probabilidades bajas para más clientes buenos que malos, revelando cierto grado de coherencia, la distinción pareciera un poco difusa en torno al punto 0.5. Además, pareciera que se están aceptando muchos malos con el punto de corte definido en la calibración anterior.

Como una forma de medir más precisamente la capacidad discriminante del modelo, se calculó el *C-Index*. Su valor es **0.671**.

Como no se cuenta con el valor del índice de concordancia de la calibración del *scoring*, éste será utilizado para comparar su capacidad discriminante con la de un eventual modelo re-calibrado.

4. Capacidad Predictiva del Modelo

En el cuadro 16, se muestra la matriz de confusión del modelo actual, al predecir los registros del set de datos ingresado:

Cuadro 16: Matriz de Confusión e Indicadores (Modelo Actual)

	Pred. = 0	Pred = 1	Total
Obs. = 0	773	175	948
Obs. = 1	241	138	379

	Pred. = 0	Pred = 1	Total
Obs. = 0	81,5%	18,5%	100%
Obs. = 1	63,6%	36,4%	100%

Acierto Global	68,7%
Efectividad	59,0%
Cobertura	76,4%

En la matriz se puede apreciar que el modelo acepta un 81.5% de los clientes buenos, y rechaza un 36.4% de los malos.

El acierto global refiere a qué porcentaje de los clientes el modelo acertó en su predicción, y su valor alcanza el 68.7%. La efectividad es un 10% menor,

dado que existe una mayor proporción de clientes buenos en el set de datos. Finalmente, la cobertura muestra que, esta configuración del modelo, permitiría atender a un 76.4% del total de clientes que soliciten un crédito.

La efectividad de este mismo modelo, al momento de calibración, fue de 65%, lo que releja una disminución considerable de su capacidad predictiva.

5. Alternativas de Punto de Corte

Ahora que se sabe cuál es el desempeño de la política de corte actual, es necesario compararla con cortes alternativos, mediante una función de utilidad, que considere el beneficio de aceptar un cliente bueno y costo de aceptar uno malo. Es decir, se verá si otro punto de corte aumentaría la utilidad de la empresa.

El cuadro 17, que contiene el desempeño de los distintos puntos de corte, se muestra a continuación:

Cuadro 17: Desempeño de distintos Puntos de Corte (Modelo Actual)

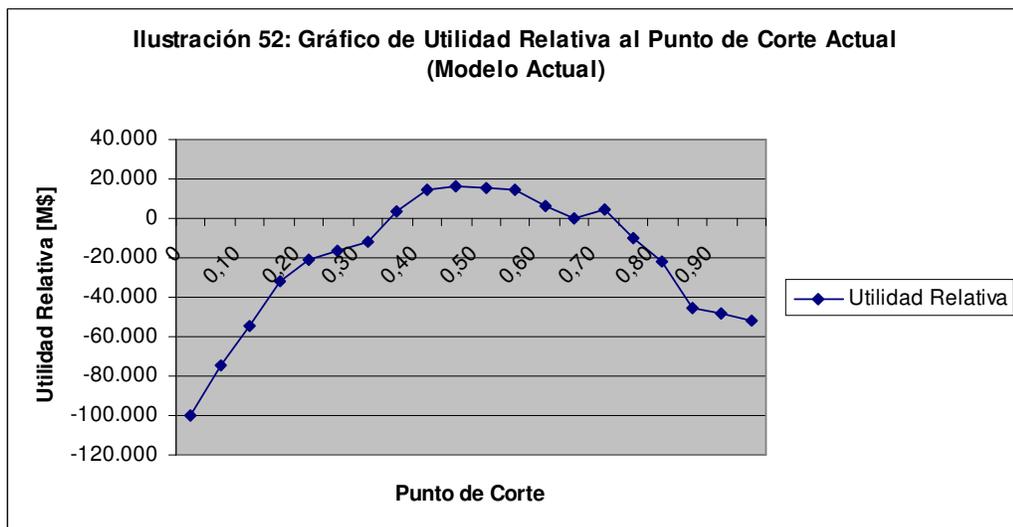
Pbb Corte	% Buenos Aceptados	% Malos Rechazados	% Efectividad	% Acierto Global	% Cobertura Total	Utilidad [M\$]	Utilidad Relativa [M\$]
0,05	10,1%	96,3%	53,2%	34,7%	8,3%	31.964	-99.660
0,10	18,9%	92,3%	55,6%	39,9%	15,7%	56.760	-74.863
0,15	25,1%	90,2%	57,7%	43,7%	20,7%	76.997	-54.626
0,20	32,1%	87,9%	60,0%	48,0%	26,4%	99.585	-32.038
0,25	37,3%	84,4%	60,9%	50,8%	31,1%	110.637	-20.987
0,30	42,9%	78,9%	60,9%	53,2%	36,7%	115.273	-16.350
0,35	48,3%	73,6%	61,0%	55,5%	42,0%	119.938	-11.685
0,40	54,7%	69,9%	62,3%	59,1%	47,7%	135.245	3.621
0,45	60,2%	66,2%	63,2%	61,9%	52,7%	146.268	14.644
0,50	65,2%	60,7%	62,9%	63,9%	57,8%	148.049	16.425
0,55	70,4%	54,1%	62,2%	65,7%	63,4%	146.860	15.236
0,60	74,1%	49,3%	61,7%	67,0%	67,4%	145.871	14.247
0,65	77,3%	43,3%	60,3%	67,6%	71,4%	138.076	6.452
0,70	81,5%	36,4%	59,0%	68,7%	76,4%	131.624	0
0,75	86,7%	31,4%	59,1%	70,9%	81,5%	136.317	4.693
0,80	91,6%	21,6%	56,6%	71,6%	87,8%	121.936	-9.688
0,85	96,6%	12,1%	54,4%	72,5%	94,1%	109.487	-22.136
0,90	99,3%	2,6%	51,0%	71,7%	98,7%	86.092	-45.531
0,95	99,7%	1,5%	50,6%	71,5%	99,2%	83.840	-47.784
1,00	100,0%	0,0%	50,0%	71,4%	100,0%	79.620	-52.004

Para la confección del cuadro anterior, se utilizó el beneficio y costo de un crédito de monto, tasa y plazo promedio, del periodo en estudio. Como resultado, se tiene un beneficio de M\$ 340 y un costo de M\$ 280.

Si el criterio de cambio de punto de corte, fuese exclusivamente la maximización de utilidad, entonces convendría “hacer más exigente” la política de selección, reduciendo el punto a 0.5. Si se toma esa decisión, se tendrá una ganancia, adicional a la actual, de 16.4 millones por cada 1330 clientes, aproximadamente, siempre y cuando la proporción de buenos se mantenga.

Pero para tomar esta decisión, es muy importante tener en consideración la estrategia de la empresa. Por ejemplo, si ésta apuntara al aumento de participación de mercado, entonces podría convenir mover el corte a 0.75, y obtener una utilidad menos alta, pero aún positiva (4.69 millones), incrementando la cantidad de clientes. Por otro lado, si la estrategia fuese la generación de una cartera rentable, con clientes de mejor calidad, puede ser conveniente reducir el corte a 0.5. Es más, puede darse el caso que la empresa desee aumentar la rentabilidad, pero que considere que el corte 0.5 excluye muchos clientes buenos, por lo que estaría dispuesta a ganar un menos y utilizar un corte de 0.6.

De forma complementaria, es posible ver la utilidad relativa al punto de corte actual, en el gráfico de la ilustración 52:



Independiente de la estrategia de la empresa, el análisis de los 5 aspectos del modelo actual, permite concluir que el modelo es coherente en la discriminación de clientes, y que todas las variables que utiliza, le aportan para predecir su comportamiento de pago. Sin embargo, la estrategia de “masividad” del año 2006 y 2007, ha aumentado significativamente la proporción de clientes nuevos de la cartera, repercutiendo además en la deuda promedio. Aún así, es posible lograr una mayor utilidad re-calibrar los coeficientes de modelo, mediante la variación de su punto de corte.

Por lo tanto, dado el cambio poblacional en las variables antes mencionadas, se procederá a re-calibrar del modelo, utilizando las mismas variables del modelo actual, esperando mejorar su desempeño.

9.3.2 Seleccionar Criterios de Re-calibración

Aunque hacer una re-calibración sólo del punto de corte, es una alternativa que genera utilidad a la empresa, es necesario volver a ajustar los coeficientes de las variables, por el cambio que ha experimentado la población. Debido a esto, se procederá a realizar una re-calibración con selección total de variables, utilizando las mismas variables del modelo actual, en el set de datos

completo (con los 2816 registros). Ingresando los parámetros $\alpha = 0.05$, para el análisis univariado, y $n = 10$, para la estabilidad del modelo nuevo, se procede el proceso de re-calibración.

9.3.3 Comparar Desempeño de los Modelos

1. Análisis Univariado

Primero, se realizó el análisis univariado de todas las variables incluidas en el modelo actual. Los resultados se muestran en el cuadro 18:

Cuadro 18: Análisis Univariado de Variables del Modelo Actual (con resultados reales)

Variable	Tipo de Test	Valor	Sig.
Casaprop	Chi Pearson	74,382	0,000
Edad	K-S	4,618	0,000
Cump_Imp	Chi Pearson	44,955	0,000
N_Cred1	K-S	5,148	0,000
Vehic	Chi Pearson	35,503	0,000
Reg_M	Chi Pearson	0,502	0,478
SerBas	Chi Pearson	12,101	0,001
SiNuevo	Chi Pearson	86,916	0,000
LN(Deuda Com+Cons)	K-S	1,661	0,008

La única variable que pareciera no ser relevante para predecir si un cliente es bueno o malo, es Reg_M. Es decir, la ubicación geográfica del cliente podría ya no tener incidencia en su riesgo de no pago. Para tener mayor

seguridad de este fenómeno, deberá analizarse si el coeficiente beta que la acompaña es significativo.

2. Entrenamiento del Nuevo Modelo

Antes de proceder con esta parte, se mostrará que el set de datos efectivamente fue dividido en un set de entrenamiento y de prueba, con un 80% y 20% de los casos, respectivamente. El cuadro 19, extraído de SPSS®, ratifica lo recién mencionado.

Cuadro 19: Porcentaje de casos para entrenar y probar el Modelo Nuevo

Case Processing Summary		N	Percent
Unweighted Cases ^a			
Selected Cases	Included in Analysis	2267	80,5
	Missing Cases	0	,0
	Total	2267	80,5
Unselected Cases		549	19,5
Total		2816	100,0

a. If weight is in effect, see classification table for the total number of cases.

En el cuadro anterior, los casos no seleccionados corresponden al set de prueba, y los seleccionados, al set de entrenamiento.

Ahora que es claro que la calibración del modelo nuevo fue realizada de forma adecuada, en el cuadro 20 se muestran sus coeficientes beta (B), calculados según el método de máxima verosimilitud, para un corte de 0.6:

Cuadro 20: Coeficientes Beta del Modelo Nuevo
Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,445	,131	11,542	1	,001	,641
	Edad	-,027	,005	25,722	1	,000	,973
	Cump_Imp			22,620	2	,000	
	Cump_Imp(1)	,520	,228	5,225	1	,022	1,682
	Cump_Imp(2)	,580	,125	21,692	1	,000	1,786
	NCred1	-,089	,044	4,034	1	,045	,915
	Vehic	-,664	,143	21,654	1	,000	,515
	Reg_M	,246	,137	3,226	1	,072	1,279
	SerBas	-,159	,121	1,714	1	,190	,853
	SiNuevo	,722	,169	18,128	1	,000	2,058
	LNDCCM1	,088	,020	18,547	1	,000	1,092
	Constant	,568	,304	3,502	1	,061	1,765

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, NCred1, Vehic, Region, SerBas, SiNuevo, LNDCCM1.

Todas las variables resultaron significativas al 95%, salvo la variable Región (Reg_M) y SerBas. Luego, es probable que en su interacción con las otras variables, estén aportando menos que antes en la discriminación clientes.

Para proseguir con la re-calibración, es necesario analizar el grado de ajuste del modelo a los datos (*chi-square*), más la fortaleza de relación entre las variables independientes y la variable objetivo (*Nagelkerke R square*). Estos resultados, se muestran en los cuadros 21 y 22, correspondientemente:

Cuadro 21: Ajuste del Modelo Nuevo

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	183,204	10	,000
	Block	183,204	10	,000
	Model	183,204	10	,000

Cuadro 22: R² de Nagelkerke Modelo Nuevo

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1758,410 ^a	,123	,163

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

El primer recuadro, corresponde al estadístico del *likelihood ratio* (G), que sigue una chi-cuadrado con 10 grados de libertad, y cuyo valor (183.204) muestra que el modelo se ajusta a los datos con un nivel de significancia del 99,9%. En el recuadro siguiente, se encuentra el R² de Nagelkerke, cuyo valor es igual a 0.163. Aunque dicho valor pareciera no ser suficientemente alto, muestra que existe algún grado de relación entre las variables independientes y la variable de salida, cuya fortaleza será cuantificada en la matriz de confusión.

3. Comparación entre el Modelo Actual y Nuevo

Para ilustrar cómo se determinó el punto de corte óptimo para el modelo nuevo, en el cuadro 23 se mostrará el desempeño de los distintos puntos de corte:

Cuadro 23: Desempeño de distintos Puntos de Corte (Modelo Nuevo)

Pbb Corte	% Buenos Aceptados	% Malos Rechazados	% Efectividad	% Acierto Global	% Cobertura Total	Utilidad [M\$]
0,05	0,0%	100,0%	50,0%	29,3%	0,0%	0
0,10	1,0%	100,0%	50,5%	30,1%	0,7%	1.924
0,15	2,6%	99,4%	51,0%	31,0%	2,0%	3.856
0,20	7,2%	99,4%	53,3%	34,2%	5,3%	12.516
0,25	13,7%	94,4%	54,0%	37,3%	11,3%	16.904
0,30	21,1%	89,4%	55,3%	41,2%	18,0%	23.217
0,35	31,2%	85,1%	58,1%	47,0%	26,4%	35.296
0,40	40,7%	78,3%	59,5%	51,7%	35,2%	42.594
0,45	52,1%	71,4%	61,7%	57,7%	45,2%	53.259
0,50	62,1%	65,2%	63,7%	63,0%	54,1%	62.473
0,55	70,4%	54,0%	62,2%	65,6%	63,2%	60.682
0,60	83,8%	42,2%	63,0%	71,6%	76,1%	67.557
0,65	87,1%	30,4%	58,8%	70,5%	82,0%	55.670
0,70	93,8%	19,9%	56,8%	72,1%	89,8%	51.947
0,75	95,9%	11,8%	53,8%	71,2%	93,6%	43.384
0,80	98,7%	3,1%	50,9%	70,7%	98,2%	35.309
0,85	99,7%	1,9%	50,8%	71,0%	99,3%	35.323
0,90	100,0%	0,0%	50,0%	70,7%	100,0%	32.940
0,95	100,0%	0,0%	50,0%	70,7%	100,0%	32.940
1,00	100,0%	0,0%	50,0%	70,7%	100,0%	32.940

Tal como se destaca en el cuadro anterior, el punto de corte que genera la utilidad máxima es 0.6. Para comparar más claramente el desempeño del modelo nuevo con el actual, se expondrán en los cuadros 24 y 35 las matrices de confusión e índices de concordancia de ambos modelos en el set de prueba, con sus respectivos puntos de corte.

Cuadro 24: Matrices de Confusión de Modelos Actual y Nuevo

Matriz de Confusión Modelo Actual

	Pred. = 0	Pred. = 1
Obs. = 0	81,1%	18,9%
Obs. = 1	63,5%	36,5%

Matriz de Confusión Modelo Nuevo

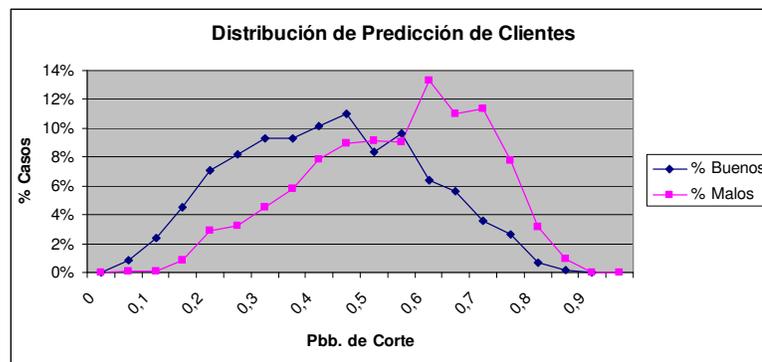
	Pred. = 0	Pred. = 1
Obs. = 0	82,7%	17,3%
Obs. = 1	56,9%	43,1%

Cuadro 25: Índices de Concordancia de Modelos Actual y Nuevo

	Modelo Actual	Modelo Nuevo
C - Index	0,677	0,697

De forma complementaria, se mostrará en la ilustración 53, el gráfico de distribución de buenos y malos que generan las predicciones del modelo nuevo:

Ilustración 53: Gráfico de Distribución de Predicción de Clientes (Modelo Nuevo)



La mayor efectividad y acierto global del modelo nuevo, también se ve reflejada en un mayor valor del índice de concordancia y distinción más clara

entre las distribuciones de clientes buenos y malos. Es decir, si se utilizara el modelo nuevo, se lograría mantener el porcentaje de acierto de los clientes buenos (pequeño aumento del 1%), pero *incrementando* el acierto de los malos, en cerca de un 7%. Para poder realizar dicha afirmación con mayor seguridad, es necesario mostrar que este resultado es estable, y no producto de la elección del *set* de prueba.

4. Evaluar Sobreajuste y Verificar Estabilidad del Modelo Propuesto

La salida para evaluar el sobreajuste del modelo, ha sido tomado directamente de SPSS®, y corresponde al cuadro 26:

Cuadro 26: Predicción del Modelo Nuevo en Sets de Entrenamiento y Prueba

Classification Table^d

		Predicted						
		Selected Cases ^a			Unselected Cases ^{b,c}			
		Var_Obj		Percentage Correct	Var_Obj		Percentage Correct	
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	562	135	80,6	140	29	82,7
		1,00	368	318	46,4	103	78	43,1
	Overall Percentage				63,6			62,2

a. Selected cases Filter2 EQ 1

b. Unselected cases Filter2 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Dado que la efectividad global del modelo varía en sólo un 0.5%, es posible afirmar que éste no se encuentra sobreajustado a la muestra. Sólo falta determinar si éste predice de forma estable. La estabilidad del modelo, puede ser analizada en el cuadro 27, que contiene la efectividad de predicción para el número de particiones ingresadas por el usuario (que son 10, en este caso):

Cuadro 27: Efectividad de Predicción del Modelo Nuevo, en 10 Particiones distintas

Estabilidad Modelo	% Efect. Buenos	% Efect. Malos	% Efect. Total
Partición 1	83,8%	42,2%	63,7%
Partición 2	82,7%	43,1%	62,2%
Partición 3	78,5%	41,1%	58,9%
Partición 4	82,2%	44,6%	63,5%
Partición 5	77,0%	48,2%	62,7%
Partición 6	82,7%	41,8%	61,3%
Partición 7	82,7%	38,7%	59,3%
Partición 8	80,6%	44,6%	62,4%
Partición 9	79,6%	41,5%	60,0%
Partición 10	81,9%	43,9%	63,2%
Promedio	81,2%	43,0%	61,7%
Desv. Estánd.	2,2%	2,6%	1,8%

Como la desviación estándar de la efectividad, para clientes buenos y malos, es pequeña, es posible concluir que la predicción del modelo recalibrado es estable, y mejor que la del actual. Es más, es posible establecer un intervalo de confianza al 95% para la efectividad en los clientes malos (E_M), y que tendrá la siguiente forma:

$$38.74\% \leq E_M \leq 47.26\% \quad [15]$$

En la expresión [15], las cotas inferior y superior fueron calculadas como $E_M \pm 1,64*\sigma$.

A través del establecimiento de este intervalo, y recordando que la efectividad actual en clientes malos es 36.5%, se puede afirmar con un 95% de seguridad, que el modelo nuevo mejorará dicha efectividad, si la población se mantiene constante en sus características.

Antes concluir el análisis, falta determinar si el modelo es estable en los valores de sus coeficientes. Para esto, se construyó el cuadro 28, que contiene el cálculo de coeficientes de las variables, para cada partición:

Cuadro 28: Cálculo de Coeficientes del Modelo Nuevo, en 10 Particiones distintas

Valores Beta Variable	Particiones										Promedio	Desv. Estánd.
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10		
Casaprop	-0,445	-0,382	-0,336	-0,364	-0,353	-0,385	-0,516	-0,4	-0,382	-0,38	-0,394	0,052
Edad	-0,027	-0,025	-0,025	-0,028	-0,028	-0,023	-0,028	-0,029	-0,027	-0,027	-0,027	0,002
Cump_imp (1)	0,52	0,448	0,521	0,303	0,539	0,426	0,416	0,521	0,559	0,512	0,477	0,078
Cum_imp (2)	0,58	0,633	0,649	0,572	0,608	0,579	0,529	0,612	0,587	0,566	0,592	0,035
N_Cred	-0,089	-0,155	-0,117	-0,127	-0,082	-0,134	-0,115	-0,099	-0,15	-0,176	-0,124	0,030
Vehic	-0,664	-0,64	-0,751	-0,595	-0,584	-0,593	-0,567	-0,648	-0,587	-0,677	-0,631	0,057
Reg_M	0,246	0,212	0,13	0,18	0,172	0,259	0,208	0,24	0,222	0,36	0,223	0,062
SerBas	-0,159	-0,213	-0,119	-0,172	-0,182	-0,164	-0,159	-0,086	-0,164	-0,107	-0,153	0,038
SiNuevo	0,722	0,565	0,716	0,593	0,721	0,57	0,624	0,625	0,507	0,548	0,619	0,078
LN(Deuda Com+Cons)	0,088	0,084	0,085	0,078	0,09	0,076	0,084	0,087	0,071	0,078	0,082	0,006

Siguiendo el raciocinio del cuadro 27, como la desviación estándar de los coeficientes que acompañan a cada variable es pequeña, se puede concluir que el modelo es además estable en sus coeficientes.

Por lo tanto, se ha logrado re-calibrar el modelo, encontrando una solución estable, que predice más efectivamente el riesgo de los clientes.

9.3.4 Solución Alternativa

Pese a que la re-calibración logra mejorar el desempeño del modelo, ahora se busca responder si existe alguna variable adicional, que no estuviera contemplada actualmente, que le ayude a aumentar aún más su efectividad. Para descubrirlo, se realizará una re-calibración con selección estadística de variables, que encuentre, de existir, alguna relación interesante entre las variables ingresadas.

El set de datos contendrá, adicionales a las variables actuales del modelo, un conjunto de variables “candidatas”, con las que competirán directamente para mantenerse en la regresión logística. El método de selección de variables escogido es *forward*, con un $\alpha=0.05$ de entrada y $\alpha'=0.1$ de salida. Las nuevas variables a incluir en esta re-calibración, se describen a continuación:

- *Ini_Act*: Vale 1 si el cliente tiene iniciación de actividades, y 0 sino.

- *Pat_Com*: Vale 1 si el cliente tiene patente comercial, y 0 sino.
- *Perm_M*: Vale 1 si el cliente tiene permiso municipal, y 0 sino.
- *dia*: Variable continua, que corresponde al porcentaje de las cuentas que paga al día.
- *Antig*: Variable que indica el número de años de antigüedad del cliente en el Banco del Desarrollo.
- *Ahorro*: Vale 1 si el cliente tiene libreta de ahorro, y la ocupa de forma mensual, trimestral o sin periodicidad. Si no tiene libreta, la variable es 0.
- *Tuvo_Cred*: Vale 1 si el cliente tuvo un crédito en cualquier institución financiera, y 0 sino.
- *Otros_Act*: Vale 1 si el cliente tiene activos adicionales al vehículo o casa (como maquinaria, por ejemplo), y 0 sino.
- *Si_Hijos*: Vale 1 si el cliente tiene hijos, y 0 sino.

1. *Análisis Univariado*

Como primer filtro de ingreso de las nuevas variables al método stepwise, se realiza el análisis univariado de las variables candidatas, con los resultados que se exponen en el cuadro 29:

Cuadro 29: Análisis Univariado de Variables Candidatas (con Resultado Real de Pago)

Variable	Tipo de Test	Valor	Sig.
Ini_Acti	Chi Pearson	23,447	0,000
Pat_Com	Chi Pearson	5,836	0,016
Perm_M	Chi Pearson	0,743	0,389
dia	K-S	1,899	0,001
Antig	K-S	5,482	0,000
Ahorro	Chi Pearson	8,450	0,004
Tuvo_Cred	Chi Pearson	67,662	0,000
Otros_Act	Chi Pearson	10,996	0,001
Si_Hijos	Chi Pearson	0,277	0,599

Se observa que las variables Perm_M y Si_Hijos no tienen mayor relación con el comportamiento de pago de un cliente, por lo que sólo ingresarán a la re-calibración el resto de las variables candidatas, más las actuales.

2. Análisis Multivariado

Las primeras salidas han sido tomadas de SPSS® y corresponden los cuadros 30 y 31, que contienen el ajuste del modelo (*likelihood ratio*) y al R² de Nagelkerke, para cada uno de los pasos de selección de variables:

Cuadro 30: Ajuste del Modelo *Stepwise*

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	64,375	1	,000
	Block	64,375	1	,000
	Model	64,375	1	,000
Step 2	Step	36,451	1	,000
	Block	100,826	2	,000
	Model	100,826	2	,000
Step 3	Step	22,823	1	,000
	Block	123,649	3	,000
	Model	123,649	3	,000
Step 4	Step	17,567	1	,000
	Block	141,217	4	,000
	Model	141,217	4	,000
Step 5	Step	15,821	2	,000
	Block	157,038	6	,000
	Model	157,038	6	,000
Step 6	Step	17,756	1	,000
	Block	174,794	7	,000
	Model	174,794	7	,000
Step 7	Step	5,389	1	,020
	Block	180,182	8	,000
	Model	180,182	8	,000
Step 8	Step	4,378	1	,036
	Block	184,561	9	,000
	Model	184,561	9	,000

Cuadro 31: R² Nagelkerke del Modelo *Stepwise*

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1860,779 ^a	,045	,060
2	1824,328 ^a	,070	,093
3	1801,505 ^b	,085	,114
4	1783,938 ^b	,097	,129
5	1768,117 ^b	,107	,143
6	1750,361 ^b	,118	,158
7	1744,972 ^b	,122	,162
8	1740,594 ^b	,124	,166

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

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

En ambas salidas, se observa que el método *forward* ha seleccionado 8 variables para el modelo. Además, el alto valor de G, indica que modelo escogido se ajusta a los datos, y el valor del R² de Nagelkerke es levemente superior al de la re-calibración anterior (0.163).

El resultado de las variables ingresadas en el último paso del método, se muestra en el cuadro 32, donde además se aprecia el orden de incorporación al modelo:

Cuadro 32: Variables ingresadas al Modelo por el Método *Stepwise*
Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step	CasaProp	-,476	,130	13,352	1	,000	,621
8	Edad	-,027	,005	26,550	1	,000	,973
	Cump_Imp			20,223	2	,000	
	Cump_Imp(1)	,501	,228	4,818	1	,028	1,651
	Cump_Imp(2)	,552	,126	19,337	1	,000	1,737
	Vehic	-,669	,143	21,916	1	,000	,512
	SiNuevo	,587	,181	10,513	1	,001	1,799
	LNDCCM1	,115	,026	20,057	1	,000	1,122
	Antig	-,048	,023	4,394	1	,036	,953
	Tuvo_Cred	-,457	,219	4,341	1	,037	,633
	Constant	,729	,309	5,570	1	,018	2,073

- a. Variable(s) entered on step 1: Edad.
- b. Variable(s) entered on step 2: SiNuevo.
- c. Variable(s) entered on step 3: Vehic.
- d. Variable(s) entered on step 4: CasaProp.
- e. Variable(s) entered on step 5: Cump_Imp.
- f. Variable(s) entered on step 6: LNDCCM1.
- g. Variable(s) entered on step 7: Antig.
- h. Variable(s) entered on step 8: Tuvo_Cred.

En el cuadro anterior se puede observar que las variables SerBas, Reg_M y NCred1, no fueron seleccionadas por el método, en desmedro de

Antig y Tuvo_Cred. Esto puede responder al cambio de la composición de la cartera, donde otras variables comenzarían a ser relevantes para predecir el riesgo del crédito. La inclusión de las variables candidatas tiene sentido, puesto que sus coeficientes negativos indican que un cliente antiguo sería mejor pagador que uno nuevo, al igual que un cliente que ya haya tenido un crédito en el sistema financiero.

Antes de proseguir con el entrenamiento del modelo, se debe revisar si existe alguna correlación entre todas las variables que han sido sometidas al método *stepwise*, que pueda “ensuciar” el resultado predictivo. En el cuadro 33, sólo se mostrará la parte de la matriz que se considera relevante para el análisis. La matriz completa, se encuentra en Anexos.

Llama la atención la alta correlación (positiva) entre ambas variables que ingresaron al modelo, y la excluida variable NCred1. Tiene sentido pensar que, a mayor antigüedad del cliente, mayor número de créditos posee, pero la comparación no es tan directa con el hecho que haya tenido crédito en alguna institución financiera. Es posible, entonces, que ambas variables contengan y/o complementen la información de NCred1.

Cuadro 33: Matriz de Correlación de Variables seleccionadas por Método Stepwise

Correlations

		CasaProp	Edad	Cump Imp	NCred1	Vehic	SiNuevo	LNDCDCm1	Antig	Tuvo Cred
CasaProp	Pearson Correlation	1	,277**	,102**	,217**	,071**	-,145**	,163**	,204**	,163**
	Sig. (2-tailed)		,000	,000	,000	,000	,000	,000	,000	,000
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816
Edad	Pearson Correlation	,277**	1	,016	,246**	-,005	-,207**	,146**	,198**	,202**
	Sig. (2-tailed)	,000		,394	,000	,779	,000	,000	,000	,000
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816
Cump Imp	Pearson Correlation	,102**	,016	1	,067**	,085**	-,031	,134**	,104**	,086**
	Sig. (2-tailed)	,000	,394		,000	,000	,097	,000	,000	,000
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816
NCred1	Pearson Correlation	,217**	,246**	,067**	1	,030	-,657**	,371**	,594**	,540**
	Sig. (2-tailed)	,000	,000	,000		,108	,000	,000	,000	,000
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816
Vehic	Pearson Correlation	,071**	-,005	,085**	,030	1	-,009	,136**	,045*	,068**
	Sig. (2-tailed)	,000	,779	,000	,108		,638	,000	,017	,000
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816
SiNuevo	Pearson Correlation	-,145**	-,207**	-,031	-,657**	-,009	1	-,470**	-,383**	-,757**
	Sig. (2-tailed)	,000	,000	,097	,000	,638		,000	,000	,000
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816
LNDCDCm1	Pearson Correlation	,163**	,146**	,134**	,371**	,136**	-,470**	1	,201**	,679**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000		,000	,000
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816
Antig	Pearson Correlation	,204**	,198**	,104**	,594**	,045*	-,383**	,201**	1	,353**
	Sig. (2-tailed)	,000	,000	,000	,000	,017	,000	,000		,000
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816
Tuvo Cred	Pearson Correlation	,163**	,202**	,086**	,540**	,068**	-,757**	,679**	,353**	1
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

3. Entrenamiento del Nuevo Modelo

Al igual que en la re-calibración anterior, se ratificará que se utilizó un set de entrenamiento con el 80% de los casos y un set de prueba con el 20% restante. La salida de SPSS® que muestra esta división, se encuentra en el cuadro 34.

Cuadro 34: Porcentaje de casos para entrenar y probar el Modelo *Stepwise*

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	2267	80,5
	Missing Cases	0	,0
	Total	2267	80,5
Unselected Cases		549	19,5
Total		2816	100,0

a. If weight is in effect, see classification table for the total number of cases.

Nuevamente, los casos seleccionados corresponden a los del set de entrenamiento, y los no seleccionados, al set de prueba.

En el cuadro 35 se muestran los coeficientes beta (B) asociados a las variables que entraron al modelo, calculados a partir del set de entrenamiento:

Cuadro 35: Coeficientes beta del Modelo *Stepwise*

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	
Step 1 ^a	CasaProp	-,466	,130	12,871	1	,000	,628
	Edad	-,027	,005	25,606	1	,000	,974
	Cump_Imp			21,692	2	,000	
	Cump_Imp(1)	,499	,228	4,800	1	,028	1,646
	Cump_Imp(2)	,572	,125	20,913	1	,000	1,771
	Vehic	-,667	,142	21,971	1	,000	,513
	SiNuevo	,558	,180	9,584	1	,002	1,747
	LNDCDCm1	,119	,026	21,599	1	,000	1,126
	Tuvo_Cred	-,490	,218	5,046	1	,025	,612
	Antig	-,052	,023	5,102	1	,024	,949
	Constant	,721	,307	5,506	1	,019	2,056

a. Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, Vehic, SiNuevo, LNDCDCm1, Tuvo_Cred, Antig.

Todos los coeficientes beta son significativos al 95%.

Dado que en el análisis multivariado ya se concluyó sobre el ajuste del modelo y la fortaleza de relación entre la variable dependiente y las predictoras, no es necesario mostrar sus valores para el set de entrenamiento. Sólo se mencionará que son prácticamente iguales.

4. Comparar Modelo Actual y Nuevo

El punto de corte óptimo del modelo, se determinó bajo el concepto de utilidad máxima, tal como se destaca en el cuadro 36:

Cuadro 36: Desempeño de distintos Puntos de Corte (Modelo Stepwise)

Pbb Corte	% Buenos Aceptados	% Malos Rechazados	% Efectividad	% Acierto Global	% Cobertura Total	Utilidad [M\$]
0,05	0,0%	100,0%	50,0%	29,3%	0,0%	0
0,10	0,5%	100,0%	50,3%	29,7%	0,4%	962
0,15	2,3%	100,0%	51,2%	31,0%	1,6%	4.330
0,20	7,7%	99,4%	53,6%	34,6%	5,6%	13.478
0,25	14,7%	95,7%	55,2%	38,4%	11,7%	20.738
0,30	21,6%	88,2%	54,9%	41,2%	18,8%	22.270
0,35	30,7%	85,1%	57,9%	46,6%	26,0%	34.334
0,40	41,0%	80,7%	60,9%	52,6%	34,6%	46.894
0,45	51,3%	73,3%	62,3%	57,7%	44,1%	54.680
0,50	61,1%	65,2%	63,1%	62,3%	53,4%	60.549
0,55	71,4%	56,5%	64,0%	67,0%	63,2%	66.425
0,60	79,4%	41,0%	60,2%	68,1%	73,4%	57.469
0,65	86,3%	31,7%	59,0%	70,3%	81,1%	56.137
0,70	93,6%	16,8%	55,2%	71,0%	90,5%	46.692
0,75	95,9%	13,0%	54,5%	71,6%	93,3%	45.293
0,80	99,0%	3,7%	51,3%	71,0%	98,2%	36.744
0,85	99,7%	1,9%	50,8%	71,0%	99,3%	35.301
0,90	100,0%	0,0%	50,0%	70,7%	100,0%	32.940
0,95	100,0%	0,0%	50,0%	70,7%	100,0%	32.940
1,00	100,0%	0,0%	50,0%	70,7%	100,0%	32.940

El criterio de utilidad máxima sugiere utilizar el corte de 0.55. Dado que la proporción de buenos que se acepta con esta configuración del modelo, es muy distinta a la que se acepta con la estructura del modelo actual, la comparación entre ambos es menos directa que en el caso anterior. Si bien es cierto que se acepta casi un 10% menos de clientes pagadores, también se rechaza un 20% más de no pagadores.

Por lo tanto, hace más sentido comparar esta efectividad con el corte que acepte “similar” porcentaje de buenos en el modelo actual. De esta forma, se podrá evaluar si el modelo elegido por el método *stepwise*, rechaza o no mayor cantidad de clientes malos. Dicho punto de corte del modelo actual también corresponde a 0.55, y permite aceptar al 70.4% de los clientes buenos y rechazar a un 54.1% de los malos.

En el cuadro 37, se muestra la matriz de confusión para el modelo encontrado con el método *stepwise*, con el punto de corte de utilidad máxima:

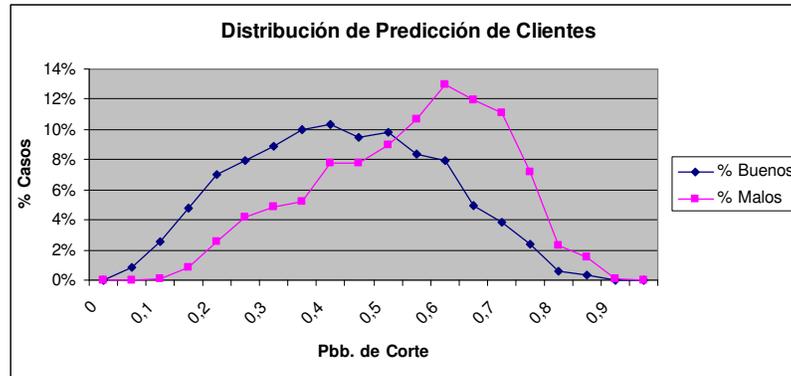
Cuadro 37: Matriz de Confusión Modelo *Stepwise*

	Pred. = 0	Pred = 1
Obs. = 0	71,4%	28,6%
Obs. = 1	43,5%	56,5%

Por otro lado, el valor del *C-Index* el modelo *stepwise* es **0.697**.

El gráfico de la ilustración 54, muestra la distribución de predicción de probabilidades del modelo *stepwise*.

Ilustración 54: Gráfico de Distribución de Predicción de Clientes (Modelo *Stepwise*)



La mayor efectividad y acierto global del modelo encontrado por el método *stepwise*, también se ve reflejada en un mayor valor del índice de concordancia y distinción más clara entre las distribuciones de clientes buenos y malos. Es decir, si se utilizara este modelo, se lograría aumentar en un 1% la aceptación de clientes buenos, e incrementar el acierto de los malos en cerca de un 2.4%. Para poder realizar dicha afirmación con mayor seguridad, es necesario mostrar que este resultado es estable, y no producto de la elección del set de prueba.

5. Evaluar Sobreajuste y Verificar Estabilidad del Modelo Propuesto

Cuadro 38: Predicción del Modelo *Stepwise* en Sets de Entrenamiento y Prueba

Classification Table^d

Observed		Predicted					
		Selected Cases ^a			Unselected Cases ^{b,c}		
		Var. Obj		Percentage Correct	Var. Obj		Percentage Correct
,00	1,00	,00	1,00				
Step 1	Var_Obj	,00					
		1,00	502	192	72,3	123	49
	Overall Percentage		295	411	58,2	70	91
					65,2		64,2

a. Selected cases Filter1 EQ 1

b. Unselected cases Filter1 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,550

Como lo especifica el cuadro 38, dado que la efectividad global del modelo varía en un 0.9%, es posible afirmar que éste no se encuentra sobreajustado a la muestra. La estabilidad de predicción del modelo puede ser concluida a partir del cuadro 39:

Cuadro 39: Efectividad de Predicción del Modelo *Stepwise*, en 10 Particiones distintas

Efectividad Modelo	% Efect. Buenos	% Efect. Malos	% Efect. Total
Partición 1	71,4%	56,5%	64,2%
Partición 2	73,8%	50,3%	61,6%
Partición 3	69,3%	55,6%	62,1%
Partición 4	73,2%	57,1%	65,2%
Partición 5	71,0%	61,3%	66,2%
Partición 6	76,7%	54,6%	65,1%
Partición 7	72,9%	48,2%	59,7%
Partición 8	69,8%	61,0%	65,4%
Partición 9	73,4%	56,3%	64,6%
Partición 10	75,1%	54,4%	64,9%
Promedio	72,7%	55,5%	63,9%
Desv. Estánd.	2,3%	4,1%	2,1%

El bajo valor de la desviación estándar de la efectividad, permite concluir que el modelo predice establemente. En este caso, no es tan claro que el nuevo modelo prediga mejor que el actual, ya que el promedio es sólo un 1.4% superior, y la desviación es de 4.1%. La mayor volatilidad en la predicción, puede deberse al ruido que introduce la alta correlación de las variables ingresadas al modelo. De hecho, si se quisiera establecer un intervalo de confianza al 95%, como se hizo con la re-calibración anterior, la cota inferior sería 48.76%, muy por debajo del 54,1% del modelo actual. Por otro lado, en los clientes buenos la diferencia es de sólo un 2.3%. Luego, la mejoría del nuevo modelo no es suficientemente significativa, como para recomendar que reemplace al actual.

En el cuadro 40 se expondrán los coeficientes de las variables, para los modelos generados en las distintas particiones. La baja desviación estándar de los coeficientes, permite concluir que el modelo *stepwise* sí es estable en sus coeficientes.

Cuadro 40: Cálculo de Coeficientes del Modelo Stepwise, en 10 Particiones distintas

Valores Beta	Particiones										Promedio	Desv. Estánd.
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10		
Casaprop	-0,466	-0,405	-0,363	-0,389	-0,382	-0,408	-0,535	-0,42	-0,425	-0,4	-0,419	0,049
Edad	-0,027	-0,026	-0,026	-0,028	-0,028	-0,023	-0,027	-0,03	-0,027	-0,027	-0,027	0,002
Cump_Imp (1)	0,499	0,438	0,494	0,29	0,537	0,402	0,394	0,504	0,537	0,491	0,459	0,078
Cum_Imp (2)	0,572	0,62	0,634	0,566	0,602	0,553	0,519	0,596	0,578	0,544	0,578	0,035
Vehic	-0,667	-0,648	-0,758	-0,608	-0,59	-0,589	-0,569	-0,644	-0,594	-0,662	-0,633	0,056
SiNuevo	0,558	0,531	0,63	0,535	0,601	0,486	0,49	0,491	0,521	0,666	0,551	0,062
LN(Deuda Com+Cons)	0,119	0,112	0,114	0,104	0,115	0,102	0,117	0,117	0,096	0,096	0,109	0,009
Antig	-0,49	-0,463	-0,482	-0,428	-0,406	-0,45	-0,508	-0,486	-0,409	-0,312	-0,443	0,058
Tuvo_Cred	-0,052	-0,07	-0,047	-0,062	-0,048	-0,076	-0,068	-0,052	-0,056	-0,057	-0,059	0,010

10. GENERALIZACIÓN DE LA EXPERIENCIA

La aplicación de patrones de procesos de negocio, a un dominio particular, permite formalizar y detallar los procesos de la empresa, considerando no sólo las actividades de los procesos diseñados, sino también los flujos de información que hay entre ellas (Barros, 2004). La aplicación de ingeniería de negocio al rediseño de procesos, además permite estructurarlos en una arquitectura empresarial y apoyarlos adecuadamente con tecnología.

Cada proceso diseñado posee lógicas de negocio, que explicitan cómo las actividades deben ser realizadas. Por ello, la definición y formalización de dichas lógicas, es trascendental para comprender el rol que cumplen en la generación del producto o servicio. Adicionalmente, el diseño de software debe estar alineado con los procesos que soporta, de tal forma que se integre sin problemas con la tecnología existente y además responda a las necesidades del negocio.

El *Framework* es una estructura genérica de clases, que sirve como base común para el desarrollo de software en empresas de un dominio particular, pero que se puede adaptar a las características y necesidades propias de cada una de ellas. Este esquema permite que los esfuerzos se centren en las especificaciones de la aplicación, reduciendo los costos y tiempos asociados a

su desarrollo. Barros (2004), explica que este concepto de *Framework* “difiere del tradicional, en que se encuentra orientado a una lógica de negocio compleja, y está basado en métodos analíticos avanzados, provenientes de Estadística e Inteligencia de Negocios”.

Adicionalmente, Barros (2004) muestra que los BOF (*Business Objects Frameworks*), pueden ser derivados de los patrones de procesos y lógicas de negocio. En consecuencia, es posible proponer una estructura genérica de objetos de negocio que incorpore el conocimiento del diseño de procesos de este trabajo, para facilitar el desarrollo de aplicaciones que resuelvan el mismo problema en un dominio similar.

10.1 Elección y Definición del *Framework*

En este trabajo se ha rediseñado dos procesos fundamentales para la relación del Banco con sus clientes: la mantención del modelo de comportamiento de clientes y la evaluación de clientes. Para este último, es posible proponer un *Framework* que responda a la necesidad de atender, evaluar y ofrecer productos (o servicios) a los clientes de forma segmentada y estandarizada, así como realizar un control del desempeño del negocio por segmento atendido.

Respecto al otro proceso, tal como se mencionó en el Marco Teórico, pocos autores se han referido a la importancia de actualizar los modelos, y menos aún, han formalizado un proceso para hacerlo. Por ello, la generación de un *Framework*, derivado de un diseño detallado de procesos y realización de una prueba de concepto, que permita a una empresa mantener un modelo predictivo existente, constituye una propuesta novedosa que une la Ingeniería de Negocios con la Inteligencia de Negocios, concibiendo los modelos de forma dinámica. Por esta razón, se realizará un *Framework* para este último proceso.

Ahora que se ha seleccionado el *Framework*, es necesario definir el dominio para el cual será confeccionado. El dominio está constituido por las empresas que poseen modelos predictivos y que requieran actualizarlos en el tiempo. Las razones por las que este *Framework* genera valor a las empresas del dominio, son principalmente dos: asegura la utilización de un modelo predictivo actualizado, y permite prescindir, por un período más prolongado, de especialistas en desarrollo de modelos.

Particularmente, se propone un *Framework* para los modelos de regresión logística y arboles de decisión. Lo anterior se justifica en que su estructura es “transparente”, es decir, se puede traducir en reglas de negocio claras y entendibles por los ejecutivos de las empresas. Esto hace que los

modelos mencionados sean más fáciles de vender, por cuanto confirman o descubren relaciones interesantes del fenómeno en estudio. Por otro lado, métodos como redes neuronales, son considerados “cajas negras”, pues es difícil descifrar cuáles son las consideraciones del modelo para realizar la predicción. Esto ha traído como consecuencia que sea utilizado más bien con fines académicos que de negocios.

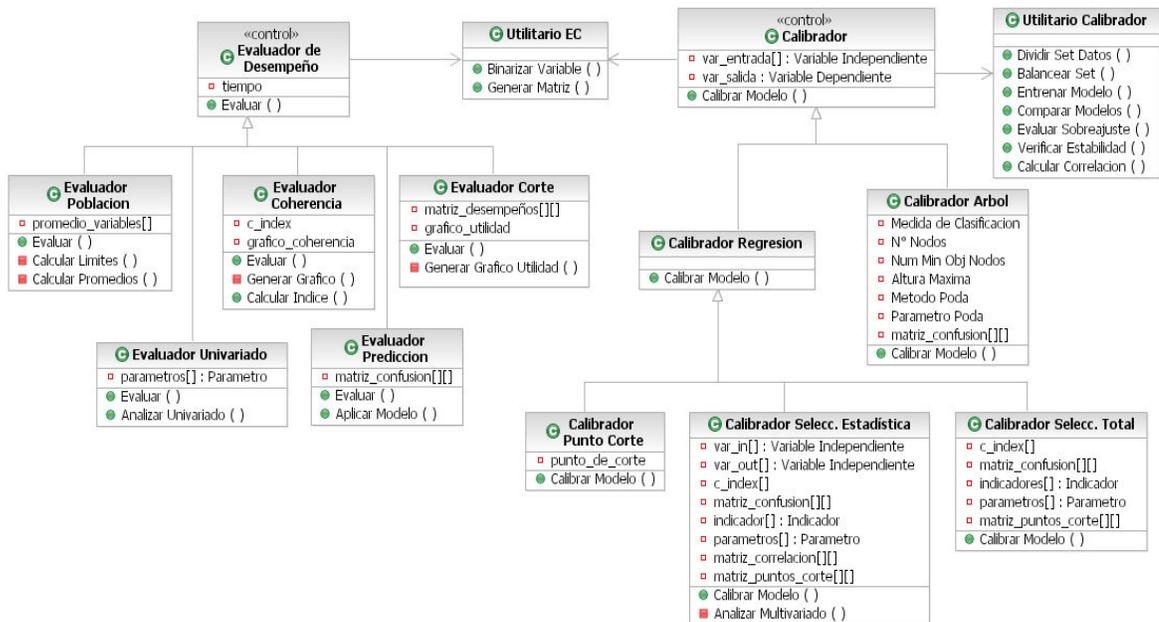
10.2 Construcción del *Framework*

Para construir el *Framework* es necesario identificar cuáles son los elementos que debe poseer, para cumplir el objetivo de medir el desempeño del modelo predictivo y actualizarlo.

Según lo expuesto en el diseño de procesos, la estructura de clases del *Framework* debe permitir evaluar si el modelo debe ser re-calibrado y ejecutar el tipo de re-calibración que se requiera, en caso que sea necesario. Es decir, se necesita una estructura de clases de control que orqueste el procesamiento de la información para las funciones mencionadas, coordinando el acceso a los datos, utilizando lógicas de negocio propias y requiriendo al paquete estadístico.

Por otro lado, el *Framework* debe contener clases que permitan hacer un diseño orientado a objetos del modelo y que se integre con los sistemas de la empresa. Luego, se necesita un conjunto de clases *entity* que contenga la información de la estructura y desempeño del modelo, en cualquier periodo de tiempo. Para satisfacer lo mencionado anteriormente, se construye el diagrama de clases de control, que se muestra en la ilustración 55.

Ilustración 55: Diagrama de Clases de Control del Framework



- **Clases de Control Utilitarias**

Las clases utilitarias poseen los métodos utilizados por varias clases de control, pero que no son propios de ninguna de ellas. La razón para

ubicar estos métodos comunes en clases utilitarias, está fundada en ahorro de código y facilidad de mantención. Éstas son:

- Utilitario EC: Clase que posee métodos accedidos por las Clases Evaluadoras y Re-calibradoras.
- Utilitario Calibrador: Clase que posee métodos accedidos sólo por las Clases Re-calibradoras, salvo por “Calibrador Punto de Corte”.

- ***Clases de Control para Evaluación***

- Evaluador de Desempeño: Clase abstracta, definida para evaluar distintos aspectos del desempeño del modelo. La evaluación posee una componente temporal y puede utilizar los métodos de la clase Utilitario EC.
- Evaluador Población: Clase que evalúa si la población ha cambiado de forma significativa. Incluye el cálculo de los promedios de cada variable que describe a la población, así como sus límites asociados.
- Evaluador Univariado: Clase que evalúa si cada variable aporta, por sí sola, a predecir la variable dependiente. A través del método “Analizar Univariado”, solicita el análisis univariado de cualquier

variable al paquete estadístico. Como atributo posee los parámetros asociados a cada variable, con su estadístico correspondiente.

- Evaluador Predicción: Clase que evalúa la calidad predictiva del modelo. Para predecir el comportamiento de cada cliente de un set de datos, utiliza el método “Aplicar Modelo”. Para generar la matriz de confusión, este método requiere a “Generar Matriz”, de la clase Utilitario EC.

- Evaluador de Corte: Clase sólo para regresión logística, que evalúa la política de corte del modelo. El método “Aplicar Modelo”, es llamado al Evaluador Predicción, para predecir los registros con distintos puntos de corte. Adicionalmente, se requiere “binarizar” las predicciones anteriores, para lo cual utiliza el método “Binarizar Variable”, disponible en la clase Utilitario EC. Con los datos anteriores, genera y grafica la utilidad relativa al modelo actual, y la matriz de desempeños.

- Evaluador de Coherencia: Clase sólo para la regresión logística, que evalúa la coherencia del modelo, a través del cálculo del índice de concordancia. Para esa función específica, requiere el método “Calcular Índice”. Además, utiliza el método “Generar Gráfico”, para mostrar gráficamente su capacidad discriminante.

- ***Clases de Control para Re-Calibración***

- **Calibrador:** Clase abstracta, definida para realizar el tipo de re-calibración del modelo que se desee. Para ello, utiliza las variables independiente y variable dependiente de los registros incluidos. El método de calibración es distinto para la regresión logística y el árbol de decisión. Todas sus clases derivadas tienen acceso a los métodos públicos de las Clases Evaluadoras y de los Utilitarios.
- **Calibrador Árbol:** Clase que asiste la re-calibración de un árbol de decisión, en base a los criterios especificados por el usuario, como por ejemplo, la altura máxima. Esto es solicitado al paquete estadístico, a través del método “Calibrar Modelo”.
- **Calibrador Regresión:** Clase abstracta, definida para asistir los distintos tipos de re-calibración que se pueden hacer a la regresión logística.
- **Calibrador Punto de Corte:** Clase que redefine el punto de corte que utilizará el modelo en adelante, a través del método “Calibrar Modelo”.
- **Calibrador con Selección Total:** Clase que asiste la re-calibración del modelo, utilizando sólo las variables independientes ingresadas por el usuario. Los métodos necesarios para entrenar el modelo,

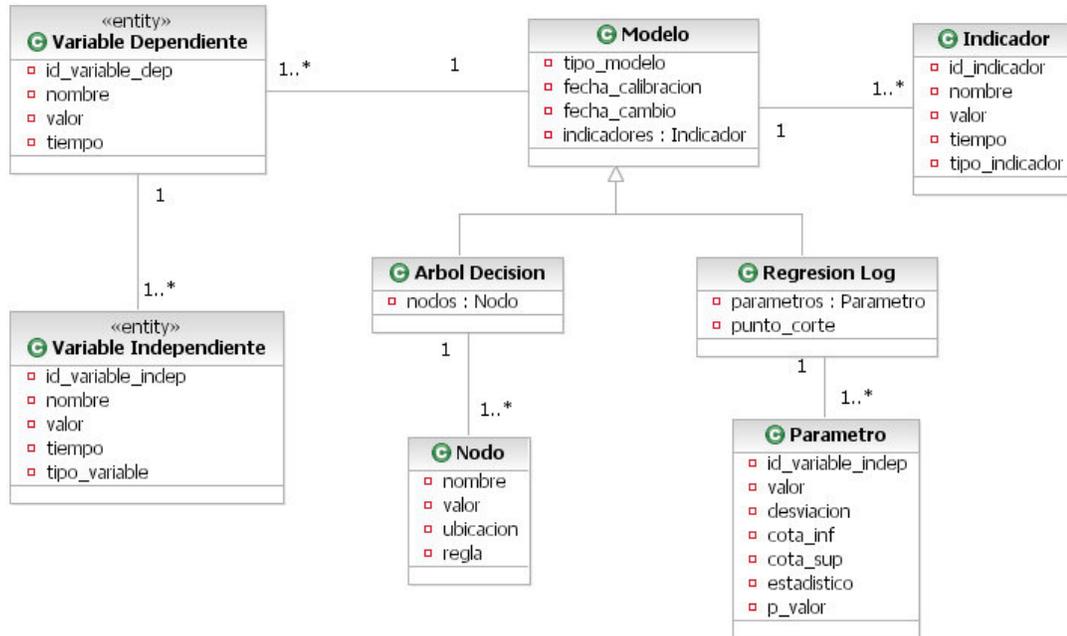
conectándose con el paquete estadístico, y calcular los indicadores de desempeño, ajuste y estabilidad del nuevo modelo, se encuentran en la Clase Utilitario Calibrador. Éstos son llamados por el método “Calibrar Modelo”.

- Calibrador con Selección Estadística: Clase que asiste la recalibración de la regresión, utilizando un set de variables candidatas. Éstas últimas son escogidas a través del método stepwise, que se realiza en el análisis multivariado. Los demás indicadores de desempeño, ajuste y estabilidad del nuevo modelo, son solicitados por el método “Calibrar Modelo”, a la Clase Utilitario Calibrador.

Los métodos de las clases de control que refieren al procesamiento de la información, deben basarse en los seudocódigos expuestos en este trabajo. Los métodos que requieren conexión al paquete estadístico, así como solicitud y recepción de información desde el mismo, deben ser implementados de acuerdo al software que se utilice. El acceso a los datos dependerá de la estructura de las bases de datos de la empresa.

El diagrama de clases *entity*, que incluye indicadores del modelo en el tiempo, se encuentra en la ilustración 56,

Ilustración 56: Diagrama de Clases *Entity* del Framework



- **Clases Entity**

- Variable Dependiente: Clase que almacena el valor de la predicción del modelo en un instante de tiempo. Tienen asociadas las variables explicativas y el modelo.
- Variable Independiente: Clase que almacena los datos de la variable explicativa para un instante de tiempo. Por ejemplo, si el tipo de variable es continua, ordinal o categórica.
- Modelo: Clase abstracta definida para los dos tipos de modelos en cuestión. Su periodo de validez termina cuando es re-calibrado.

- **Árbol Decisión:** Clase que contiene uno o más Nodos, cada uno vinculado a una variable independiente. Para decidir cómo manejar cada variable, cada Nodo posee un valor y una regla. Para almacenar el resultado del modelo, se asocia con la variable dependiente.
- **Regresión Log:** Clase que contiene uno o más parámetros, cada uno vinculado a una variable independiente. Las características de cada parámetro, como su valor y rango permitido, se encuentran en su clase respectiva. Adicionalmente, la regresión posee el punto de corte utilizado en el periodo relevante.
- **Indicador:** Clase que almacena los datos de los indicadores asociados a cada modelo, en un instante de tiempo. Por ejemplo, el índice de concordancia, el acierto global, efectividad, cobertura, R^2 de Nagelkerke, log-likelihood ratio G, etc.

11. CONCLUSIONES

11.1 Del Proceso de Evaluación de Clientes

La importancia de conocer a los clientes y brindarles un trato segmentado, con productos atractivos para cada perfil, es hoy en día un requisito para las empresas que se desenvuelven en entornos altamente competitivos y desean obtener rentabilidades sobre el promedio.

El diseño del proceso de evaluación de clientes, no sólo se remite a proponer una mejora y automatización de las evaluaciones de clientes, que reduce el uso de la tecnología sólo a la virtualización de los papeles físicos, sino que induce transversalmente a la empresa a concebir el negocio de forma segmentada. La definición de distintos perfiles de clientes materializa y enfatiza la existencia de microempresas de distinta naturaleza y funcionamiento, cuyos dueños tienen características propias y que requieren de productos específicos. Pero el diseño más allá, por cuanto permite controlar constantemente el desempeño de los segmentos, facilitando la gestión comercial y desarrollo de nuevos productos para cada uno de ellos.

Evidentemente, también es posible segmentar los clientes apoyándose en técnicas de Minería de Datos, para lo cual será trascendental estructurar los

procesos para mantener el modelo de clasificación, tal como se ha mostrado en este trabajo para el modelo de *scoring*.

Este enfoque es extrapolable a cualquier empresa que tenga o desee realizar algún tipo de segmentación de sus clientes, ya sea para su evaluación, control o generación de nuevos productos. Sin embargo, debe tenerse en cuenta que requiere de constante supervisión y evaluación, con el costo de tiempo y recursos que esto conlleva. Por eso, es muy importante contar con algún proceso de mantención del modelo de clasificación (segmentación), independiente si éste ha sido generado de forma cualitativa o cuantitativa.

11.2 Del Proceso de Mantención del Modelo de *Scoring*

La realización de la prueba de concepto permite concluir tanto respecto de la actualización de modelos, como de los resultados obtenidos en este problema y set de datos en particular. Las conclusiones son las siguientes:

1. Los modelos predictivos se desactualizan. La concepción estática que tienen las empresas de su funcionamiento, genera espacios para que un

mismo modelo sea utilizado por un periodo de tiempo muy prolongado, dentro del cual su calidad predictiva puede verse seriamente mermada.

Prueba de lo anterior, es el cambio de la población atendida por el Banco, consecuencia de su nueva estrategia en los últimos dos años. El aumento de la proporción de clientes nuevos en la cartera y la disminución de la deuda promedio, ha generado una baja en la efectividad del modelo a 59%, siendo que su efectividad de calibración bordeaba el 65%. Por eso, es importante que la re-calibración del modelo siempre tenga una relación uno a uno con los lineamientos estratégicos. Un claro ejemplo de esto, es que la política de punto de corte no siempre responderá al criterio de maximización de una función de utilidad, sino a la dirección que la empresa desea tomar en el futuro (participación de mercado, disminución del riesgo, etc.)

Pero el cambio poblacional no necesariamente está ligado a un cambio en la estrategia. Actualmente, el volátil escenario financiero mundial ha tenido repercusiones en la economía chilena, principalmente en el aumento de precios y presiones inflacionarias. Evidentemente, esto tiene un impacto negativo en la capacidad de pago de los microempresarios, generando un cambio en su perfil de riesgo. Por lo tanto, la re-calibración del modelo tendrá relación con la posición que desea tomar la empresa frente a este escenario y no a una estrategia distinta. Es probable que el perfil de riesgo actual de los clientes sea más difuso, lo que permitiría explicar por qué el

modelo propuesto en este trabajo haya recuperado sólo parcialmente la capacidad predictiva del modelo actual (62% global aproximadamente).

2. Es posible mejorar el desempeño del modelo. Independiente de la causa del cambio poblacional y aunque éste no sea significativo, es posible realizar algún tipo de re-calibración que aumente su efectividad, utilizando el proceso propuesto. En este caso, se logró recuperar un 3% del acierto global del modelo.

3. En caso que se realice alguna re-calibración (distinta del punto de corte), se debe analizar la estabilidad de predicción y de los coeficientes del modelo nuevo. En esta prueba de concepto, en la re-calibración con selección estadística de variables, la mayor desviación estándar de la efectividad del modelo (4.1%), permitió concluir que no mejoraría significativamente el desempeño del modelo re-calibrado con las mismas variables actuales. Es decir, la estabilidad puede ser decisiva para cambiar o no el modelo. Por otro lado, en la re-calibración con selección total de variables, la menor desviación estándar permite asegurar con un 95% de confianza, que el nuevo modelo mejorará la predicción de clientes malos entre un 2% y un 11%, manteniendo el acierto de clientes buenos.

4. La inclusión de nuevas variables debe ser rigurosamente sometida a juicio experto, para determinar si tiene sentido desde el punto de vista del negocio y si podría generar mucho ruido en el modelo. En esta prueba de concepto, se pudo corroborar que si las variables nuevas tienen una correlación muy alta con las actuales, aumenta la volatilidad de la efectividad de predicción, generando modelos menos estables.

11.3 De la Ingeniería de Negocios

La metodología de diseño formal de procesos, propuesta en la Ingeniería de Negocios, constituye una guía fundamental para el diseño de procesos de cualquier empresa, partiendo del alineamiento con la estrategia hasta el diseño de los apoyos computacionales de los procesos de último nivel.

Lo innovador de este enfoque es que concibe las empresas como un conjunto de procesos que generan el producto o servicio, por lo que el énfasis se hace en el mejoramiento del proceso completo, independiente del número de actores de la empresa que participen en él. Además, para ordenar y coordinar la gran cantidad de procesos existentes, propone una arquitectura de procesos, que a su vez permite a cada persona comprender en qué parte del proceso se encuentra y en qué medida contribuye a la producción del bien o servicio.

Aún quedan muchos desafíos para la Ingeniería de Negocios en Chile. La mayoría de las empresas tiene un diseño muy precario de sus procesos, bajo distintas notaciones para el modelamiento (que dependen del encargado de turno), que generan confusión en los que deseen tanto comprenderlos como mejorarlos. Más aún, el “diseño” sólo se remite a la existencia de muchos diagramas de pista aislados, frecuentemente desactualizados y carentes de estructura alguna, que no brindan ningún valor ni son propicios para realizar una reingeniería del negocio. Esta disciplina no sólo responde a la estandarización del modelamiento de procesos, sino también a estructurarlos de una manera organizada, creativa y novedosa, para hacer frente a un entorno cada vez más competitivo.

12. BIBLIOGRAFÍA

1. Álvarez, R. 1994. *Estadística multivariante y no paramétrica con SPSS: Aplicación a las ciencias de la salud*. Díaz de Santos.
2. BancoEstado Microempresas premió a los más emprendedores [en línea]. 2007. BancoEstado. <http://www.bancoestado.cl/0C86BFFAF1A04F6383AEB91B656CC29/508C54C26B541D4B36B2C316C03BC71/articulo/11878.asp>. [consulta: 23 Mayo 2008].
3. BARROS, O. 2000. *Rediseño de Procesos de Negocios mediante el uso de Patrones*. Dolmen.
4. BARROS, O. 2004. *Business Process Patterns and Frameworks: Reusing Knowledge in Process Innovation*. Working paper 56. DII, Universidad de Chile.
5. CADARSO-SUÁREZ, C. [s.a]. [en línea]. *Test Chi-Cuadrado*. Universidad de Santiago de Compostela. <http://eio.usc.es/eipc1/MATERIALES/331102884.pdf> [consulta: 16 Septiembre 2008].
6. *Clasificación de Créditos mediante Técnicas Avanzadas de Análisis de Datos para Banco Desarrollo*. [no publicado]. 2007. DII, Universidad de Chile.
7. *FAQ: What are pseudos R-squared?* [en línea]. [s.a]. UCLA: Academic Technology Service, Statistical Consulting Group. http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm. [consulta: 08 Septiembre 2008].

8. GARSON, D. 2008. *Logistic Regression* [en línea]. University of North Carolina. <http://faculty.chass.ncsu.edu/garson/PA765/logistic.htm> [consulta: 22 Agosto 2008].
9. *Gestión Comercial con Impacto Social* [en línea]. 2007. BancoEstado. <http://www.bancoestado.cl/imagenes/corporativo/PresentacionCajaVecina2008.pdf>. [consulta: 06 Mayo 2008].
10. HARRELL, F. 2001. *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York, Springer.
11. HOSMER, D. y LEMESHOW, S. 1989. *Applied Logistic Regression*. New York, Wiley & Sons.
12. *Informe Estado de las Microfinanzas en Chile* [en línea]. 2006. Red para el Desarrollo de las Microfinanzas en Chile. http://www.redmicrofinanzas.cl/index_files/Informe%20Red%20Microfinanzas%20al%2031%2012%202006.pdf. [consulta: 06 Mayo 2008].
13. *La Microempresa en Chile* [en línea]. 2005. Red para el Desarrollo de las Microfinanzas en Chile. http://www.redmicrofinanzas.cl/index_files/MICROEMPRESA.pdf. [consulta: 23 Mayo 2008].
14. MCGEE, L. 1990. R^2 Measures based on Wald and Likelihood Ratio Joint Significance Tests. *The American Statistician*, Vol. 44, N°3.
15. *Memoria de sostenibilidad* [en línea]. 2007. Banco Santander http://www.santander.cl/contenidos/infocorporativa/Resp_social_2005/pdf/memoriarss2007.pdf [consulta: 22 Mayo 2008].

16. MENARD, S. 2002. *Applied Logistic Regression Analysis*. 2a edición. Thousand Oaks, CA, Sage Publications.
17. MORALES, L. y YÁÑEZ, A. 2007. *Serie técnica de estudios N°11: Microfinanzas en Chile* [en línea]. Departamento de estudios. SBIF. Santiago, Chile. http://www.sbif.cl/sbifweb/internet/archivos/publicacion_6536.pdf [consulta: 20 Mayo 2008].
18. PORTER, M. 1997. *¿Qué es Estrategia?* INCAE. Vol. 10(1); 35-52.
19. *Registro Pesquero Artesanal (RPA)* [en línea]. 2007. Servicio Nacional de Pesca. Gobierno de Chile. http://www.sernapesca.cl/index.php?option=com_content&task=view&id=84&Itemid=222. [consulta: 23 Mayo 2008].
20. *Valoración de Atributos e Impacto de Microempresas*. [no publicado]. 2004. Gerencia de Marketing. Banco del Desarrollo.
21. WEISS, J. 2006. *Lecture 37* [en línea]. University of North Carolina <http://www.unc.edu/courses/2006spring/ecol/145/001/docs/lectures/lecture37.htm> [consulta: 06 Agosto 2008].

13. ANEXOS: SALIDAS DE SPSS

13.1 Análisis del Modelo Actual

13.1.1 Análisis Univariado

Frequencies

	VO_gorro	N
Edad	,00	1014
	1,00	313
	Total	1327

Test Statistics^a

		Edad
Most Extreme Differences	Absolute	,285
	Positive	,001
	Negative	-,285
Kolmogorov-Smirnov Z		4,414
Asymp. Sig. (2-tailed)		,000

a. Grouping Variable: VO_gorro

Frequencies

	VO_gorro	N
Ncred1	,00	1014
	1,00	313
Total		1327

Test Statistics^a

		Ncred1
Most Extreme Differences	Absolute	,568
	Positive	,000
	Negative	-,568
Kolmogorov-Smirnov Z		8,782
Asymp. Sig. (2-tailed)		,000

a. Grouping Variable: VO_gorro

Frequencies

	VO_gorro	N
LNDCDCm1	,00	1014
	1,00	313
Total		1327

Test Statistics^a

		LNDCDCm1
Most Extreme Differences	Absolute	,311
	Positive	,000
	Negative	-,311
Kolmogorov-Smirnov Z		4,809
Asymp. Sig. (2-tailed)		,000

a. Grouping Variable: VO_gorro

CasaProp * VO_gorro Crosstabulation

Count		VO_gorro		Total
		,00	1,00	
CasaProp	,00	567	255	822
	1,00	447	58	505
Total		1014	313	1327

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	66,246 ^b	1	,000		
Continuity Correction ^a	65,166	1	,000		
Likelihood Ratio	71,614	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	66,196	1	,000		
N of Valid Cases	1327				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 119,11.

Cump_Imp * VO_gorro Crosstabulation

Count

		VO_gorro		Total
		,00	1,00	
Cump_Imp	,00	39	61	100
	1,00	377	250	627
	2,00	598	2	600
Total		1014	313	1327

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	349,932 ^a	2	,000
Likelihood Ratio	445,946	2	,000
Linear-by-Linear Association	339,255	1	,000
N of Valid Cases	1327		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 23,59.

Vehic * VO_gorro Crosstabulation

Count

		VO_gorro		Total
		,00	1,00	
Vehic	,00	723	272	995
	1,00	291	41	332
Total		1014	313	1327

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	31,024 ^a	1	,000		
Continuity Correction ^b	30,198	1	,000		
Likelihood Ratio	34,294	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	31,000	1	,000		
N of Valid Cases	1327				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 78,31.

Region * VO_gorro Crosstabulation

Count

		VO_gorro		Total
		,00	1,00	
Region	,00	810	269	1079
	1,00	204	44	248
Total		1014	313	1327

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5,782 ^b	1	,016		
Continuity Correction ^a	5,390	1	,020		
Likelihood Ratio	6,081	1	,014		
Fisher's Exact Test				,016	,009
Linear-by-Linear Association	5,777	1	,016		
N of Valid Cases	1327				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 58,50.

SerBas * VO_gorro Crosstabulation

Count

		VO_gorro		Total
		,00	1,00	
SerBas	,00	573	232	805
	1,00	441	81	522
Total		1014	313	1327

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	31,091 ^b	1	,000		
Continuity Correction ^a	30,357	1	,000		
Likelihood Ratio	32,387	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	31,067	1	,000		
N of Valid Cases	1327				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 123,12.

SiNuevo * VO_gorro Crosstabulation

Count

		VO_gorro		Total
		,00	1,00	
SiNuevo	,00	619	21	640
	1,00	395	292	687
Total		1014	313	1327

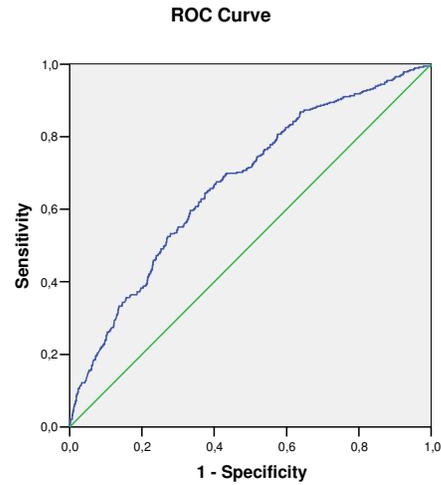
Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	282,809 ^a	1	,000		
Continuity Correction ^b	280,637	1	,000		
Likelihood Ratio	328,110	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	282,596	1	,000		
N of Valid Cases	1327				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 150,96.

13.1.2 Curva ROC y C- Index



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s): VAR00001

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,671	,016	,000	,639	,703

The test result variable(s): VAR00001 has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

13.2 Re-calibración con Selección Total de Variables

13.2.1 Análisis Univariado

Frecuencias

	Var_Obj	N
NCred1	,00	1949
	1,00	867
	Total	2816
LNDCDCm1	,00	1949
	1,00	867
	Total	2816
Edad	,00	1949
	1,00	867
	Total	2816

Test Statistics^a

		NCred1	LNDCDCm1	Edad
Most Extreme Differences	Absolute	,210	,068	,189
	Positive	,001	,006	,000
	Negative	-,210	-,068	-,189
Kolmogorov-Smirnov Z		5,148	1,661	4,618
Asymp. Sig. (2-tailed)		,000	,008	,000

a. Grouping Variable: Var_Obj

Crosstab

Count

		Var_Obj		Total
		,00	1,00	
CasaProp	,00	1113	643	1756
	1,00	836	224	1060
Total		1949	867	2816

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	74,382 ^b	1	,000		
Continuity Correction ^a	73,657	1	,000		
Likelihood Ratio	76,903	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	74,356	1	,000		
N of Valid Cases	2816				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 326,36.

Crosstab

Count

		Var Obj		Total
		,00	1,00	
Cump_Imp	,00	123	73	196
	1,00	885	493	1378
	2,00	941	301	1242
Total		1949	867	2816

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	44,955 ^a	2	,000
Likelihood Ratio	45,539	2	,000
Linear-by-Linear Association	39,536	1	,000
N of Valid Cases	2816		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 60,35.

Crosstab

Count

		Var. Obj		Total
		,00	1,00	
Vehic	,00	1413	719	2132
	1,00	536	148	684
Total		1949	867	2816

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	35,503 ^b	1	,000		
Continuity Correction ^a	34,938	1	,000		
Likelihood Ratio	37,196	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	35,490	1	,000		
N of Valid Cases	2816				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 210,59.

Crosstab

Count

		Var. Obj		Total
		,00	1,00	
Region	,00	1503	658	2161
	1,00	446	209	655
Total		1949	867	2816

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,502 ^b	1	,478		
Continuity Correction ^a	,436	1	,509		
Likelihood Ratio	,500	1	,479		
Fisher's Exact Test				,499	,254
Linear-by-Linear Association	,502	1	,478		
N of Valid Cases	2816				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 201,66.

Crosstab

Count

		Var Obj		Total
		,00	1,00	
SerBas	,00	1144	569	1713
	1,00	805	298	1103
Total		1949	867	2816

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	12,101 ^b	1	,001		
Continuity Correction ^a	11,812	1	,001		
Likelihood Ratio	12,217	1	,000		
Fisher's Exact Test				,001	,000
Linear-by-Linear Association	12,097	1	,001		
N of Valid Cases	2816				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 339,60.

Crosstab

Count

		Var Obj		Total
		,00	1,00	
SiNuevo	,00	973	269	1242
	1,00	976	598	1574
Total		1949	867	2816

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	86,916 ^b	1	,000		
Continuity Correction ^a	86,151	1	,000		
Likelihood Ratio	88,800	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	86,885	1	,000		
N of Valid Cases	2816				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 382,39.

13.2.2 Desempeño de Puntos de Corte del Modelo Nuevo

En este anexo, se mostrarán las matrices de confusión para los distintos puntos de corte del modelo, y que fueron utilizadas para generar la tabla de desempeños.

Classification Table^d

		Predicted					
		Selected Cases ^a			Unselected Cases ^{b,c}		
		Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00				
Observed							
Step 1	Var_Obj	,00					
		1,00					
	Overall Percentage						
		0	695	,0	0	173	,0
		0	706	100,0	0	161	100,0
				50,4			48,3

a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1

b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,050

Classification Table^d

		Predicted					
		Selected Cases ^a			Unselected Cases ^{b,c}		
		Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00				
Observed							
Step 1	Var_Obj	,00					
		1,00					
	Overall Percentage						
		5	689	,8	2	171	1,0
		2	704	99,7	0	161	100,0
				50,6			48,8

a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1

b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,100

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
			,00	1,00		,00	1,00	
Observed								
Step 1	Var_Obj	,00	23	672	3,3	4	168	2,6
		1,00	2	704	99,7	1	160	99,4
	Overall Percentage				51,9			49,3

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,150

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
			,00	1,00		,00	1,00	
Observed								
Step 1	Var_Obj	,00	54	640	7,8	12	160	7,2
		1,00	9	697	98,7	1	160	99,4
	Overall Percentage				53,6			51,7

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,200

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
			,00	1,00		,00	1,00	
Observed								
Step 1	Var_Obj	,00	105	590	15,1	24	149	13,7
		1,00	26	680	96,3	9	152	94,4
	Overall Percentage				56,0			52,6

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,250

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
			,00	1,00		,00	1,00	
Observed								
Step 1	Var_Obj	,00	163	531	23,5	36	136	21,1
		1,00	46	660	93,5	17	144	89,4
	Overall Percentage				58,8			54,1

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,300

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
			,00	1,00		,00	1,00	
Observed								
Step 1	Var_Obj	,00	227	468	32,7	54	119	31,2
		1,00	78	628	89,0	24	137	85,1
	Overall Percentage				61,0			57,2

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,350

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
			,00	1,00		,00	1,00	
Observed								
Step 1	Var_Obj	,00	292	403	42,0	70	102	40,7
		1,00	118	588	83,3	35	126	78,3
	Overall Percentage				62,8			58,8

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,400

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	360	334	51,9	90	83	52,1
		1,00	175	531	75,2	46	115	71,4
	Overall Percentage				63,6			61,4

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,450

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	438	256	63,1	107	65	62,1
		1,00	244	462	65,4	56	105	65,2
	Overall Percentage				64,3			63,6

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,500

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	497	198	71,5	121	51	70,4
		1,00	305	401	56,8	74	87	54,0
	Overall Percentage				64,1			62,5

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,550

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	557	138	80,1	145	28	83,8
		1,00	364	342	48,4	93	68	42,2
	Overall Percentage				64,2			63,7

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,600

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	606	89	87,3	150	22	87,1
		1,00	460	246	34,8	112	49	30,4
	Overall Percentage				60,8			59,8

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,650

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	643	51	92,6	162	11	93,8
		1,00	538	168	23,8	129	32	19,9
	Overall Percentage				57,9			58,1

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,700

Classification Table^d

		Predicted						
		Selected Cases ^a			Unselected Cases ^{b,c}			
		Var. Obj		Percentage Correct	Var. Obj		Percentage Correct	
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	671	24	96,6	166	7	95,9
		1,00	623	83	11,8	142	19	11,8
	Overall Percentage				53,8			55,3

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,750

Classification Table^d

		Predicted						
		Selected Cases ^a			Unselected Cases ^{b,c}			
		Var. Obj		Percentage Correct	Var. Obj		Percentage Correct	
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	689	5	99,2	170	2	98,7
		1,00	676	30	4,2	156	5	3,1
	Overall Percentage				51,4			52,6

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,800

Classification Table^d

		Predicted						
		Selected Cases ^a			Unselected Cases ^{b,c}			
		Var. Obj		Percentage Correct	Var. Obj		Percentage Correct	
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	689	5	99,2	170	2	98,7
		1,00	676	30	4,2	156	5	3,1
	Overall Percentage				51,4			52,6

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,800

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	695	0	100,0	173	0	100,0
		1,00	706	0	,0	161	0	,0
	Overall Percentage				49,6			51,7

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,900

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	695	0	100,0	173	0	100,0
		1,00	706	0	,0	161	0	,0
	Overall Percentage				49,6			51,7

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,950

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	695	0	100,0	173	0	100,0
		1,00	706	0	,0	161	0	,0
	Overall Percentage				49,6			51,7

- a. Selected cases Approximately 80 % of cases (SAMPLE) EQ 1
- b. Unselected cases Approximately 80 % of cases (SAMPLE) NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,990

13.2.3 Estabilidad del Modelo Nuevo

En este anexo, se mostrarán tanto las matrices de confusión de cada partición, como los coeficientes Beta asociados a las variables de cada modelo calibrado, con corte 0.6. Una forma de corroborar que las particiones eran distintas, es mediante la variable Filter_i, a través de la cual se seleccionaron los casos.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj		Percentage Correct	Var. Obj		Percentage Correct
,00	1,00	,00	1,00					
Step 1	Var. Obj	,00	557	138	80,1	145	28	83,8
		1,00	364	342	48,4	93	68	42,2
Overall Percentage					64,2			63,7

a. Selected cases Filter1 EQ 1

b. Unselected cases Filter1 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,445	,131	11,542	1	,001	,641
	Edad	-,027	,005	25,722	1	,000	,973
	Cump_Imp			22,620	2	,000	
	Cump_Imp(1)	,520	,228	5,225	1	,022	1,682
	Cump_Imp(2)	,580	,125	21,692	1	,000	1,786
	NCred1	-,089	,044	4,034	1	,045	,915
	Vehic	-,664	,143	21,654	1	,000	,515
	Region	,246	,137	3,226	1	,072	1,279
	SerBas	-,159	,121	1,714	1	,190	,853
	SiNuevo	,722	,169	18,128	1	,000	2,058
	LNDCDCm1	,088	,020	18,547	1	,000	1,092
	Constant	,568	,304	3,502	1	,061	1,765

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, NCred1, Vehic, Region, SerBas, SiNuevo, LNDCDCm1.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj.		Percentage Correct	Var. Obj.		Percentage Correct
			,00	1,00		,00	1,00	
Step 1	Var. Obj.	,00	562	135	80,6	140	29	82,7
		1,00	368	318	46,4	103	78	43,1
Overall Percentage					63,6			62,2

a. Selected cases Filter2 EQ 1

b. Unselected cases Filter2 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,382	,130	8,673	1	,003	,683
	Edad	-,025	,005	21,679	1	,000	,976
	Cump_Imp			25,482	2	,000	
	Cump_Imp(1)	,448	,225	3,954	1	,047	1,565
	Cump_Imp(2)	,633	,126	25,253	1	,000	1,882
	NCred1	-,155	,046	11,113	1	,001	,857
	Vehic	-,640	,144	19,743	1	,000	,527
	Region	,212	,139	2,329	1	,127	1,237
	SerBas	-,213	,121	3,065	1	,080	,808
	SiNuevo	,565	,171	10,954	1	,001	1,760
	LNDCDCm1	,084	,020	17,044	1	,000	1,088
	Constant	,672	,310	4,708	1	,030	1,959

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, NCred1, Vehic, Region, SerBas, SiNuevo, LNDCDCm1.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Step 1	Var_Obj	,00	575	129	81,7	129	35	78,5
		1,00	366	321	46,7	106	74	41,1
Overall Percentage					64,4			58,9

a. Selected cases Filter3 EQ 1

b. Unselected cases Filter3 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)						
CasaProp	-,336	,130	6,694	1	,010	,715
Edad	-,025	,005	22,857	1	,000	,975
Cump_Imp			27,739	2	,000	
Cump_Imp(1)	,521	,231	5,093	1	,024	1,683
Cump_Imp(2)	,649	,125	27,161	1	,000	1,913
NCred1	-,117	,045	6,729	1	,009	,890
Vehic	-,751	,144	27,386	1	,000	,472
Region	,130	,138	,887	1	,346	1,139
SerBas	-,119	,122	,962	1	,327	,887
SiNuevo	,716	,171	17,570	1	,000	2,047
LNDCDCm1	,085	,020	17,453	1	,000	1,089
Constant	,488	,309	2,488	1	,115	1,629

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, NCred1, Vehic, Region, SerBas, SiNuevo, LNDCDCm1.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Step 1	Var_Obj	,00	564	134	80,8	139	30	82,2
		1,00	374	325	46,5	93	75	44,6
Overall Percentage					63,6			63,5

a. Selected cases Filter4 EQ 1

b. Unselected cases Filter4 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,364	,129	7,970	1	,005	,695
	Edad	-,028	,005	27,738	1	,000	,973
	Cump_Imp			21,209	2	,000	
	Cump_Imp(1)	,303	,235	1,669	1	,196	1,354
	Cump_Imp(2)	,572	,124	21,208	1	,000	1,771
	NCred1	-,127	,047	7,419	1	,006	,881
	Vehic	-,595	,141	17,811	1	,000	,551
	Region	,180	,136	1,761	1	,184	1,198
	SerBas	-,172	,121	2,011	1	,156	,842
	SiNuevo	,593	,171	11,968	1	,001	1,810
	LNDCDCm1	,078	,020	14,782	1	,000	1,081
	Constant	,783	,307	6,524	1	,011	2,189

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, NCred1, Vehic, Region, SerBas, SiNuevo, LNDCDCm1.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj.		Percentage Correct	Var. Obj.		Percentage Correct
			,00	1,00		,00	1,00	
Step 1	Var. Obj.	,00	573	126	82,0	130	39	77,0
		1,00	372	327	46,8	87	81	48,2
Overall Percentage					64,4			62,7

a. Selected cases Filter5 EQ 1

b. Unselected cases Filter5 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,353	,128	7,560	1	,006	,703
	Edad	-,028	,005	28,470	1	,000	,972
	Cump_Imp			25,509	2	,000	
	Cump_Imp(1)	,539	,231	5,414	1	,020	1,714
	Cump_Imp(2)	,608	,123	24,520	1	,000	1,838
	NCred1	-,082	,043	3,684	1	,055	,921
	Vehic	-,584	,141	17,286	1	,000	,558
	Region	,172	,135	1,616	1	,204	1,188
	SerBas	-,182	,121	2,258	1	,133	,834
	SiNuevo	,721	,167	18,746	1	,000	2,057
	LNDCDCm1	,090	,020	19,899	1	,000	1,094
	Constant	,562	,299	3,537	1	,060	1,755

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, NCred1, Vehic, Region, SerBas, SiNuevo, LNDCDCm1.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var Obj		Percentage Correct	Var Obj		Percentage Correct
			,00	1,00		,00	1,00	
Step 1	Var_Obj	,00	568	122	82,3	147	31	82,7
		1,00	371	300	44,7	114	82	41,8
Overall Percentage					63,8			61,3

a. Selected cases Filter6 EQ 1

b. Unselected cases Filter6 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,385	,131	8,650	1	,003	,680
	Edad	-,023	,005	18,888	1	,000	,977
	Cump_Imp			21,501	2	,000	
	Cump_Imp(1)	,426	,231	3,422	1	,064	1,532
	Cump_Imp(2)	,579	,126	21,209	1	,000	1,784
	NCred1	-,134	,046	8,452	1	,004	,875
	Vehic	-,593	,142	17,313	1	,000	,553
	Region	,259	,138	3,502	1	,061	1,295
	SerBas	-,164	,123	1,799	1	,180	,848
	SiNuevo	,570	,173	10,853	1	,001	1,769
	LNDCCM1	,076	,020	13,760	1	,000	1,079
	Constant	,565	,310	3,333	1	,068	1,760

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, NCred1, Vehic, Region, SerBas, SiNuevo, LNDCCM1.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj		Percentage Correct	Var. Obj		Percentage Correct
			,00	1,00		,00	1,00	
Step 1	Var_Obj	,00	581	119	83,0	138	29	82,7
		1,00	367	309	45,7	117	74	38,7
Overall Percentage					64,7			59,3

a. Selected cases Filter7 EQ 1

b. Unselected cases Filter7 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,516	,131	15,491	1	,000	,597
	Edad	-,028	,005	26,708	1	,000	,973
	Cump_Imp			18,225	2	,000	
	Cump_Imp(1)	,416	,231	3,238	1	,072	1,516
	Cump_Imp(2)	,529	,125	17,878	1	,000	1,697
	NCred1	-,115	,046	6,384	1	,012	,891
	Vehic	-,567	,143	15,783	1	,000	,567
	Region	,208	,139	2,233	1	,135	1,232
	SerBas	-,159	,123	1,684	1	,194	,853
	SiNuevo	,624	,172	13,194	1	,000	1,866
	LNDCDCm1	,084	,020	17,023	1	,000	1,088
	Constant	,721	,309	5,429	1	,020	2,056

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, NCred1, Vehic, Region, SerBas, SiNuevo, LNDCDCm1.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj.		Percentage Correct	Var. Obj.		Percentage Correct
,00	1,00	,00	1,00					
Step 1	Var. Obj.	,00	566	127	81,7	140	34	80,6
		1,00	375	315	45,7	98	79	44,6
Overall Percentage					63,7			62,4

a. Selected cases Filter8 EQ 1

b. Unselected cases Filter8 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,400	,129	9,531	1	,002	,671
	Edad	-,029	,005	31,102	1	,000	,971
	Cump_Imp			24,966	2	,000	
	Cump_Imp(1)	,521	,234	4,941	1	,026	1,683
	Cump_Imp(2)	,612	,124	24,280	1	,000	1,844
	NCred1	-,099	,044	5,044	1	,025	,906
	Vehic	-,648	,142	20,704	1	,000	,523
	Region	,240	,136	3,105	1	,078	1,271
	SerBas	-,086	,122	,500	1	,479	,918
	SiNuevo	,625	,169	13,651	1	,000	1,868
	LNDCDCm1	,087	,020	18,259	1	,000	1,091
	Constant	,687	,303	5,127	1	,024	1,988

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, NCred1, Vehic, Region, SerBas, SiNuevo, LNDCDCm1.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj.		Percentage Correct	Var. Obj.		Percentage Correct
,00	1,00	,00	1,00					
Step 1	Var. Obj.	,00	570	125	82,0	137	35	79,6
		1,00	372	312	45,6	107	76	41,5
Overall Percentage					64,0			60,0

a. Selected cases Filter9 EQ 1

b. Unselected cases Filter9 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,382	,131	8,524	1	,004	,683
	Edad	-,027	,005	25,652	1	,000	,974
	Cump_Imp			23,346	2	,000	
	Cump_Imp(1)	,559	,237	5,553	1	,018	1,749
	Cump_Imp(2)	,587	,124	22,246	1	,000	1,799
	NCred1	-,150	,046	10,659	1	,001	,861
	Vehic	-,587	,141	17,259	1	,000	,556
	Region	,222	,138	2,590	1	,108	1,248
	SerBas	-,164	,122	1,793	1	,181	,849
	SiNuevo	,507	,169	8,972	1	,003	1,661
	LNDCDCm1	,071	,020	12,410	1	,000	1,074
Constant	,814	,308	6,999	1	,008	2,256	

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, NCred1, Vehic, Region, SerBas, SiNuevo, LNDCDCm1.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj		Percentage Correct	Var. Obj		Percentage Correct
			,00	1,00		,00	1,00	
Step 1	Var_Obj	,00	551	139	79,8	145	32	81,9
		1,00	358	338	48,6	96	75	43,9
Overall Percentage					64,1			63,2

a. Selected cases Filter10 EQ 1

b. Unselected cases Filter10 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,380	,130	8,583	1	,003	,684
	Edad	-,027	,005	24,987	1	,000	,974
	Cump_Imp			21,303	2	,000	
	Cump_Imp(1)	,512	,226	5,118	1	,024	1,668
	Cump_Imp(2)	,566	,125	20,413	1	,000	1,762
	NCred1	-,176	,048	13,596	1	,000	,838
	Vehic	-,677	,144	22,012	1	,000	,508
	Region	,360	,138	6,864	1	,009	1,434
	SerBas	-,107	,123	,752	1	,386	,899
	SiNuevo	,548	,172	10,094	1	,001	1,730
	LNDCCM1	,078	,020	14,930	1	,000	1,082
	Constant	,813	,312	6,808	1	,009	2,255

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, NCred1, Vehic, Region, SerBas, SiNuevo, LNDCCM1.

13.3 Solución Alternativa

13.3.1 Análisis Univariado

En este anexo, sólo se mostrará el análisis univariado de las variables candidatas, ya que el de las variables del modelo actual, fue expuesto en el ítem 10.2.1.

Frecuencias

	Var_Obj	N
Antig	,00	1949
	1,00	867
	Total	2816
dia	,00	1949
	1,00	867
	Total	2816

Test Statistics^a

		Antig	dia
Most Extreme Differences	Absolute	,224	,078
	Positive	,000	,000
	Negative	-,224	-,078
Kolmogorov-Smirnov Z		5,482	1,899
Asymp. Sig. (2-tailed)		,000	,001

a. Grouping Variable: Var_Obj

Crosstab

Count		Var Obj		Total
		,00	1,00	
Ini_Act	,00	794	436	1230
	1,00	1142	421	1563
Total		1936	857	2793

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	23,447 ^b	1	,000		
Continuity Correction ^a	23,049	1	,000		
Likelihood Ratio	23,360	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	23,439	1	,000		
N of Valid Cases	2793				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 377,41.

Crosstab

Count		Var Obj		Total
		,00	1,00	
Pat_	,00	1237	588	1825
Com	1,00	699	269	968
Total		1936	857	2793

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5,836 ^b	1	,016		
Continuity Correction ^a	5,630	1	,018		
Likelihood Ratio	5,889	1	,015		
Fisher's Exact Test				,016	,009
Linear-by-Linear Association	5,834	1	,016		
N of Valid Cases	2793				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 297,02.

Crosstab

Count		Var Obj		Total
		,00	1,00	
Perm_M	,00	1501	677	2178
	1,00	435	180	615
Total		1936	857	2793

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,743 ^b	1	,389		
Continuity Correction ^a	,660	1	,417		
Likelihood Ratio	,748	1	,387		
Fisher's Exact Test				,400	,209
Linear-by-Linear Association	,743	1	,389		
N of Valid Cases	2793				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 188,71.

Crosstab

Count		Var Obj		Total
		,00	1,00	
Ahorro	,00	1824	835	2659
	1,00	125	32	157
Total		1949	867	2816

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8,450 ^a	1	,004		
Continuity Correction ^b	7,940	1	,005		
Likelihood Ratio	9,088	1	,003		
Fisher's Exact Test				,003	,002
Linear-by-Linear Association	8,447	1	,004		
N of Valid Cases	2816				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 48,34.

Crosstab

Count		Var Obj		Total
		,00	1,00	
Tuvo_Cred	,00	720	464	1184
	1,00	1229	403	1632
Total		1949	867	2816

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	67,662 ^b	1	,000		
Continuity Correction ^a	66,983	1	,000		
Likelihood Ratio	67,203	1	,000		
Fisher's Exact Test				,000	,000
Linear-by-Linear Association	67,638	1	,000		
N of Valid Cases	2816				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 364,53.

Crosstab

Count		Var Obj		Total
		,00	1,00	
Otros_Act	,00	202	56	258
	1,00	1747	811	2558
Total		1949	867	2816

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	10,996 ^b	1	,001		
Continuity Correction ^a	10,532	1	,001		
Likelihood Ratio	11,667	1	,001		
Fisher's Exact Test				,001	,000
Linear-by-Linear Association	10,992	1	,001		
N of Valid Cases	2816				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 79,43.

Crosstab

Count

		Var. Obj		Total
		,00	1,00	
Si_	,00	631	272	903
Hijos	1,00	1318	595	1913
Total		1949	867	2816

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,277 ^b	1	,599		
Continuity Correction ^a	,233	1	,629		
Likelihood Ratio	,278	1	,598		
Fisher's Exact Test				,630	,315
Linear-by-Linear Association	,277	1	,599		
N of Valid Cases	2816				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 278,02.

13.3.2 Análisis Multivariado

En este anexo, se muestra la Matriz de Correlación completa, que contiene todas las variables ingresadas al método *stepwise*.

Correlations

		CasaProp	Edad	Cump_ Imp	NCred1	Vehic	Region	SerBas	SiNuevo	LNDCDCm1	Ini Act	Pat Com	dia	Antig	Ahorro	Tuvo Cred	Otros Act
CasaProp	Pearson Correlation	1	.277**	.102**	.217**	.071**	.020	.184**	-.145**	.163**	.107**	.067**	.177**	.204**	.048*	.163**	-.129**
	Sig. (2-tailed)		.000	.000	.000	.000	.292	.000	.000	.000	.000	.000	.000	.000	.012	.000	.000
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
Edad	Pearson Correlation	.277**	1	.016	.246**	-.005	.011	.065**	-.207**	.146**	-.034	.003	.052**	.198**	.020	.202**	-.046*
	Sig. (2-tailed)	.000		.394	.000	.779	.544	.001	.000	.000	.073	.854	.006	.000	.294	.000	.015
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
Cump_ Imp	Pearson Correlation	.102**	.016	1	.067**	.085**	.038*	.090**	-.031	.134**	.514**	.310**	.142**	.104**	.052**	.086**	.040*
	Sig. (2-tailed)	.000	.394		.000	.000	.043	.000	.097	.000	.000	.000	.000	.000	.005	.000	.034
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
NCred1	Pearson Correlation	.217**	.246**	.067**	1	.030	.037	.045*	-.657**	.371**	.031	.085**	.037	.594**	.110**	.540**	-.002
	Sig. (2-tailed)	.000	.000	.000		.108	.050	.017	.000	.000	.103	.000	.050	.000	.000	.000	.913
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
Vehic	Pearson Correlation	.071**	-.005	.085**	.030	1	.064**	.080**	-.009	.136**	.181**	-.119**	.056**	.045*	.050**	.068**	-.437**
	Sig. (2-tailed)	.000	.779	.000	.108		.001	.000	.638	.000	.000	.003	.017	.008	.000	.000	.000
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
Region	Pearson Correlation	.020	.011	.038*	.037	.064**	1	.092**	-.005	.048*	.053**	.188**	.078**	.032	-.031	.018	-.026
	Sig. (2-tailed)	.292	.544	.043	.050	.001		.000	.780	.010	.005	.000	.000	.089	.098	.348	.165
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
SerBas	Pearson Correlation	.184**	.065**	.090**	.045*	.080**	.092**	1	.004	.021	.083**	.073**	.871**	.067**	.036	.019	-.058**
	Sig. (2-tailed)	.000	.001	.000	.017	.000	.000		.847	.271	.000	.000	.000	.000	.053	.318	.002
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
SiNuevo	Pearson Correlation	-.145**	-.207**	-.031	-.657**	-.009	-.005	.004	1	-.470**	.024	-.043*	-.010	-.383**	-.099**	-.757**	.005
	Sig. (2-tailed)	.000	.000	.097	.000	.638	.780	.847		.000	.198	.025	.599	.000	.000	.000	.771
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
LNDCDCm1	Pearson Correlation	.163**	.146**	.134**	.371**	.136**	.048*	.021	-.470**	1	.174**	.174**	.065**	.201**	.076**	.679**	-.038*
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.010	.271	.000		.000	.000	.001	.000	.000	.000	.046
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
Ini Act	Pearson Correlation	.107**	-.034	.514**	.031	.181**	.053**	.083**	.024	.174**	1	.434**	.172**	.076**	.029	.057**	-.029
	Sig. (2-tailed)	.000	.073	.000	.103	.000	.005	.000	.198	.000		.000	.000	.000	.123	.002	.121
	N	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793
Pat Com	Pearson Correlation	.087**	.003	.310**	.085**	-.119**	.188**	.073**	-.043*	.174**	.434**	1	.177**	.075**	.021	.093**	.164**
	Sig. (2-tailed)	.000	.854	.000	.000	.000	.000	.000	.025	.000	.000		.000	.000	.276	.000	.000
	N	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793	2793
dia	Pearson Correlation	.177**	.052**	.142**	.037	.056**	.078**	.871**	-.010	.065**	.172**	.177**	1	.059**	.038*	.041*	-.031
	Sig. (2-tailed)	.000	.006	.000	.050	.003	.000	.000	.599	.001	.000	.000		.002	.044	.031	.098
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
Antig	Pearson Correlation	.204**	.198**	.104**	.594**	.045*	.032	.067**	-.383**	.201**	.076**	.075**	.059**	1	.085**	.353**	-.031
	Sig. (2-tailed)	.000	.000	.000	.000	.017	.089	.000	.000	.000	.000	.000	.002		.000	.000	.106
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
Ahorro	Pearson Correlation	.048*	.020	.052**	.110**	.050**	-.031	.036	-.099**	.076**	.029	.021	.038*	.085**	1	.088**	-.019
	Sig. (2-tailed)	.012	.294	.005	.000	.008	.098	.053	.000	.000	.123	.276	.044	.000		.000	.303
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
Tuvo Cred	Pearson Correlation	.163**	.202**	.086**	.540**	.068**	.018	.019	-.757**	.679**	.057**	.093**	.041*	.353**	.088**	1	-.031
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.348	.318	.000	.000	.002	.000	.031	.000	.000		.099
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816
Otros Act	Pearson Correlation	-.129**	-.046*	.040*	-.002	-.437**	-.026	-.058**	.005	-.038*	-.029	.164**	-.031	-.031	-.019	-.031	1
	Sig. (2-tailed)	.000	.015	.034	.913	.000	.165	.002	.771	.046	.121	.000	.098	.106	.303	.099	
	N	2816	2816	2816	2816	2816	2816	2816	2816	2816	2793	2793	2816	2816	2816	2816	2816

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

13.3.3 Desempeño de Puntos de Corte del Modelo Alternativo

En este anexo, se mostrarán las matrices de confusión para los distintos puntos de corte del modelo propuesto por el método *stepwise*, y que fueron utilizadas para generar la tabla de desempeños.

Classification Table^d

Observed		Predicted					
		Selected Cases ^a			Unselected Cases ^{b,c}		
		Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00				
Step 1	Var_Obj	,00	1,00				
		0	695	,0	0	173	,0
		0	706	100,0	0	161	100,0
	Overall Percentage			50,4			48,3

a. Selected cases Filter1 EQ 1

b. Unselected cases Filter1 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,050

Classification Table^d

Observed		Predicted					
		Selected Cases ^a			Unselected Cases ^{b,c}		
		Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00				
Step 1	Var_Obj	,00	1,00				
		7	688	1,0	1	172	,5
		0	706	100,0	0	161	100,0
	Overall Percentage			50,9			48,5

a. Selected cases Filter1 EQ 1

b. Unselected cases Filter1 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,100

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	25	670	3,6	4	169	2,3
		1,00	1	705	99,9	0	161	100,0
	Overall Percentage				52,1			49,5

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,150

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	56	639	8,1	13	159	7,7
		1,00	9	697	98,7	1	160	99,4
	Overall Percentage				53,8			52,0

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,200

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	104	591	14,9	25	147	14,7
		1,00	24	682	96,6	7	154	95,7
	Overall Percentage				56,1			53,8

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,250

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	162	533	23,3	37	135	21,6
		1,00	50	656	92,9	19	142	88,2
	Overall Percentage				58,4			53,8

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,300

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	226	469	32,5	53	120	30,7
		1,00	81	625	88,5	24	137	85,1
	Overall Percentage				60,7			56,9

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,350

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	299	396	43,0	71	102	41,0
		1,00	122	584	82,7	31	130	80,7
	Overall Percentage				63,0			60,2

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,400

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	362	332	52,1	89	84	51,3
		1,00	180	526	74,5	43	118	73,3
	Overall Percentage				63,4			61,9

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,450

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	431	263	62,1	105	67	61,1
		1,00	244	462	65,4	56	105	65,2
	Overall Percentage				63,8			63,1

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,500

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	502	192	72,3	123	49	71,4
		1,00	295	411	58,2	70	91	56,5
	Overall Percentage				65,2			64,2

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,550

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	558	137	80,3	137	36	79,4
		1,00	367	339	48,0	95	66	41,0
	Overall Percentage				64,0			60,9

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,600

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	609	86	87,6	149	24	86,3
		1,00	452	254	36,0	110	51	31,7
	Overall Percentage				61,6			60,0

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,650

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	644	51	92,7	162	11	93,6
		1,00	538	168	23,8	134	27	16,8
	Overall Percentage				58,0			56,5

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,700

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	669	25	96,3	166	7	95,9
		1,00	617	89	12,6	140	21	13,0
	Overall Percentage				54,1			55,9

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,750

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	689	6	99,2	171	2	99,0
		1,00	680	26	3,7	155	6	3,7
	Overall Percentage				51,0			53,0

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,800

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	693	2	99,7	172	0	99,7
		1,00	697	9	1,3	158	3	1,9
	Overall Percentage				50,1			52,5

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,850

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	695	0	100,0	173	0	100,0
		1,00	706	0	,0	161	0	,0
	Overall Percentage				49,6			51,7

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,900

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	695	0	100,0	173	0	100,0
		1,00	706	0	,0	161	0	,0
	Overall Percentage				49,6			51,7

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,950

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	695	0	100,0	173	0	100,0
		1,00	706	0	,0	161	0	,0
	Overall Percentage				49,6			51,7

- a. Selected cases Filter1 EQ 1
- b. Unselected cases Filter1 NE 1
- c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.
- d. The cut value is ,990

13.3.4 Estabilidad del Modelo Alternativo

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Step 1	Var_Obj	,00	558	137	80,3	137	36	79,4
		1,00	367	339	48,0	95	66	41,0
Overall Percentage					64,0			60,9

a. Selected cases Filter1 EQ 1

b. Unselected cases Filter1 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,466	,130	12,871	1	,000	,628
	Edad	-,027	,005	25,606	1	,000	,974
	Cump_Imp			21,692	2	,000	
	Cump_Imp(1)	,499	,228	4,800	1	,028	1,646
	Cump_Imp(2)	,572	,125	20,913	1	,000	1,771
	Vehic	-,667	,142	21,971	1	,000	,513
	SiNuevo	,558	,180	9,584	1	,002	1,747
	LNDCCM1	,119	,026	21,599	1	,000	1,126
	Tuvo_Cred	-,490	,218	5,046	1	,025	,612
	Antig	-,052	,023	5,102	1	,024	,949
	Constant	,721	,307	5,506	1	,019	2,056

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, Vehic, SiNuevo, LNDCCM1, Tuvo_Cred, Antig.

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var_Obj		Percentage Correct	Var_Obj		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	562	136	80,5	139	31	81,9
		1,00	371	315	45,9	107	74	40,9
	Overall Percentage				63,3			60,7

a. Selected cases Filter2 EQ 1

b. Unselected cases Filter2 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,405	,128	9,940	1	,002	,667
	Edad	-,026	,005	23,358	1	,000	,975
	Cump_Imp			24,569	2	,000	
	Cump_Imp(1)	,438	,225	3,786	1	,052	1,550
	Cump_Imp(2)	,620	,126	24,347	1	,000	1,860
	Vehic	-,648	,144	20,307	1	,000	,523
	SiNuevo	,531	,182	8,562	1	,003	1,701
	LNDCCm1	,112	,026	19,162	1	,000	1,118
	Tuvo_Cred	-,463	,219	4,475	1	,034	,630
	Antig	-,070	,024	8,912	1	,003	,932
	Constant	,646	,312	4,291	1	,038	1,907

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, Vehic, SiNuevo, LNDCCm1, Tuvo_Cred, Antig.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj		Percentage Correct	Var. Obj		Percentage Correct
,00	1,00	,00	1,00					
Step 1	Var_Obj	,00	573	130	81,5	126	38	76,9
		1,00	376	311	45,3	98	82	45,6
Overall Percentage					63,6			60,5

a. Selected cases Filter3 EQ 1

b. Unselected cases Filter3 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,363	,128	8,057	1	,005	,695
	Edad	-,026	,005	23,761	1	,000	,975
	Cump_Imp			26,341	2	,000	
	Cump_Imp(1)	,494	,231	4,581	1	,032	1,638
	Cump_Imp(2)	,634	,125	25,886	1	,000	1,886
	Vehic	-,758	,143	27,971	1	,000	,468
	SiNuevo	,630	,180	12,230	1	,000	1,878
	LNDCCM1	,114	,025	20,311	1	,000	1,120
	Tuvo_Cred	-,482	,215	5,049	1	,025	,617
	Antig	-,047	,023	4,335	1	,037	,954
	Constant	,551	,311	3,134	1	,077	1,734

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, Vehic, SiNuevo, LNDCCM1, Tuvo_Cred, Antig.

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj.		Percentage Correct	Var. Obj.		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var. Obj.	,00	559	139	80,1	136	34	80,1
		1,00	384	315	45,1	94	74	44,0
	Overall Percentage				62,6			62,1

a. Selected cases Filter4 EQ 1

b. Unselected cases Filter4 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,389	,127	9,360	1	,002	,678
	Edad	-,028	,005	28,326	1	,000	,973
	Cump_Imp			20,698	2	,000	
	Cump_Imp(1)	,290	,235	1,518	1	,218	1,336
	Cump_Imp(2)	,566	,124	20,694	1	,000	1,762
	Vehic	-,608	,141	18,717	1	,000	,544
	SiNuevo	,535	,181	8,752	1	,003	1,708
	LNDCCM1	,104	,025	16,900	1	,000	1,110
	Tuvo_Cred	-,428	,218	3,853	1	,050	,652
	Antig	-,062	,023	7,324	1	,007	,940
	Constant	,787	,306	6,617	1	,010	2,196

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, Vehic, SiNuevo, LNDCCM1, Tuvo_Cred, Antig.

Classification Table^d

Observed		Predicted						
		Selected Cases ^a			Unselected Cases ^{b,c}			
		Var. Obj		Percentage Correct	Var. Obj		Percentage Correct	
,00	1,00	,00	1,00					
Step 1	Var_Obj	,00	566	132	81,1	132	37	78,1
		1,00	384	315	45,1	87	81	48,2
Overall Percentage					63,1			63,2

a. Selected cases Filter5 EQ 1

b. Unselected cases Filter5 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,382	,127	9,095	1	,003	,682
	Edad	-,028	,005	28,579	1	,000	,972
	Cump_imp			24,917	2	,000	
	Cump_imp(1)	,537	,232	5,365	1	,021	1,710
	Cump_imp(2)	,602	,123	23,910	1	,000	1,825
	Vehic	-,590	,140	17,715	1	,000	,555
	SiNuevo	,601	,177	11,562	1	,001	1,824
	LNDCCM1	,115	,025	21,055	1	,000	1,122
	Tuvo_cred	-,406	,215	3,557	1	,059	,666
	Antig	-,048	,023	4,505	1	,034	,953
	Constant	,652	,302	4,672	1	,031	1,919

a Variable(s) entered on step 1: CasaProp, Edad, Cump_imp, Vehic, SiNuevo, LNDCCM1, Tuvo_cred, Antig.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj		Percentage Correct	Var. Obj		Percentage Correct
,00	1,00	,00	1,00					
Step 1	Var_Obj	,00	562	127	81,5	147	30	83,0
		1,00	371	300	44,7	114	82	41,8
Overall Percentage					63,4			61,4

a. Selected cases Filter6 EQ 1

b. Unselected cases Filter6 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,408	,130	9,894	1	,002	,665
	Edad	-,023	,005	18,177	1	,000	,978
	Cump_Imp			19,585	2	,000	
	Cump_Imp(1)	,402	,231	3,020	1	,082	1,495
	Cump_Imp(2)	,553	,126	19,347	1	,000	1,738
	Vehic	-,589	,142	17,122	1	,000	,555
	SiNuevo	,486	,185	6,941	1	,008	1,626
	LNDCCM1	,102	,025	16,304	1	,000	1,107
	Tuvo_Cred	-,450	,218	4,263	1	,039	,637
	Antig	-,076	,024	10,052	1	,002	,927
	Constant	,601	,310	3,762	1	,052	1,823

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, Vehic, SiNuevo, LNDCCM1, Tuvo_Cred, Antig.

Classification Table^d

			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj.		Percentage Correct	Var. Obj.		Percentage Correct
,00	1,00	,00	1,00					
Observed								
Step 1	Var_Obj	,00	573	127	81,8	136	32	81,1
		1,00	378	298	44,1	114	77	40,3
	Overall Percentage				63,3			59,4

a. Selected cases Filter7 EQ 1

b. Unselected cases Filter7 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,535	,130	16,956	1	,000	,586
	Edad	-,027	,005	26,288	1	,000	,973
	Cump_Imp			17,435	2	,000	
	Cump_Imp(1)	,394	,231	2,908	1	,088	1,484
	Cump_Imp(2)	,519	,125	17,171	1	,000	1,681
	Vehic	-,569	,143	15,896	1	,000	,566
	SiNuevo	,490	,182	7,224	1	,007	1,632
	LNDCCM1	,117	,026	21,069	1	,000	1,125
	Tuvo_Cred	-,508	,220	5,343	1	,021	,602
	Antig	-,068	,023	8,560	1	,003	,934
	Constant	,810	,308	6,939	1	,008	2,249

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, Vehic, SiNuevo, LNDCCM1, Tuvo_Cred, Antig.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj		Percentage Correct	Var. Obj		Percentage Correct
,00	1,00	,00	1,00					
Step 1	Var_Obj	,00	562	132	81,0	138	36	79,5
		1,00	375	315	45,7	96	81	45,8
Overall Percentage					63,4			62,5

a. Selected cases Filter8 EQ 1

b. Unselected cases Filter8 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,420	,128	10,800	1	,001	,657
	Edad	-,030	,005	31,201	1	,000	,971
	Cump_Imp			23,506	2	,000	
	Cump_Imp(1)	,504	,234	4,622	1	,032	1,656
	Cump_Imp(2)	,596	,125	22,889	1	,000	1,816
	Vehic	-,644	,142	20,573	1	,000	,525
	SiNuevo	,491	,181	7,342	1	,007	1,634
	LNDCCM1	,117	,026	21,071	1	,000	1,124
	Tuvo_Cred	-,486	,219	4,910	1	,027	,615
	Antig	-,052	,023	5,257	1	,022	,949
	Constant	,838	,306	7,493	1	,006	2,313

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, Vehic, SiNuevo, LNDCCM1, Tuvo_Cred, Antig.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj.		Percentage Correct	Var. Obj.		Percentage Correct
,00	1,00	,00	1,00					
Step 1	Var_Obj	,00	565	130	81,2	137	35	79,6
		1,00	382	302	44,2	101	82	44,8
Overall Percentage					62,8			61,7

a. Selected cases Filter9 EQ 1

b. Unselected cases Filter9 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,425	,128	10,936	1	,001	,654
	Edad	-,027	,005	26,924	1	,000	,973
	Cump_Imp			22,546	2	,000	
	Cump_Imp(1)	,537	,237	5,146	1	,023	1,712
	Cump_Imp(2)	,578	,124	21,583	1	,000	1,782
	Vehic	-,594	,141	17,839	1	,000	,552
	SiNuevo	,521	,179	8,502	1	,004	1,684
	LNDCCM1	,096	,025	14,335	1	,000	1,101
	Tuvo_Cred	-,409	,217	3,550	1	,060	,664
	Antig	-,056	,023	5,793	1	,016	,945
	Constant	,745	,306	5,914	1	,015	2,106

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, Vehic, SiNuevo, LNDCCM1, Tuvo_Cred, Antig.

Classification Table^d

Observed			Predicted					
			Selected Cases ^a			Unselected Cases ^{b,c}		
			Var. Obj.		Percentage Correct	Var. Obj.		Percentage Correct
,00	1,00	,00	1,00					
Step 1	Var_Obj	,00	549	142	79,5	143	33	81,1
		1,00	372	324	46,6	96	75	43,9
Overall Percentage					63,0			62,8

a. Selected cases Filter10 EQ 1

b. Unselected cases Filter10 NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or categorical variables with values out of the range of the selected cases.

d. The cut value is ,600

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	CasaProp	-,400	,128	9,809	1	,002	,670
	Edad	-,027	,005	26,343	1	,000	,973
	Cump_Imp			19,835	2	,000	
	Cump_Imp(1)	,491	,226	4,721	1	,030	1,634
	Cump_Imp(2)	,544	,125	19,004	1	,000	1,722
	Vehic	-,662	,143	21,357	1	,000	,516
	SiNuevo	,666	,181	13,578	1	,000	1,947
	LNDCCM1	,096	,025	14,778	1	,000	1,101
	Tuvo_Cred	-,312	,215	2,106	1	,147	,732
	Antig	-,057	,023	6,128	1	,013	,944
	Constant	,658	,308	4,573	1	,032	1,931

a Variable(s) entered on step 1: CasaProp, Edad, Cump_Imp, Vehic, SiNuevo, LNDCCM1, Tuvo_Cred, Antig.