



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
INGENIERÍA DE SISTEMAS – INVESTIGACIÓN DE OPERACIONES

**CLASIFICACIÓN DE CUENTAS DE TARJETAS DE CRÉDITO CON BASE EN
ANÁLISIS DE RIESGO PARA OPTIMIZAR EL COSTO DE LA COBRANZA EN UN
SISTEMA FINANCIERO**

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN INGENIERÍA

PRESENTA:
ACT. KATHYA LEDESMA AMAYA

TUTOR PRINCIPAL
DRA. MAYRA ELIZONDO CORTÉS, FACULTAD DE INGENIERÍA

MÉXICO, D. F. JUNIO 2013

JURADO ASIGNADO:

Presidente: DR. JOSÉ JESÚS ACOSTA FLORES

Secretario: DRA. IDALIA FLORES DE LA MOTA

Vocal: DRA. MAYRA ELIZONDO CORTÉS

1^{er.} Suplente: DR. FEDERICO HERNÁNDEZ ÁLVAREZ

2^{do.} Suplente: MTRO. JORGE ELIECER SÁNCHEZ CERÓN

Lugar o lugares donde se realizó la tesis: UNAM, POSGRADO DE INGENIERÍA

TUTOR DE TESIS:

DRA. MAYRA ELIZONDO CORTÉS

FIRMA



ÍNDICE

<i>Resúmen</i>	5
<i>Abstract</i>	6
<i>Objetivo</i>	7
<i>Introducción</i>	8
I. Problemática en la cobranza de créditos	11
1.1 Análisis de factibilidad	13
1.2 Propósito del desarrollo	17
1.3 Portafolios impactados	18
1.4 Metodología para el desarrollo	18
II. Instrumentos de análisis	20
2.1 Análisis de Riesgo	20
2.2 Descripción de las fuentes de información	29
2.3 Exclusiones	32
2.4 Definición de la probabilidad de incumplimiento	34
2.4.1 Factores que afectan las definiciones de desempeño	35
2.4.2 Definiciones de desempeño y volúmenes	36
2.5 Análisis exploratorio de los datos	37
III. Desarrollo de la segmentación	39
3.1 Desarrollo	39
3.2 Tamaño de la muestra	40
3.3 Diseño de las muestras	40
3.4 Metodología de la selección de variables independientes	41
3.5 Resultados de la segmentación	42
3.6 Consideraciones para el desarrollo del modelo	43
3.7 Modelo de segmentación	44
3.8 Resultados del modelo final	46
3.9 Segmentos de riesgo	54
3.10 Estrategia de Cobranza	58
3.11 Validación fuera de la muestra	61
3.11.1 Estabilidad de la población	62
3.12 Indicadores de la eficiencia del modelo	63
3.13 Implementación	64



3.14	Monitoreo del modelo de segmentación	65
IV.	Optimización de Costos	66
4.1	Costos asociados a la operación de cobro	67
4.2	Planteamiento del problema a minimizar	69
4.3	Solución al problema de reducción de costos	71
	Conclusiones	76
	Referencias bibliograficas.....	78
	Apéndice A. Instrumentos de Análisis.....	80
A.1	Ejemplos de definiciones de desempeño	80
	Apéndice B. Instrumentos de Análisis.....	82
B.1	<i>Software</i> en línea	82
	Apéndice C. Desarrollo del modelo de segmentación	83
C.1	Diseño de las muestras	83
C.2	Análisis de correlación	83
	Apéndice D. Resultados.....	85
D.1	Prueba de homogeneidad	85
	Apéndice E. Optimización de Costos	87



Resumen

Este trabajo de tesis tiene como objetivo presentar una propuesta para definir perfiles de riesgo de crédito de cuentas de Tarjeta de Crédito, que permitan establecer tratamientos de cobro diferenciados en la operación de cobranza, utilizando árboles de decisión y probabilidades de incumplimiento de pago, con el fin de priorizar la toma de decisiones en el saneamiento del portafolio, enfocándose en grupos de clientes con mayor probabilidad de morosidad, y minimizar los costos en la operación de cobranza.

Este texto va dirigido a estudiantes y profesionistas interesados en la optimización de sistemas bancarios, tales como el costo de la operación de cobranza, es decir, realizar llamadas telefónicas y visitas físicas para solicitar el pago de su deuda a los clientes. Su contenido muestra al lector un trabajo con una propuesta para el desarrollo e implementación de estrategias de cobranza, partiendo de un análisis de factibilidad para el caso Tarjeta de Crédito, así como instrumentos de análisis, tales como fuentes de información, metodología estadística empleada, aplicación de probabilidades de incumplimiento, que van relacionadas con el pago o no pago por parte de los clientes, además de cómo desarrollar un modelo de segmentación que será utilizado en las estrategias de cobranza, dentro del cual deberán hacerse una selección de características relacionadas con el comportamiento crediticio de los clientes. Se presentan resultados finales del modelo de segmentación, grupos de riesgo determinados por este modelo, estrategias de cobranza, una validación con población actual, sugerencias para implementación y una propuesta de seguimiento de estas estrategias. Se concluye con una propuesta para minimizar costos en la operación de cobranza, a través de una función objetivo cuyas restricciones están basadas en la clasificación de las cuentas por riesgo obtenidas a través del modelo de segmentación.



Abstract

This thesis has as its main objective to put forward a proposal that outlines the procedure to define business strategies based on risk profiles for loan portfolios, in specific to Credit Card; such procedures allow establishing differentiated collection strategies within a collections operation. The methodology applied makes use of decision trees together with default probabilities, the objective is to prioritize the decision making process when looking to achieve “healthy” portfolios, it focuses on sets of clients with the highest probability of delinquency, at the same time it minimizes the running cost of the operation.

The target audience of this material is university students and professionals interested on optimization techniques applied to certain operational areas within a given banking system, such as the optimization of running costs of the collections operation, i.e., the running costs of making follow-up calls and personal visits to collect outstanding loan debts. The content of this material provides the reader a proposal to develop and implement collection strategies, it starts with a feasibility analysis on Credit Card, then it describes techniques for analysis, such as analysis of data sources, applied statistical methods, which involve defaulting probabilities related to patterns of payment and defaulting associated to clients. Furthermore, the proposal provides a segmentation model applied to collection strategies, which requires a segmentation exercise for key behavioral indicators associated to the clients in the portfolio of analysis. Final results show the segmentation model, a set of risks obtained with the model, a set of strategies for collection, results for the back-testing exercise, a set of recommendations and a proposal to follow-up the given strategies. The conclusion gives a strategy that minimizes the running costs of the collections operation, based on a target function where the boundaries are the key risks of the selected accounts obtained by the segmentation model.



Objetivo

El presente trabajo de tesis tiene como objetivo definir perfiles de riesgo de crédito de cuentas de Tarjeta de Crédito, que permitan establecer tratamientos de cobro diferenciados en la operación de cobranza, utilizando árboles de decisión y probabilidades de incumplimiento de pago, con el fin de priorizar la toma de decisiones en el saneamiento del portafolio, enfocándose en grupos de clientes con mayor probabilidad de morosidad, y minimizar los costos en la operación de cobranza.



Introducción

Este trabajo surge de la implementación de un sistema sofisticado para la cobranza, del cual derivó la necesidad de desarrollar segmentaciones para clasificar las tarjetas de crédito por nivel de riesgo, y considerar esta clasificación para definir estrategias para priorizar las cuentas para realizar el cobro, así como optimizar el costo operativo de cobranza.

Para llevar a cabo el desarrollo de estas segmentaciones, fue necesario considerar información actualizada, de acuerdo con las condiciones del nuevo sistema y con el nivel de morosidad del portafolio.

El diseño de las estrategias consiste en definir acciones de cobro diferentes de acuerdo con la segmentación obtenida y realizando particiones en la población, es decir, al 50% de la población se le asigna una acción y al otro restante una diferente. Estas estrategias se les denomina “Campeonas y Retadoras”, estos nombres son utilizados para describir las estrategias de cobranza existentes, la estrategia “Campeona” es la que actualmente se aplica a la mayor parte de la población de cuentas, y esta a su vez se pone a prueba con otra estrategia llamada “Retadora”, la cual es una estrategia que se aplica a un segmento muy pequeño de la población de cuentas. Las estrategias “Retadoras” se aplican en ambientes productivos sobre muestras pequeñas estadísticamente robustas, manteniendo control para evitar pérdidas; a través de la definición de segmentos de riesgo (riesgo muy alto, alto, medio, etc.) basados en árboles de decisión que permiten mejorar y optimizar el tratamiento de las cuentas en cobranza para prevención de mora (0 pagos vencidos pero con alta probabilidad de incumplimiento), 1, 2, 3, 4 o más pagos vencidos.

Hoy en día, es muy importante fortalecer las áreas críticas de cobranza en una Institución de crédito e incrementar la productividad en la recuperación de cartera vencida o prevención de ésta, a través de la identificación, optimización y automatización de la administración de la cobranza, las cuales son algunos de los factores críticos que las empresas enfrentan en sus procesos de negocio.



La cobranza es un proceso de negociación integral, a través del cual se proporciona servicio al cliente. La empresa le vende una solución al cliente para resolver el problema que lo aqueja: su deuda. Sin embargo, es fundamental definir una estrategia de cobranza, la cual se pueda orientar por segmentos definidos por su perfil de riesgo dentro del portafolio al corriente (cobranza preventiva “*Early collection*”) o con 1 ó más pagos vencidos, con el fin de tener un direccionamiento preciso de las acciones de cobranza con cada cliente o deudor; “La cobranza de calidad es aquella que deja al cliente satisfecho y logra la recuperación del adeudo”, afirmó Javier Loíacono¹.

Para llevar a cabo una mejor labor de cobranza, es necesario segmentar la cartera asignada a la cobranza de acuerdo con las características comunes de los clientes y/o las cuentas. Por ejemplo, puntajes obtenidos de “*scorecard*” de comportamiento, que incluyen porcentajes de utilización de las tarjetas de créditos (saldo pendiente entre el límite de crédito), porcentajes de pago respecto al total de la deuda, pago vencido al momento de la deuda, número de veces que ha caído en mora y como parte sustancial e importante de mencionar, ya que esta forma parte de los modelos de segmentación aquí mostrados, es el comportamiento crediticio con el resto de la banca, el cual proviene de fuentes externas como Buró de Crédito². La identidad privada recibe información oportuna, confiable y segura de personas y empresas que han tenido o tienen algún tipo de crédito, cuya misión es proveer soluciones a otorgantes de crédito y consumidores para administrar el riesgo de crédito, que faciliten la toma de decisiones para la generación de negocios. Con lo anterior mencionado se hace posible que las Instituciones de crédito puedan cobrar más rápido y a menor costo, ya que no sólo realizan llamadas, también manejan otros sistemas de comunicación como: correo electrónico, mensajes por celular, correo convencional (estrategias de medios masivos), etc. Es a través de una estrategia que considere una combinación de canales, cuando se logra la recuperación de las deudas en mora o prevención de mora.

Es importante mencionar que este trabajo contiene dos partes importantes, las cuales son, diseño de la estrategia, la primera será desarrollar el modelo de segmentación y definir los grupos de riesgo de crédito con respecto a la probabilidad de incumplimiento o la llamada “tasa de malos”. En la segunda parte, se definen las estrategias y tratamientos de cobranza, por último se realiza una propuesta de asignación de cuentas para reducir los costos en la operación. Los apartados que se incluyen son los siguientes:

En el capítulo I se detalla la problemática del caso en estudio, análisis de factibilidad, que incluye la justificación de por qué la necesidad de desarrollar un modelo de segmentación para tarjeta de crédito y los datos disponibles.

El capítulo II contiene los instrumentos de análisis, en los cuales se describe el concepto de riesgo y la descripción de las fuentes de información. Adicionalmente, se detalla la metodología utilizada para el desarrollo y segmentación del modelo, una breve introducción a los árboles de decisión y se define la probabilidad de incumplimiento.

¹Fuente: <http://www.cash-pot.com.ar/ejemplo-de-estrategia-de-cobranza.aspx>, Javier Loíacono, Gerente de Sistemas de la Agencia de Cobranzas *Cashcollector*

² <http://www.burodecredito.com.mx>



Dentro del capítulo III se detalla cómo se lleva a cabo el desarrollo del modelo de segmentación, el tamaño de la población que se utilizó para éste trabajo. Se describe brevemente la metodología para la selección de las variables independientes. Por otro lado, se presentan los resultados del modelo de segmentación, la definición de segmentos de riesgo, a partir del modelo y la aplicación de estos en el diseño de las estrategias de cobranza. Este capítulo incluye la validación del modelo en una muestra más reciente, es decir, con una muestra de créditos fuera de la muestra de desarrollo, se valida la eficiencia del modelo. Se sugiere una implementación de las estrategias y un procedimiento de monitoreo, que permitirá tomar decisiones a corto plazo con base al desempeño de las cuentas.

El capítulo IV, se plantea un problema de optimización, con el fin de obtener una alternativa de distribución de cuentas, que serán asignadas para realizar llamadas telefónicas de cobro, de acuerdo a la clasificación por riesgo planteada en el capítulo III y el costo asociado por horario laboral, con el fin de minimizar el costo total esta operación. Para la solución del problema se utiliza el “Método Simplex” y se hace uso de “*solvers*” disponibles en la red.



I. Problemática en la cobranza de créditos

El incremento sustancial en la crisis crediticia como parte de una recesión global, ha tenido un impacto sobre la operación en la cobranza y por lo tanto de las Instituciones de crédito, lo que implica una dificultad para hacer frente a este cambio dramático en la dinámica de la economía.

De acuerdo con lo publicado por la Revista Fortuna, el investigador Alejandro López Bolaños del Instituto de Investigaciones Económicas de la UNAM, advierte que: “El endeudamiento de las familias con la banca se ha incrementado por encima del nivel de su ingreso y, de continuar esta tendencia, podría repercutir a nivel macroeconómico y en la estabilidad financiera, e incluso desplomar los ingresos ante una variación abrupta en las tasas de interés”. El investigador estima que en 2012, las familias ya tenían comprometido el 35% de sus ingresos con pagos por los financiamientos adquiridos. Para López Bolaños, existen indicios de que la situación podría agravarse más, por el hecho de que la banca ha lanzado campañas para ofrecer tarjetas de crédito de una manera muy fácil, incluso con sólo pedir una identificación oficial en la calle, en los centros comerciales, en el metro, en los aeropuertos.

En la práctica existe una revisión constate y desarrollo de estrategias para ofertar créditos, es decir, las Instituciones prestatarias aplican herramientas analíticas que permiten otorgar créditos mitigando el riesgo. Además, realizan monitoreos para revisar que el desempeño de éstas sea correcto o en su caso reemplazar las herramientas. Para el caso de las estrategias de cobranza, este proceso no es aplicado, siendo que el éxito de las organizaciones depende de la capacidad que tienen las mismas para la recuperación de la cartera en un momento dado, aplicando la estrategia óptima para cada grupo de deudores. Por lo tanto, una labor fundamental que requieren realizar las Instituciones de crédito después del otorgamiento de créditos (líneas de crédito), es el seguimiento de las cuentas para su recuperación, con el fin de minimizar riesgos de posibles pérdidas. Ésto acompañado de una política de crédito y cobranza, con la cual se pretende hacer consciente a los clientes de las fechas en que debe realizar el pago correspondiente a su saldo pendiente, cuándo se considera un saldo vencido y por lo tanto cuándo un crédito se considera vencido.



Una problemática que algunas Instituciones de crédito tienen de forma recurrente es la falta de seguimiento de las cuentas, es decir, contactar de forma oportuna a los clientes que incurren en falta de pago hasta que se ponga al corriente. Sin embargo, es importante identificar a qué clientes se debe de contactar y en qué momento, con el fin de que se pongan al corriente. Para llevar a cabo esta tarea el primer paso es contar con bases de datos actualizadas que incluyan datos demográficos, comportamiento interno y externos de referencias crediticias, esto con el fin de tener los datos correctos que permitan el envío de correspondencia oportuna.

El segundo paso, que incluye el seguimiento de las cuentas para su recuperación y que posiblemente llega ser de gran impacto sobre costos y beneficios en el proceso de cobranza, y que resulta ser una problemática dentro de las Instituciones de crédito ya que en su mayoría no cuentan con ello, es una definición de estrategias de cobranza a partir de una adecuada segmentación de la cartera, es decir, clasificación de las cuentas por pago vencidos, tales como 0, 1, 2, 3 o 4 ó mas pagos vencidos; posteriormente determinar el perfil de cuentas haciendo uso de variables internas y externas de comportamiento crediticio, que permitan clasificar dentro de cada grupo de pago vencido por nivel de riesgo y con esto se pueda realizar el direccionamiento preciso de las acciones de cobranza con cada cliente o deudor, que permita fortalecer el proceso de atención al cliente, porque entre más claros sean los argumentos y exista un buen trato con el cliente, se va a lograr una cobranza de mayor calidad, favoreciendo al mismo tiempo la lealtad del cliente. Dentro de este paso es posible considerar una reducción en los costos, ya que no será necesario realizar estrategias de cobro sobre el total de la población.

El segmentar la cartera asignada a la cobranza, permite agrupar de acuerdo a las características comunes de los clientes y/o las cuentas. Por ejemplo: puntajes obtenidos de modelos de comportamiento, que incluyen porcentaje de utilización de las tarjetas de crédito (saldo pendiente entre el límite de crédito), porcentaje de pago respecto al total de la deuda, pago vencido al momento de la deuda, número de veces que ha caído en mora y como parte sustancial e importante de mencionar, ya que esta forma parte de los modelos de segmentación aquí desarrollados, es el comportamiento crediticio con el resto de la banca, el cual proviene de fuentes externas como Buró de Crédito. Esta tarea puede ser llevada a cabo identificando estrategias óptimas de cobranza a través del uso de herramientas analíticas.

Para el caso particular que se desarrollará a continuación, se definen perfiles de riesgo de cuentas de Tarjeta de Crédito, que permitan establecer tratamientos diferenciados en la cobranza de una Institución bancaria, utilizando árboles de decisión y probabilidades de incumplimiento (donde probabilidad de incumplimiento baja se denominará cuenta buena y caso contrario cuenta mala) con el fin de optimizar la toma de decisiones en el saneamiento del portafolio de Tarjeta de crédito, enfocándose en grupos de clientes con mayor probabilidad de incumplimiento.



1.1 Análisis de factibilidad

Previo al desarrollo del modelo de segmentación, se realizó un análisis para evaluar la factibilidad de desarrollar un modelo de segmentación para las estrategias de cobranza para el portafolio de Tarjeta de Crédito en 0, 1, 2, 3 y 4 o más pagos vencidos.

Por otro lado, se pueden observar cambios en el perfil de riesgo del producto de Tarjeta de Crédito, principalmente originados por el entorno económico. Así como, importantes cambios en políticas de crédito, subproductos y canales de originación. Esta es una práctica, comúnmente realizada por las Instituciones de crédito, como se mencionó anteriormente, con el fin de mantener el nivel competitivo y niveles de pérdida esperados. Dentro de estas prácticas también se encuentran: cierre y apertura de productos, políticas de ofertas de créditos, productos de recuperación, entre otros. Estas acciones se llevan a cabo como medida de remediación en caso de crisis económicas, para reducir cambios negativos en los portafolios o que se incrementen los niveles de cartera vencida y por ende las pérdidas de las organizaciones.

Con el fin de ilustrar el entorno bancario de las Tarjetas de Crédito en países como México y su relevancia para la recuperación debido a las líneas de crédito otorgadas (exposición al riesgo), a continuación se anexa un cuadro, con el número de Tarjetas de Crédito por entidad financiera y tipo de pago, es decir, totalero o no totalero, donde totalero son aquellos clientes que realizan el pago total de su deuda, es decir, no pagan intereses, y no totalero son aquellos clientes que realizan el pago parcial o mínimo del saldo deudor, es decir, pagan intereses, estas cifras fueron obtenidas del Banco de México a abril de 2012³.

Cuadro 1.1. Información básica para clientes totaleros

	Número de Tarjetas		Saldo de crédito otorgado (millones de pesos)		Tasa efectiva promedio ponderada (%)		Tasa efectiva mediana (%)	
	abr-11	abr-12	abr-11	abr-12	abr-11	abr-12	abr-11	abr-12
Sistema	13,469,978	14,871,672	172,242	200,740	25.1	24.1	24.6	22.9
BBVA Bancomer	4,760,286	5,041,750	54,713	63,919	24.4	23.8	21.3	20.3
Banamex	3,434,301	3,960,626	53,077	62,639	24.9	23.8	30.2	25.8
Santander	1,690,470	1,956,974	25,416	31,893	23	22	23.6	22.9
Banorte-Ixe Tarjetas	n.a.	876,869	n.a.	12,039	n.a.	24.3	n.a.	23.5
BanCoppel	688,611	823,141	2,116	3,012	52.5	52.8	65	65
HSBC	798,286	805,920	13,136	13,141	27.3	24.5	31.3	25.8
American Express	362,721	342,229	6,250	6,458	24.9	23.1	26.1	24.9
Scotiabank	288,097	311,317	3,284	3,290	27.6	27.5	37.1	34.8
Inbursa	297,001	274,844	1,762	1,654	25.2	24.1	24	24
Banco Walmart	110,081	271,116	266	1,184	18.2	18.8	0	0
SF Soriana	32,198	44,676	289	459	28.3	25.3	39.4	30
Banco Invex	41,443	37,308	324	315	45.8	48.8	59.5	61.9
CrediScotia	38,627	35,596	219	187	48.1	47.7	44.7	49.8
BNP Paribas	54,095	34,869	204	131	38.7	54.3	51.6	62.8
Banregio	15,895	19,082	170	195	25.8	24.4	24	24
Banco Fácil	25,866	16,527	64	37	55.4	54	58	58
Banco del Bajío	9,092	10,648	87	101	17.2	17.6	32.5	32.9
Banco Afirme	7,101	8,180	74	83	31.8	35.1	39.6	42.3
Banorte	689,833	n.a.	8,739	n.a.	25.7	n.a.	28.7	n.a.
Ixe tarjetas	125,974	n.a.	2,050	n.a.	20.5	n.a.	17.4	n.a.

³ Fuentes: Banxico, elaborado con datos proporcionados por las instituciones de crédito, cifras sujetas a revisión



Cuadro 1.2. Información básica para clientes no totaleros

	Número de Tarjetas		Saldo de crédito otorgado (millones de pesos)		Tasa efectiva promedio ponderada (%)		Tasa efectiva mediana (%)	
	abr-11	abr-12	abr-11	abr-12	abr-11	abr-12	abr-11	abr-12
Sistema	8,148,057	8,687,363	142,477	162,422	30.4	29.8	30.3	28.3
BBVA Bancomer	3,156,712	3,307,896	51,350	59,881	26	25.4	22.4	21.6
Banamex	1,841,890	2,041,307	38,783	44,538	34.1	33.4	39.5	39.6
Santander	980,433	1,056,502	21,271	25,754	27.5	27.2	28	26.3
Banorte-Ixe Tarjetas	n.a.	505,364	n.a.	9,311	n.a.	31.5	n.a.	34.5
BanCoppel	470,868	560,845	1,710	2,446	65	65	65	65
HSBC	467,679	452,465	10,853	10,307	33	31.2	35.8	32.5
American Express	215,313	192,012	4,712	4,620	33	32.3	35.8	35.8
Scotiabank	143,182	143,272	2,389	2,390	37.9	37.9	37.8	39.8
Inbursa	199,299	170,343	1,555	1,410	28.6	28.3	24	24
Banco Walmart	41,949	127,794	121	570	40	39	38.7	38.7
SF Soriana	17,291	23,839	194	290	42.3	40	48.9	47
Banco Invex	30,037	23,935	306	287	48.5	53.7	60.7	65.4
CrediScotia	26,931	23,940	218	179	48.4	50	44.7	49.8
BNP Paribas	39,093	29,555	160	124	49.3	57.2	62.4	62.8
Banregio	7,784	8,820	133	148	32.9	32.1	37	24
Banco Fácil	18,713	9,830	64	37	55.4	54	58	58
Banco del Bajío	4,299	4,993	46	55	32.5	32.7	32.9	32.9
Banco Afirme	3,477	4,651	70	75	33.4	38.4	39.6	42.3
Banorte	404,115	n.a.	6,858	n.a.	32.8	n.a.	36.6	n.a.
Ixe tarjetas	78,992	n.a.	1,683	n.a.	25	n.a.	20	n.a.

De acuerdo con las cifras de Banco de México, se observa que el número de tarjetas de crédito en abril de 2012, otorgadas en México por el sistema bancario es de 23 millones, donde 37% paga de forma parcial su deuda y 63% paga su deuda total. Por lo tanto es importante que las estrategias de cobranza deban dirigirse a este 37% de la población y de forma preventiva a un porcentaje menor de los totaleros. Cabe señalar que el saldo promedio de un cliente totalero es 74 mil pesos, mientras que para los no totaleros son 53 mil pesos. Estos últimos podrían presentar una mayor exposición al incumplimiento de pago, ya que al no pagar el total de su deuda, se realiza un cargo de intereses cuyo incremento es sustancial en su deuda, por la tasa de interés que hoy en día determinan los bancos (29.8% promedio).

1.1.1 Descripción del producto bancario

Una tarjeta de crédito es una tarjeta de plástico numerada, que presenta una banda magnética o microchip, y que permite realizar compras que se pagan a futuro. Las tarjetas de crédito, también conocidas como “dinero plástico”, reciben este nombre ya que, cuando la persona realiza una compra, la entidad que administra las tarjetas le está concediendo un crédito al usuario, el cual está asociado a una línea de crédito revolviente, otorgado a personas físicas y personas físicas con actividad empresarial. Revolviente indica que el cliente puede no pagar el total de su deuda, si no un pago mínimo según lo determine la entidad financiera, y al saldo restante se le calcula un interés de acuerdo a la tasa de interés acordada. Este tipo de crédito no tiene un destino en específico, y se subdivide en diferentes segmentos dependiendo de las características de lo solicitado por el cliente, es decir, “Visa”, “Oro”, “Platimun”, “VIP”, etc. Este tipo de tarjetas son



otorgadas con base en los ingresos o inversiones del cliente y por lo tanto podrán ser un factor a considerar para determinar una estrategia de cobranza.

1.1.2 Razones para desarrollar un Modelo de segmentación

Este tipo de modelos de segmentación se desarrollan esencialmente por las siguientes causas:

- Para mejorar la estrategias de cobranza, situando los recursos de manera adecuada al perfil de riesgo de cada segmento del portafolio y adecuarlos a los sistemas que manejan las Instituciones de crédito.
- Mejorar la recuperación con relación a los costos de cada tipo de tratamiento.
- Otorgar un mejor servicio a los clientes de Tarjetas de Crédito, contactándolos en el momento correcto de acuerdo a su morosidad y probabilidad de incumplimiento. En particular, una de las estrategias a desarrollar está enfocada a clientes en cero pagos vencidos con el fin de evitar que migren a los siguientes niveles de morosidad, sus características principales es ser clientes, cuya fecha de exigibilidad de pago ha sido rebasada y no han realizado su pago mínimo y su perfil es más riesgoso.

1.1.3 Justificación

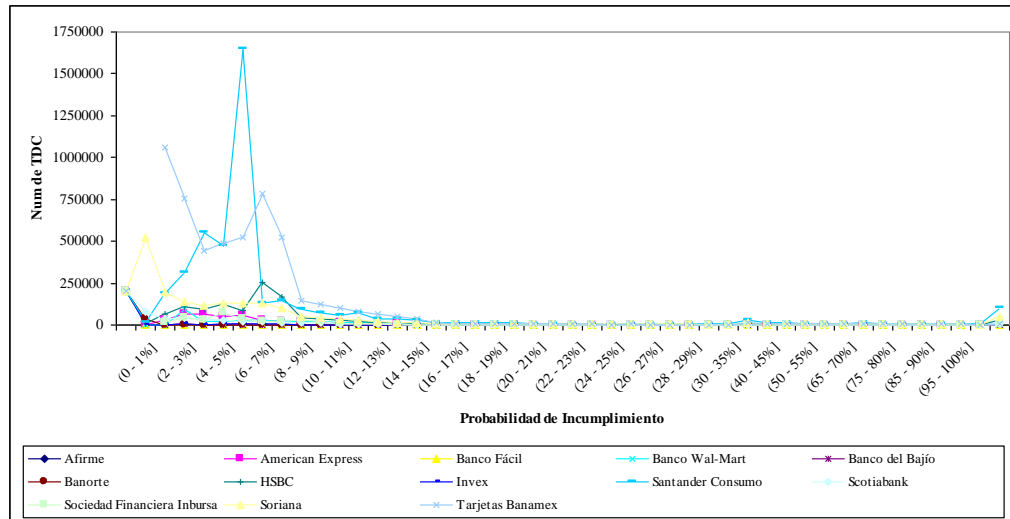
La justificación para desarrollar un modelo de segmentación se basa en lo siguiente:

- Diversos cambios en las políticas de ofertas de créditos que define una Institución de crédito, lo que implica un portafolio vigente con distinto perfil de riesgo, por lo que es necesario utilizar información más reciente y actualizar la clasificación del riesgo y por lo tanto las acciones de cobro. Para medir estos cambios en el perfil de los clientes, y justificar una actualización de la información y por lo tanto en la segmentación y estrategias, debe calcularse el índice de estabilidad de la población e indicadores que midan la eficiencia del modelo de segmentación actual. Estas métricas serán definidas en el capítulo 2.
- El entorno macroeconómico, por ejemplo, crisis económicas, asociado a mayores niveles de desempleo y baja de liquidez en el mercado, afectando de manera importante el comportamiento de pagos de los clientes. A continuación, se presentan unas gráficas del panorama general de la Cartera de Tarjeta de Crédito que sirven para ejemplificar eventos macroeconómicos como la crisis del 2008 y su impacto en las Instituciones de crédito. Las gráficas se presentan por entidad financiera y probabilidad de incumplimiento, su definición es que el cliente alcance 3 o más pagos vencidos, por número de tarjetas y por saldo deudor. Esta

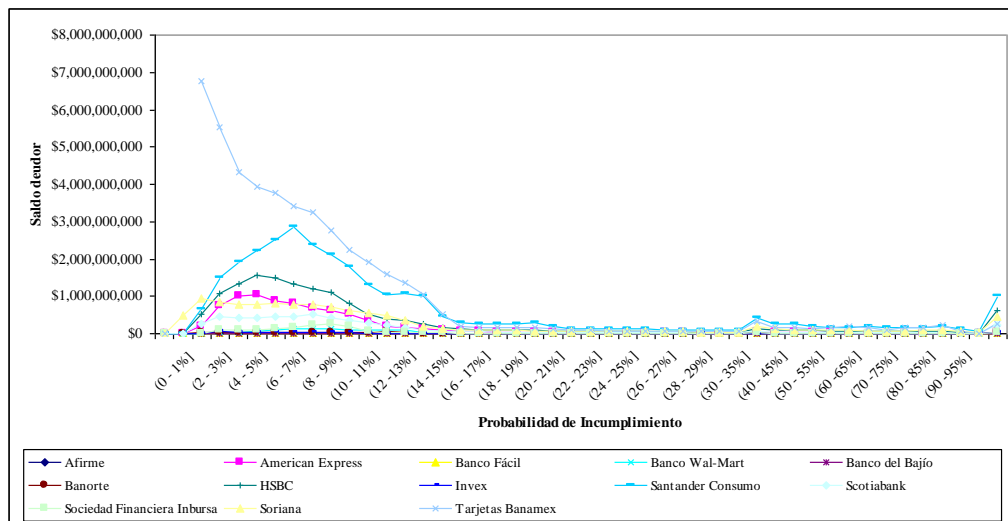


información es publicada por la Comisión Nacional Bancaria y de Valores (CNBV) a abril de 2012⁴.

Gráfica 1.1. Número de Tarjetas de Crédito vs. Probabilidad de Incumplimiento



Gráfica 1.2. Saldo deudor vs. Probabilidad de Incumplimiento



⁴ CNBV, abril 2012, <http://www.cnbv.gob.mx/Bancos/Paginas/Informacion-Estad%C3%ADstica.aspx>.



Como se puede observar en la gráfica 1.1, el porcentaje de incumplimiento por nivel de entidad financiera, que presenta mayor volumen en número de tarjetas y saldo deudor, oscila entre 5 y 9 por ciento y de acuerdo a la gráfica 1.2, esto equivale a una deuda que oscila entre 1,000 y 3,000 mil millones de pesos.

1.1.4 Mejoras al modelo de segmentación y su aplicación

El nuevo modelo de segmentación incluirá las siguientes mejoras:

- Para el caso de cartera en 0 pagos vencidos se utilizará información de diciembre de 2009 y enero 2010, mientras que para cartera vencida se utilizará población en la ventana de observación que corresponde a mayo 2010 y junio 2010. La ventana de desempeño utilizada es de 6 meses y se describe en los siguientes apartados.
- Los modelos de segmentación se utilizarán como una herramienta para la toma de decisiones sobre las estrategias de cobranza, así como para el soporte en la definición de políticas de productos de recuperación, tales como reestructuras de crédito, descuentos para poner al corriente o descuentos (quitas) para liquidar el total de la deuda.

1.2 Propósito del desarrollo

Cabe señalar que hoy en día existen tratamientos de cobranza, tales como llamadas telefónicas, mensajes o vistas físicas. Sin embargo, el propósito es clasificar por riesgo, de acuerdo a su probabilidad de incumplimiento, cuentas de Tarjeta de Crédito, que permitan diseñar estrategias para aplicar estos tratamientos ya existentes, es decir, de acuerdo a su nivel de morosidad y segmento de riesgo, priorizar la toma de decisiones en el saneamiento del portafolio, por ejemplo, aplicar solo llamadas, mensajes o ninguna acción de cobro, a ciertos segmentos de riesgo, ya que si estos clientes no presentan un perfil de riesgo alto, es posible solo llamarles, enviarles un mensaje o dejarles de llamar y enfocarse en aquellos que si lo requieran.

El propósito de diseñar estrategias en la cobranza, nos permite determinar si las acciones de cobro son las adecuadas, es decir, si llamarle a todos los clientes es lo mejor para mantener un nivel de morosidad y una recuperación monetaria adecuada o llamarle a los clientes con perfil mas riesgoso, nos ayuda a mantener el mismo nivel de morosidad y misma recuperación, pero una reducción en costos en la operación.



1.3 Portafolios impactados

Este caso práctico se desarrolló para una estrategia de cobranza que aplica para productos de crédito, tales como líneas revolventes como Tarjetas de Crédito, cuyos pagos vencidos son 1, 2, 3 y 4 o más, o cobranza preventiva.

1.4 Metodología para el desarrollo

Se utilizará información histórica del comportamiento de créditos, para determinar el comportamiento de clientes actuales, es decir, utilizando el pasado se hará un pronóstico del futuro comportamiento de tarjetahabientes, es decir, se clasificarán las cuentas por nivel de riesgo y posteriormente se determinarán estrategias de cobro, cuya intensidad dependerá de la clasificación de riesgo.

La información histórica, contiene variables como utilización de las tarjetas internas y con otras Instituciones de crédito, morosidad (cuántas veces ha dejado de realizar pagos) en diferentes ventanas de tiempo, por ejemplo, en los últimos 6 ó 12 meses. Por otro lado, se determinará la probabilidad de incumpliendo, la cual considera los últimos 6 o 12 meses de los pagos vencidos que ha presentado el cliente, por ejemplo, un cliente que ha alcanzado 3 o mas pagos vencidos en los últimos 6 meses, se establecerá como clientes “malo”, de lo contrario un cliente “bueno”. La probabilidad de incumplimiento o variables dependiente, se definirá como el total de malos entre el total de la población (cuentas buenas mas cuentas malas).

Para segmentar la población, se utilizarán árboles de decisión, se obtendrá un árbol de decisión por pago vencido, que a través de las variables que anteriormente se mencionaron (variables independientes), permiten explicar las características de los clientes con diferentes probabilidades de incumplimiento (variable dependiente).

Una vez segmentada la población por el árbol de decisión, los nodos finales de árbol se agruparan por su probabilidad de incumplimiento y se clasificarán por nivel de riesgo, donde riesgo crítico estará definido por la mayor probabilidad de incumplimiento y así de forma descendiente hasta riesgo bajo, cuya probabilidad es menor.

La clasificación de las cuentas será utilizada para asignar las acciones de cobro, es decir, se determinaran estrategias para cada pago vencido (0, 1, 2, 3 y 4 o más pagos vencidos), llamadas campeonas y retadoras, donde la estrategia campeona pudiera asignar un tratamiento a cada grupo de riesgo, que pudiera ser llamada, mensajes, visita física o ninguna acción y la retadora lo opuesto, por ejemplo, si al segmento de riesgo bajo se determina realizar una acción de llamada como estrategia campeona y como estrategia retadora no llamarle.



Con base a las estrategias diseñadas, se propondrá una distribución de cuentas por horario laboral del centro de telefónico de cobranza, en donde se considerará el costo unitario asociado a cada cuenta por su clasificación de riesgo y el total de cuentas que deben de gestionarse al mes, con el fin de reducir el costo total que se deriva de esta operación.

La modelación con datos históricos reales para determinar comportamientos actuales, es metodología utilizada en varios campos y particularmente la industria financiera y sus fundamentos son estadísticos.



II. Instrumentos de análisis

Una de las principales causas de los problemas que enfrentan las Instituciones de crédito está directamente relacionada con los estándares que determinan el otorgamiento de créditos, es decir, las políticas de crédito que incluyen los requisitos mínimos (edad de cliente, ingresos, buró de crédito, etc.) para ser acreedor de una Tarjeta de Crédito, entre otros productos bancarios. Estas pueden ser laxas y tienen como consecuencia una probabilidad baja de recuperación. Por otro lado, la falta de atención a los cambios económicos puede conducir a un deterioro de la situación del crédito, por ejemplo, pérdida del empleo del acreditado debido al cierre de empresas, entre otros. Este entorno está determinado por el riesgo del crédito y se define como la probabilidad que un acreditado no pueda cumplir sus obligaciones de acuerdo con los términos acordados.

El objetivo de la gestión del riesgo, es maximizar las tasa de rendimiento que pudiera tener un banco sujeto a mantener su exposición al riesgo de crédito dentro de los parámetros aceptables, es decir, toda Institución de crédito esta expuesta a una determinada pérdida, sin embargo ésta debe de fijar parámetros de riesgo que hagan posible hacer frente a esta pérdida crediticia. Como parte de la gestión, está determinar estrategias de Cobranza que permitan mantener una cartera con baja morosidad considerando los niveles de riesgo crediticio.

2.1 Análisis de Riesgo

El análisis de riesgo permite definir los niveles de riesgo asociados, éste análisis es una técnica que se basa en la probabilidad de incumplimiento, la cual tiene relacionada una serie de eventos secuenciales que podrían dar como resultado un riesgo potencial, que se traduce en una pérdida monetaria. Con esta técnica es posible identificar los eventos indeseables y analizarlos de forma individual, como es en el caso de las estrategias de cobranza, cuyo fin es determinar estos eventos indeseables en el comportamiento de los clientes y aplicar un tratamiento de cobranza con mayor efectividad.



En términos generales, el análisis de riesgo proporciona una base para determinar los grupos de riesgo que deben de tener un seguimiento exhaustivo y revisar si los tratamientos son adecuados. Dentro de éste análisis se incluye el uso de una serie de herramientas estadísticas cuyo nivel de complejidad varía de acuerdo a las necesidades de cada entidad financiera y recursos disponibles. Estas herramientas están disponibles en “softwares” especializados, cuyo costo es alto, cuya inversión puede ser recuperada a largo plazo, ya que permite diseñar estrategias para mitigar el riesgo y evitar una inversión mayor para enfrentar las pérdidas monetarias. Un ejemplo de estas herramientas, son programas de minería de datos, que proporcionan módulos para explotación de datos, desarrollo de “Scorecards”, árboles de decisión, análisis multivariados, entre otros; los cuales son utilizados en la industria para la gestión del riesgo de crédito. Como anteriormente se mencionó para este problema práctico, se hizo uso de árboles de decisión, cuya técnica se describirá en el siguiente apartado.

2.1.1 Árboles de decisión

De acuerdo a la literatura la construcción de árboles de decisión forma parte de una serie de métodos de aprendizaje inductivo (Michalski, Car-Carbonell y Mitchell, 1983, 1986; Bratko y Lavrac, 1987). Este método fue desarrollado inicialmente por Hunt, Marin y Stone (1966) y modificado posteriormente por Quinlan (1979, 1983). Sus aplicaciones abarcan un gran número de industrias, entre ellas la industria crediticia, para la cual se aplicó.

El análisis de decisiones con cierto nivel de complejidad que conlleva incertidumbre, tiene como consecuencia que el resultado de la selección de cualquier decisión no puede predecirse con certeza. Más aún, si no se consideran una serie de factores que pueden apoyar en reducir la incertidumbre en la decisión, por ejemplo, el determinar acciones diferenciadas en cobranza con base únicamente a la probabilidad de tener pago o no pago de un grupo de clientes, tiene como implicación el no poder determinar el perfil de cliente (cuántas veces ha incumplido en los últimos 3, 6, 12 meses etc.) y predecir con mayor certeza el futuro comportamiento del cliente, es decir, la decisión de cobrar o no cobrar un crédito debería estar determinada por una combinación de características de comportamiento crediticio histórico, y no solamente su probabilidad de pago o no pago, todo lo anterior con el fin de poder tomar una decisión con mayor certeza.

Para llevar a cabo este tipo de decisiones con mayor certidumbre, son utilizados los árboles de decisión, los cuales son diagramas de decisión, que dependiendo de la orientación se leen de izquierda a derecha (figura 1) o de arriba hacia abajo (figura 2), como se ilustra a continuación:



Figura 2.1. Árbol horizontal

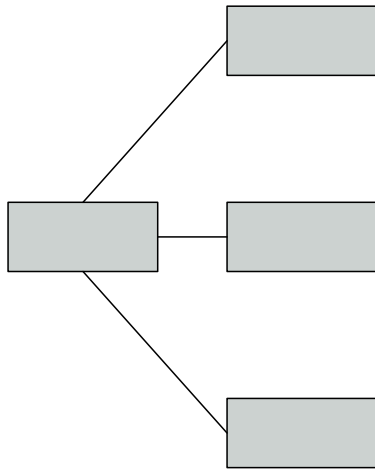
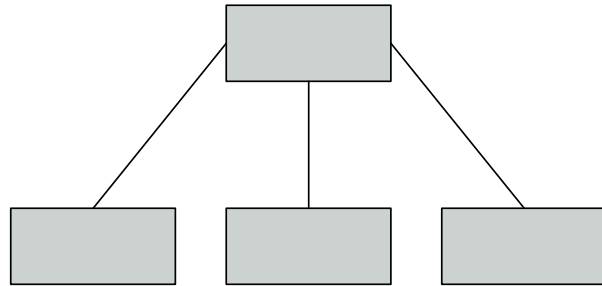


Figura 2.2. Árbol vertical

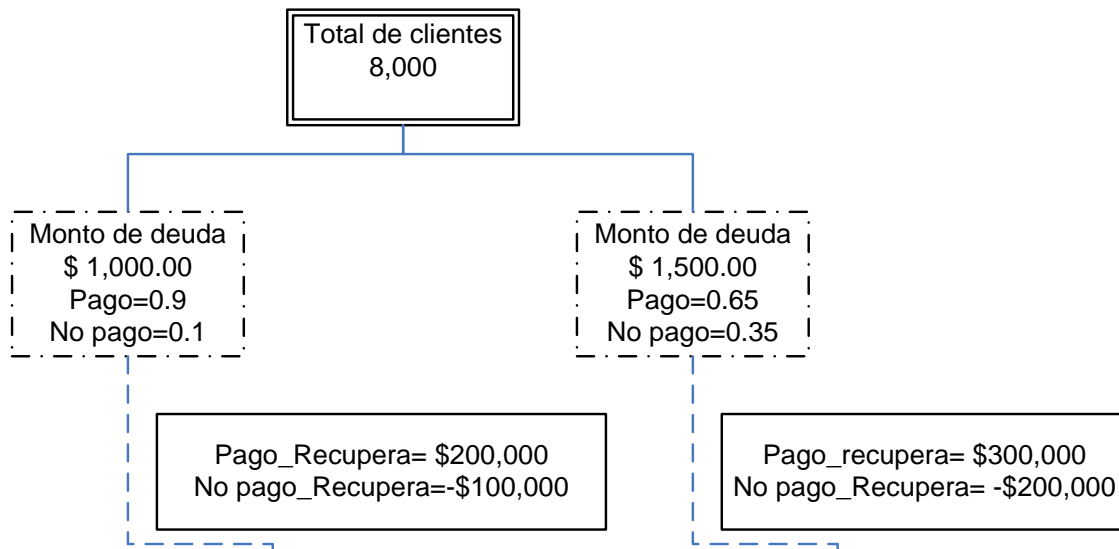


Cada uno de los rectángulos determina un nodo, en caso de la Figura 2.1.1.1. Árbol horizontal, el nodo más a la izquierda se determina como nodo raíz, y para el caso de la Figura 2.1.1.2. Árbol vertical, el primer nodo es el nodo raíz. Estos a su vez derivan ramas, las cuales determinan un conjunto de alternativas de decisión posibles, estos nodos son llamados nodos “chance”. Cada uno de estos nodos tiene una probabilidad de ocurrencia, que para el caso de estudio, se les asigna la probabilidad de pago o no pago (cuenta buena o mala, cuya definición se detalla en las siguientes secciones). Los rectángulos finales, son llamados nodos finales y representa el resultado final de seguir un camino desde el nodo raíz del árbol de decisión hasta ese punto final.

Posterior a la modelación del árbol, es necesario determinar un criterio para seleccionar una alternativa de decisión, es decir, definir una regla para determinar una decisión. Esta regla será determinada por el problema que se busca resolver. Por ejemplo, un criterio puede ser el valor esperado, el cual es usado para tomar una decisión que toma en cuenta tanto los posibles resultados para cada alternativa de decisión y la probabilidad de que cada resultado se producirá. El valor esperado se calcula multiplicando cada resultado posible de la alternativa por su probabilidad, y sumando los resultados. El valor esperado permite seleccionar la alternativa, de acuerdo al problema que se espera resolver. Por ejemplo, en situaciones de participación de las ganancias se elige que a "mayor valor esperado es mejor", y en situaciones que afectan a los costos, en la que "menor valor esperado es mejor". Para ilustrar lo anterior se presenta el siguiente ejemplo:



Figura 2.3. Ejemplo



De acuerdo al ejemplo mostrado en la figura 2.1.1.3, el problema consiste en decidir que segmentación por monto de la deuda nos lleva a una mayor recuperación de acuerdo a la probabilidad que pago o no pago. El valor esperado se calcula para cada decisión a continuación:

1) $EV = (200000 * 0.9) + (-100000) * 0.1 = \$170,000$

2) $EV = (300000 * 0.65) + (-200000) * 0.35 = \$125,000$

De acuerdo a los resultado del valor esperado para cada decisión del árbol, la mejor opción en términos de mayor recuperación lo tiene la opción uno, ya que se obtendría mayor recuperación sobre aquellos clientes cuyo monto de deuda es \$1,000.00.

Los árboles de decisión permiten tomar decisiones más complejas que el ejemplo anterior, puesto que pueden intervenir más variables de incertidumbre y factores que determinan acciones diferenciadas y que llevan a diseñar estrategias de decisión como es el caso de este trabajo.



2.1.2 Clasificación de las cuentas

Como se mencionó en la sección anterior, los árboles de decisión que se desarrollarán en este trabajo de tesis, parten de una población en donde cada registro tiene asociada una bandera (variable dependiente) que determina el pago o no pago de esa cuenta, es decir, permite clasificar cuentas entre buenas o malas, a las que en lo sucesivo nos referiremos con ese nombre.

A continuación se presenta la clasificación de las cuentas que se proponen para este trabajo:

- Se determina un punto de observación o mes de observación, para éste caso se seleccionaron los meses de diciembre 2009 y enero de 2010 en caso de las cuentas en 0 pagos vencidos, y para el caso de cuentas en mora (1 o más pagos vencidos) se toman mayo y junio de 2010.
- Se determina un periodo de desempeño, el cual llamaremos ventana de desempeño (siguientes 6 ó 12 meses posteriores al mes de observación), es decir, se observa el comportamiento de pagos de las cuentas contenidas en el mes de observación durante los siguientes meses. Para este trabajo se considera una ventana de 6 meses.
- Realizar un análisis de migración de las cuentas y determinar cuándo se trata de una cuenta buena o mala. Por ejemplo, cuando la cuenta se mantiene al corriente (0 pagos vencidos) o el máximo pago vencido fue 2 en la ventana de desempeño, podría considerarse como una cuenta buena, de lo contrario una cuenta mala.

La siguiente figura ilustra los pasos anteriores:

Figura 2.4. Definición de cuentas

Mes de observación							
Diciembre 2009	Pago_ven_ene10	Pago_ven_feb10	Pago_ven_mar10	Pago_ven_abr10	Pago_ven_may10	Pago_ven_jun10	Bandera
Tarjeta 1	0	0	0	0	0	0	Bueno
Tarjeta 2	0	1	1	0	1	0	Bueno
Tarjeta 3	0	1	2	3	2	1	Malo
Tarjeta 4	1	2	3	2	1	0	Malo
Tarjeta 5	1	1	1	1	1	1	Bueno
Tarjeta 6	1	1	1	2	1	1	Bueno
.
.
.
Tarjeta n	1	2	3	4	5	4	Malo

Donde el mes de observación es diciembre 2009, es decir, se tomarán todas las tarjetas de este mes (Tarjeta₁, Tarjeta₂,..., Tarjeta_n) y se observó el pago vencido de los siguientes 6 meses, los cuales están determinados por la variable Pago_ven_ene10, Pago_ven_feb10, etc., los cuales son los meses de la ventana de desempeño. Al final del



periodo de desempeño, se determina si la cuenta es buena o mala y está definido en el campo de Bandera.

2.1.3 Indicadores estadísticos para medir la eficiencia del modelo

La eficiencia de los modelos de segmentación (árboles de decisión), se mide a través del cálculo de indicadores estadísticos, los cuales determinan qué tan robusto es el árbol en términos de separar cuentas buenas de las cuentas malas, a través de las variables independientes o cambios en el comportamiento de la población, es decir, si el árbol se aplica a una población mas reciente, esta se distribuya a los largo de los nodos, igual o similar que la población con la que se desarrolló. Los índices para medir todas estas características de eficiencia de los árboles son:

- a) Kolmogorov-Smirnov (K-S)
- b) Coeficiente “Gini”
- c) Valor de la información (VI)
- d) Índice de estabilidad de la población (IEP).

2.1.3.1 Prueba Kolmogorov –Smirnov (K-S)

La estadística K-S consiste en calcular la separación máxima entre las distribuciones acumuladas de dos grupos diferentes de poblaciones, cuya clasificación puede ser pago y no pago (cuenta buena o mala), es decir, es una medida porcentual en valor absoluto. Esto implica que esta medida puede tomar valores entre 0 y 100 %. Generalmente valores grandes de K-S indican una buena separación. Sin embargo, dado que la K-S es una medida puntual, considera sólo un punto de la distribución, es necesario que esta estadística sea complementada con otras. El cálculo es de la siguiente forma, más adelante se muestra un cuadro y gráfica para ilustrar con mayor detalle (cuadro 2.1.3.1 y gráfica 2.1.3.1):

$$K - S = \max |X_i - Y_i|, i = 1, \dots, n$$

Donde X_i es la distribución acumulada sobre el renglón i de las cuentas buenas y Y_i es la distribución acumulada sobre el renglón i de las cuentas malas.

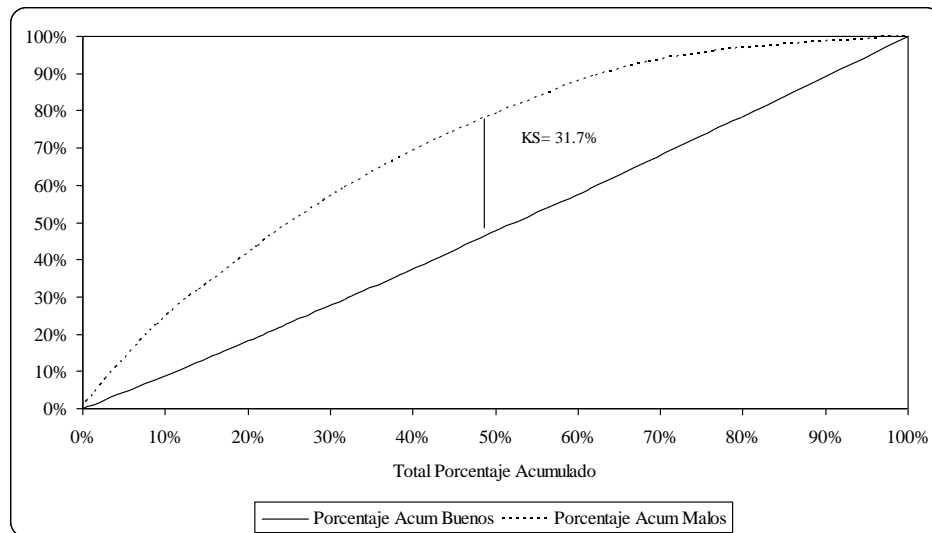
Procedimiento para la prueba K-S.

Como primer paso es necesario ordenar los nodos finales del árbol de decisión de acuerdo a la probabilidad de incumplimiento de mayor a menor o lo que también se llama clientes con pago o no pago o clientes buenos o malos. Dentro del siguiente cuadro 2.1.3.2 se define como “Prob Malos”.



Posteriormente es necesario calcular el porcentaje acumulado del grupo de cuentas Malos y Buenas, por último se calcula la diferencia en valor absoluto de X_i e Y_i y se obtiene la máxima diferencia. Este cálculo se ilustra con la siguiente gráfica 2.1, en donde se gráfica la distribución acumulada de cuentas buenas y malas:

Gráfica 2.1. Distribución acumulada de cuentas (K-S)



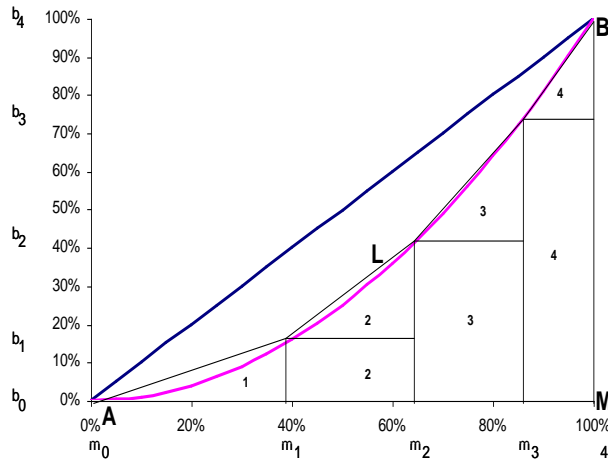
El valor máximo entre los dos puntos se alcanza en 31.7%, es decir $KS=31.7\%$.

2.1.3.2 Coeficiente “Gini”

El coeficiente de “Gini” es una medida de inequidad de una distribución, y se define como el porcentaje de área que existe entre la curva de una distribución acumulada y la curva de la distribución uniforme. El coeficiente de “Gini” es un número entre 0 y 1, donde 0 significa que existe una perfecta equidad, es decir, el modelo no diferencia entre clientes que pagan y no pagan, y 1 significa que hay una perfecta separación de entre clientes que pagan y no pagan.



Gráfica 2.2. Distribución acumulada de cuentas (K-S)



El coeficiente de Gini se calcula de la siguiente manera:

$$IG = \text{área ALB} / \text{área AMB}$$

Donde $AMB = \frac{1}{2}$

Entonces

$$IG = 2 \text{ área ALB}$$

$$IG = 2 [1/2 - \text{área ALBM}]$$

$$IG = 1 - 2 \text{ área ALBM}$$

El área bajo la curva ALBM

$$\frac{1}{2} \sum_{i=1}^J (P_{mi} - P_{mi-1}) * (P_{bi} + P_{bi-1}) \dots\dots\dots (1)$$

Donde

i = Renglón del nodo ordenado de acuerdo a la probabilidad de incumplimiento (mayor a menor)

P_{mi} = Porcentaje de cuentas malas acumuladas en el renglón del nodo

P_{bi} = Porcentaje de cuentas buenas acumuladas en el renglón del nodo

Finalmente el índice de “Gini” se calcula como:

$$IG = 1 - \sum_{i=1}^J (P_{mi} - P_{mi-1}) * (P_{bi} + P_{bi-1}) \dots\dots\dots (2)$$



2.1.3.3 Valor de la información (VI)

El valor de la información se define como la suma de las diferencias entre las distribuciones buenas y malas, multiplicadas por el logaritmo natural del cociente entre cuentas buenas y malas. Entre mayor sea el VI, es más eficiente el modelo de segmentación, de acuerdo con las mejores prácticas en la industria, valores mayores a 1 indican que el modelo es robusto en términos de diferenciar clientes con pago y no pago o buenos y malos.

$$VI = \sum_{i=1}^n (P_{bi} - P_{mi}) * \ln\left(\frac{P_{bi}}{P_{mi}}\right) \dots\dots\dots (3)$$

Donde

- i = Renglón del nodo ordenado de acuerdo a la probabilidad de incumplimiento
- P_{mi} = Porcentaje de cuentas malas en el renglón del nodo
- P_{bi} = Porcentaje de cuentas buenas en el renglón del nodo

2.1.3.4 Índice de estabilidad de la población (IE)

El índice de estabilidad de población es utilizado para medir los cambios en las condiciones normales de operación del mercado y de la Institución de crédito. El propósito de este índice es detectar cualquier cambio significativo entre la distribución de la población por cada nodo y la distribución de población estimada en cada nodo, que se utilizó en la construcción del modelo. Al igual que en los índices anteriores, los nodos deberán ordenarse de acuerdo a la probabilidad de incumplimiento de mayor a menor. Este índice se calcula de la siguiente manera:

$$IE = \sum_{i=1}^n (P_{Mi} - P_{Ai}) * \ln\left(\frac{P_{Mi}}{P_{Ai}}\right) \dots\dots\dots (4)$$

Donde

- P_{Mi} = porcentaje de cuentas en el renglón Mi de la población del modelo
- P_{Ai} = porcentaje de cuentas en el renglón Ai de la población actual



Ya que se obtiene el índice de estabilidad, se dará la interpretación del valor de acuerdo a la siguiente tabla:

Cuadro 2.1. Interpretación del índice de estabilidad

1. Si $IE = 0$	entonces las distribuciones son iguales
2. Si $0 < IE \leq 0.10$	entonces son similares o la diferencia es muy pequeña
3. Si $0.10 < IE \leq 0.25$	entonces existen algunos pequeños cambios entre dichas distribuciones por lo tanto es necesario realizar un análisis de características
4. Si $IE > 0.25$	entonces existe un cambio significativo entre las distribuciones

De acuerdo al cuadro anterior, si el valor del IE, se encuentra en el punto 3, es necesario realizar un análisis de características, que consiste en analizar las variables que tiene el árbol y verificar si ha cambiado su composición en términos del cantidad de cuentas que contiene.

En caso de que el valor de IE, se encuentre en el punto 4, es necesario revisar las estadísticas anteriores, ya que si estas también presentan un deterioro, se deberá determinar si se debe de hacer un nuevo desarrollo con una muestra diferente a la seleccionada.

2.2 Descripción de las fuentes de información

La información utilizada usualmente proviene de sistemas internos de las Instituciones de crédito, que tienen como función la administración de cuentas revolventes, particularmente de tarjetas de crédito. Dentro de estos sistemas, normalmente se encuentran diferentes módulos que permiten implementar estrategias para aumentar y disminuir líneas de crédito o estrategias de cobranza, objetivo de este trabajo.

Parte de la información que puede ser extraída de estos módulos, contiene las transacciones de las tarjetas de crédito, tales como saldos, fechas de corte, fecha de exigibilidad de pago, límite de crédito, pago vencido, puntaje proveniente de los modelos de comportamiento, por mencionar algunas. Por otro lado, también contiene una serie de campos con información de buró de crédito.



La población para el desarrollo del modelo de segmentación de este trabajo, incluye datos relacionados con el desempeño del producto de Tarjeta de Crédito para 0, 1, 2, 3 y 4 o más pagos vencidos.

Las fuentes de información para la construcción de los modelos de segmentación se describen a continuación:

- Para 0 pagos vencidos, la base de datos histórica contiene diciembre 2009 y enero 2010;
- Para 1 o más pagos vencidos, la base de datos histórica contiene mayo y junio de 2010;

Como anteriormente se mencionó, los modelos de segmentación deben ser validados con una muestra de población, cuyas características provengan de una fecha posterior a las fechas en que se toman los datos para el desarrollo, a esta etapa se le llama “Validación fuera de la muestra”. Para el caso de la población de clientes en 0 pagos vencidos, se tomó mayo de 2011 y para cuentas en mora, 1, 2 y 3 o más pagos vencidos, agosto y noviembre de 2011, ambas con 6 meses de desempeño.

2.2.1 Bases de datos

Una parte fundamental para el desarrollo de modelos es la construcción de las bases de datos, ya que éstas forman parte del insumo principal de los árboles de decisión.

Para estos modelos de segmentación, la información proviene de sistemas internos de alguna Institución de crédito, la cual fue relacionada con información anterior (ventana de observación) y posterior a los meses de observación (ventana de desempeño).

La información interna contiene registros relacionados con el comportamiento crediticio de las tarjetas de crédito, estas bases deberán tener asociado un identificador único, el cual debe ser generado o en caso de existir dentro de las bases, deberá ser considerado. Con esta llave se relacionó la información histórica interna y externa, así como el comportamiento de pagos de las tarjetas de crédito.

La relación de las bases comienza tomando como pivote los meses de observación, es decir, se toman todas aquellas cuentas que en estos meses de observación (diciembre 2009 y enero 2010) están en cero pagos vencidos, pero pagaron unos días después de su fecha de exigibilidad de pago.

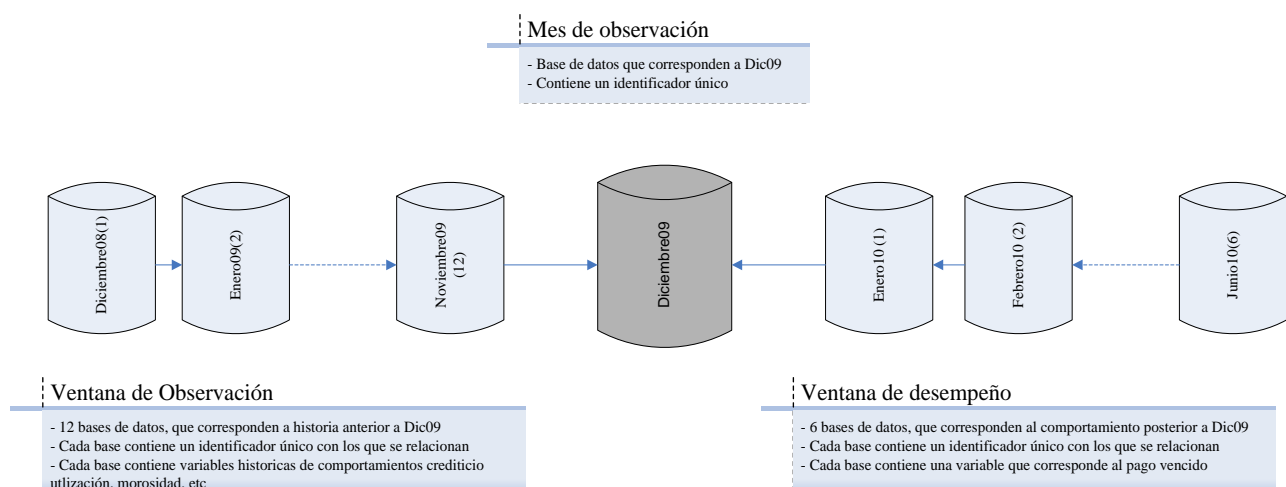
Lo mismo se realiza con la población de créditos en 1, 2 y 3 o mas pagos vencidos, se seleccionan todas las cuentas que al momento del mes de observación (mayo y junio 2010) están en 1, 2 y 3 o más pagos vencidos. Estas bases se relacionan de forma horizontal y se eliminan registros repetidos considerando el identificador único.



Seguido de esto, se relaciona información histórica de meses anteriores, para este caso se tomaron 12 meses de historia hacia atrás, esto se denomina “Ventana de observación”. Finalmente se relaciona la información del comportamiento, la cual está determinada por los meses posteriores a los meses de observación, para este caso se relacionaron 6 meses (ventana de desempeño).

A continuación se muestra un diagrama que ilustra como se relacionó la información:

Figura 2.5. Relación de bases



Como se ilustra en la figura 2.4, la ventana de observación contiene 12 meses anteriores a diciembre 2009, es decir, diciembre 2008, enero 2009, febrero 2009, etc., hasta noviembre 2009, y se relacionan a través de un identificador único asociado a la tarjeta de crédito. Ya relacionada esta información, se toma el mes de observación, el cual fue sombreado en la figura, y se pega con estos 12 meses anteriores, con el fin de tener la información histórica.

Finalmente se relaciona la información de desempeño, que contiene el pago vencido de cada mes de la ventana de desempeño, tal como se ilustró en la figura 2.5. El mismo proceso se repite el mes de observación de enero 2009.



2.2.2 Descripción de la población

A partir de la información descrita en el apartado anterior, se seleccionan y calculan variables, que de acuerdo con la experiencia del desarrollador y de las áreas de negocio involucradas indirectamente o directamente en el desarrollo, pudieran tener alguna correlación con el perfil de riesgo de una cuenta y buró de crédito, incluyendo aquellas que resultaron estadísticamente significativas en desarrollos previos.

Las bases de datos contienen indicadores para relacionar las bases, variables crudas provenientes del sistema, variables construidas a partir de las crudas, variables de comportamiento histórico e información de buró de crédito. Las cuales se describe de manera general a continuación:

- a) Información de la ventana de observación, creados a partir de la información cruda
 - Máximo pago vencido en los últimos 6 y 12 meses;
 - Número de veces en 1, 2 o 3 pagos vencidos en los últimos 6 y 12 meses.
- b) La información de buró de crédito se obtiene de las consultas mensuales y disponibles en el módulo de cobranza:
 - Edad de la cuenta más antigua de todas las referencias crediticias;
 - Meses desde la mora de 30 o más días;
 - Número de cuentas revolventes utilizadas al 100% o más de su línea de crédito autorizada;
 - Peor *MOP*⁵ reportado en los 6 meses en otras Instituciones de crédito.

2.3 Exclusiones

Posterior a la construcción de las bases de datos es necesario seleccionar aquellos registros de tarjetas de crédito que son cobrables, es decir, que se les pueda asignar alguna acción de cobranza.

Como parte de las exclusiones, deberán ser aquellas cuentas cuyas líneas de crédito estén bloqueadas, canceladas o determinadas como una pérdida (también denominadas créditos castigados) al momento de la observación, es decir, todo aquello que no tenga un saldo pendiente y no sea una cuenta activa al momento de la observación. Por otro lado, es

⁵ La definición de MOP es *manner of payment* o forma de pago, donde MOP=1 es al corriente, 2 un pago vencido, etc.



importante mencionar que todas aquellas tarjetas de crédito que hayan sido reestructuradas⁶ serán clasificadas como de alto riesgo.

Existen exclusiones relacionadas a las políticas de crédito, por lo que, cualquier valor fuera de los parámetros establecidos será una exclusión por política. Por ejemplo, los clientes clasificados como VIP.

2.3.1 Exclusiones para el desarrollo de los modelos

Las exclusiones para el desarrollo de los modelos de segmentación son definidas de acuerdo al pago vencido como se muestra en el ejemplo siguiente:

Para el caso de cuentas en 0 pagos vencidos:

- Cuentas con saldo pendiente menor al monto definido por la Institución de crédito, debido a que el costo asociado es mayor a la recuperación esperada;
- Cuentas cuyo saldo pendiente pudiera ser bajo y ser operado por agencias externas a la Institución financiera;
- Cuentas relacionadas con empleados de la Institución de crédito;
- Cuentas clasificadas como VIP (*very important people*).

Para el caso de cuentas en 1 o más pagos vencidos:

- Cuentas con dos meses de apertura y que han alcanzado 1 pago vencido (FPD, *first payment default*);
- Cuentas con saldo pendiente bajo (BB, balance bajo);
- Cuentas con saldo pendiente alto (BA, balance alto);
- Cuentas clasificadas como VIP.

En el cuadro 2.2 se ilustra como puede ser clasificada la población, de acuerdo a lo mencionado en párrafos anteriores, este proceso es recomendado con el fin de poder clasificar la población e identificar aquella que usaremos para el desarrollo de modelos de segmentación:

⁶ Cuenta reestructurada es aquella cuyas condiciones iniciales han sido modificadas, tales como la tasa y plazo.



Cuadro 2.2. Clasificación de la población

Población

Población Total
Población para desarrollar el modelo de segmentación
Población Excluida
<ul style="list-style-type: none">• Cuentas no activas al momento de la observación• Exclusiones del Score
Segmentos Especiales

2.4 Definición de la probabilidad de incumplimiento

Posterior a la creación del campo relacionado con el desempeño de las cuentas, es necesario definir a partir de este histórico, cuándo una cuenta será mala o buena, tal como se mencionó en la [sección 2.1.2](#). Para realizar esta definición se analizan movimientos de morosidad para cada población, es decir, 0, 1, 2, 3 y 4 o más pagos vencidos.

Este análisis consiste en obtener el máximo pago vencido en la ventana de desempeño, es decir, durante los siguientes 6 meses y observar el porcentaje de migración en los 3 y 6 meses posteriores.

Cabe mencionar que en la industria de crédito, la cartera vencida se considera cuando la cuenta alcanza 90 o más días de vencida (3 o más pagos vencidos).

Para ilustrar el análisis de migración por pago vencido se muestra el cuadro 2.3, en donde es necesario considerar la muestra aleatoria definida como “Población para desarrollar el modelo de segmentación” en la sección 2.3.

Cuadro 2.3. Porcentajes de migración para 0 pago vencido

Máximo pv en 6 meses	Portfolio Porcentaje	3+ pv en 9 meses Porcentaje	Castigo Porcentaje sobre 12 meses
3+ pv	6.2%	100.0%	34.5%
2 pv	4.5%	50.6%	0.5%
1 pv	11.9%	16.9%	0.1%
0 pv	77.4%	1.2%	0.0%
Total	100%	11.4%	2.2%



Como se puede observar en el cuadro anterior, se distribuye la población por pago máximo vencido en 6 meses, es decir, 77.4% tuvo cero pagos vencidos como máximo en los siguientes 6 meses, así sucesivamente, hasta que 6.2% obtuvo 3 o más pagos vencidos como máximo en una ventana de 6 meses.

Si observamos esta última población en los siguientes tres meses continúa en mora 100% de la población, y 34.5% se determinan como pérdidas tres meses después.

Es posible concluir que una definición consistente y preventiva para este portafolio, es que las cuentas al momento que alcancen 3 o más pagos vencidos se definan como cuentas malas, dado que en un periodo de 6 meses se deterioraron y tres meses después continuaron en el mismo pago vencido y posteriormente 34.5% fueron una pérdida para la institución financiera.

Este análisis debe de realizarse para las poblaciones de 1, 2 y 3 pagos vencidos, con el fin de terminar cual es la mejor definición de malo, de acuerdo a estos porcentajes de migración.

Para el caso de la población de 4 o más pagos vencidos, se determinó usar una definición en la cual nos permita observar el porcentaje de pago, es decir, la definición de una cuenta buena será aquella cuyo pago sea al menos el mínimo⁷, sobre el total de su deuda y de lo contrario será una cuenta mala. Esto debido a que en alta mora tiene como principal misión recuperar y no poner al corriente estas cuentas, ya que en su mayoría el saldo deudor podría ser mayor a los ingresos mensuales de los clientes.

Con base a los anteriores análisis de migración fue determinada la definición de desempeño, sin embargo se recomienda explorar más definiciones más por pago vencido con el fin de obtener la más adecuada por nivel de migración. En [Apéndice A](#), se presenta un ejemplo de este análisis.

2.4.1 Factores que afectan las definiciones de desempeño

Las cuentas identificadas como castigadas y reestructuradas en la ventana de desempeño se consideran como malas, debido a que las primeras generaron una pérdida monetaria y las segundas están relacionadas con programas cuyo fin es reducir tasa de interés y fijar plazos, debido a la disminución en la capacidad de pago que manifiestan tener los clientes.

⁷ Pago mínimo: es el importe que se solicita al titular de la tarjeta de crédito, de acuerdo a lo estipulado por Banco de México, circular 13/ 2011, ver referencias.



Por otro lado, debido al deterioro en las cuentas con 1, 2, 3 y 4 o más pagos vencidos, se decidió considerar una ventana de desempeño de 6 meses, con el fin de actuar de manera inmediata y no esperar a que éstas cuentas lleguen a pérdidas. En la industria crediticia, normalmente se asume la pérdida de un crédito si alcanza 8 o más pagos vencidos.

2.4.2 Definiciones de desempeño y volúmenes

La definición de desempeño de una cuenta buena o una cuenta mala, para el desarrollo del modelo es la siguiente:

- i. Para 0, 1 y 2 pagos vencidos:
 - Mala: cuando el crédito al momento de la observación esté en 0, 1 ó 2 pagos vencidos, y la cuenta alcanza 3 o más pagos vencidos (incluyendo castigadas o reestructuras) en la ventana de desempeño, es decir, en los siguientes 6 meses.
 - Buena: la cuenta que presenta cualquier otra situación.
- ii. Para 3 pagos vencidos:
 - Mala: cuando el crédito al momento de la observación esté en 3 pagos vencidos, y la cuenta que alcanza 4 o más pagos vencidos (incluyendo castigadas o reestructuras) en la ventana de desempeño, es decir, en los siguientes 6 meses.
 - Buena: la cuenta que presenta cualquier otra situación
- iii. Para 4+ pagos vencidos:
 - Mala: cuando el crédito al momento de la observación esté en 4 ó más pagos vencidos, la cuenta haya sido reestructurada, castigada o el % de pago respecto al total adeudo sea menor al pago mínimo en la ventana de desempeño, es decir, en los siguientes 6 meses
 - Buena: la cuenta que presenta cualquier otra situación

Además de considerar el comportamiento del portafolio que se estudia, es importante considerar la operación de cobranza para baja mora, ya que ésta tiene como principal misión mantener una cuenta vigente, o sea en un nivel máximo de 2 pagos vencidos, mientras que para alta mora tiene como principal misión detener el deterioro, por lo que es necesario incluir estrategias que prevengan estas migraciones.

De acuerdo con la clasificación de las cuentas en buenas y malas, se muestra la distribución por pago vencido del caso de estudio, en el siguiente cuadro:



Cuadro 2.4. Distribución de la población por clasificación de cuentas⁸

Modelo	Buenos (%)	Malos (%)	Total (%)
Modelo de segmentación de TDC baja mora (0 PV)	92.47%	7.53%	100%
Modelo de segmentación de TDC baja mora (1 PV)	56.84%	43.16%	100%
Modelo de segmentación de TDC baja mora (2 PV)	15.94%	84.06%	100%
Modelo de segmentación de TDC alta mora (3 PV)	13.31%	86.69%	100%
Modelo de segmentación de TDC alta mora (4+PV)	9.88%	90.12%	100%

2.5 Análisis exploratorio de los datos

Ya que se cuenta con la base de datos integrada con las variables dependientes (información histórica de 12 meses) y la bandera de cuenta buena o mala (considerando 6 meses posteriores al mes de observación), tal como se describió en la sección 2.2.1., es necesario realizar un análisis exploratorio de todas las variables (independientes), el cual consiste en el cálculo de algunas de las medidas descriptivas de posición y variabilidad, el cual no es más que el análisis univariado, que contiene mínimo, máximo, media, cuantiles, etc., reflejando la distribución de la población. A continuación se presenta un extracto de algunos descriptivos que se obtuvieron de la base de datos que se utilizará para el desarrollo de los modelos de segmentación.

Cuadro 2.5. Resultados del análisis descriptivo

Field Name	Data Type	Cardinality	# of Missing Values	Minimum	Maximum	Mean	Standard Deviation	Unique Count
Var_1	Number	500	0	-1	11,753.00	95.76	103.82	46
Var_2	Number	72	0	-2	103	2.69	6.84	1
Var_3	Number	15	0	-1	13	0.69	1.07	2
Var_4	Number	12	0	-1	99	9.71	27.76	0

Este análisis descriptivo se puede obtener a través del programa estadístico disponible o en Excel, si la base de datos es pequeña, donde considera las variables dependientes (en el ejemplo nombrada como var_1, var_2, etc.), las cuales pueden ser, porcentaje de utilización promedio, cuantas veces ha estado en 1 pago vencido, etc. El número de

⁸ Fuente: Información de la Institución financiera, la cual es una muestra aleatoria, no contiene información relacionada al cliente



variables, depende de la información disponible que tiene cada Institución y de las que pueda crear el desarrollador.

La finalidad de este análisis es detectar aquellas variables que pudieran presentar un número significativo (90% o más) de valores “*missings*”, o valores fuera de catálogos, es decir no existentes. Por ejemplo, en caso de antigüedad de la cuenta en la Institución crediticia, sea un número negativo o exceda el máximo o errores de formato en las variables. Además, de asegurarse que en caso de una revisión externa (por ejemplo, auditoría por parte de la CNBV⁹), los datos no sean una causa de no factibilidad del modelo. De acuerdo a las mejores prácticas normalmente se recomienda revisar los siguientes aspectos:

- i. Lógica entre los registros que pudieron ser relacionados entre las bases históricas. Ejemplo, duplicidad de cuentas.
- ii. Nombre de las variables, formatos o posible sobre escrituras.
- iii. Para el caso de tener variables nuevas generadas, revisar si no existe un error de sintaxis en los códigos.
- iv. De acuerdo con las exclusiones, revisar que las variables tengan la etiqueta correspondiente, Bueno, Malo, VIP, Balance bajo, etc., así como que el volumen sea el correcto.
- v. Valores “*Missing*” o vacíos.
- vi. Valores de acuerdo con los catálogos.
- vii. Valores especiales, por ejemplo, existen valores predeterminados, como -1, -99, etc., que sean los correctos.
- viii. Revisar todos los campos de acuerdo al diccionario de datos.

Una vez que son detectadas las variables con las características antes mencionadas, deberán ser excluidas del desarrollo del modelo, sin que sean borradas de la base de datos, esto con el fin de poder llevar un seguimiento adecuado de la información.

Por otro lado, es necesario excluir del desarrollo, aquellas variables de control, tales como número de identificador o cualquier otro campo que no determine alguna característica de comportamiento crediticio.

Cabe mencionar que este no será el único método para la selección de variables, ya que es necesario analizar el nivel de significancia estadística y la correlación que existiera entre ellas, esto se explicará en el siguiente capítulo.

⁹ CNBV: Comisión Nacional Bancario y de Valores



III. Desarrollo de la segmentación

Como se mencionó en el capítulo anterior, para el desarrollo del modelo de segmentación se utilizó un programa para construcción de árboles, que permite utilizar diferentes métodos, tales como “*Clusters o Entropy Ratio*”. Este último fue el que se seleccionó como método, el cual permite agrupar poblaciones de acuerdo con el nivel de significancia por cada variable (Valor de la información). Es importante mencionar, que es posible realizar segmentaciones con “*Clusters*”, sin embargo, las conocidas como mejores prácticas en la industria financiera, recomiendan la aplicación del método de Entropía.

3.1 Desarrollo

Posterior al análisis exploratorio, que se realizó con las variables independientes, disponibles e identificadas las variables que quedarán fuera del desarrollo del modelo de segmentación, fue necesario dividir la población en partes, las cual llamaremos “Población para el Desarrollo” y “Población para la Validación”.

Para realizar este último paso, es necesario utilizar un proceso aleatorio. De acuerdo a las mejores prácticas en los modelos de cobranza o de portafolio, no es necesario ponderar las bases para obtener muestras balanceadas 50% buenos y 50% malos. Tal como realiza en los modelos de originación de créditos, ya que con este esquema de pesos, se permite que el modelo final sea igualmente sensible a las poblaciones de buenos y malos, lo que permite recabar características de ambas.

Para el caso de modelos de segmentación, se requiere observar la población con su distribución natural en el portafolio, ya que permite definir un perfilamiento con base en su comportamiento actual.



De acuerdo al tamaño de la población, esta se puede dividir en 70% para desarrollar y 30% para validar u 80% y 20% respectivamente. Este proceso se realiza con el objetivo de validar que los resultados que se obtienen con la población de desarrollo, sean similares a los de validación, es decir, se desarrolla el modelo con 70% de la población, el modelo resultante se aplica a la población de validación (30% del total).

Este tipo de pruebas se realiza con el fin de evitar un sobre ajuste del modelo, es decir, que las variables que se incluyeron en el modelo, tengan un nivel de predictividad alto debido a las peculiaridades de la base que se construyó, y al ser aplicadas a una muestra diferente (30%) no se comporten de la misma forma.

Los indicadores de la eficiencia del modelo de segmentación (KS, “*Gini*” y VI), deben ser comparados como parte de la prueba de validación. Se recomienda que la diferencia en puntos porcentuales sea $\pm 3\%$, entre los resultados de desarrollo y validación. En la siguiente sección se describe a detalle el proceso de construcción de estas bases.

3.2 Tamaño de la muestra

Hay ocasiones en las que se cuenta con muestras muy grandes, por ejemplo, más de 600 mil registros, por lo que es recomendable obtener una muestra aleatoria representativa, y posteriormente dividir la población en desarrollo y validación.

Es posible desarrollar con muestras muy grandes, sin embargo, algunos programas pierden eficiencia cuando el volumen de registros es grande y regularmente los resultados no presentan diferencias significativas, para éste trabajo no fue necesario realizar este proceso de muestreo ya que el volumen era suficiente para el desarrollo (tamaño y representatividad).

En caso de disponer de bases pequeñas en términos de registros, se recomienda tomar bases completas.

3.3 Diseño de las muestras

Como se mencionó anteriormente en el proceso de segmentar las poblaciones, se debe asegurar que sea de forma aleatoria, ya sea asignando un número aleatorio a cada registro o usando comandos ya incluidos en la mayoría de los programas, para este caso de estudio se utilizó la siguiente función de SAS, “RANUNI” (ver [Apéndice C](#) para mayor detalle), la cual crea números entre 0 y 1, a través de una distribución uniforme sobre un intervalo (0,1). Esto se codificó de la siguiente manera:

```
Data DEV VAL;  
set basedatos;
```




```
if ranuni (8) <= 0.7 then split="DEV";  
else split= "VAL";  
run;
```

El resultado de ejecutar esta función son dos bases de datos, “DEV” y contiene 70% de la población y “VAL” 30% de la población. Este proceso se realizó para cada población, es decir, para las bases de 0, 1, 2, 3 y 4 o más pagos vencidos.

3.4 Metodología de la selección de variables independientes

La selección de variables independientes puede ser un proceso difícil, si se cuenta con un número considerable de variables, cientos o miles, incluso después de la primera etapa de depuración que se describió en el capítulo dos. Para este caso de estudio, se consideraron los siguientes pasos, como segunda etapa del proceso de selección:

- i. Selección de variables bajo un criterio definido por el desarrollador, por ejemplo:
 - La calidad de cada variable: se sugiere revisar que las variables seleccionadas provengan de datos que fueron validados con documentación proveniente de los clientes, por ejemplo, edad, ingresos, etc.
 - Variables robustas con respecto a cambios económicos: por ejemplo, no considerar variables como ingresos extras, bonos, gratificaciones, etc.
 - Valor predictivo de las variables: después de obtener el Valor de la información, considerar aquellas variables con mayor valor.
 - De fácil interpretación: evitar usar variables contraídas cuya interpretación no es inmediata, por ejemplo, desviación estándar de los ingresos, etc.
 - Fácil captura e implementación: por ejemplo evitar construir variables combinadas con otras.
- ii. Análisis de correlación entre las variables independientes:

De acuerdo con la práctica estándar de desarrollo de modelos, el análisis de correlación tiene como fin poder incluir o excluir variables dentro de los modelos, basado en el índice “Pearson” y/o “Spearman” (ver detalle [Apéndice C](#)), se sugiere excluir aquellas variables cuyo valor de “Pearson” es mayor a ± 0.60 o valor de la información es mayor igual a 4.

Para realizar este análisis se utilizó la siguiente función en “SAS”:

```
PROC CORR BEST=20 NOPROB DATA=basedatos;  
WHERE split="DEV";  
VAR %VARLIST;  
RUN;
```



El resultado de ejecutar esta función es un archivo, en donde muestra la matriz de correlaciones del total de variables independientes seleccionadas y que están correlacionadas con las 20 primeras, considerando la base de desarrollo. A continuación se muestra un extracto de la matriz de correlaciones con dos variables y las variables con las que se encuentran correlacionadas.

Cuadro 3.1. Matriz de correlaciones

Var_1	Var_1	Var_2	Var_3	Var_4	Var_5	Var_6	Var_7	Var_8	Var_9	Var_10
Pearson	1	0.99981	0.54879	0.53074	-0.50169	-0.50137	0.49029	-0.38958	0.38502	-0.34347
	68889	68889	68889	68889	68889	68889	68889	68889	68889	68889
Var_2	Var_2	Var_1	Var_3	Var_4	Var_5	Var_6	Var_7	Var_8	Var_9	Var_10
Pearson	1	0.99981	0.54905	0.53089	-0.5009	-0.50065	0.49046	-0.38869	0.38563	-0.34276
	68889	68889	68889	68889	68889	68889	68889	68889	68889	68889

Con esta matriz de correlaciones, se deben de excluir aquellas variables cuyo índice de *Person* es mayor a ± 0.60 , como se ilustra en el cuadro 3.4.1, la variable, Var_1 se encuentra altamente correlacionada con la variable Var_2, ya que su coeficientes es 0.99981. En este caso se determina excluir la variable Var_2, éste proceso de debe de continuar con el resto de las variables, por ejemplo, Var_1 con Var_3, Var_4, etc. Como se observa, el coeficiente esta entre ± 0.60 , lo cual nos indica que estas variables son independientes, es decir, no son variables cuya definición describe lo mismo de un cliente.

Por ejemplo, si la variable Var_1 se refiere a la utilización de la línea de crédito cuando es más de 70% o más y la variable Var_2 determina la utilización de la línea de crédito de 100% o más, se concluye que están correlacionadas, debido a que la segunda variable esta contenida en la Var_1, puesto que la población de clientes que tengan una utilización de su línea de 100% o más esta contenida en la población que determina la Var_1.

3.5 Resultados del la segmentación

Posterior a la construcción y procesamiento de la información histórica, tal como se describió en los capítulos anteriores, se utilizó el programa especializado en la construcción de árboles de decisión, con el fin de diseñar los modelos de segmentación.

Ésta herramienta permite diseñar árboles de decisión, con variables de respuesta continua o dicotómica¹⁰, para este trabajo la variable de respuesta (probabilidad de incumplimiento

¹⁰ Variable dicotómica: que sólo puede tomar dos valores, los valores son complementarios y dichos valores no son comparables.



o variable dependiente, como la hemos llamado en los capítulos anteriores) es dicotómica y está definida como cuenta buena o mala, paga o no paga.

El desarrollo del modelo comienza con la carga de datos al paquete estadístico, el cual puede leer datos con formatos txt, dbf (FOX pro), sas7bdat (SAS), entre otros. Este último fue utilizado en el procesamiento de la información.

El siguiente paso, es hacer la selección de variables a utilizar, así como un análisis descriptivo, el cual incluye el conteo de valores “missing”, mínimos, máximos, media, desviación estándar y conteo de valores únicos por variable, proceso descrito anteriormente y que forma parte de la etapa de la selección de variables.

La siguiente parte es obtener el nivel de significancia de cada una de las variables, es decir, la razón de entropía, la cual nos permitirá escoger las variables que formarán parte de los modelos de segmentación, con este valor es posible determinar qué variables segmentan mejor la población.

Algunos programas, para construir árboles de decisión, permiten diseñar árboles de forma automática o manual, la primera opción permite obtener un árbol de decisión con las variables más predictivas, sin importar el sentido del negocio o significado de la variable.

Mientras el diseño manual, permite seleccionar las variables que por sentido del negocio o experiencia del modelador, deberían ser parte del modelo de segmentación, además de considerar el nivel de significancia de cada una de las variables. Esta última opción fue la que se consideró para este trabajo, ya que permitió incluir variables que además de tener un nivel predictivo estadísticamente significativo, aportaban sentido lógico de acuerdo al objetivo de la segmentación, tal es el caso de variables de buró de crédito.

3.6 Consideraciones para el desarrollo del modelo

Las consideraciones dentro del diseño del árbol fueron, incluir variables con sentido lógico para la cobranza, es decir, variables que determinen morosidad de forma histórica o relacionada con el historial crediticio fuera de la Institución de crédito, entre otras descritas en los resultados finales.

Las variables utilizadas para estos modelos son en su mayoría variables independientes continuas, por lo que éstas deberán ser definidas como ordinarias dentro del árbol de segmentación, además de conservar un orden respecto a la tasa de malos. Por ejemplo, si la variable “número de veces que ha alcanzado 2 pagos vencidos”, forma parte del árbol, se espera que entre mayor sea el número de veces que ha alcanzado 2 pagos vencidos, está relacionada con una tasa de malos mayor y viceversa, entre menor sea el número de veces que ha alcanzado, la tasa de malos será menor.



Otra de las consideraciones que permite la estabilidad del modelo, es el porcentaje de población que estará contenida en cada uno de los nodos finales. Como parte de las mejores prácticas en el desarrollo de modelos, se sugiere que éstos contengan al menos 5% de la población total. Cabe mencionar, que este criterio dependerá del modelador, ya que algunas ocasiones es necesario conservar nodos con menos población, debido a las tasas de malos que éstos presentan.

Lo anterior está relacionado con el número de ramas, y niveles que deberá contener el árbol, ya que en caso de que los nodos finales, contengan menos del 5% del total de la población, se recomienda no segmentar más a partir de éstos nodos, con la finalidad de no perder estabilidad y por lo tanto nivel de predicción.

Es importante mencionar, que un modelo de segmentación no resulta robusto por el número de variables, es decir, el incluir un número grande de variables no siempre da como resultado un modelo más robusto y sí puede complicar la implementación.

Finalmente, los árboles son aplicados a las poblaciones de validación (30% del total), población total y una población actual fuera de la muestra, con el fin de evitar sobre ajustes.

3.7 Modelo de segmentación

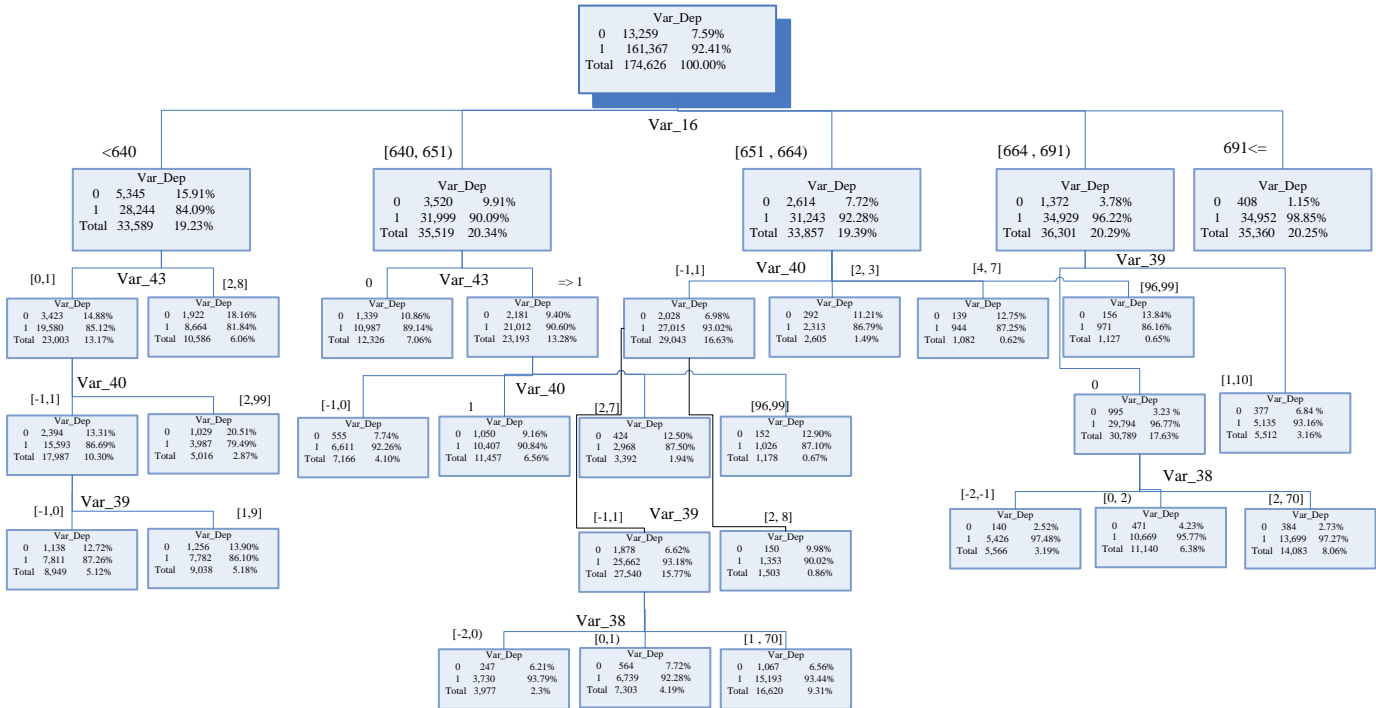
A continuación, se muestran uno de los modelos de segmentación (árboles de decisión), que se obtuvieron para la población de desarrollo (70%) por pago vencido.

La variable dependiente o probabilidad de incumplimiento, está determinada por la Var_Dep, donde 0 corresponde a las cuentas malas y 1 a las cuentas buenas.

Posteriormente, cada nivel presenta la variable con la que se segmentó, para este caso están nombradas como Var_16, Var_43, Var_40, etc. Es importante considerar, que el número de variables y cuáles deberán ser parte del modelo, dependerá del criterio del desarrollador y la disponibilidad de datos, que puede variar de una Institución de crédito a otra. Por lo que, se consideró que no es relevante el detalle de cada variables, sin embargo, se recomienda considerar información tal como se describió en la sección 2.2.2.



Figura 3.1 Modelo de segmentación para la población de cero pagos vencidos



Con el fin de obtener los indicadores de eficiencia, es necesario numerar los nodos finales, con el fin de capturar los datos de cada uno de éstos, dentro de una tabla de Excel, y calcular las métricas, las cuales se mostrarán en las siguientes secciones.

Para este modelo de segmentación únicamente se utilizaron 5 variables independientes, entre ellas, el valor del modelo de comportamiento (Var_16), el resto de las variables están relacionadas con utilización y morosidad.

El desarrollar un árbol de decisión, es relativamente simple, sin embargo se recomienda, crear diferentes árboles por cada población, con el fin de obtener un árbol robusto y con el menor número de variables, y disminuir la complejidad de la implementación. Nuevamente este es un criterio que puede ser establecido por el desarrollador.

El árbol anterior fué seleccionado de entre 3 y 4 opciones, debido a que mostraron mejores resultados en términos de KS.



3.8 Resultados del modelo final

Ya contruidos los árboles finales, se obtuvieron los indicadores que miden la eficiencia de éstos árboles, para la población de desarrollo (70%), validación (30%) y la población total (100%).

Como se explicó al inició de este capítulo, es importante realizar una validación exhaustiva sobre los árboles, con el objetivo de verificar si se obtendrá un desempeño similar sobre poblaciones diferentes a las utilizadas en el desarrollo. Esto ayuda a evitar resultados no esperados post implementación.

Debido a lo anterior, se sugiere observar el desempeño sobre las poblaciones antes descritas y sobre una población actual fuera de la muestra de desarrollo, cuyos resultados son mostrados en la [sección 4.6](#) de este capítulo. Es importante señalar, que cambios en indicadores que miden la eficiencia de los modelos es inevitable. Sin embargo, un cambio considerable hacia la baja, indica un sobre ajuste en la población de desarrollo y las recomendaciones serían desarrollar un nuevo modelo de segmentación, con menos variables, y ajustando los parámetros estadísticos, es decir, considerar niveles diferentes en la razón de entropía. Los cambios en los índices estadísticos hacia el alta, pueden ser considerados como un mejor desempeño del modelo, por lo que es posible conservar modelos de segmentación con estas características.

A continuación se muestran los resultados obtenidos para las diferentes muestras:

- i. Modelo de segmentación para la población con 0 pagos vencidos

Población Tarjeta de Crédito 0 PV	GINI	K-S	IV
Desarrollo	43%	31.7%	0.71
Validación	43%	32.3%	0.71
Desarrollo + Validación	43%	31.8%	0.71



ii. Modelo de segmentación para la población de 1 pago vencido

Población Tarjeta de Crédito 1 PV	GINI	K-S	IV
Desarrollo	46%	30.8%	0.46
Validación	45%	29.5%	0.45
Desarrollo + Validación	45%	30.4%	0.46

Los resultados en el índice de KS para la población de 0 y 1 pago vencido, muestran que los árboles son robustos, considerando que son mayores o iguales a 30%. En la actualidad, no existe literatura que marque cuales son los niveles que se debe obtener para concluir que el árbol es robusto, sin embargo, si se considera que éste es un indicador que muestra la separación entre cuentas buenas y malas, se espera que este sea lo más grande posible.

iii. Modelo de segmentación para la población de 2 pagos vencidos

Población Tarjeta de Crédito 2 PV	GINI	K-S	IV
Desarrollo	34%	23.6%	0.23
Validación	34%	24.1%	0.23
Desarrollo + Validación	34%	23.8%	0.23

iv. Modelo de segmentación para la población de 3 pagos vencidos

Población Tarjeta de Crédito 3 PV	GINI	K-S	IV
Desarrollo	39%	27.9%	0.3
Validación	36%	24.7%	0.23
Desarrollo + Validación	38%	26.5%	0.27

El índice de KS que se obtuvo para las poblaciones de 2 y 3 pagos vencidos, muestran que los árboles son medianamente robustos, considerando que son mayores o iguales a 20%. Este resultado se considera aceptable, ya que en cuentas que están en 2 ó 3 pagos vencidos, la probabilidad de incumplimiento es mayor, y las cuentas buenas son semejantes a las cuentas malas, es decir, tienen un perfil similar, y por lo tanto la separación no es significativa.



v. Modelo de segmentación para la población de 4 o más pagos vencidos

Población Tarjeta de Crédito 4 + PV	GINI	K-S	IV
Desarrollo	54%	43.4%	0.79
Validación	55%	45.2%	0.84
Desarrollo + Validación	54%	44.0%	0.8

Para el caso de la población con 4 o más pagos vencidos, existe una diferencia, significativa entre las cuentas buenas y cuentas malas, las cuales están relacionadas con el pago que realizan los clientes, es decir, las características generales que determinan el perfil de los clientes que pagan un porcentaje de su deuda, y aquellos que no pagan o realizan un pago menor a lo que se estimó con esta definición, son diferentes, y esta separación esta dada por el indicador KS.

En términos generales, se observan diferencias en los indicadores estadísticos Gini, KS y VI, de las muestras de desarrollo (70%), validación (30%) y total (100%), las cuales no son representativas y guardan una diferencia razonable en términos de puntos porcentuales, es decir, los árboles construidos con el 70% de la población, muestran el mismo desempeño en la muestra de 30 y 100%.

3.8.1 Cálculo de indicadores estadísticos

El cálculo de los índices estadísticos mostrados en la sección anterior fueron formulados y obtenidos en Excel, nodos finales se ordenaron de acuerdo a las tasas de malos (columna Malos por interv%), de mayor a menor. Posterior a esto, se determinaron los grupos de riesgo, cuya definición se basó en la tasa de malos, es decir, nodos con tasas de malos similares (diferencial no significativo de acuerdo al criterio del modelador), permanecerán en el mismo grupo de riesgo. A continuación, se muestran los resultados por pago vencido para la muestra de desarrollo, donde la descripción por columnas es la siguiente:

1) Grupos de Riesgo:

Grupo de Riesgo	Valor	Descripción
1	RC	Riesgo Crítico
2	RMA	Riesgo muy alto
3	RA	Riesgo alto
4	RM	Riesgo medio
5	RB	Riesgo bajo

2) Nodo: Se número cada nodo final de derecha a izquierda dentro del árbol



- 3) Total%: Porcentaje de la población total que contiene cada nodo
- 4) Malos%: Porcentaje de cuentas malas que contiene el nodo de la población total de cuentas malas
- 5) Buenos%: Porcentaje de cuentas buenas que contiene el nodo de la población total de cuentas buenas
- 6) Acum% Total, Acum% Malos y Acum% Buenos: son los porcentajes acumulados de la población total, malos y buenos.
- 7) Tasa de Malos por intervalo: Es la tasa de malos por cada nodo, es decir, son las cuentas malas entre el total de cuentas contenidas en el nodo final
- 8) Difer%: es la diferencia en valor absoluto, entre el porcentaje Acum% Malos y Acum% Buenos

Posteriormente se muestra una gráfica para ilustrar el índice KS obtenido por cada árbol, en donde se grafican las frecuencias acumuladas de cuentas malas (Acum% Malos) contra las frecuencias acumuladas de cuentas buenas (Acum% Buenos).

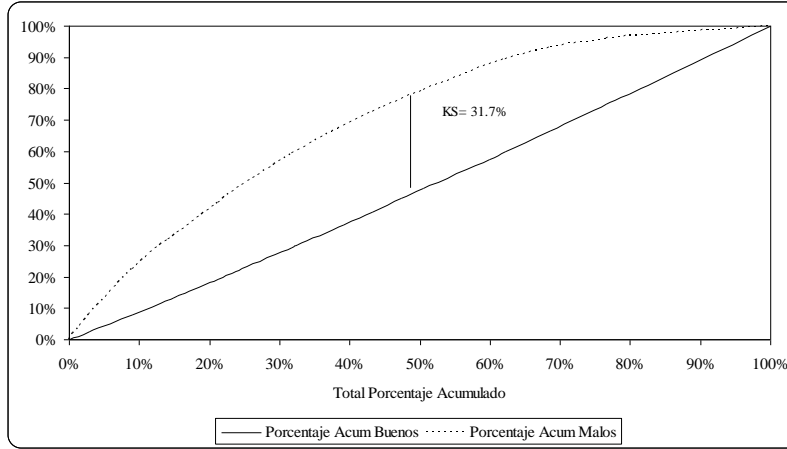
i. Población de desarrollo con 0 pagos vencidos

Cuadro 3.2. Distribución de la población por grupo de riesgo y nodo

									Valor de Información:	0.71	
									K-S:	31.7%	
									GINI:	0.43	
Grupo de Riesgo	Nodo	Malos %	Buenos %	Acum % Total	Acum % Malos	Acum % Buenos	Malos por interv %	Difer %			
1	3	7.76%	2.47%	2.87%	7.76%	2.47%	20.51%	5.3%			
	4	14.50%	5.37%	8.93%	22.26%	7.84%	18.16%	14.4%			
	2	9.47%	4.82%	14.11%	31.73%	12.66%	13.90%	19.1%			
	16	1.18%	0.60%	14.76%	32.91%	13.26%	13.84%	19.6%			
2	9	1.15%	0.64%	15.43%	34.05%	13.90%	12.90%	20.2%			
	15	1.04%	0.59%	16.05%	35.09%	14.48%	12.75%	20.6%			
	1	8.58%	4.84%	21.17%	43.68%	19.33%	12.72%	24.4%			
3	8	3.20%	1.84%	23.12%	46.87%	21.16%	12.50%	25.7%			
	14	2.20%	1.43%	24.61%	49.08%	22.60%	11.21%	26.5%			
	5	10.10%	6.81%	31.67%	59.17%	29.41%	10.86%	29.8%			
	13	1.13%	0.84%	32.53%	60.31%	30.25%	9.98%	30.1%			
	7	7.92%	6.45%	39.09%	68.23%	36.69%	9.16%	31.5%			
4	6	4.19%	4.10%	43.19%	72.41%	40.79%	7.74%	31.6%			
	11	4.25%	4.18%	47.37%	76.66%	44.97%	7.72%	31.7%			
	20	2.84%	3.18%	50.53%	79.51%	48.15%	6.84%	31.4%			
	12	8.05%	9.42%	59.84%	87.56%	57.57%	6.56%	30.0%			
5	10	1.86%	2.31%	62.12%	89.42%	59.88%	6.21%	29.5%			
	18	3.55%	6.61%	68.50%	92.97%	66.49%	4.23%	26.5%			
	19	2.90%	8.49%	76.56%	95.87%	74.98%	2.73%	20.9%			
5	17	1.06%	3.36%	79.75%	96.92%	78.34%	2.52%	18.6%			
	21	3.08%	21.66%	100.00%	100.00%	100.00%	1.15%	0.0%			
		100.00%	100.00%						31.7%		



Gráfica 3.1. Gráfica de KS



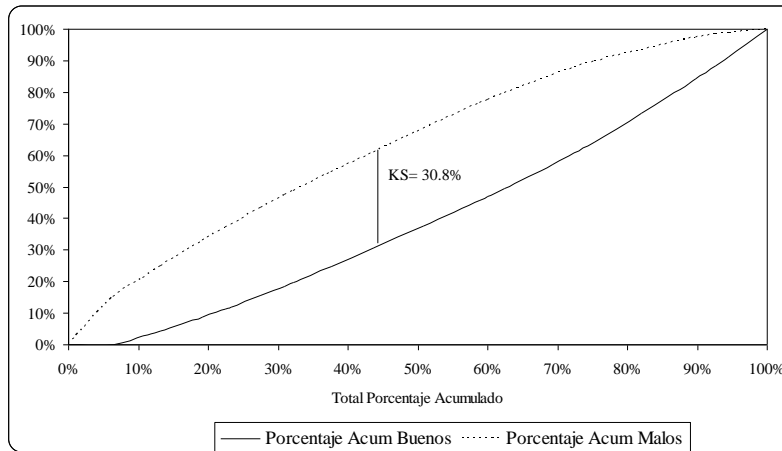
ii. Población de desarrollo con 1 pago vencido

Cuadro 3.3. Distribución de la población por grupo de riesgo y nodo

										Valor de Información:	0.46	
										K-S:	30.8%	
										GINI:	0.46	
Grupo de Riesgo	Nodo	Total %	Malos %	Buenos %	Acum % Total	Acum % Malos	Acum % Buenos	Malos por interv %	Difer %			
1	1	6.43%	14.94%	0.00%	6.43%	14.94%	0.00%	100.00%	14.9%			
	4	4.20%	6.13%	2.75%	10.63%	21.08%	2.75%	62.78%	18.3%			
	7	2.54%	3.54%	1.79%	13.17%	24.61%	4.53%	59.87%	20.1%			
	3	6.02%	8.28%	4.30%	19.19%	32.89%	8.84%	59.22%	24.1%			
	8	3.27%	4.23%	2.55%	22.45%	37.12%	11.38%	55.61%	25.7%			
	11	0.76%	0.94%	0.62%	23.21%	38.06%	12.01%	53.35%	26.1%			
2	6	4.35%	5.24%	3.68%	27.56%	43.31%	15.68%	51.85%	27.6%			
	2	3.31%	3.89%	2.88%	30.88%	47.20%	18.56%	50.48%	28.6%			
	17	1.26%	1.44%	1.13%	32.14%	48.63%	19.69%	48.97%	28.9%			
	23	5.38%	5.88%	5.00%	37.52%	54.51%	24.70%	47.00%	29.8%			
	22	6.16%	6.44%	5.96%	43.69%	60.95%	30.66%	44.92%	30.3%			
	5	7.14%	7.45%	6.90%	50.82%	68.40%	37.56%	44.91%	30.8%			
3	18	8.07%	7.93%	8.18%	58.90%	76.33%	45.73%	42.27%	30.6%			
	14	1.13%	1.11%	1.14%	60.02%	77.44%	46.88%	42.16%	30.6%			
	13	2.38%	2.09%	2.59%	62.40%	79.53%	49.47%	37.80%	30.1%			
	10	6.83%	5.83%	7.58%	69.23%	85.36%	57.05%	36.76%	28.3%			
4	9	3.12%	2.41%	3.66%	72.35%	87.77%	60.71%	33.19%	27.1%			
	20	2.93%	1.89%	3.71%	75.28%	89.66%	64.42%	27.75%	25.2%			
	12	6.71%	3.69%	8.98%	81.98%	93.36%	73.40%	23.70%	20.0%			
	21	6.84%	3.38%	9.44%	88.82%	96.74%	82.84%	21.29%	13.9%			
	19	0.94%	0.41%	1.34%	89.76%	97.15%	84.18%	18.83%	13.0%			
	16	2.91%	1.21%	4.19%	92.67%	98.36%	88.37%	17.90%	10.0%			
5	15	1.78%	0.43%	2.80%	94.45%	98.80%	91.17%	10.48%	7.6%			
	24	5.55%	1.20%	8.83%	100.00%	100.00%	100.00%	9.33%	0.0%			
			100.00%	100.00%						30.8%		



Gráfica 3.2. Gráfica de KS



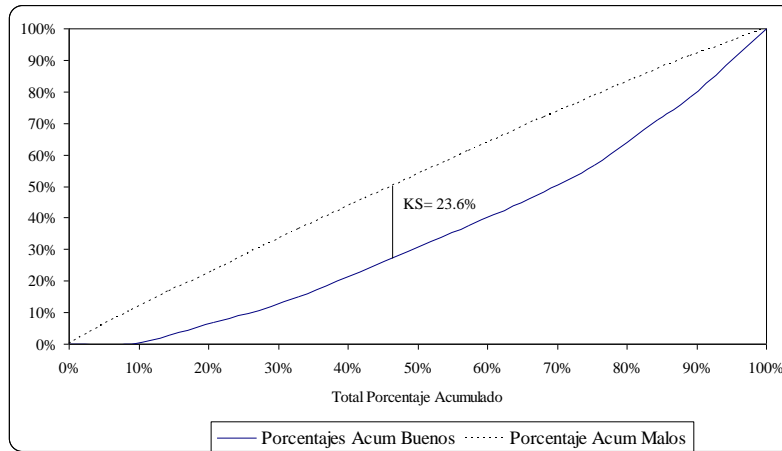
iii. Población de desarrollo con 2 pagos vencidos

Cuadro 3.4. Distribución de la población por grupo de riesgo y nodo

										Valor de Información:	0.23
										K-S:	23.6%
										GINI:	0.34
Grupo de Riesgo	Nodo	Total %	Malos %	Buenos %	Acum % Total	Acum % Malos	Acum % Buenos	Malos por interv %	Difer %		
1	1	8.9%	10.6%	0.0%	8.9%	10.6%	0.0%	100.0%	10.6%		
2	2	5.5%	5.9%	3.0%	14.4%	16.6%	3.0%	91.3%	13.6%		
1	6	6.3%	6.8%	4.0%	20.7%	23.3%	7.0%	89.8%	16.3%		
2	5	12.3%	13.1%	8.1%	33.0%	36.4%	15.1%	89.4%	21.3%		
3	4	24.3%	24.6%	22.6%	57.3%	61.0%	37.7%	85.1%	23.4%		
	3	10.8%	10.9%	10.6%	68.1%	71.9%	48.3%	84.3%	23.6%		
	8	7.1%	6.9%	8.4%	75.3%	78.8%	56.7%	81.1%	22.1%		
4	7	9.2%	8.2%	14.3%	84.5%	87.0%	71.0%	75.0%	16.0%		
	9	4.7%	4.2%	7.3%	89.1%	91.2%	78.3%	75.0%	12.9%		
	10	10.9%	8.8%	21.7%	100.0%	100.0%	100.0%	68.0%	0.0%		
			100.0%	100.0%					23.6%		



Gráfica 3.3. Gráfica de KS



iv. Población de desarrollo con 3 pagos vencidos

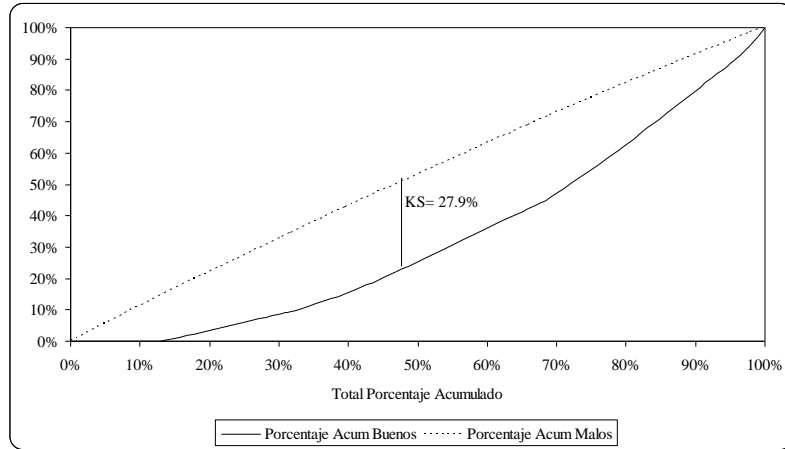
Cuadro 3.5. Distribución de la población por grupo de riesgo y nodo

Valor de Información: 0.30
K-S: 27.9%
GINI: 0.39

Grupo de Riesgo	Nodo	Total %	Malos %	Buenos %	Acum % Total	Acum % Malos	Acum % Buenos	Malos por interv %	Difer %
1	14	11.85%	13.34%	0.00%	11.85%	13.34%	0.00%	100.00%	13.3%
	4	2.48%	2.73%	0.51%	14.33%	16.07%	0.51%	97.69%	15.6%
	3	16.09%	17.07%	8.24%	30.41%	33.15%	8.75%	94.26%	24.4%
2	2	1.32%	1.40%	0.74%	31.74%	34.54%	9.50%	93.70%	25.0%
	9	7.76%	8.01%	5.77%	39.50%	42.55%	15.27%	91.67%	27.3%
	7	7.27%	7.34%	6.70%	46.77%	49.90%	21.97%	89.68%	27.9%
	8	19.11%	18.97%	20.25%	65.88%	68.87%	42.23%	88.14%	26.6%
	1	3.09%	3.04%	3.45%	68.97%	71.91%	45.67%	87.50%	26.2%
3	6	9.10%	8.50%	13.87%	78.07%	80.41%	59.54%	82.93%	20.9%
	5	5.84%	5.33%	9.87%	83.91%	85.74%	69.41%	81.07%	16.3%
	11	8.38%	7.59%	14.62%	92.29%	93.33%	84.03%	80.46%	9.3%
1	12	2.53%	2.30%	4.42%	94.82%	95.62%	88.45%	80.45%	7.2%
3	10	3.10%	2.74%	5.91%	97.92%	98.36%	94.37%	78.62%	4.0%
4	13	2.08%	1.64%	5.63%	100.00%	100.00%	100.00%	69.75%	0.0%
			100.00%	100.00%					27.9%



Gráfica 3.4. Gráfica de KS



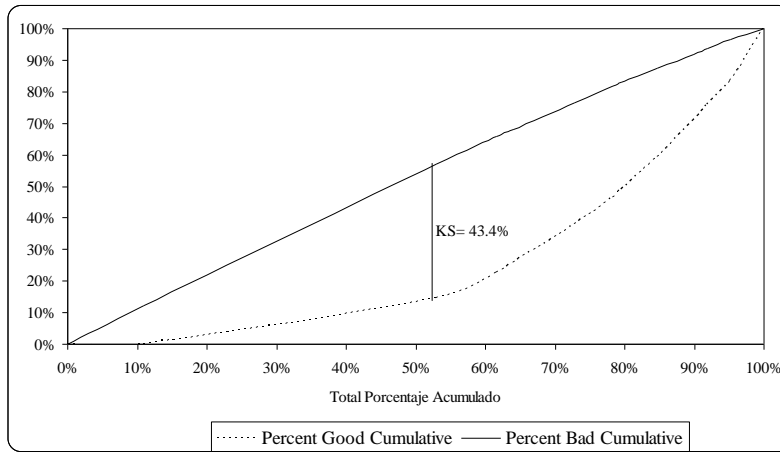
v. Población de desarrollo con 4 pagos vencidos

Cuadro 3.6. Distribución de la población por grupo de riesgo y nodo

										Valor de Información:	0.79
										K-S:	43.4%
										GINI:	0.54
Grupo de Riesgo	Nodo	Total %	Malos %	Buenos %	Acum % Total	Acum % Malos	Acum % Buenos	Malos por interv %	Difer %		
1	1	10.96%	12.16%	0.00%	10.96%	12.16%	0.00%	100.00%	12.2%		
	2	41.89%	44.89%	14.45%	52.85%	57.05%	14.45%	96.60%	42.6%		
2	9	8.74%	8.82%	8.01%	61.58%	65.87%	22.46%	90.95%	43.4%		
1	12	2.33%	2.25%	3.13%	63.92%	68.11%	25.59%	86.76%	42.5%		
	8	8.87%	8.50%	12.30%	72.79%	76.61%	37.89%	86.31%	38.7%		
	11	5.47%	5.15%	8.43%	78.26%	81.76%	46.31%	84.80%	35.4%		
2	6	1.38%	1.26%	2.44%	79.64%	83.02%	48.75%	82.51%	34.3%		
	10	5.08%	4.48%	10.61%	84.72%	87.50%	59.36%	79.40%	28.1%		
3	7	5.18%	4.48%	11.57%	89.90%	91.97%	70.94%	77.94%	21.0%		
	4	2.07%	1.76%	4.84%	91.97%	93.74%	75.78%	76.89%	18.0%		
	5	3.20%	2.71%	7.66%	95.16%	96.45%	83.44%	76.35%	13.0%		
4	3	4.84%	3.55%	16.56%	100.00%	100.00%	100.00%	66.21%	0.0%		
			100.00%	100.00%					43.4%		



Gráfica 3.5. Gráfica de KS



De acuerdo al tamaño de la población se obtuvieron árboles entre 10 y 24 nodos, donde algunos de ellos, tienen menos de 5% de la población, debido a que su tasa de malos fue significativamente mayor y representan un riesgo mayor, sin embargo este criterio, dependerá del desarrollador. El cálculo del valor de la información, K-S y Gini, fueron de acuerdo a lo descrito en el capítulo II.

3.9 Segmentos de riesgo

Los niveles de riesgo mencionados, fueron definidos por la agrupación de nodos finales de los árboles de decisión resultantes del proceso de segmentación. Los criterios de agrupación fueron:

- i. Tasas de malos: los nodos finales con tasas de malos similares constituyen un grupo de riesgo.
- ii. Volúmenes: cada grupo de riesgo debe presentar un volumen significativo de tal manera que se justifique aplicar acciones de cobranza específicas.

Los segmentos utilizados fueron definidos en la sección anterior, y están determinados por un número al que se le asoció una etiqueta de riesgo, que a continuación se describen:

- 1. Riesgo Crítico (RC)
- 2. Riesgo Muy Alto (RMA)
- 3. Riesgo Alto (RA)
- 4. Riesgo Medio (RM)
- 5. Riesgo Bajo (RB)



Por otro lado, se tienen los segmentos especiales asociados a las necesidades del negocio, para este caso se tomaron el balance bajo (BB) y balance alto (BA).

Con base a la definición descrita, esta indica que toda aquella cuenta que cumpla con las características de cierto nodo, pertenece a alguna de las agrupaciones antes señaladas, y tendrá un tratamiento diferenciado respecto a los otros segmentos. A continuación se presentan las agrupaciones por pago vencido, información que se extrajo de los cuadros descritos en la sección anterior:

Cuadro 3.7. Segmentos de riesgo para 0 pagos vencidos

Segmento de Riesgo	Nodo final	Tasa de Malos	% total	Etiqueta
1	3	7.59%	15.00%	RC
	4	7.21%		
	2	6.48%		
	16	6.04%		
	9	5.98%		
2	15	5.92%	8.00%	RMA
	1	5.87%		
	8	5.43%		
3	14	5.25%	24.00%	RA
	5	5.13%		
	13	4.54%		
	7	4.47%		
	6	3.96%		
4	11	3.69%	21.00%	RM
	20	3.37%		
	12	3.15%		
	10	2.35%		
5	18	2.12%	32.00%	RB
	19	1.69%		
	17	1.34%		
	21	1.15%		



Cuadro 3.8. Segmentos de riesgo para 1 pago vencido

Segmento de Riesgo	Nodo final	Tasa de Malos	% total	Etiqueta
1	1	43.01%	23%	RC
	4	39.10%		
	7	37.98%		
	3	37.34%		
	8	35.72%		
	11	34.88%		
2	6	34.70%	37%	RMA
	2	33.66%		
	17	32.86%		
	23	32.56%		
	22	31.32%		
	5	29.83%		
3	18	27.64%	12%	RA
	14	24.76%		
	13	24.27%		
	10	23.42%		
4	9	20.46%	20%	RM
	20	19.02%		
	12	17.99%		
	21	15.86%		
5	19	12.54%	7%	RB
	16	11.96%		
	15	9.61%		
	24	9.33%		

Cuadro 3.9. Segmentos de riesgo para 2 pagos vencidos

Segmento de Riesgo	Nodo final	Tasa de Malos	% total	Etiqueta
1	1	83.95%	15%	RC
2	2	82.37%	18%	RMA
1	6	81.81%	15%	RC
2	5	81.17%	18%	RMA
3	4	79.66%	42%	RA
	3	76.58%		
	8	73.96%		
4	7	71.90%	25%	RM
	9	70.07%		
	10	67.98%		



Cuadro 3.10. Segmentos de riesgo para 3 pagos vencidos

Segmento de Riesgo	Nodo final	Tasa de Malos	% total	Etiqueta
1	14	88.80%	16%	RC
	4	87.30%		
	3	87.00%		
	2	85.32%		
2	9	85.16%	55%	RMA
	7	84.32%		
	8	83.59%		
	1	81.04%		
3	6	80.40%	26%	RA
	5	79.34%		
	11	78.72%		
1	12	76.82%	16%	RC
3	10	75.05%	26%	RA
4	13	69.75%	2%	RM

Cuadro 3.11. Segmentos de riesgo para 4 o más pagos vencidos

Segmento de Riesgo	Nodo final	Tasa de Malos	% total	Etiqueta
1	1	90.13%	55%	RC
	2	88.92%		
2	9	82.09%	24%	RMA
1	12	80.08%	55%	RC
	8	79.65%		
2	11	77.47%	24%	RMA
	6	75.63%		
	10	75.16%		
3	7	73.75%	16%	RA
	4	71.60%		
	5	70.24%		
	3	66.21%		
4	3	66.21%	5%	RM

Cabe señalar, que además de los segmentos de riesgo descritos anteriormente, cada modelo de segmentación debe tener definido un nodo cuyo tratamiento será “No acción de cobro”, cuando el saldo pendiente sea menor a cierto monto poco significativo y que el costo operativo es mucho mayor que lo que se pudiera recuperar. Por otro lado, también se determinó saldo bajo para moras de 0 a 2 pagos vencidos, cuando el saldo pendiente este entre ciertos montos, que han resultado históricamente mayor eficiencia en la recuperación, si son enviadas a una agencia externa de cobranza.

En la siguiente sección se describirá con mayor detalle las acciones que se pueden seguir.



3.10 Estrategia de Cobranza

Como se mencionó en el primer capítulo, derivado de la implementación de un nuevo sistema para administrar las cuentas en cobranza, fue necesario desarrollar estrategias que reten a las actualmente existentes para buscar mejores resultados, sin poner en riesgo al portafolio.

Las estrategias de cobranza implementadas en sistemas de decisión, pueden asignar de forma aleatoria las cuentas a los ejecutivos en centros de cobro. Estas a su vez, pueden tener asociado un periodo de tiempo para comenzar el cobro, regularmente días, para que un cliente pague después de su fecha de vencimiento. Esto se le conoce como “*Holdout*”.

Este tipo de estrategias se utilizan para aquellos clientes cuyo riesgo es bajo, pero que no pagaron en la fecha límite, ya que muchas de las razones pudieran ser, un olvido, y este pagará en los siguientes días, para aquellas cuentas que no paguen en este periodo, el día siguiente se realiza una acción de cobro.

Dentro de las mejores prácticas en la industria, que aplican equipos que se dedican a diseñar estrategias de cobranza en Latinoamérica, recomienda aplicar 7 días de espera antes de realizar una acción de cobro, con el fin de reducir costos en la operación y evitar llamar a clientes que regularmente pagan, además no generar un descontento por parte de los clientes. Regularmente, se aplica para aquellas carteras al corriente, que no pagaron en su fecha límite de pago (como es el caso del modelo de segmentación de 0 pagos vencidos) o en aquellas cuentas en 1 pago vencido.

Las estrategias campeonas y retadoras son definidas de acuerdo al tamaño de la población a la que serán asignadas. A continuación se enlistan una serie de pasos que se sugiere dentro de éste trabajo, para acciones de cobro:

- a) Con base en el modelo de segmentación, se deberán determinar los segmentos de riesgo y obtener un estimado de cuentas por segmento y pago vencido. Este volumen de cuentas puede ser utilizado como una estimación para calcular el número de ejecutivos, que se requieren para gestionar las cuentas. Este análisis se le conoce como el “Plan de capacidad”, el cual incluye la estimación de cuentas, ejecutivos y costos asociados.
- b) Para determinar si la población puede ser dividida en 50% para aplicar la estrategia campeona, y 50% para la retadora, se realiza un análisis para que estas poblaciones sean aleatorias y significativas, así como homogéneas. En el [Apéndice D](#), se detalla la prueba de homogeneidad.



Es importante mencionar, que además de utilizar estas herramientas estadísticas, la determinación de estrategias campeonas y retadoras están basada en los comportamientos históricos y experiencia en la cobranza, con base en mediciones mensuales de por lo menos 6 meses. Con estas mediciones, es posible determinar qué estrategias pueden ser campeonas y cuáles retadoras, así como el volumen se pudiera incluirse en éstas.

- c) Una vez que la población es dividida en estrategias campeonas y retadoras, se determina qué segmentos de riesgo serán gestionados y que tipo de acción se realizará. Comúnmente, se realizan gestiones telefónicas para los clientes en baja mora, es decir, aquellas cuentas cuyo pago vencido es 0, 1 y 2. Gestiones con visitas físicas a su domicilio, son aplicadas para moras altas, 3 y 4 o más pagos vencidos. Ya determinados los segmentos que serán gestionados, se ajustan los costos asociados, los volúmenes de cuentas a gestionar y los ejecutivos requeridos, con el fin de terminar el costo beneficio por tipo de estrategia. Dentro de este análisis, se verá reflejado el impacto que tienen los modelos de segmentación, al clasificar las cuentas por riesgo, los esfuerzos en la cobranza se vuelven más específicos y es posible reducir el costo operacional.

A continuación, se plantean las estrategias para realizar acciones de cobro con base en los modelos de segmentación que se desarrollaron en este trabajo.

Se propone dividir la población en 50% para la estrategia campeona y 50% para la estrategia retadora, esto con el fin de determinar que acción de cobro pudiera ser mejor, estando en condiciones de grupos de riesgo y tamaño de población iguales, debido a que no se tiene certeza alguna que alguna de las estrategias es mejor.

Cuadro 3.12. Estrategia de cobranza para 0 pagos vencidos

Tipo de Segmento	Segmento de Riesgo	Campeona (50% población)					Retadora (50% población)				
		Llamada	Externo	Holdout	No Acción	Mensajes	Llamada	Externo	Holdout	No acción	Mensajes
Riesgo	Riesgo Crítico	√				■	√				■
	Riesgo Muy Alto	√				■	√				■
	Riesgo Alto	√				■	√				■
	Riesgo Medio	√							√		■
	Riesgo Bajo	√							√		■
No Riesgo	Bajo Balance				√				√		
Especial	Empleados VIP				√				√		
					√				√		



El cuadro anterior muestra que se determinaron dos estrategias campeona y retadora, cada una contiene 50% de la población de cero pagos vencidos, cuya característica se mencionó anteriormente. La estrategia retadora será “no llamar” a clientes de “Riesgo medio y bajo”, únicamente se le enviarán mensajes, por otro lado a “Riesgos Crítico, Muy alto y Alto”, se les hará una llamada telefónica, además de enviar una carta vía correo convencional.

Cuadro 3.13. Estrategia de cobranza para 1 pago vencido

Tipo de Segmento	Segmento de Riesgo	Campeona			
		Llamada	Externo	Holdout	Mensajes
Riesgo	Riesgo Crítico	√			■
	Riesgo Muy Alto	√			■
	Riesgo Alto	√			■
	Riesgo Medio	√		7 días	
	Riesgo Bajo	√		7 días	
No Riesgo	Bajo Balance		√	7 días	
	FPD	√			■
Especial	Empleados	√			
	VIP	√		7 días	

Para un pago vencido, se determinó hacer una sólo estrategia campeona, en donde para los “Riesgos Medio y Bajo” se aplicara un periodo de 7 días de no acción de cobro, en caso de no realizarse el pago durante este periodo, se asignará la cuenta a cobranza. Bajo balance se enviará a agencias externas, pero solo aquellas que no paguen durante el periodo de 7 días. Al igual que la estrategia de cero pago vencido, riesgo crítico, muy alto, alto y FPD se les enviará una carta vía correo convencional. La estrategia completa tiene como objetivo asignar todas las cuentas pero en diferentes periodos de tiempo.

Cuadro 3.14. Estrategia de cobranza para 2 pagos vencidos

Tipo de Segmento	Segmento de Riesgo	Campeona			
		Llamada	Externo	Holdout	Mensajes
Riesgo	Riesgo Crítico	√			■ ■
	Riesgo Muy Alto	√			■ ■
	Riesgo Alto	√			■ ■
	Riesgo Medio	√			■ ■
	Riesgo Bajo	√			■ ■
No Riesgo	Bajo Balance		√		■ ■
Especial	Empleados	√			■ ■
	VIP	√			■ ■



Como se puede observar en el cuadro anterior, todas cuentas en cualquier nivel de riesgo son asignadas y se les realiza una llamada telefónica, además de enviar una carta vía correo convencional y un mensaje a celular, en caso de contar con ese dato.

Cuadro 3.15. Estrategia de cobranza para 3 y 4 o más pagos vencidos

Tipo de Segmento	Segmento de Riesgo	Campeona (50% población)				Retadora (50% población)			
		Llamada	Físico	Externo	Mensajes	Llamada	Físico	Externo	Mensaje
Riesgo	Riesgo Crítico		√		■ ■		√		■ ■
	Riesgo Muy Alto		√		■ ■		√		■ ■
	Riesgo Alto		√		■ ■		√		■ ■
	Riesgo Medio	√			■ ■		√		■ ■
	Riesgo Bajo	√			■ ■		√		■ ■
No Riesgo	Bajo Balance			√	■ ■	√		√	■ ■
Especial	Empleados	√			■ ■		√		■ ■
	VIP	√			■ ■		√		■ ■

Para el caso de 3 y 4 o más pagos vencidos, se aplicará la estrategia anterior con el fin de incrementar las cuentas asignadas por ejecutivo físico, ya que éste resulta mucho más eficiente que un ejecutivo en un centro telefónico.

La estrategia está determinada por una campeona con 50% de la población, y retadora con el resto, los segmentos de riesgo que se pondrán a competir serán el “medio” y “bajo”, y de “no riesgo”, “Empleados” y “VIP. En la estrategia campeona se les realizará una llamada telefónica y en la retadora se les realizará una visita física. Para el caso de “balance bajo”, en la estrategia campeona se enviará a agencias externas y en la retadora se realizara una llamada.

3.11 Validación fuera de la muestra

Como parte del desarrollo del modelo de segmentación, se lleva a cabo la validación en una muestra diferente para observar el desempeño de las segmentaciones mediante el uso de información reciente y posterior a la ventana de desempeño.

Para el caso del modelo de segmentación de 0 pagos vencidos, se tomó como mayo 2011 con 6 meses de desempeño y para el resto de las cubetas la muestra es agosto y noviembre del 2011, con 6 meses de desempeño.



3.11.1 Estabilidad de la población

Parte de esta validación es verificar si existen cambios en la población de acuerdo con la segmentación de los modelos, ésto se realiza verificando cambios nodo a nodo. Los cambios que pudieran existir en la población se cuantifican con base en lo explicado en el capítulo II:

Cuadro 3.16 Resultados del índice de estabilidad de la población

1. Si $IE = 0$	entonces las distribuciones son iguales
2. Si $0 < IE \leq 0.10$	entonces son similares o la diferencia es muy pequeña
3. Si $0.10 < IE \leq 0.25$	entonces existen algunos pequeños cambios entre dichas distribuciones por lo tanto es necesario realizar un análisis de características
4. Si $IE > 0.25$	entonces existe un cambio significativo entre las distribuciones

Cuadro 3.17 Resultados del índice de estabilidad de la población

Población	Estabilidad	Indice
0 pagos vencidos	Si es estable	0.03
1 pago vencido	Si es estable	0.01
2 pagos vencidos	Si es estable	0.01
3 pagos vencidos	Si es estable	0.01
4 o más pagos vencidos	Si es estable	0.01

De acuerdo con los resultados observados, la población con la que se desarrollaron los modelos de segmentación muestra una distribución similar a lo largo de todos los nodos.



3.12 Indicadores de la eficiencia del modelo

La siguiente parte de la validación, tiene como objetivo comparar los indicadores estadísticos que miden la eficiencia de los modelos de segmentación. Estos indicadores son obtenidos en la muestra total (desarrollo más validación), desarrollo y validación, y son comparados con los índices obtenidos con las muestras recientes, para cada pago vencido.

i. Modelo de segmentación para la población con cero pagos vencidos

Tipo de muestra	Meses de desarrollo	Ventana de desempeño	KS	Gini	VI
Total	Diciembre 09- Enero 2010	6 meses	32%	43%	0.71
Desarrollo	Diciembre 09- Enero 2010	6 meses	32%	43%	0.71
Validación	Diciembre 09- Enero 2010	6 meses	32%	43%	0.71
Fuera de la muestra	Mayo 2011	6 meses	37%	49%	0.93

ii. Modelo de segmentación para la población con 1 pago vencido

Tipo de muestra	Meses de desarrollo	Ventana de desempeño	KS	Gini	VI
Total	Mayo y Junio 2010	6 meses	30%	45%	0.46
Desarrollo	Mayo y Junio 2010	6 meses	31%	46%	0.46
Validación	Mayo y Junio 2010	6 meses	30%	45%	0.45
Fuera de la muestra	Agosto y Nov 2011	6 meses	29%	44%	0.38

iii. Modelo de segmentación para la población con 2 pagos vencidos

Tipo de muestra	Meses de desarrollo	Ventana de desempeño	KS	Gini	VI
Total	Mayo y Junio 2010	6 meses	24%	34%	0.23
Desarrollo	Mayo y Junio 2010	6 meses	24%	34%	0.23
Validación	Mayo y Junio 2010	6 meses	24%	34%	0.23
Fuera de la muestra	Agosto y Nov 2011	6 meses	23%	33%	0.20



viii. Modelo de segmentación para la población con 3 pagos vencidos

Tipo de muestra	Meses de desarrollo	Ventana de desempeño	KS	Gini	VI
Total	Mayo y Junio 2010	6 meses	27%	38%	0.27
Desarrollo	Mayo y Junio 2010	6 meses	28%	39%	0.30
Validación	Mayo y Junio 2010	6 meses	25%	36%	0.23
Fuera de la muestra	Agosto y Nov 2011	6 meses	28%	39%	0.30

ix. Modelo de segmentación para la población con 4 o más pagos vencidos

Tipo de muestra	Meses de desarrollo	Ventana de desempeño	KS	Gini	VI
Total	Mayo y Junio 2010	6 meses	44%	54%	0.80
Desarrollo	Mayo y Junio 2010	6 meses	43%	54%	0.79
Validación	Mayo y Junio 2010	6 meses	45%	55%	0.84
Fuera de la muestra	Agosto y Nov 2011	6 meses	39%	51%	0.64

Como se puede observar, los resultados de los indicadores de eficiencia muestran pequeñas diferencias respecto a lo obtenido en el desarrollo, en la práctica se considera que se obtuvieron resultados semejantes en muestras recientes si las diferencias porcentuales son 6 puntos porcentuales.

Esta validación, también nos permite concluir, que no existe un sobre ajuste de los modelos de segmentación, sobre la población para la que se desarrollaron los modelos, es decir, los modelos siguen clasificando las cuentas por riesgo de acuerdo con lo esperado.

3.13 Implementación

Se sugiere que se realice la implementación de los modelos de segmentación, sobre una herramienta de decisión, que permita determinar acciones a través de estos segmentos.

Por otro lado, las cuentas deberán ser evaluadas por los árboles, cada vez que exista un movimiento financiero o cada fecha de corte.



3.14 Monitoreo del modelo de segmentación

Los modelos de segmentación deberán ser monitoreados por lo menos una vez al año, con el fin observar si el modelo sigue siendo eficiente, esto a través de obtener nuevamente los indicadores de eficiencia con una muestra mas reciente con la misma ventana de desempeño.

Por otro lado, las estrategias por segmento de riesgo, deberán ser monitoreadas de acuerdo con el pago vencido y el tratamiento definido, de forma mensual, con el fin de observar la eficiencia en los tratamientos, es decir, observar el porcentaje de cuentas que se mantienen al corriente, se mantienen en la misma mora o se deterioraron.



IV. Optimización de Costos

La clasificación por cuentas y la definición de tratamientos de cobranza, tiene como objetivo final minimizar los costos relacionados con la operación de cobranza.

La estrategia definida anteriormente, nos permite determinar qué cuentas por su nivel de riesgo de crédito, pueden ser excluidas de cobro, lo cuál implica reducir el número de cuentas para hacer llamadas o visitas físicas, y por lo tanto los gastos asociados, como por ejemplo, el sueldo del personal que realiza las llamadas, espacios dentro de las oficinas, parte tecnológica (computadores para revisar estados de cuenta) etc. Sin embargo, existe otros factores que pudieran ayudar a reducir los costos asociados a la operación en la cobranza, y es la distribución de cuentas a lo largo de las “horas trabajo”, es decir, el horario que una Institución crediticia determina para realizar llamadas para contactar a los clientes y solicitarles el pago de su deuda.

Como se mencionó en capítulos pasados, hacer llamadas telefónicas a clientes, son acciones que se utilizan en aquellas cuentas con morosidad baja, es decir, aquellos clientes con 0, 1 y 2 pagos vencidos, mientras que en morosidades altas (3 o más pagos vencidos), se utilizan acciones, que requieren visitas al domicilio del cliente. La mayor parte de la población se concentra en bajas moras, para la mayoría de las Instituciones de crédito, lo que significa, que cualquier cambio con el fin de reducir costos, pueden ser enfocado en esta población de clientes. Por lo tanto, esta sección se enfocara en la población de clientes con cero pagos vencidos, que requieren un recordatorio para realizar el pago de su deuda, con el fin de proponer una reducción de costos en la operación.

Cabe mencionar, que el objetivo de este capítulo es proponer una alternativa de distribución de cuentas de tarjetas de crédito, por horas de trabajo y segmento de riesgo para realizar llamadas telefónicas y que minimicen el costo asociado a la operación, utilizando algoritmos disponibles en la red.



4.1 Costos asociados a la operación de cobro

Las Instituciones de crédito requieren de sistemas sofisticados y una logística adecuada, para llevar a cabo un proceso de cobro. Dentro de los sistemas se encuentran módulos internos que administran los créditos, los cuales contienen información como pagos vencidos, monto de la deuda, utilización entre otros datos. Existen otros módulos que permiten que la operación dentro de la cobranza sea efectiva y las llamadas telefónicas a los clientes sean de forma aleatoria, estos módulos tienen un costo asociado, ya que regularmente son adquiridos y requieren actualizaciones de licencias. Este tipo de costos son difícilmente reducidos, ya que dependen de proveedores externos. Sin embargo, existen costos asociados a la planeación de la operación, por ejemplo, el plan de capacidad “*Capacity Plan*”, el cual incluye el personal requerido para hacer llamadas telefónicas y por otro lado, el horario laboral “*working days*”.

A continuación se muestran características de la planeación en la operación de cobranza:

Cuadro 4.1. Logística de la operación de cobranza¹¹

Horas laborales por día	15 horas	
Turnos por día	3	
	Horario 1 : 5 hrs.	Mañana
	Horario 2 : 4 hrs.	Tarde
	Horario 3 : 6 hrs.	Tarde- Noche
Promedio de días laborales por mes	22	

Regularmente las Instituciones o agencias de cobro tienen en promedio 15 horas al día, para realizar llamadas telefónicas. El horario comienza desde las 7 am y termina 10 pm, dependiendo de la Institución o agencia, y lo dividen en tres turnos, con el fin de emplear personal de medio tiempo y la eficiencia sea mayor.

De acuerdo a los datos proporcionados por la Institución financiera de estudio, la efectividad en el contacto con los clientes, depende del horario, por lo que se considera lo siguiente:

- a) Horario 1: La efectividad del contacto es medianamente superior al horario 2, por lo que se contacta al 30% de la población de clientes.
- b) Horario 2: La efectividad en el contacto es menor al promedio y se contacta al 20% de los clientes

¹¹ Fuente: Horas promedio por horario laboral establecidos en Institución financiera de estudio



- c) Horario 3: La efectividad en el contacto es mayor al promedio y se contacta al 50% de los clientes.

Estos eventos ocurren debido al perfil de la población que cuenta con tarjeta de crédito, es decir, son clientes que se encuentran empleados y su horario laboral en promedio es de 9am a 7pm, y no es posible contactarlos en sus domicilios dentro de este horario. Por lo tanto, se considera que el horario 1 y horario 3 son los más efectivos, y los costos asociados pueden variar. A continuación, se muestra una tabla con ejemplos de costos asociados al cobro de una cuenta (éste incluye personal e instalaciones), dependiendo del horario en que se realice la llamada, y el segmento de riesgo de crédito en que está clasificada la tarjeta de crédito.

Cuadro 4.2 Tablero de costo unitario por segmento y horario¹²

Costo por cuenta	Horario_1	Horario_2	Horario_3
Riesgo Crítico	\$10.0	\$7.0	\$15.0
Riesgo Muy Alto	\$9.0	\$6.3	\$13.5
Riesgo Alto	\$8.0	\$5.6	\$12.0
Riesgo Medio	\$6.0	\$4.2	\$9.0
Riesgo Bajo	\$4.0	\$2.8	\$6.0

Como se observa en el cuadro 4.2, los costos por cobrar una cuenta, se incrementan si el segmento es más riesgoso, así como el horario en que se realiza la llamada, considerando que el mejor horario es el tercer turno.

Estos datos pueden variar y son determinados por cada Institución y de acuerdo al mercado, por lo que pueden ser ajustados de acuerdo a las necesidades de cada empresa.

¹² Fuente: Costo unitario promedio provenientes de la Institución financiera de estudio



4.2 Planteamiento del problema a minimizar

El objetivo es minimizar un rubro de los costos que forman parte de la operación de cobranza, los cuales están asociados a los sueldos, gastos de instalaciones y tecnológicos, que se incurren de forma mensual para gestionar tarjetas de clientes en cero pagos vencidos, pero que requieren un recordatorio de pago (“*Early Collection*”), de acuerdo con la estrategia planteada en el capítulo anterior. Lo que se espera obtener es una distribución óptima de cuentas en los tres horarios, que permitan disminuir los costos, es decir nuestra variable de decisión es “cuentas”. A continuación se detallan las condiciones:

- a. En promedio las tarjetas de crédito que se deben de gestionar de forma mensual del segmento de 0 pagos vencidos de la población es 1, 661,250¹³ tarjetas, si no existiera ninguna clasificación por riesgo, es decir, si le hablaran a todas las cuentas.

Aplicando la estrategia propuesta en el capítulo III, para el tratamiento de tarjetas de crédito, es sólo llamarle al 50% del segmento Riesgo Medio y %50 del Riesgo bajo, capítulo 3, sección 3.5. En el cuadro 4.3 siguiente, se muestra la distribución por segmento de riesgo, del total de la población y aplicando la estrategia propuesta.

Cuadro 4.3. Cuentas 0 pagos vencidos por segmento de riesgo

Segmento de Riesgo	% Tarjetas	Total de cuentas	Total propuesto
Riesgo Crítico	15%	249,188	249,188
Riesgo Muy Alto	8%	132,900	132,900
Riesgo Alto	24%	398,700	398,700
Riesgo Medio	21%	348,863	174,431
Riesgo Bajo	32%	531,600	265,800
Total a gestionar		1,661,250	1,221,019

Como se puede observar en el cuadro anterior existe una reducción de cuentas en casi un 26%, aplicando el tratamiento propuesto. Esto representa una reducción de costos, pero que será sujeta al desempeño de la estrategia, es decir que la migración a 1 pago vencido de esta población se mantenga en los niveles esperados, de lo contrario se tendrán que volver a incluir estas cuentas a la gestión de cobranza.

Los porcentajes por segmento de riesgo mostrados en el anterior cuadro, se determinan de acuerdo al modelo de segmentación, es decir, con base a lo

¹³ Fuente: Cuentas promedio mensual de tarjetas de crédito que tiene que gestionar de la Institución financiera



obtenido en el desarrollo, se espera una distribución similar en poblaciones actuales de tarjetas de crédito (porcentajes por segmento de riesgo observados en la población de desarrollo se aplican en la población actual):

- b. Por otro lado, con base en los porcentajes de tarjetas de crédito que se deben de gestionar por turno, la distribución queda de la siguiente manera:

Cuadro 4.4. Cuentas 0 pagos vencidos por segmento de riesgo

Turno	% Tarjetas	Total de cuentas	Propuesta
Horario_1	30%	498,375	366,306
Horario_2	20%	332,250	244,204
Horario_3	50%	830,625	610,509
Total a gestionar		1,661,250	1,221,019

- c. De acuerdo a los puntos a y b, el problema a minimizar es el siguiente:

Función objetivo:

$$\min imizar \sum_i^H \sum_j^R \cos t[i, j] * cuentas[i, j]$$

Donde la variable de decisión, es el número de cuentas (*cuentas[i,j]*).

Donde *i* en *H* y *j* en *R* son los siguientes conjuntos

$H = \{Horario_1, Horario_2, Horario_3\}$

$R = \{RC, RMA, RA, RM, RB\}$

RC= Riesgo crítico

RMA= Riesgo muy alto

RA= Riesgo alto

RM= Riesgo medio

RB= Riesgo bajo

Cost [i,j] = Costos por cuenta asociado al *horario_i* del segmento de riesgo *j*, de acuerdo al cuadro 4.2



Cuentas $[i,j]$ = Número de cuentas en del *horario* $_i$ y segmento de riesgo j

Las restricciones del problema son las siguientes:

Sujeto a:

$$\begin{cases} \text{Horario}_1(\text{RC}) + \text{Horario}_1(\text{RMA}) + \text{Horario}_1(\text{RA}) + \text{Horario}_1(\text{RM}) + \text{Horario}_1(\text{RB}) = 366306 \\ \text{Horario}_2(\text{RC}) + \text{Horario}_2(\text{RMA}) + \text{Horario}_2(\text{RA}) + \text{Horario}_2(\text{RM}) + \text{Horario}_2(\text{RB}) = 244204 \\ \text{Horario}_3(\text{RC}) + \text{Horario}_3(\text{RMA}) + \text{Horario}_3(\text{RA}) + \text{Horario}_3(\text{RM}) + \text{Horario}_3(\text{RB}) = 610509 \end{cases}$$

Dentro del horario 1, se deben de gestionar el 30% de las cuentas (366,306 cuentas), el horario 2 el 20% de las cuentas (244,204 cuentas), y el horario 3, 50% de las cuentas (610,509), considerando que dentro de ese porcentaje de cuentas, pudieran cobrar cuentas de los diferentes segmentos de riesgo.

Por otro lado, es necesario que todas las cuentas por segmento de riesgo se gestionen, dentro de cualquiera de los tres horarios laborales:

$$\begin{cases} \text{Horario}_1(\text{RC}) + \text{Horario}_2(\text{RC}) + \text{Horario}_3(\text{RC}) = 249188 \\ \text{Horario}_1(\text{RMA}) + \text{Horario}_2(\text{RMA}) + \text{Horario}_3(\text{RMA}) = 132900 \\ \text{Horario}_1(\text{RA}) + \text{Horario}_2(\text{RA}) + \text{Horario}_3(\text{RA}) = 398700 \\ \text{Horario}_1(\text{RM}) + \text{Horario}_2(\text{RM}) + \text{Horario}_3(\text{RM}) = 174431 \\ \text{Horario}_1(\text{RB}) + \text{Horario}_2(\text{RB}) + \text{Horario}_3(\text{RB}) = 265800 \end{cases}$$

Con base en el cuadro 4.3, el total de cuentas que se tienen en el grupo de riesgo crítico es 249,188, riesgo muy alto es 132,900, y así sucesivamente.

4.3 Solución al problema de reducción de costos

Para obtener la solución del problema a minimizar, se tomó una alternativa de programación y paquete de optimización, planteados y explicados en mi trabajo de tesis de licenciatura “Solución de problemas de optimización a través del servidor de Internet NEOS (*Network-Enabled Optimization System*)” [13].

Para la programación del problema, se consideró el lenguaje de modelación “*AMPL (A modeling language for mathematical programming)*” [14], el cuál esta diseñado para ayudar al usuario a desarrollar y aplicar modelos de programación matemática. El lenguaje de “*AMPL*” entro hace unos años en el campo algebraico de lenguajes de programación matemática, se caracteriza por la semejanza de sus expresiones aritméticas a la notación algebraica y por la generalidad de su aplicación, éste es un lenguaje específicamente para problemas de optimización. La codificación es muy similar a la



forma en la que se describe un problema matemáticamente, lo permite pasar muy fácilmente de una descripción algebraica a “AMPL”.

Existe la opción de enviar el problema codificado en “AMPL” al “Servidor de Internet NEOS”, y obtener la solución del problema utilizando una gran variedad de “*solvers*” o obtener una versión libre para estudiante de “AMPL”, el cual utiliza una variedad de “*solvers*”, en particular “*Mínos 5.51*” [16], cuyo método es el “*Método Simplex*”, también disponible en el servidor “NEOS”. La versión estudiante de “AMPL” tiene una vigencia y no requiere Internet.

Para cualquiera de las alternativas seleccionadas es necesario crear dos archivos, el primero será la codificación del problema a minimizar y el segundo será los datos, a continuación se muestra ambos archivos:

Figura 4.1. Codificación en AMPL de la función objetivo y restricciones

```
set HOR;      # horarios
set RISK;     # segmentos de riesgo

param acct    {HOR} >= 0; # cuentas por horario que deben ser gestionadas
param acctasg {RISK}>= 0; # cuentas por segmento de riesgo asignadas a la cobranza

check: sum {i in HOR} acct[i] = sum {j in RISK} acctasg [j];

param cost {HOR,RISK} >=0; # costo por gestion por horario y segmento de riesgo
var cuentas {HOR,RISK} >=0; # cuentas que deben de ser gestionadas

minimize total_cost:
    sum {i in HOR, j in RISK} cost[i,j] * cuentas[i,j];

subject to Acct {i in HOR}:
    sum {j in RISK} cuentas[i,j]= acct[i];

subject to Acctasg {j in RISK}:
    sum {i in HOR} cuentas[i,j]= acctasg[j];
```

Este problema deberá ser guardado con la extensión “.mod”, para este caso se nombró “**problema.mod**”. A continuación se muestra la codificación de los datos, cuya extensión es “.dat” cuyo nombre es “**datos.dat**”:

Figura 4.2. Codificación en AMPL de los datos



```
param : HOR:      acct      := # cuentas por horario que puede ser gestionadas
      horario_1  366306
      horario_2  244204
      horario_3  610509;

param : RISK:      acctasg   := # cuentas por grupo de riesgo

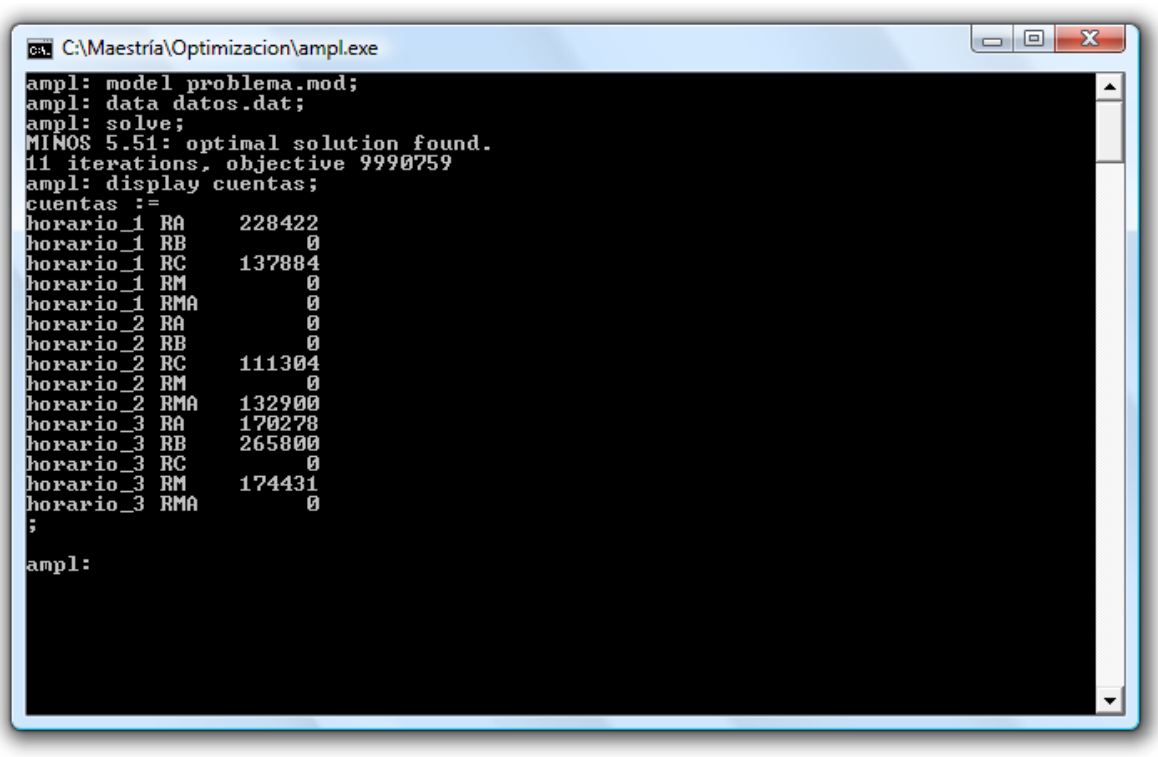
      RC        249188
      RMA       132900
      RA        398700
      RM        174431
      RB        265800;

param cost:

horario_1  RC    RMA   RA    RM    RB :=
horario_2  7     6     6     4     3
horario_3  15    14    12    9     6;
```

A continuación se muestra la solución del problema utilizando la versión de estudiante de “AMPL”:

Figura 4.3. Resultados utilizando versión estudiante AMPL





Como se muestra en la figura 4.3, posterior al “*prompt ampl:*” es necesario escribir las siguientes instrucciones:

ampl: model nombre del archivo.mod;	Con esta instrucción se llama al archivo que contiene la función objetivo y las restricciones
ampl: data nombre del archivo.dat;	Con esta instrucción se llama al archivo que contiene la datos de costos y número de cuentas por segmento y horario
ampl: solve;	Esta instrucción llama al “ <i>solver</i> ” que buscara la solución al problema
MINOS 5.51: optimal solution found 11 iterations, objective 9990759	Indica el “ <i>solver</i> ” que utilizó, si fue posible encontrar la solución y el valor de la función objetivo
ampl: display cuentas;	Despliega el valor asociado a las variables, tal como horario y segmento de riesgo

De acuerdo a los resultados, podemos determinar que la forma más óptima para distribuir las cuentas por horario y segmento de riesgo, que minimiza los costos es:

Cuadro 4.5 Cuentas 0 pagos vencidos por segmento de riesgo

Turno	Horario_1	Horario_2	Horario_3
Riesgo Crítico	137,884	111,304	0
Riesgo Muy Alto	0	132,900	170,278
Riesgo Alto	228,422	0	0
Riesgo Medio	0	0	174,431
Riesgo Bajo	0	0	265,800
Total a gestionar	366,306	244,204	610,509

El costo mensual total de acuerdo a esta distribución es: **\$ 9,990, 759 MxN.**

Si solo se considera el número de cuentas que debieron gestionarse por horario laboral, es decir, tomar 366,306 (30%) y que incluyan de todos los riesgos, 244,204 (20%) y que se abarquen todos los riesgos y los mismo para el horario 3, entonces el costo total sería **\$10, 954,790 MxN.** Lo que implica un ahorro de **\$964,031 MxN.**

Los costos en la gestión de cobranza pueden ser reducidos por dos caminos, el primero fue planteado en el capítulo III, donde se plantean estrategias campeonas y retadoras, que



reducen el número de cuentas en 27% (total mensual 1, 661,250 y con la nueva propuesta 1, 221,019 cuentas).

Si estas cuentas que fueron excluidas del segmento bajo y medio, se gestionarán, en el mismo horario que propone el “*solver*” (horario 3, donde una cuenta de Riesgo bajo cuesta \$6 pesos y de Riesgo medio cuesta \$9 pesos), el total sería **\$3, 164,681 MxN. Por lo tanto, la reducción en el costo de la operación de la cobranza, sería \$4, 128,712 MxN.**

Dentro del [Apéndice E](#) se muestra los resultados con dos diferentes “*solvers*”, con el fin de observar que el valor de la función objetivo es el óptimo utilizando cualquier otro algoritmo.

Estos resultados, se obtuvieron a través de “*NEOS Server*” utilizando “*OOQP with AMPL*” y “*AMPL/Gurobi*”, la diferencia en el resultado consistió en el número de iteraciones realizadas por estos “*solvers*”, para obtener el objetivo.

El optimizador “*OOQP*” es un paquete hecho en C++, basado en el método de mínimos cuadrados [17] y [18]. Para el caso de optimizar “*Gurobi*”, esta basado en el método *simplex* y de punto interior [19].



Conclusiones

En la actualidad el acceso a créditos bancarios es mucho más sencillo que hace 6 u 8 años. Los requisitos mínimos que deben de cubrir los solicitantes han sido mucho más relajados que años anteriores, particularmente en Tarjetas de Crédito, lo que ha impactado de manera significativa a las Instituciones de crédito, directamente su exposición al riesgo. Por lo tanto, las Instituciones de crédito, deberán definir políticas e implementar herramientas que permitan controlar el riesgo en la apertura de créditos. Dentro de la definición de políticas, es necesario que definan una sección enfocada a la recuperación de la cartera, tal como la implementación de estrategias de cobranza, con el objetivo de fortalecer las áreas críticas de cobranza e incrementar la productividad en la recuperación de cartera vencida o prevención de ésta. También se requieren acciones para identificar, optimizar y automatizar la administración de la cobranza, los cuales son algunos de los factores críticos que las Instituciones financieras enfrentan en sus procesos de negocio.

Para llevar a cabo una mejor labor de cobranza, es necesario segmentar la cartera asignada al cobro, de acuerdo a las características comunes de los clientes y/o las cuentas. Por ejemplo, puntajes obtenidos de modelos de comportamiento, que incluyen porcentaje de utilización de las tarjetas de créditos (saldo pendiente entre el límite de crédito), porcentaje de pago respecto al total de la deuda, pago vencido al momento de la deuda, número de veces que ha caído en mora y como parte sustancial e importante el incorporar información proveniente de buró de crédito. Este último factor es de suma relevancia, ya que el incorporar información proveniente de la industria financiera, permite determinar el nivel de endeudamiento de los clientes dentro y fuera de una Institución de crédito.

Las estrategias de cobranza deberán ser diseñadas considerando mejorar las actuales, situando los recursos de manera adecuada al perfil de riesgo de cada segmento del portafolio, es decir, optimizar la recuperación en relación a los costos de cada tipo de tratamiento, así como otorgar un mejor servicio a los clientes de Tarjetas de Crédito, contactándolos en el momento correcto de acuerdo a su morosidad y probabilidad de incumplimiento. Tal como se hizo en la estrategia enfocada a clientes en cero pagos vencidos, se mantuvo los niveles esperados de morosidad, es decir se mantuvo el porcentaje de migración a los siguientes niveles de morosidad.



Un punto importante para considerar construir o ajustar una estrategia de cobranza, es el cambio sustancial en las políticas de ofertas de créditos, ya que esto implica un portafolio vigente con distinto perfil de riesgo.

El desarrollo de modelos de segmentación deberá incluir un análisis de riesgo, ya que este proporciona una base para determinar los grupos que deben de tener un seguimiento más cercano con acciones diferenciadas. Existe una variedad de herramientas que permiten llevar a cabo este tipo de desarrollos, cuya complejidad en el uso, es considerablemente bajo y pueden ser utilizados sin previo entrenamiento, en particular aquellas que permiten construir árboles de decisión. Los árboles de decisión permiten tomar decisiones más complejas, puesto que pueden intervenir más variables de incertidumbre y factores que determinan acciones diferenciadas y que llevan a diseñar estrategias de decisión como las que se desarrollaron en este trabajo.

La aplicación de métodos estadísticos para medir la eficiencia de los modelos de segmentación, permite asegurar la robustez y la estabilidad de las estrategias, y por lo tanto la efectividad de acciones de cobro por nivel de riesgo. Por lo tanto, el obtener medidas como el índice de estabilidad, Kolmogorov-Smirnov, “*Gini*” y Valor de la información será fundamental para la justificación del buen desempeño de este tipo de modelos. Considerando que la información con la que se desarrolló, fue previamente analizada y cumple con los criterios de calidad en las variables que forman parte del modelo de segmentación, así como el nivel estadístico.

El obtener bases de datos para desarrollar modelos con niveles de calidad altos, tales como volumen bajo de vacíos en las variables, información confiable cien por ciento verificable, entre otros, dependerá de los sistemas implementados en las Instituciones financieras, así como de las políticas de originación. Por ejemplo, el capturar información de la recuperación de la cartera fuera de un sistema, es decir, un archivo de Excel, permite elaborar una base de datos, pero sin controles de calidad, con lo que se pudieran dar con mayor facilidad errores humanos, mientras que en un sistema de administración de cartera, existen una serie de controles que permiten minimizar de forma considerable los errores operativos y humanos, y por lo tanto, se podrían generar bases de datos mucho más confiables para desarrollos de modelos de segmentación.

Al poder cumplir con los criterios antes mencionados, fue posible desarrollar modelos de segmentación que permitieran la clasificación de las cuentas para optimizar las acciones de cobro en un sistema financiero. Las estrategias están actualmente productivas, y su desempeño es medido de acuerdo a porcentajes de migración por pago vencido, tipo de estrategia y nivel de riesgo.

Como parte del diseño de estrategias, es posible optimizar la asignación de las cuentas, esto aplica para cualquier producto crediticio, y así reducir costos en la operación de cobranza.

En el programa de estudios de la Maestría de Ingeniería, en la especialidad de Investigación de Operaciones, es posible aprender una diversidad de técnicas que permiten desarrollar este tipo de modelos de segmentación, tales como Simulación, Optimización, Teoría de Juegos, Estadística, entre otros.



Referencias bibliograficas

Albright, C., Winston, W., & Zappe, C. (2011,2009). *Data Analysis and Decision Making*. Mason, OH: South- Western, Cengage Learning.

Almuallim, H., Kaneda, S., & Akiba, Y. (2002). Development and Applications of decision tress. In C. Leondes, *Expert Systems: The technology of knowledge management and decision making for the 21st century* (p. 21). Orlando, Florida: Harcourt, Inc.

Bañuelos, S. (2012, Agosto 9). *T21mx*. Retrieved Septiembre 2012, 10, from <http://t21.com.mx/opinion/artes-cobrar/2012/08/08/cartas-que-cobran>

Barrera, P., & Ledesma, K. (2003). *Solución de problemas de optimización a través del servidor de Internet NEOS*. México, DF: Facultad de Ciencias, UNAM, Tesis de licenciatura.

Bratko, I. (1987). *Machine Learning: Proceedings of the Sixteenth International Conference (ICML '99)*. Bled, Slovenia,: Morgan Kaufmann Publishers.

Corporation, A. S. (2012). *Angoss Predict Perform*. Retrieved Agosto 28, 2012, from <http://www.angoss.com/files/docs/KnowledgeSTUDIO.pdf>

Discovery, W. I. (2002). *NEOS server*. Retrieved Marzo 3, 2013, from The NEOS (Network-Enabled Optimization System) Server: <http://www.neos-server.org/neos/>

Divanci, A. (2007, Noviembre 19 al 21). *Sistema de Gestión de Cobranzas Cash-pot*. Retrieved Agosto 28, 2012, from Cash-pot business software: <http://www.cash-pot.com.ar/ejemplo-de-estrategia-de-cobranza.aspx>

Fourer, R., Gay, D., & Kernigham, B. (1993). *AMPL A modeling language for mathematical programming*. Danvers, Massachusetts: Boyd & fraser publishing company.

Gertz, E. M., & Wright, S. (2004). OOQP User Guide. *Mathematics and Computer Science Division, Technical Memorandum No. 252* , 1-61.

Gertz, E. M., & Wright, S. (2003). Oriented Software for Quadratic Programming. *ACM Transactions on Mathematical Software* , 58-81.

Holmström, K., & M.M., E. (2008). *USER'S GUIDE FOR TOMLAB /MINOS*. Pullman,WA: Tomlab Optimization Inc.



Hunt, E. B., Marin, J., & Stone, P. J. (1966). *Experiments in induction*. New York, NY: Academic Press.

Inc., G. O. (2013). *Gurobi Optimization*. Retrieved Marzo 2013, 3, from Quick Start Guide version 5.5: <http://www.gurobi.com/documentation/5.1/quick-start-guide/>

Keyworth, M. (2012). *Experian A world of insight*. Retrieved Septiembre 2012, 6, from <http://www.experian.ie/assets/decision-analytics/white-papers/experian>

Lavrac, N. (1987). *Machine Learning for Data Mining in Medice*. Ljubljana, Slovenia: Subseries of Lecture Notes in Compute Sciencie.

México, B. d. (2013, Febrero). *Banco de México*. Retrieved Marzo 30, 2013, from Banco de

México publicaciones: <http://www.banxico.org.mx/sistema-financiero/publicaciones/reporte-de-tasas-de-interes-efectivas-de-tarjetas-%7B7D00598A-FCB0-23B9-79E4-597324729F0F%7D.pdf>

México, B. d. (2012). *Indicadores básicos de tarjetas de crédito*. México, DF: Banco de México.

Montgomery, D., & Runger, G. (2011). *Applied Statistic and Probability for Engeniers*. Danvers, MA: John Wiley, Inc.

Pang-Ning, T. (2007). *Introduction To Data Mining*. Boston: Pearson Addison Wesley.

Pyle, D. (1999). *Data Preparation for Data Mining*. San Francisco, California: Morgan Kaufmann Publisher, Inc.

Quinlan, J. R. (1975). *C4.5: programs for machine learning*. Sidney: Morgan Kaufmann.

Reincke, K. (2006-2009). *myPmps-0.8.7.1* . Retrieved Agosto 28, 2012, from <http://www.mypmps.net/en/mypmps/knowledgeareas/risk/qualitative-risk-analysis.html>

Ryszard Michalski, J. C.-C., & Mitchell, T. (1983). Machine Learning: A Historical and Methodological Analysis. *The AI Magazie* , 1-11.

Ville, B. d. (2006). *Decision Tree for Business Intelligence and Data Mining: Using SAS Enterprise Miner*. Cary, NC: SAS Institute Inc.



Apéndice A. Instrumentos de Análisis

A.1 Ejemplos de definiciones de desempeño

i. Población 0 pagos vencidos

Máximo pv en 6 meses	Portfolio Porcentaje	5+ pv en 9 meses Porcentaje	Castigo Porcentaje sobre 12 meses
5+ pv	1.7%	505.6%	84.6%
4 pv	2.1%	63.5%	34.5%
3 pv	2.5%	52.6%	0.6%
2 pv	4.5%	13.3%	0.5%
1 pv	11.9%	0.2%	0.1%
0 pv	77.4%	0.0%	0.0%
Total	100%	11.6%	2.2%

Máximo pv en 6 meses	Portfolio Porcentaje	4+ pv en 9 meses Porcentaje	Castigo Porcentaje sobre 12 meses
4+ pv	3.8%	100.0%	56.7%
3 pv	2.5%	69.1%	0.6%
2 pv	4.5%	38.2%	0.5%
1 pv	11.9%	10.1%	0.1%
0 pv	77.4%	0.0%	0.0%
Total	100%	8.4%	2.2%

ii. Población 1 pagos vencidos

Máximo pv en 6 meses	Portfolio Porcentaje	5+ pv en 9 meses Porcentaje	Castigo Porcentaje sobre 12 meses
5+ pv	21.2%	100.0%	83.6%
4 pv	6.2%	54.5%	34.9%
3 pv	11.8%	27.8%	0.8%
2 pv	14.6%	10.5%	0.0%
1 pv	23.5%	0.2%	0.0%
0 pv	23%	0.1%	0.0%
Total	100.0%	29.5%	20.0%

Máximo pv en 6 meses	Portfolio Porcentaje	4+ pv en 9 meses Porcentaje	Castigo Porcentaje sobre 12 meses
4+ pv	27.4%	100.0%	72.6%
3 pv	11.8%	37.5%	0.8%
2 pv	14.6%	17.0%	0.0%
1 pv	23.5%	3.7%	0.0%
0 pv	22.7%	0.2%	0.0%
Total	100%	35.2%	20.0%



iii. Población 2 pagos vencidos

Máximo pv en 6 meses	Portfolio	5+ pv en 9 meses	Castigo
	Porcentaje	Porcentaje	Porcentaje sobre 12 meses
5+ pv	53.7%	100.0%	83.2%
4 pv	10.8%	40.4%	23.3%
3 pv	17.2%	15.2%	0.6%
2 pv	6.4%	11.3%	0.0%
1 pv	6.0%	0.4%	0.0%
0 pv	6%	0.1%	0.0%
Total	100.0%	61.4%	47.3%

Máximo pv en 6 meses	Portfolio	4+ pv en 9 meses	Castigo
	Porcentaje	Porcentaje	Porcentaje sobre 12 meses
4+ pv	64.5%	100.0%	73.2%
3 pv	17.2%	22.2%	0.6%
2 pv	6.4%	17.9%	0.0%
1 pv	6.0%	6.6%	0.0%
0 pv	5.9%	0.3%	0.0%
Total	100%	69.9%	47.3%

iv. Población 3 pagos vencidos

Máximo pv en 6 meses	Portfolio	5+ pv en 9 meses
	Porcentaje	Porcentaje
5+ pv	74.1%	100.0%
4 pv	10.1%	21.5%
3 pv	4.5%	18.4%
2 pv	2.6%	10.7%
1 pv	3.1%	0.5%
0 pv	6%	0.2%
Total	100.0%	77.4%



Apéndice B. Instrumentos de Análisis

B.1 *Software* en línea

<http://www.kdnuggets.com/software/classification-decision-tree.html>

- [C4.5](#), the "classic" decision-tree tool, developed by J. R. Quinlan, (restricted distribution).
- [GAtree](#), genetic induction and visualization of decision trees (free and commercial versions available).
- [Mangrove](#), a tool for visualization of decision tree learning,
- [OC1](#), decision tree system continuous feature values; builds decision trees with linear combinations of attributes at each internal node; these trees then partition the space of examples with both oblique and axis-parallel hyperplanes.
- [PC4.5](#), a parallel version of C4.5 built with Persistent Linda (PLinda) system.
- [SMILES](#), advanced decision tree learner, with new splitting criteria, non-greedy search, extraction of different solutions, boosting, cost-sensitive learning, and more.
- [Random forests from Leo Breiman](#), a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest.
- [YaDT: Yet another Decision Tree builder](#), a new implementation of the C4.5 decision tree induction algorithm designed and implemented in standard C++ from scratch.



Apéndice C. Desarrollo del modelo de segmentación

C.1 Diseño de las muestras

De acuerdo a la descripción disponible en el programa SAS, la función “RANUNI” regresa un número, que es generado de una distribución uniforme entre los número 0 y 1, utilizando un generador multiplicativo de módulo primo con 2^{31} y multiplicador 397204094, ver biografía. Se puede utilizar este multiplicador para cambiar la longitud del intervalo y una constante agregada para mover el intervalo. Por ejemplo:

```
random_variate=a*ranuni(seed)+b;
```

Regresa un número que es generado por una distribución uniforme sobre un intervalo (b, a+b).

C.2 Análisis de correlación

De acuerdo a la definición que se encuentra Wikipedia, el coeficiente de correlación de Pearson es un índice que mide la relación lineal entre dos variables aleatorias cuantitativas. A diferencia de la covarianza, la correlación de Pearson es independiente de la escala de medida de las variables.

De manera menos formal, podemos definir el coeficiente de correlación de Pearson como un índice que puede utilizarse para medir el grado de relación de dos variables siempre y cuando ambas sean cuantitativas.

Para el caso de estudio de este trabajo, este coeficiente se obtiene entre las variables independientes y la variable dependiente de cuenta buena o mala.

En el caso de que se esté estudiando dos variables aleatorias x e y sobre una población estadística; el coeficiente de correlación de Pearson se simboliza con la letra, siendo la expresión que nos permite calcularlo:

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y},$$

Donde:



σ_{XY} es la covarianza de (X, Y)
 σ_X es la desviación típica de la variable X
 σ_Y es la desviación típica de la variable Y

De manera análoga podemos calcular este coeficiente sobre un estadístico muestral, denotado como r_{xy} a:

$$r_{xy} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{n s_x s_y} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

Interpretación

El valor del índice de correlación varía en el intervalo $[-1, 1]$:

Si $r = 1$, existe una correlación positiva perfecta. El índice indica una dependencia total entre las dos variables denominada relación directa: cuando una de ellas aumenta, la otra también lo hace en proporción constante.

Si $0 < r < 1$, existe una correlación positiva.

Si $r = 0$, no existe relación lineal. Pero esto no necesariamente implica que las variables son independientes: pueden existir todavía relaciones no lineales entre las dos variables.

Si $-1 < r < 0$, existe una correlación negativa.

Si $r = -1$, existe una correlación negativa perfecta. El índice indica una dependencia total entre las dos variables llamada relación inversa: cuando una de ellas aumenta, la otra disminuye en proporción constante.



Apéndice D. Resultados

D.1 Prueba de homogeneidad

El objetivo es comprobar si en k poblaciones (A_1, \dots, A_k) , es idéntica la distribución de probabilidad de una variable cualitativa con m posibles resultados (B_1, \dots, B_m) . Es decir, si se verifica que

$$P(B_j/A_1) = P(B_j/A_2) = \dots = P(B_j/A_k) = P(B_j) \quad \text{per a tot } j = 1, \dots, m$$

La diferencia respecto del caso de independencia es que los **totales de las filas** n_i son valores fijos y **no aleatorios** y corresponden al número de individuos seleccionados en la muestra que pertenecen a la población i . La estimación de los valores $P(B_j)$ bajo la hipótesis de homogeneidad se obtiene a partir de la tabla de contingencia a través de n_j/N .

Hipótesis nula de homogeneidad: dadas dos poblaciones cualesquiera i y i' se verifica que

$$H_0: p_{ij} = p_{i'j} \quad \text{para todo } j = 1, \dots, m$$

donde p_{ij} es la probabilidad del resultado j en la población i .

La hipótesis alternativa (alguna igualdad no es cierta) implica la no-homogeneidad de las poblaciones.

Los valores observados son n_{ij} . Los valores esperados bajo la hipótesis nula de homogeneidad se calculan de la manera siguiente:

$$e_{ij} = n_i \cdot P(B_j) = n_i \cdot (n_{.j}/N) = (n_i \cdot n_{.j})/N$$

Este resultado es idéntico al obtenido en el caso del contraste de independencia.



El estadístico de contraste se calcula de manera análoga:

$$\chi^2_{\text{exp}} = \sum_{i=1}^k \sum_{j=1}^m \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

La distribución asintótica bajo la hipótesis nula es una χ^2 con $(k - 1) \cdot (m - 1)$ grados de libertad. Los grados de libertad pueden entenderse de manera intuitiva entendiendo que el número de parámetros que se estiman son $(m - 1)$, ya que queda fijada la probabilidad de la última clase B_j una vez estimadas las restantes y al considerar que existen k restricciones de los valores esperados debido a que tenemos fijados los totales de cada población. Por tanto, aplicando la fórmula para los grados de libertad, se obtiene:

grados de libertad = número de clases – número de parámetros estimados – número de restricciones

$$\text{grados de libertad} = (k \cdot m) - (m - 1) - k = (k - 1) \cdot (m - 1)$$

El criterio de decisión es el mismo que en el contraste de independencia.

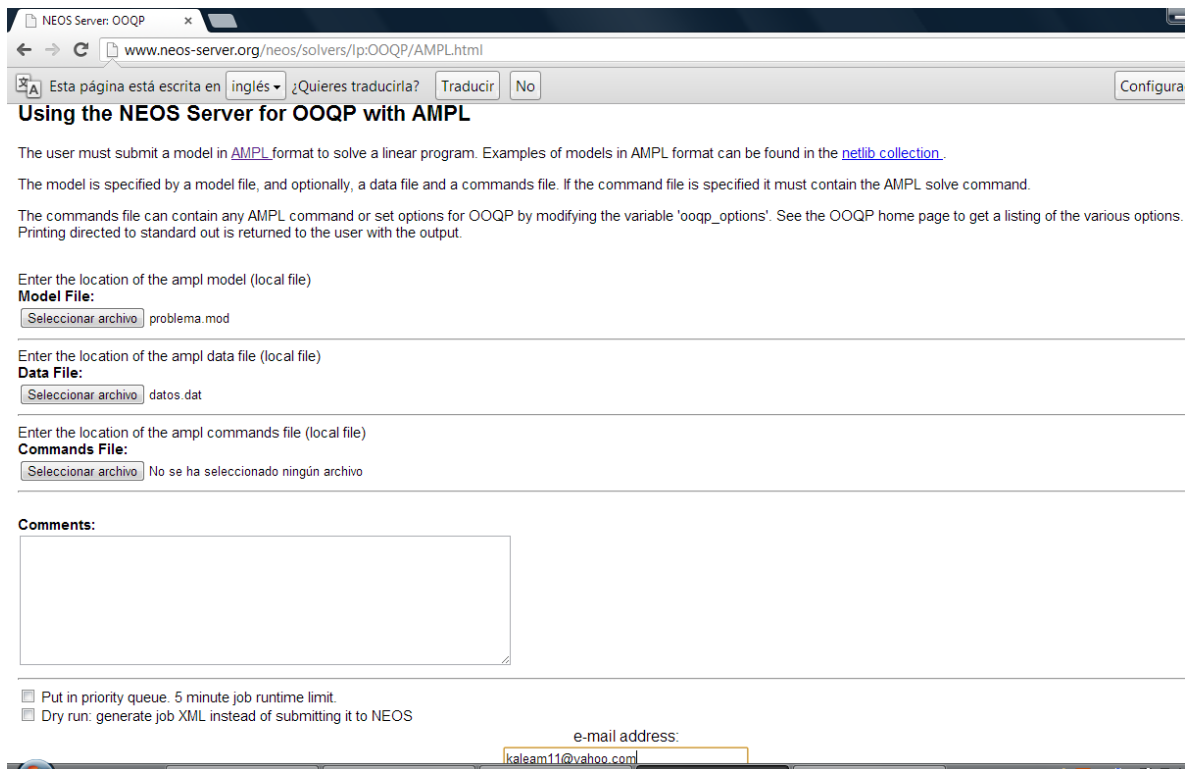
La condición de validez es que las frecuencias esperadas e_{ij} sean mayores que 5.

La fórmula que incluye Excel es **PRUEBA.T**.



Apéndice E. Optimización de Costos

E.1 Pantallas en donde se alimenta el solver con los archivos del modelo a resolver y los datos



The screenshot shows a web browser window with the URL `www.neos-server.org/neos/solvers/lp:OOQP/AMPL.html`. The page title is "Using the NEOS Server for OOQP with AMPL". The content includes instructions on how to submit a model in AMPL format, specifying the model file, data file, and commands file. There are three input fields for these files, each with a "Seleccionar archivo" button. The "Model File" field contains "problema.mod", the "Data File" field contains "datos.dat", and the "Commands File" field is empty. Below these fields is a "Comments:" section with a text area. At the bottom, there are two checkboxes: "Put in priority queue. 5 minute job runtime limit." and "Dry run: generate job XML instead of submitting it to NEOS". An "e-mail address:" field contains "kaleam11@yahoo.com".

NEOS Server: OOQP

www.neos-server.org/neos/solvers/lp:OOQP/AMPL.html

Esta página está escrita en inglés ¿Quieres traducirla? Traducir No Configura

Using the NEOS Server for OOQP with AMPL

The user must submit a model in [AMPL](#) format to solve a linear program. Examples of models in AMPL format can be found in the [netlib collection](#).

The model is specified by a model file, and optionally, a data file and a commands file. If the command file is specified it must contain the AMPL solve command.

The commands file can contain any AMPL command or set options for OOQP by modifying the variable 'ooqp_options'. See the OOQP home page to get a listing of the various options. Printing directed to standard out is returned to the user with the output.

Enter the location of the ampl model (local file)
Model File:
 problema.mod

Enter the location of the ampl data file (local file)
Data File:
 datos.dat

Enter the location of the ampl commands file (local file)
Commands File:
 No se ha seleccionado ningún archivo

Comments:

Put in priority queue. 5 minute job runtime limit.
 Dry run: generate job XML instead of submitting it to NEOS

e-mail address:



NEOS Job #512422 x
www.neos-server.org/neos/jobs/512422.html
Esta página está escrita en inglés ¿Quieres traducirla? Traducir No

```
NEOS Server Version 5.0
Job#      : 512422
Password  : vI2lLAc0
Solver    : lp:OOQP:AMPL
Start     : 2013-03-07 13:27:32
End       : 2013-03-07 13:27:32
Host      : neos-3.chtc.wisc.edu
```

Disclaimer:

This information is provided without any express or implied warranty. In particular, there is no warranty of any kind concerning the fitness of this information for any particular purpose.

```
Job 512422 sent to neos-3.chtc.wisc.edu
password: vI2lLAc0
----- Begin Solver Output -----
Executing /opt/neos/Drivers/ooqp-ampl/ooqp-driver.py at time: 2013-03-07 13:28:25.412689
File exists
You are using the solver ooqp-ampl.
Executing AMPL.
processing data.
processing commands.
```

```
15 variables, all linear
8 constraints, all linear; 30 nonzeros
    8 equality constraints
1 linear objective; 15 nonzeros.
```

```
OOQP: OOQP completed successfully. Objective value: 9990759.
```