

ESCUELA SUPERIOR POLITECNICA DEL LITORAL



**FACULTAD DE CIENCIAS NATURALES Y
MATEMÁTICAS
DEPARTAMENTO DE MATEMATICAS**

**TESIS DE GRADUACION
PREVIO A LA OBTENCIÓN DEL TÍTULO DE:
“MAGISTER EN CONTROL DE OPERACIONES Y
GESTIÓN LOGÍSTICA”**

TEMA

**Diseño de un modelo matemático para optimizar la gestión de
un sistema de cobranza en una empresa comercial.**

AUTOR:

HUGO RENAN RUIZ LOPEZ

Guayaquil – Ecuador

2013

AGRADECIMIENTOS

A mi Dios por haber permitido terminar este trabajo. A mi familia por el apoyo brindado en todo momento. A todas las personas que colaboraron con este trabajo, en especial a los ingenieros Carlos Cepeda y Guillermo Baquerizo, y a todos los profesores y compañeros que con quienes he compartido esta dura experiencia de estudiar.

DEDICATORIA

Dedico este trabajo a toda mi familia, en especial a mi Huguito Sebastián que es mi razón ser y vivir, a mi Esposa por su cariño y fuerzas en momentos de cansancio, a mis Padres por el apoyo brindado a lo largo de mi vida, sin ellos jamás hubiese podido conseguir lo que hasta ahora, a mi Hermanita y Sobrinas por su cariño, y una dedicatoria especial a la señora Graciela que desde el cielo me ayudó en esta difícil tarea.

TRIBUNAL DE GRADUACIÓN

MS.c. Carlos Cepeda de La Torre
DIRECTOR DE LA TESIS

ME.d Janet Valdiviezo
PRESIDENTE DEL TRIBUNAL

MS.c. Erwin Delgado Bravo
VOCAL DEL TRIBUNAL

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Graduación, me corresponden exclusivamente; el patrimonio intelectual del mismo, corresponde exclusivamente a la **Facultad de Ciencias Naturales y Matemáticas, Departamento de Matemáticas** de la Escuela Superior Politécnica del Litoral.

Ing. Hugo Ruíz López

Resumen

El objetivo de este trabajo fue el diseño de modelo matemático para optimizar la gestión de un sistema de cobranzas en una empresa comercial. Para ello se desarrollaron varios modelos bajo diferentes metodologías tradicionales como Árboles de Decisión, Regresión Logística y Análisis Discriminante, los cuales resultaron significativos y estables. Adicionalmente se desarrolló un modelo bajo la metodología de Programación Matemática.

Mediante la utilización de los métodos estadísticos tradicionales se pudo conocer que las variables que más discriminan a los clientes morosos de los no morosos son: la mora máxima en los últimos 6 meses y número de pagos en los últimos 6 meses. De acuerdo a un análisis comparativo de los tres métodos tradicionales, el modelo realizado bajo regresión logística clasificó mejor a los clientes morosos de los no morosos.

Se puede comprobar que el problema de planificación de cobranza puede ser modelado mediante un enfoque de programación matemática tal como se lo demuestra en el presente trabajo, el cual permite proyectar la recuperación de cartera y a planificar las estrategias de cobro, adicionalmente el modelo desarrollado bajo esta técnica asigna los clientes vencidos a los gestores de cobranzas, permitiendo de esta manera asignar metas de recuperación de cartera a cada uno de ellos.

Se recomienda profundizar en la automatización del modelo bajo programación matemática desarrollando un sistema computacional que permita a los encargados de la cobranza utilizar la herramienta para planificación.

TABLA DE CONTENIDOS

	Pag.
Índice de Tablas	I
Índice de Figuras	II
1. Objetivos y Generalidades	
1.1 Introducción.....	1
1.2 Riesgo Crediticio.....	1
1.3 Planteamiento del Problema.....	2
1.4 Justificación del Problema.....	3
1.5 Objetivo General.....	4
1.6 Objetivo Específico.....	4
2 Definición del Sistema	
2.1 Introducción.....	5
2.2 Conceptos Básicos de Cobranzas.....	5
2.3 Políticas del departamento.....	6
2.4 Procedimientos.....	7
3 Construcción del Modelo	
3.1 Introducción.....	9
3.2 Construcción del Modelo mediante Técnicas Estadísticas.....	9
3.2.1 Definición de Clientes Bueno y Cliente Malo.....	9
3.2.2 Muestras.....	10
3.2.3 Estadístico de Kolmogorov-Smirnov (K-S).....	10
3.2.4 Modelo Bajo Árboles de Decisión.....	11
3.2.5 Modelo Bajo Análisis Discriminante.....	14
3.2.6 Modelo Bajo Regresión Logística.....	20
3.3 Construcción del Modelo mediante Programación Matemática.....	29
3.3.1 Introducción.....	29
3.3.2 Programación Matemática.....	30
3.3.3 El problema de la mochila.....	32
3.3.4 Diseño del Modelo.....	33
4 Validación, Comparación y Aplicación del Modelo	
4.1 Introducción.....	38
4.2 Validación del Modelo.....	38
4.3 Comparación de Resultados.....	43

4.4 Aplicación del Modelo bajo Programación Matemática.....	46
5 Conclusiones y Recomendaciones	
5.1 Conclusiones.....	49
5.2 Recomendaciones.....	50
Bibliografía	51
Anexos	52

ÍNDICE DE TABLAS

	Pag.
Tabla 1.1 Distribución de Cartera.....	3
Tabla 2.1 Porcentajes de Provisión.....	6
Tabla 3.1 Segmentos Modelo Árbol de Decisión.....	13
Tabla 3.2 Variables Modelo Análisis Discriminante.....	17
Tabla 3.3 Coeficientes Modelo Análisis Discriminante.....	17
Tabla 3.4 Segmentos Modelo Análisis Discriminante.....	18
Tabla 3.5 Resumen Modelo Regresión Logística.....	25
Tabla 3.6 Prueba de Hosmer y Lemeshow.....	26
Tabla 3.7 Coeficientes Modelo Regresión Logística.....	27
Tabla 3.8 Segmentos Modelo Regresión Logística.....	28
Tabla 3.9 Promedio de Duración de Llamadas por edad.....	34
Tabla 3.10 Promedio de Duración de Llamadas Final.....	35
Tabla 4.1 Clasificación del Riesgo por Segmentos.....	44
Tabla 4.2 Datos Iniciales Modelo Programación Matemática.....	46
Tabla 4.3 Asignación de Clientes Modelo Programación Matemática.....	47
Tabla 4.4 Valor Vencido Promedio Asignado y no Asignado.....	47
Tabla 4.5 Recuperación de Cartera Esperada Modelo Programación Matemática...	48

ÍNDICE DE FIGURAS

	Pag.
Figura 3.1 K-S Modelo General.....	10
Figura 3.2 Ejemplo Árbol de Decisión.....	11
Figura 3.3 Tasa de Malos Árbol de Decisión.....	14
Figura 3.4 % de Población Árbol de Decisión.....	14
Figura 3.5 Tasa de Malos Análisis Discriminante.....	19
Figura 3.6 % de Población Análisis Discriminante.....	19
Figura 3.7 K-S Análisis Discriminante.....	19
Figura 3.8 Tasa de Malos Regresión Logística.....	28
Figura 3.9 % de Población Regresión Logística.....	28
Figura 3.10 K-S Regresión Logística.....	29
Figura 3.11 # Gestiones por Edad de Mora.....	35
Figura 4.1 Seguimiento Tasa de Malos Árbol de Decisión.....	39
Figura 4.2 Seguimiento % Población Árbol de Decisión.....	39
Figura 4.3 Seguimiento K-S Árbol de Decisión.....	40
Figura 4.4 Seguimiento Tasa de Malos Análisis Discriminante.....	40
Figura 4.5 Seguimiento % Población Análisis Discriminante.....	41
Figura 4.6 Seguimiento K-S Análisis Discriminante.....	41
Figura 4.7 Seguimiento Tasa de Malos Regresión Logística.....	42
Figura 4.8 Seguimiento % Población Regresión Logística.....	42
Figura 4.9 Seguimiento K-S Regresión Logística.....	43
Figura 4.10 Comparativo K-S por Modelos.....	44
Figura 4.11 % Coincidencia Análisis Discriminante vs Regresión Logística.....	45
Figura 4.12 % Coincidencia Análisis Discriminante vs Árbol de Decisión.....	45
Figura 4.13 % Coincidencia Regresión Logística vs Árbol de Decisión.....	46
Figura 4.14 % Clientes Asignados Modelo Programación Matemática.....	47

Capítulo 1

Objetivo y Generalidades

1.1.Introducción

La incertidumbre [1] que caracteriza el entorno económico actual al que deben enfrentarse las empresas en el desarrollo de su actividad ha convertido al riesgo en un elemento esencial que se debe tener muy en cuenta para la supervivencia y crecimiento de las mismas. Entendiéndose por riesgo de manera general, cualquier situación en la que no sabemos con exactitud qué ocurrirá en el futuro.

En el ámbito de la empresa se puede encontrar distintos tipos de riesgo, que se concentran en dos principales:

- Riesgo económico, que afecta a los resultados de explotación de la empresa.
- Riesgo Financiero, que resulta del diseño de la estructura del capital de la misma.

En el caso de las entidades financieras o de crédito, a los riesgos anteriores, se añade un conjunto de riesgos adicionales que podríamos calificar como propios de su actividad, y que se sintetizan en cinco grandes apartados: riesgo de mercados, riesgo de crédito, riesgo de liquidez, riesgo operacional y riesgo legal.

1.2.Riesgo Crediticio

De todos los riesgos mencionados anteriormente, el riesgo de crédito puede provocar mayores pérdidas potenciales, de ahí que sea el que mayor número de crisis financieras ha provocado y al que se le dedica mayor atención.

El riesgo está determinado principalmente por los siguientes factores:

- Factores Internos: que depende de la capacidad de los gerentes de las empresas para de alguna manera realizar estrategias que permitan de minimizar el riesgo.

- Factores Externos: que no depende de la persona que administra la empresa, por ejemplo la inflación, desastres climáticos, inestabilidad política, etc.

Un [2] aspecto de extraordinaria importancia en la gestión de riesgos crediticios, es el relativo al análisis y evaluación del riesgo, así como la clasificación de los clientes. Estos procesos de riesgos precisan de fuentes de información, tanto internas como externas.

En la gestión de riesgos puede adoptarse mecanismos preventivos o defensivos y mecanismos curativos.

- Procesos preventivos.- tiene el objetivo de evitar el ingreso de clientes con un alto riesgo de incumplimiento de pago, tienen que ver con el análisis previo de la solicitud de crédito, los sistemas de análisis financieros y Scoring que son sistemas masivos de calificación de clientes según reglas estadísticas.
- Procesos curativos.- Los procesos curativos tienen como finalidad la recuperación de la deuda, pueden incluirse las garantías del crédito, los sistemas de avisos de vencimientos, las acciones judiciales y los sistemas de gestión de cobro.

1.3.Planteamiento del Problema

El problema trata de determinar los factores de mayor influencia en el comportamiento de pagos de los clientes en una entidad del sector comercial, y que permita distinguir los clientes solventes o de menor riesgo de incumplimiento (cliente bueno), de los clientes morosos o de mayor riesgo de incumplimiento (cliente malo), es decir estamos con proceso curativo de crédito.

El objetivo se centrará en ayudar al departamento de cobranzas en implementar un sistema de gestión de cobro, que responda el pago rápido y oportuno de los créditos vencidos, que será medido de acuerdo al cumplimiento de metas de provisiones, índices de morosidad y cuentas dadas de baja.

La cartera de cobranza está distribuida de la siguiente forma:

Tabla 1.1 Distribución de Cartera

Edad	Rango	Porceso de Cobro	Evento
Vigente	0 días	Cobranza Preventiva	Contacto telefónico y Clientes si puede comprar
30 días	1 a 30 días vencido	Cobranza administrativa (interna)	Contacto telefónico y Clientes no puede comprar
60 días	31 a 60 días vencido	Cobranza administrativa (interna)	Contacto telefónico y Clientes no puede comprar
90 días	61 a 90 días vencido	Cobranza administrativa (interna)	Contacto telefónico y Clientes no puede comprar
120-180 días	91 a 180 días vencido	Cobranza Pre-juridica	Lo realizan las casas de cobro externas
210 en adelante	Desde 181 días de vencido	Cobranza judicial	se gestiona con agentes externos (abogados), si la cuenta permanece 1 año y no se ha recuperado, es dada de baja.

El modelo tratará de mejorar la estrategia de cobro de la cartera interna, es decir la cartera comprendida entre 30 y 90 días de mora, de tal manera de priorizar la gestión en función del riesgo y no de otras variables como actualmente se lo realiza.

1.4. Justificación del Problema

Actualmente la gestión de cobro a la cartera interna se la realiza en función de la edad de mora y del saldo deudor, a pesar que se tiene un modelo basado en árboles de decisión, este no ha satisfecho las necesidades actuales de la gestión, especialmente por la forma como se lo diseñó.

Además el gran crecimiento que ha tenido la empresa en los últimos años, por su política agresiva de captación de clientes, ha hecho que el número de clientes en esta etapa de cobro crezca en un 20%, y con los recursos que se tiene actualmente solo permite gestionar entre el 40% y 50% de los clientes que se encuentra en la etapa de cobro interna, por ello es importante realizar un modelo que ayude a identificar los clientes con mayor probabilidad de incumplimiento para priorizar su gestión de cobro.

1.5. Objetivo general

Proponer un modelo de gestión de cobro que ayude al departamento de cobranza a priorizar la gestión, maximizando la recuperación de cuentas vencidas en la etapa interna, optimizando el uso de los recursos con los que cuenta actualmente el departamento, para de esta manera maximizar los beneficios a la empresa.

1.6. Objetivo específico

En cuanto a los objetivos específicos se refiere, con el desarrollo de un modelo para la gestión de cobro de cartera interna se persigue lo siguiente:

- Priorizar la cobranza
- Disminuir el índice de provisiones
- Disminuir el índice de mora de la cartera
- Optimizar los recursos que dispone el departamento, ya sea estos gestores de cobranzas, envió de mensajes de texto, cartas y asignación de recaudadores de campo.

Capítulo 2

Definición del Sistema

2.1.Introducción

La labor de cobranza tendrá como objetivo general garantizar la calidad del portafolio de crédito que se medirá de acuerdo al cumplimiento de metas en índices de morosidad, provisiones y clientes dados de baja.

Su tarea será desarrollar estrategias de cobro para la compañía, así como proponer mejoras siempre con el objetivo de obtener el pago rápido y oportuno de los créditos vencidos.

Antes de empezar con la construcción del modelo matemático es importante conocer todos los detalles sobre el proceso en el cual se estará inmerso, identificando los factores que determinen su comportamiento. Considerando esto, a continuación se describe las características generales del departamento de cobranzas de una empresa comercial.

La cobranza estará organizada en función de los siguientes elementos:

- El proceso de cobro, consiste en determinar la frecuencia de cobro, así como las etapas (o instancias) a las cuales se debe recurrir en la medida en que las cuentas morosas se vayan deteriorando.
- Las técnicas de contacto, hacen referencia a los métodos que se deben utilizar en cada etapa para comunicarse con el cliente e indicar que está atrasado en sus pagos y motivarla que se ponga al día.
- La estrategia de cobranzas que se utilizará en cada etapa del proceso de cobro estará dada por la Gerencia y Jefatura del departamento.

2.2.Conceptos Básicos de Cobranzas

- Índice de Morosidad.- El índice de morosidad es la cartera de crédito vencida como proporción de la cartera total. Es uno de los indicadores más utilizados como medida de riesgo

$$\text{Índice de Morosidad} = \frac{\text{Cartera Vencida}}{\text{Cartera Total}}$$

Para clasificar los créditos como vencidos se requieren 30 días o más de vencimiento.

- Índice de Provisiones.- Las provisiones representan un porcentaje de la cartera que es calificada como incobrable. Este porcentaje es contabilizado como reserva para dicha cartera, y se incrementa de acuerdo a la edad de morosidad de la cartera, y se calcula sobre el saldo total de la cartera de cada cliente.

Cuando se presumen estafas se solicitará la provisión al 100% de la deuda, previa autorización de la Gerencia del Departamento. El porcentaje de provisión a aplicarse será:

Tabla 2.1 Porcentajes de Provisión

Edad de Morosidad	% Provisión
Vigente: 01-30 días	2 %
30 días	4 %
60 días	10 %
90 días	20 %
120 días	30 %
150 días	40 %
180 días	60 %
+ de 180 días	100 %

2.3.Políticas del departamento

- La distribución y gestión de cobranza deberá estar dada por la edad de mora de la cartera y gestionada en su totalidad en el ciclo de recuperación.
- Las cartas utilizadas deberán ser revisadas cada 3 meses por los jefes del departamento.
- Los clientes que se encuentre con edad vencida de 121 días, su cuenta será cerrada de forma automática y no podrá hacer uso de la tarjeta hasta la cancelación total de la deuda, previa una solicitud de reapertura de cuenta.
- Las cuentas que se hallen vencidas de 210 días en adelante, serán transferidas a cobranza judicial, para que se realice la acción de cobranzas

correspondiente por medio de agentes de cobro externos (abogados o casas de cobro).

- Las cuentas que presenten los siguientes motivos serán transferidas directamente a cobranza judicial.
 - Clientes desaparecidos y nadie se quiere hacer cargo de la deuda
 - Cuentas que tienen indicios de estafas
- No se podrá divulgar información sobre cliente alguno para fines personales
- Monitoreo de llamadas telefónicas a números convencionales y celulares, para asegurar su adecuado uso.
- Cuando el valor vencido de la cuenta es menor a \$ 4 dólares, no se reflejará la cuenta como vencida, por lo tanto no estará dentro de las cuentas asignadas para la gestión.

2.4.Procedimiento

La distribución por edad de mora de los gestores telefónicos puede variar de acuerdo a las necesidades que se presente y al volumen de cartera vencida en cada edad de mora.

Con el objetivo de que los gestores de cobranza interna puedan destinar el 100% de su tiempo a la labor de cobranza, las labores de apoyo operativo del área estarán a cargo de un personal diferente. El nivel de riesgo de la cartera dependerá el hacer más o menos frecuente los contactos a los clientes vencidos.

La organización de la función de cobro contará con tres elementos para desarrollar una buena gestión dentro del área.

- Especialización del cobro, para garantizar exclusivamente a la gestión de recuperación de cuentas morosas, y para crear grupo de gestores especialistas en cada edad de cobro.
- Función de soporte, para integrar e independizar de la gestión de cobro aquellas tareas operativas que compiten con el objetivo principal de la cobranza, como por ejemplo:
 - Manejo de reclamos
 - Documentación
 - Gestión de cheques devueltos

- Herramientas de control, que son todos los reportes o mediciones que se utilizan para hacer seguimiento diario a la gestión, para efectuar ajustes oportunos al proceso.

Capítulo 3

Construcción del Modelo

3.1. Introducción

Una vez presentados los riesgos a los que se enfrentan las empresas comerciales, generamos el modelo a partir de un conjunto de variables de comportamiento¹ o variables indirectas, es decir calculadas a partir del comportamiento crediticio. El objetivo principal a los que se enfrentan los departamentos de cobranza es controlar los niveles de morosidad y provisiones. El presente trabajo tratará de complementar dicha labor mediante la creación de modelos matemáticos, que tienen como objetivo medir el nivel de riesgo de los clientes morosos y predecir su comportamiento de pago en un periodo determinado.

3.2. Construcción del modelo mediante métodos estadísticos.

La construcción del modelo mediante métodos estadísticos consiste en el desarrollo de variables existentes históricamente del cliente.

Se generó un total de 199 variables indirectas², a partir de su comportamiento en compras, pagos y comportamiento crediticio durante el periodo de un año, las variables fueron construidas mediante procesamiento de comandos en SPSS.

3.2.1. Definición de Clientes Bueno y Malo.

A través de un análisis y evaluación de las necesidades del departamento de cobranzas, se decidió realizar un modelo que calcule la probabilidad de pago de un cliente moroso al periodo más corto posible, y este periodo fue de una semana, por lo tanto la definición de bueno es aquel cliente que al inicio de la semana se encuentra en mora y que al terminar la semana realizó un pago, caso contrario se lo considera como cliente malo.

¹“Comportamiento” es la conducta de pago de clientes.

² Variables Indirectas” variables creadas a partir de variables existentes.

3.2.2. Muestras

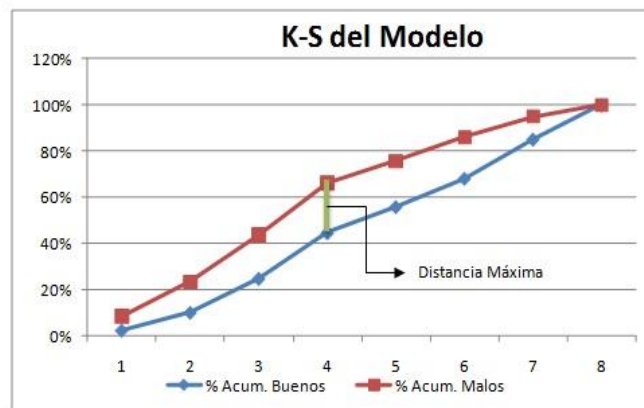
Hay que distinguir entre una muestra con la que se construye un modelo, sea esta con árboles de decisión, análisis discriminante, regresión logística, o de otra naturaleza, de muestras de testeo.

Una muestra de testeo es una muestra que es independiente de la que se usará para el de construcción y nos servirá para estimar que pasará con otros clientes.

3.2.3. Estadístico de Kolmogorov-Smirnov (Coeficiente K-S)

Una forma de medir el ajuste de un modelo es mediante las funciones de distribución de buenos y malos, en las cuales se puede observar la máxima distancia

Figura 3.1 K-S Modelo General



Para determinar si son equivalentes estadísticamente o no, de acuerdo a la siguiente prueba de hipótesis:

H_0 : La distribución de los buenos es equivalente a la de los malos

H_1 : No es verdad H_0

Se rechaza la hipótesis nula de acuerdo al valor alcanzado por la máxima distancia entre las dos distribuciones.

El estadístico K-S está dado por la siguiente ecuación:

$$K - S = \max |F_b(x) - F_m(x)|$$

Donde

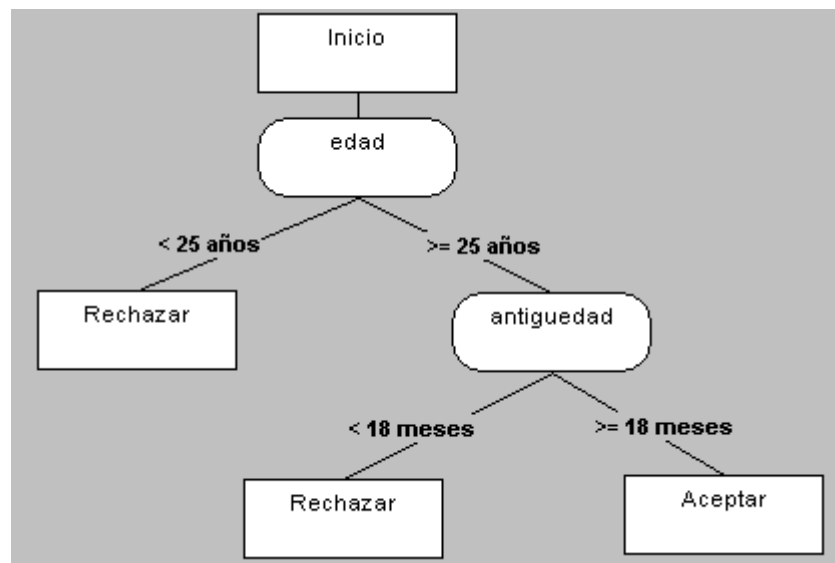
$F_b(x)$: es la función de distribución acumulada de los clientes buenos

$F_m(x)$: es la función de distribución acumulada de los clientes malos

3.2.4. Modelo Bajo Árbol de Decisión

El árbol de decisión es un tipo particular de modelo, es una excelente ayuda para la elección de varios cursos de acción. A continuación se muestra un árbol de decisión para ejemplo:

Figura 3.2 Ejemplo Árbol de Decisión



Fuente: <http://www.automind.cl>

Este árbol de decisión contiene dos variables: edad y antigüedad. Se lee de arriba hacia abajo. En el rectángulo inicio está el comienzo. Primero pregunta por la edad. Si es menor a 25 años el cliente es rechazado. Si su edad es igual o mayor a 25 años el árbol pregunta por la antigüedad. Si es menor a 18 meses el cliente es rechazado. En cambio, si su antigüedad es igual o mayor a 18 meses entonces el cliente es aceptado. En el caso anterior ambas variables eran numéricas. También puede haber variables categóricas.

Los diagramas de árbol [3] son representaciones naturales a la mente que ayudan a clasificar información, establecer regulaciones y políticas, y a seguir razonamientos. Inventados aparentemente el año 500 por lógicos sirios, fueron traídos a España cerca del año 600. En el año 1.300 fueron muy utilizados por el catalán Ramón Lull, lo que habría influido en Leibnitz, el gran matemático co-inventor del cálculo infinitesimal, quién habría utilizado estas ideas para proponer un lenguaje universal de cálculo lógico.

Construcción del Modelo.- El modelo fue construido con un escenario donde la proporción de clientes malos estuvo alrededor del 74%, fueron considerados clientes entre 30 y 90 días de mora, el tamaño de la muestra para el modelo de construcción fue de 9.540 clientes y para el de test 4.736 clientes. Las muestras fueron obtenidas de los resultados de Agosto, Septiembre y Octubre 2011. (Ver anexo I)

A continuación presentamos las 5 variables que fueron más discriminantes en la elaboración del modelo con su respectiva interpretación:

- MMORMAX6: mora máxima en los últimos 6 meses
- Tota_pr: número de promesas de pagos realizadas por el cliente la semana anterior al análisis.
- total_np: número de pagos en los últimos 6 meses.
- Contacto: número de contacto al cliente la semana anterior.
- vencido_inicio: valor vencido del cliente

Se obtuvo un modelo compuesto por 7 segmentos, a continuación se presenta la descripción de cada segmento:

Tabla 3.1 Segmentos Árbol de Decisión

Segmento	Descripción
1	mmormax6>30, contacto=0, total_np<=3
2	mmormax6>30, contacto=0, total_np>3
3	mmormax6<=30, vencido<=21, total_np<=5
4	mmormax6>30, contacto>0
5	mmormax6<=30, vencido<=21, total_np>5
6	mmormax6<=30, vencido>21, total_pr=0
7	mmormax6<=30, vencido>21, total_pr>0

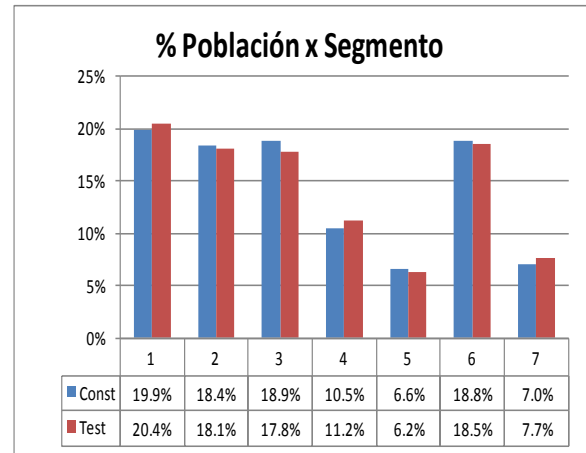
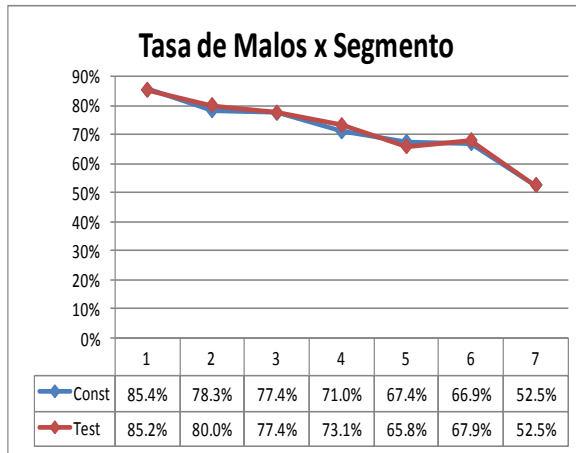
Tasa de malos y población por segmento

Los cuatro primeros segmentos contienen una tasa de malos superiores al 70%, la característica principal de estos segmentos excepto el tercero, es que la mora máxima en los últimos seis meses es superior a 30 días, es decir son clientes que en los últimos seis meses al menos se han encontrado vencidos en dos meses seguidos. El segmento con menor tasa de malos es el 7 con el 52%, la característica principal de este segmento es que realizaron una promesa de pago la semana anterior.

Como se puede apreciar en los gráficos de tasa de malos y porcentaje de población, que la tendencia se mantiene en la muestra de test, el KS del modelo en la base de construcción fue de 19.8%, mientras que en la base de test del 19.3%.

Figura 3.3 Tasa de Malos Árbol de Decisión

Figura 3.4 %Población Árbol de Decisión



3.2.5. Modelo Bajo Análisis Discriminante

El análisis [5] de la función discriminante es una técnica estadística de la rama del análisis multivariante, en la cual la variable independiente es indicadora y no numérica como en el de regresión lineal. La variable indicadora puede tomar dos o más valores por ejemplo, la variable Y es una variable indicadora, que denota la ocurrencia o no de un evento y se la define de esta forma:

$$Y = \begin{cases} 0 & \text{si es hombre} \\ 1 & \text{si es mujer} \end{cases}$$

En este contexto, el análisis de la función discriminante se emplea para determinar cuál o cuáles variables contribuyen a discriminar entre dos o más grupos que se observan en la práctica.

La Función Discriminante Lineal

El método está basado en modelos de probabilidad lineal, y se lo conoce como función discriminante lineal. Se supone que tiene n entidades para las que se le conocen k variables explicativas, y se observa que n1 de ellas pertenece a un grupo (1) y n2 a un grupo (2), donde n1+n2=n; es posible construir una función lineal de las k variables que puedan usarse para predecir si una nueva observación pertenece

a un grupo u a otro con una probabilidad determinada. La función lineal general se define como:

$$Z = \lambda_0 + \sum_{i=1}^k \lambda_i x_i$$

Donde λ_0 es el término independiente o una constante

λ_i los coeficiente de la ecuación.

x_i la variable independiente

El problema de la función de análisis discriminante desde el punto de vista del análisis de varianza consiste en responder a la pregunta de si dos o más grupos son significativamente diferentes uno de otro respecto a la media de una variable particular. Debe tenerse presente que si la media de una variable es significativamente diferente en varios grupos, puede decirse que esta variable discrimina entre grupos.

Dado que nuestro trabajo se trata de diferenciar entre dos grupos de clientes (Clientes buenos y malos), se aplica el análisis discriminante simple, en que la variable independiente es dicotómica, siendo sus modalidades cada uno de los grupos definidos a priori.

Supuestos Implícitos en el Análisis Discriminante

- Distribución normal: se asume que los datos para las variables representan una muestra proveniente de una distribución normal multivariante.
- Homogeneidad de varianzas y covarianzas: se supone que las matrices de varianzas y covarianzas son homogéneas entre grupos.
- Correlaciones entre medias y varianzas: el principal obstáculo para la validez de las pruebas de significancia se presenta cuando la media de las variables entre grupos están correlacionadas con las varianzas.
- Variables no redundantes: se supone que las variables empleadas para discriminar entre grupos no son completamente redundantes.

Lambda de Wilks

Para determinar la significancia de las variables que se introducen, se emplea el estadístico Lambda de Wilks que se obtiene de la razón entre el determinante de la matriz de varianzas y covarianzas dentro de los grupos y el determinante de matriz de varianzas y covarianzas total, este se aproxima como una F.

Construcción del Modelo.-La información fue la misma que el modelo anterior, es decir que la proporción de clientes malos estuvo alrededor del 74%, las funciones discriminantes, obtenidas como combinación lineales de las variables explicativas en el análisis discriminante permiten clasificar a los individuos de la muestra en los grupos definidos por la variable dependiente.

Para nuestro caso, se trata de una variable de respuesta con dos modalidades (clientes morosos y no morosos), por lo que se obtiene una única función discriminante, el programa utilizado para el presente estudio es el SPSS.

Para las variables seleccionadas se irán introduciendo una a una en orden creciente según los valores del estadístico asociado Lambda de Wilks, siendo eliminadas aquellas variables que estén muy correlacionadas con las introducidas en los pasos anteriores.

La función discriminante para el modelo de los datos contiene 7 variables independientes de las 199 variables disponibles, a continuación presentamos las variables que forman parte del modelo:

Tabla 3.2 Variables Modelo Análisis Discriminante

Nombre Técnico	Descripción
contacto	# Contactos semana pasada
mensaje	# Mensajes semana pasada
mancta_m	Antigüedad de la cuenta
mcmxcp12	Monto de compra máximo últimos 12 meses
mmormax6	Edad mora máxima últimos 6 meses
total_f	Número de gestiones en los últimos 6 meses
total_np	Número de pagos en los últimos 6 meses

Las ponderaciones sin tipificar y estandarizadas, así como los estadísticos que valoran su significancia individual, se recogen en la siguiente tabla:

Tabla 3.3 Coeficientes Modelo Análisis Discriminante

Variable	Coeficientes f. Discriminante	Coeficientes estandarizados	F Exacta	Sig.
contacto	0.052	0.235	151.162	0.000
mensaje	0.095	0.089	126.741	0.000
mancta_m	0.004	0.233	89.327	0.000
mcmxcp12	-0.005	-0.164	68.458	0.000
mmormax6	-0.009	-0.301	54.941	0.000
total_f	-0.036	-0.432	42.635	0.000
total_np	0.100	0.207	39.690	0.000
(Constante)	-0.427			

De las siete variables incluidas en la función discriminante una hacen referencia al comportamiento crediticio del cliente como la mora máxima en los últimos doce meses. También ha mostrado ser significativa las variables de gestión de cobro, tres de ellas también están incluidas en el modelo, estas son: contacto, mensaje, y total_f. Una variable de comportamiento de pagos forman parte del modelo, total_np. Las dos últimas variables, mancta_m “Antigüedad de la cuenta” y

mcmxcp12 “Monto de compra máximo en los últimos doce meses” completan el modelo de gestión.

Tasa de malos y población por segmento

Utilizando los coeficientes de la función discriminante obtenemos la puntuación para cada cliente, luego clasificamos en siete grupos en base a la proporción de clientes obtenidos en el modelo de árboles de decisión para su comparación posterior. La siguiente tabla muestra los segmentos con sus respectivos puntajes:

Tabla 3.4 Segmentos Modelo Análisis Discriminante

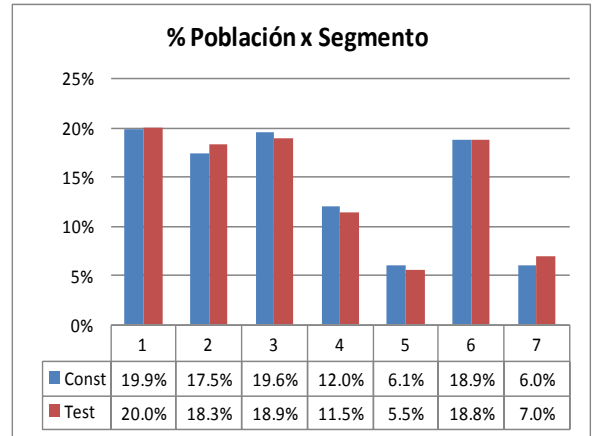
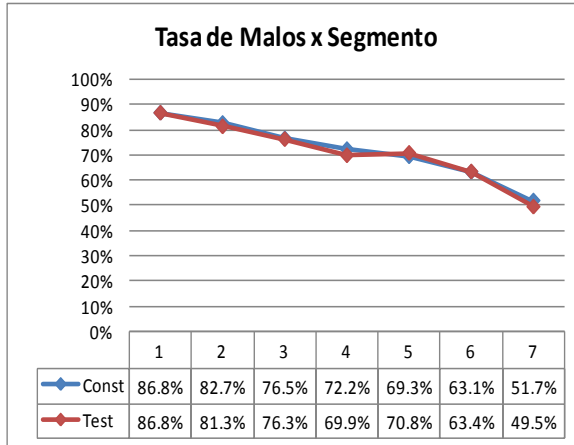
Segmento	Puntaje
1	≤ -1.16
2	> -1.16 y ≤ -0.64
3	> -0.61 y ≤ -0.15
4	> -0.15 y ≤ 0.14
5	> 0.14 y ≤ 0.31
6	> 0.31 y ≤ 1.38
7	> 1.38

Al igual que el modelo mediante árboles de decisión los cuatro primeros segmentos contienen una tasa de malos superior al 70%, a diferencia del modelo anterior aquí no podemos hablar de una característica específica. El segmento siete tiene la menor proporción de clientes malos (51%), un punto inferior a la obtenida por el segmento 8 con el modelo de árboles de decisión.

Como se puede apreciar en los gráficos la tasa de malos y porcentajes de población mantienen la misma tendencia tanto en la base de test como en la de construcción. El KS de este modelo es de 21,2%, 3 puntos superior al conseguido mediante la metodología de árboles de decisión.

Figura 3.5 Tasa de Análisis Discriminante

Figura 3.6 %Población Análisis Discriminante

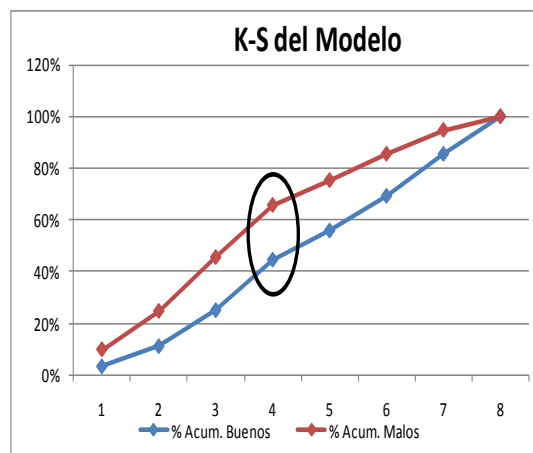


Estadístico Kolmogorov-Smirnov (estadístico K-S)

El estadístico Kolmogorov-Smirnov se contrastó la hipótesis de que las distribuciones de buenos y malos siguen la misma distribución sobre la variable Score.

Con un nivel de significancia de 0,05 rechazamos la hipótesis de que las distribuciones de buenos y malos son iguales.

Figura 3.7 K-S Análisis Discriminante



El KS del modelo alcanza una máxima distancia de 0,21, estadístico con el cual se rechazó la hipótesis de que los clientes buenos y malos son equivalentes.

3.2.6. Modelo Bajo Regresión Logística

El problema de la clasificación [5] en dos grupos puede abordarse introduciendo una variable ficticia binaria para representar la pertenencia de una observación a uno de los grupos. Para nuestro caso se desea discriminar entre créditos no cancelados o que presentan problemas para su cobro, puede añadirse a la base de datos una variable y que toma el valor de 1 cuando el crédito se cancela sin problemas y el valor de 0 en otro caso. El problema de discriminación es equivalente a la precisión del valor de la variable ficticia y .

Si el valor previsto está más cerca de 0 que a 1, clasificaremos al elemento en la primera población, caso contrario lo haremos en la segunda.

El primer enfoque es formular el siguiente modelo de regresión.

$$y = \beta_0 + \beta_1'x + u$$

Y estimar los parámetros por mínimos cuadrados de la forma habitual. Este método es equivalente a la función discriminante Fisher. Este procedimiento es óptimo para clasificar si la distribución conjunta de las variables explicativas es normal multivariante, con la misma matriz de covarianzas. Sin embargo, la discriminación lineal puede funcionar mal en otros contextos, cuando las covarianzas sean distintas o las distribuciones muy alejadas de la normal. Además, la función lineal puede encontrarse con problemas de interpretación de sus coeficientes estimados como del modelo general.

Si tomamos la esperanza en el modelo de regresión para $x = x_i$

$$E y x_i = \beta_0 + \beta_1'$$

Llamamos p_i a la probabilidad de que y tome el valor de 1 cuando $x = x_i$

$$p_i = P(y = 1|x_i)$$

Y la esperanza de y es:

$$E y x_i = P y = 1 x_i * 1 + P y = 0 x_i * 0 = p_i$$

Por tanto,

$$p_i = \beta_0 + \beta_1'x_i$$

Que es una expresión equivalente del modelo, el inconveniente principal de esta ecuación es que p_i debe estar entre cero y uno, y con el modelo anterior no hay garantía que esto se cumpla, ya que el modelo puede pronosticar probabilidades mayores a la unidad.

A pesar de los inconvenientes este simple modelo proporciona una buena regla de clasificación, ya que según la interpretación de Fisher, maximiza la separación entre grupos, para cualquier distribución de datos. Pero hay que tener presente que cuando los datos no son normales, o no tienen la misma matriz de covarianzas, la clasificación mediante la ecuación lineal no es necesariamente óptima.

El modelo de regresión logística parte de la hipótesis de que los datos siguen el siguiente modelo:

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_k * x_k + u$$

Si definimos la variable z como:

$$z = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_k * x_k$$

Por lo tanto el modelo lo podemos expresar de la siguiente forma:

$$\text{Ln} \frac{p}{1-p} = z + u$$

Donde p es la probabilidad de que ocurra el suceso de estudio.

Ahora si operamos algebraicamente sobre el modelo tenemos lo siguiente:

$$\text{Ln} \frac{p}{1-p} = z$$

$$\frac{p}{1-p} = e^z$$

$$p = 1 - p * e^z$$

$$p = e^z - p * e^z$$

$$p + p * e^z = e^z$$

$$p(1 + e^z) = e^z$$

$$p = \frac{e^z}{1 + e^z}$$

Como la función de distribución logística es:

$$F(x) = \frac{e^x}{1 + e^x}$$

Por lo tanto podemos describir el modelo de forma mucho más compacta:

$$p = \frac{e^z}{1 + e^z} = F(z) = F(x * b)$$

De donde se deduce que el modelo de regresión logística es, en principio, un modelo de regresión no lineal, pero es lineal en escala logarítmica atendiendo a su definición original:

$$\ln \frac{p}{1-p} = z$$

$$\ln p - \ln(1-p) = z$$

$$\ln p - \ln(1-p) = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_k * x_k$$

Es decir, la diferencia de la probabilidad de que ocurra un suceso respecto de que no ocurra es lineal pero en escala logarítmica. Por lo tanto, el significado de los coeficientes, aunque guardando una cierta distancia con relación al modelo lineal, va ser algo más complejo de interpretar.

La regresión logística se la puede expresar de las siguientes formas:

- $\ln p - \ln(1-p) = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_k * x_k$
- $\frac{p}{1-p} = e^{\beta_0} * e^{\beta_1 * x_1} * e^{\beta_2 * x_2} \dots e^{\beta_k * x_k}$

La primera expresión se llama Logit y la segunda Odds-ratio o cociente de probabilidades.

Interpretación del Modelo Logístico

Los parámetros del modelo son:

- β_0 la ordena en el origen.
- $\beta_1 + \beta_2 + \dots + \beta_k$ a veces también utilizados como parámetros $e^{\beta_1} * e^{\beta_2} \dots e^{\beta_k}$ que se denominan Odds ratios o ratios de probabilidades. Estos valores indican cuánto se modifican las probabilidades por unidad de cambio en las variables.

Los residuos del modelo (también llamados residuos de Pearson) se definen de la siguiente manera:

$$e = \frac{y_i - p_i}{p_i(1 - p_i)}$$

Y si el modelo es correcto, serán variables de media cero y varianza unidad que pueden servir para realizar el diagnóstico de dicho modelo. El estadístico $\chi_0^2 = \sum_i e_i^2$ que permite realizar un contraste global de la bondad del ajuste.

El estadístico de Wald: Contrasta la hipótesis de que un coeficiente aislado es distinto de 0. El valor para este coeficiente está dado por el cociente entre el valor del coeficiente y su correspondiente error estándar, además sigue una distribución chi-cuadrado con un grado de libertad.

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Construcción del Modelo.-El modelo de regresión logística permite determinar la probabilidad de que un cliente vaya o no a realizar un pago, mediante los coeficientes estimados de las variables incluidas en el modelo.

Al igual que los modelos anteriores se utilizó las 163 variables independientes disponibles, los resultados obtenidos al aplicar este modelo se muestran en la siguiente tabla, la cual contiene los coeficientes de cada variable estadísticamente significativos ($p \leq 0,05$) y la información de $\text{Exp}(B)$, que corresponde a los Odds-ratio.

Tabla 3.5 Resumen Modelo Regresión Logística

Resumen del modelo			
Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	9101.924	.021	.031
2	8971.029	.037	.054
3	8902.019	.045	.066
4	8867.135	.049	.072
5	8845.582	.052	.076
6	8809.007	.056	.082
7	8792.493	.058	.085
8	8783.362	.059	.086
9	8776.147	.060	.088
10	8770.145	.061	.089
11	8764.294	.061	.090
12	8757.887	.062	.091
13	8753.405	.063	.091
14	8749.293	.063	.092
15	8745.402	.064	.093

Se aplicó el método de máxima verosimilitud para la estimación del modelo con pasos hacia delante, va seleccionando las variables más significativas hasta no tener ninguna con un p-valor menor a 0,05. Por cada inclusión de una variable se calcula la verosimilitud del modelo con la nueva variable

El estadístico de razón de verosimilitud se calcula para el último paso con el que se contrasta la hipótesis de que el modelo no es significativo mediante la diferencias de las verosimilitud inicial y final.

$$\text{Razón de Verosimilitud} = -2 \ln \frac{L_0}{L_f} = 7.919$$

Tal como se muestra en la siguiente tabla:

Tabla 3.6 Prueba de Hosmer y Lemeshow

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	4.091	3	.252
2	6.945	6	.326
3	14.206	8	.077
4	14.536	8	.069
5	5.442	8	.709
6	10.380	8	.239
7	17.470	8	.026
8	11.176	8	.192
9	9.389	8	.311
10	9.523	8	.300
11	12.017	8	.150
12	7.634	8	.470
13	9.732	8	.284
14	6.252	8	.619
15	7.191	8	.516

Con este valor chi-cuadrado (7,191) se rechaza la hipótesis nula de que todos los coeficientes del modelo son iguales a cero con un nivel de significancia $\alpha = 0,05$

Significancia de las variables

El estadístico de Wald se utilizó para evaluar la hipótesis nula $H_0: \beta_i = 0$, a continuación se presentan la significancia estadística asociada de cada una de las variables:

Tabla 3.7 Coeficientes Modelo Regresión Logística

Variables	B	Wald	gl	Sig.	Exp(B)	Descripción Variable
gestion	-0.082	4.394	1	.036	0.92	# de llamadas perdidas en la semana anterior
ult_esta		33.649	7	.000		Estatus de la última llamada la semana anterior
ult_esta(1)	-0.198	11.293	1	.001	1.78	1 si no se realizó gestión
ult_esta(2)	0.376	12.654	1	.000	1.59	1 si fue carta
ult_esta(3)	0.267	8.822	1	.003	1.48	1 si fue contacto
ult_esta(4)	0.192	1.609	1	.205	0.74	1 si fue teléfono dañado
ult_esta(5)	-0.497	1.964	1	.161	1.49	1 si fue ilocalizado
ult_esta(6)	0.198	14.200	1	.000	1.34	1 si fue Llamada "no se encuentra"
ult_esta(7)	0.097	.521	1	.470	1.22	1 si fue mensaje
total_pr	0.352	11.399	1	.001	1.42	# Promesas la última semana
mmormax6	-0.005	8.450	1	.004	0.99	Mora máxima últimos 6 meses
total__f	-0.022	40.938	1	.000	0.98	# gestiones últimos 6 meses
total__m	0.090	3.975	1	.046	1.09	Total llamada en los último 3 meses
total_np	0.050	5.335	1	.021	1.05	# Pagos últimos 6 meses
Constante	-0.421	6.471	1	.011	0.54	

Siete variables fueron incluidas en el modelo, de las cuales cinco hacen referencia a la gestión de cobro (gestión, ult_esta, total_pr, total__f, total__m), y dos al comportamiento crediticio (mmormax6, total_np).

Tasa de Malos y Población por Segmento

Al igual que el modelo de Análisis Discriminante se tomó como referencia los resultados obtenidos con el modelo de Árboles de Decisión para la clasificación de los clientes en siete grupos, utilizando en este caso las probabilidades derivadas de la ecuación logística tenemos la siguiente clasificación:

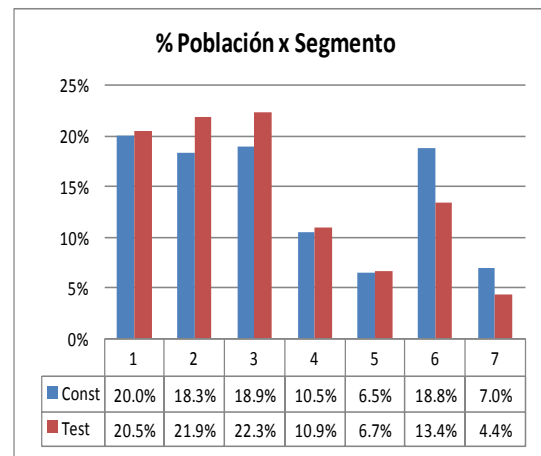
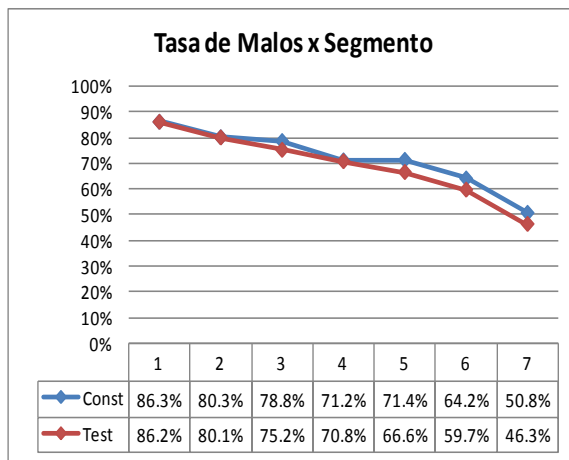
Tabla 3.8 Segmentos Modelo Regresión Logística

Segmento	Probabilidad
1	<=19.4%
2	>19.4% - <=25.5%
3	>25.5% - <=31.7%
4	>31.7% - <=35.5%
5	>35.5% - <=38.5%
6	>38.5% - <=51.1%
7	>51.1%

Los cuatro primeros segmentos tiene una tasa de malos superior al 70%, mientras que el segmento 8 tiene una proporción de malos del 50.8%, a continuación los siguientes gráficos nos presentan la tasa de malos y porcentaje de población por segmentos donde se puede observar que la tendencia es la misma tanto en la muestra de construcción como en la de test.

Figura 3.8 Tasa de Malos Regresión Logística

Figura 3.9 %Población Regresión Logística

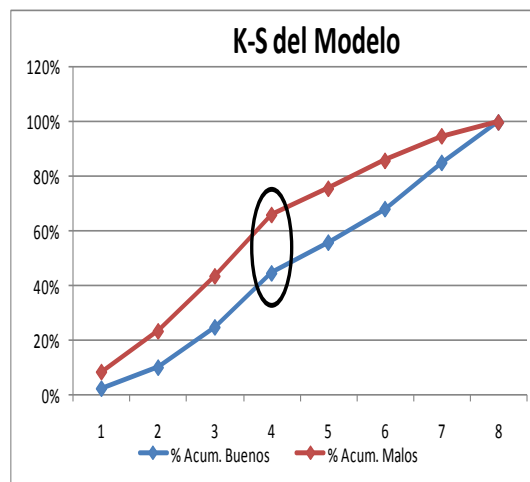


Estadístico Kolmogorov-Smirnov (estadístico K-S)

Con el estadístico Kolmogorov-Smirnov se contrastó la hipótesis de que las distribuciones de buenos y malos siguen la misma distribución sobre la variable Score.

Con un nivel de significancia de 0,05 rechazamos la hipótesis de que las distribuciones de buenos y malos son iguales.

Figura 3.10 K-S Regresión Logística



El KS del modelo alcanza una máxima distancia de 0,21.

3.3. Construcción del modelo mediante programación matemática

3.3.1. Introducción

Una vez segmentadas las cuentas vencidas ya sea por medio de un árbol de decisión, un puntaje obtenido de un análisis discriminante o una probabilidad de cumplimiento mediante regresión logística, el problema ahora consiste en determinar cuáles cuentas se debe priorizar.

Utilizando el modelo de regresión logística puedo empezar priorizando las cuentas de menor probabilidad de pago, pero dejaría de gestionar las de mayor probabilidad de cumplimiento, o viceversa empezando con las de mayor probabilidad de pago, pero dejaría de gestionar las de menor probabilidad, ya que la capacidad actual del

departamento solamente permite gestionar entre el 40% y 50% de las cuentas vencidas.

Debido a esta situación se decide elaborar un modelo mediante programación matemática que seleccione las cuentas a gestionar, de manera que se maximice la recuperación de cartera, optimizando los recursos disponibles.

En función de las descripciones realizadas de los procesos de cobranzas, a los objetivos planteados inicialmente, y al problema de capacidad para gestionar el total de cuentas vencidas, el problema se puede tratar como un problema clásico de la mochila.

3.3.2. Programación matemática

La programación matemática [6] es una potente técnica de modelado usada en el proceso de toma de decisiones. Cuando se trata de resolver un problema de este tipo, la primera etapa consiste en identificar las posibles decisiones que pueden tomarse; esto lleva a identificar las variables del problema concreto.

Dentro del campo general de la programación matemática conviene distinguir diferentes modelos de optimización. Los modelos de optimización se caracterizan por contener:

- Variables o decisiones a tomar
- Ecuaciones de restricción o limitaciones
- Función o funciones objetivo

Una de las características de los modelos de optimización es la existencia de un único decisor frente a las otras disciplinas donde puede existir más de un decisor, como por ejemplo la teoría de juegos.

El planteamiento general del problema de programación matemática es el siguiente:

Optimizar $f(x_1, x_2, \dots, x_n)$

sujeto a: $g_1(x_1, x_2, \dots, x_n) \leq b_1$

$g_2(x_1, x_2, \dots, x_n) \leq b_2$

.....

$g_m(x_1, x_2, \dots, x_n) \leq b_m$

O en forma abreviada

Opt $f(x)$

s.a. $g(x) \leq b$

Donde

$$f: R^n \rightarrow R, \quad X \in R^n, \quad g: R^n \rightarrow R^m, \quad b \in R^m$$

La función f denominada función objetivo, es una función definida de un dominio de R^n sobre R , y representa una descripción matemática de la función objetivo que se pretende alcanzar con el problema planteado.

El vector X es el vector de variables de decisión, de entre cuyos valores posibles se trata de elegir aquél o aquellos que proporcionen el valor óptimo de la función f .

Conjunto de oportunidades, denominaremos así al conjunto de puntos $X \in R^n$ que verifican todas y cada una de las restricciones.

3.3.3. El problema de la mochila

Una [6] clase importante de problemas de programación entera son aquellos en que las variables pueden tomar dos valores. Esta situación se puede modelar empleando el siguiente tipo de variables:

$$x = \begin{cases} 1 & \text{si la situación tiene lugar} \\ 0 & \text{en cualquier otra caso} \end{cases}$$

Cada valor se asocia a una de las dos posibilidades de una elección binaria.

Un problema clásico donde aparecen este tipo de variables es el problema de la mochila. Considérese un excursionista que debe preparar su mochila. Considérese a sí mismo que hay una serie de objetos de utilidad para el excursionista, pero que el excursionista solo puede llevar un número limitado de objetos. El problema consiste en elegir un subconjunto de objetos de tal forma que se maximice la utilidad que el excursionista obtiene, pero sin rebasar su capacidad de acarrear objetos.

El problema consta de los siguientes elementos:

- Datos
 - n: número de objetos
 - a_j : peso de cada objeto j
 - c_j : utilidad de cada objeto j
 - b: la capacidad máxima de la mochila
- Variables

$$x_j = \begin{cases} 1 & \text{si el objeto j se mete a la mochila} \\ 0 & \text{si no se mete} \end{cases}$$

- Restricciones.- la capacidad máxima de la mochila

$$\sum_{j=1}^n a_j x_j \leq b$$

- Función a maximizar.- El objeto de este problema es maximizar la utilidad

$$z = \sum_{j=1}^n c_j x_j$$

3.3.4. Diseño del modelo

Antes de presentar el modelo matemático desarrollado para optimizar las decisiones de cobranzas, definiremos claramente la información asociada al problema, ya que la calidad de la información con que es realizada influye directamente en los resultados esperados. Considerando estos puntos a continuación se describen la forma con que fueron estimados los parámetros que utilizará el modelo.

Saldo Vencido

Es aquella parte del saldo que es exigible debido a que venció la fecha de pago.

Probabilidad de pago

Probabilidad de pagar el valor vencido en la semana de gestión de la cuenta, esta probabilidad es calculada mediante regresión logística.

Tiempo promedio de gestión de una cuenta

Es el valor expresado en minutos del promedio de duración del conjunto de llamadas realizadas. Se tomó una muestra de 4.263 llamadas para estimar el tiempo de duración. A continuación presentamos el promedio de llamadas por edad de mora y el número de llamadas respectivamente:

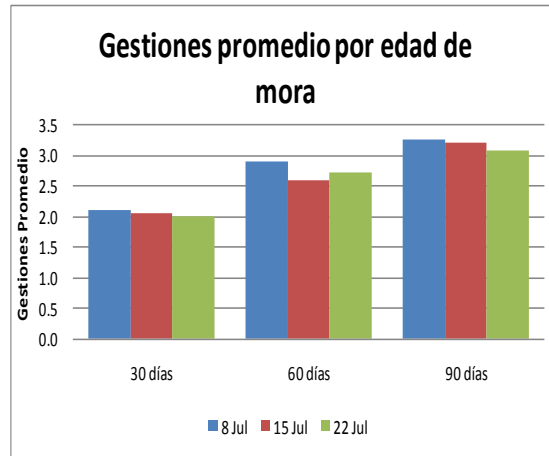
Tabla 3.9 Promedio de Duración de Llamadas por edad

Edad Mora	Promedio de Tiempo	# Llamadas
30 Días	0:01:14	2,088
60 Días	0:01:12	1,210
90 Días	0:01:19	975
Promedio General	0:01:15	4,273

Se realizó una prueba para verificar si existía o no diferencia estadística entre los promedio de tiempos por edad de mora, dando como resultado esta prueba que no existen diferencias entre edad de mora.

Adicionalmente, se determinó el número de veces que una cuenta es gestionada a la semana por edad de mora, donde se comprobó que sí existen diferencias significativas en el número de repeticiones por edad. A continuación presentamos un gráfico con el número de gestiones promedio en tres semanas diferentes, dando como resultado que una cuenta que inicia en 30 días se le llama 2.1 veces en la semana, la de 60 días 2.7 veces y la de 90 días 3.2 veces.

Figura 3.11 # Gestiones por Edad de Mora



Con un tiempo promedio de 1 minuto con 15 segundos para cada llamada, el tiempo final por edad de mora será:

Tabla 3.10 Promedio de Duración de Llamadas Final

Edad Mora	Promedio de Tiempo
30 Días	0:02:34
60 Días	0:03:25
90 Días	0:03:58
Promedio General	0:03:19

Horas laborables

Los gestores de cobranzas deberán cumplir con un total de 40 horas laborables semanalmente.

Modelo Matemático

Índices

$i =$ Conjunto de cuentas Vencidas

$j =$ Conjunto de gestores

Parámetros

$S_i =$ Saldo Vencido de la cuenta i

$P_i =$ Probabilidad de pago de la cuenta i en la semana

$T_i =$ Tiempo promedio en gestionar la cuenta i

$HL =$ Total de horas laborables de cada gestor

Variables

$X_{i,j} =$ 1 si la cuenta i es asignada al gestor j
0 si no

$Z =$ Valor de la función objetivo

Restricciones

- Cada cuenta deber ser asignada como máximo a un gestor o no asignada.

m

$$X_{i,j} \leq 1 \quad \forall i \text{ des de } 1 \text{ hasta } n$$

$j=1$

- La cantidad de cuentas asignadas al gestor J debe ser menor o igual a la disponibilidad del tiempo de trabajo semanal.

n

$$X_{i,j}T_i \leq HL \quad \forall j \text{ des de } 1 \text{ hasta } m$$

$i=1$

Función Objetivo

Maximización de la recaudación esperada.

$$z = \sum_{i=1}^n \sum_{j=1}^n X_{i,j} P_i S_i$$

Capítulo 4

Validación, Comparación y Aplicación del modelo bajo programación matemática.

4.1.Introducción

Una vez presentados los modelos bajo las tres técnicas estadísticas efectuamos una validación adicional a la de test, un análisis comparativo de los mismos y por último efectuamos una aplicación del modelo en programación matemática propuesto en el capítulo anterior.

Para el modelo de programación matemática se tomó la información únicamente de la ciudad de Quito.

4.2.Validación de los modelos.

Los modelos fueron validados con la información de comportamiento en pagos correspondiente a las tres primeras semanas de agosto 2012, a continuación se presenta la proporción de clientes malos y población por segmento en cada uno de los modelos:

Árbol Decisión

La tendencia se mantiene, el primer segmento sigue teniendo la mayor tasa de malos y el segmento siete la menor tasa de malos. Donde sí se observa una diferencia marcada es la población del segmento seis, la cual incremento su población en alrededor de un 43% con respecto a las bases de construcción y test.

Figura 4.1 Seguimiento Tasa de Malos Árbol de Decisión

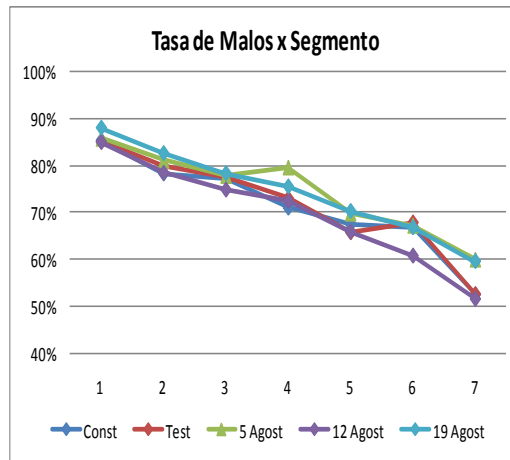
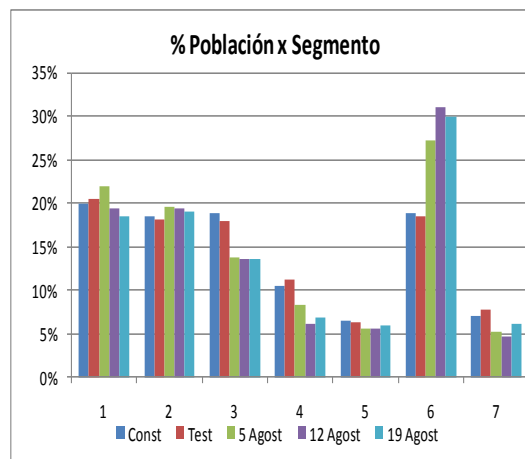
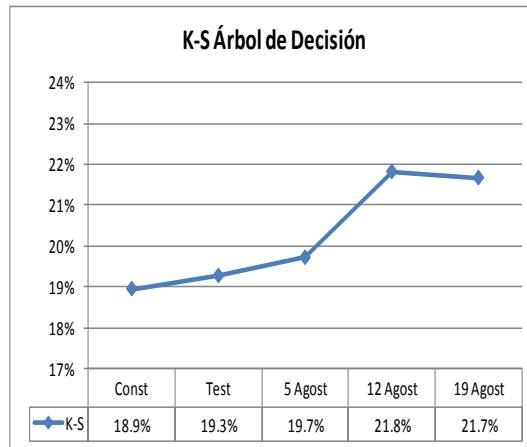


Figura 4.2 Seguimiento % Población Árbol de Decisión



El K-S del modelo se incrementó en las tres semanas de pruebas del modelo como se puede apreciar en el gráfico siguiente:

Figura 4.3 Seguimiento K-S Árbol de Decisión



Análisis Discriminante

La tendencia en tasa de malo se mantuvo al igual que el modelo de árboles de decisión, no así en población por segmento donde se observa una disminución significativa en el segmento uno, y un incremento en el segmento seis.

Figura 4.4 Seguimiento Tasa de Malos Análisis Discriminante

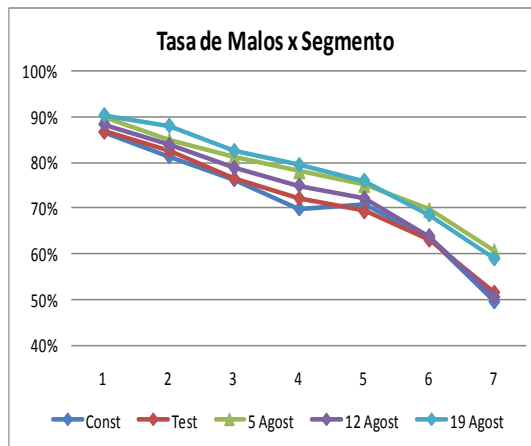
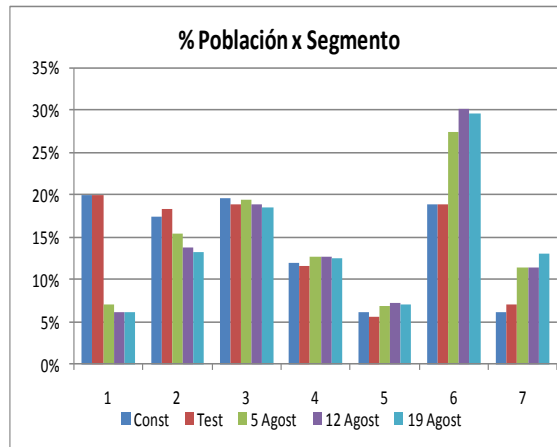
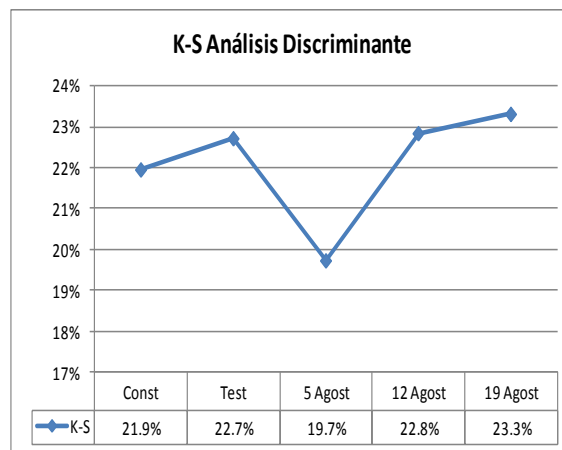


Figura 4.5 Seguimiento % Población Análisis Discriminante



El nivel del K-S se mantuvo con relación a las bases de construcción excepto en la semana del 5 de agosto que tuvo su valor más bajo “19.7%”.

Figura 4.6 Seguimiento K-S Análisis Discriminante



Regresión Logística

La tendencia en tasa de malos se mantiene al igual que los modelo anteriores. La población a diferencia de los anteriores tuvo un decrecimiento a partir del segmento seis.

Figura 4.7 Seguimiento Tasa de Malos Regresión Logística

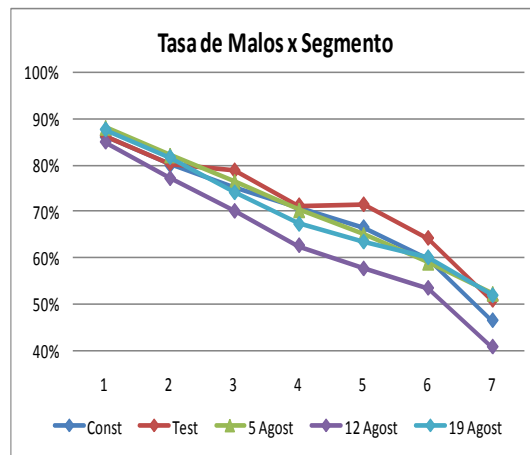
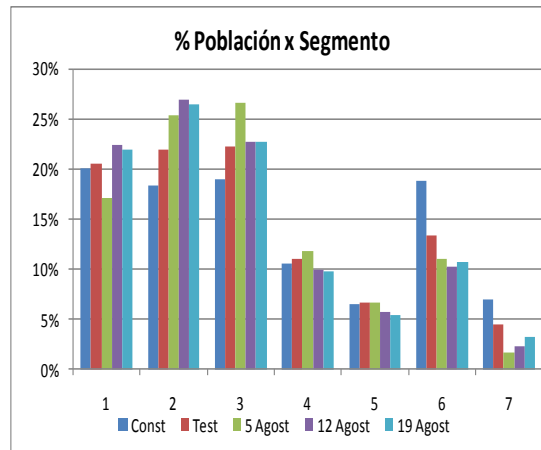
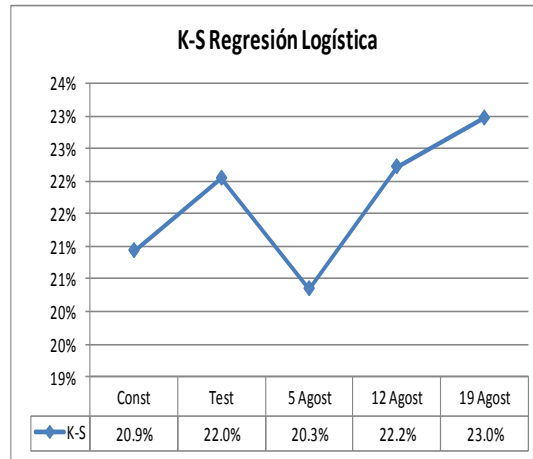


Figura 4.8 Seguimiento % Población Regresión Logística



Al igual que el modelo de Análisis Discriminante el 5 de agosto obtuvo el valor más bajo.

Figura 4.9 Seguimiento K-S Regresión Logística

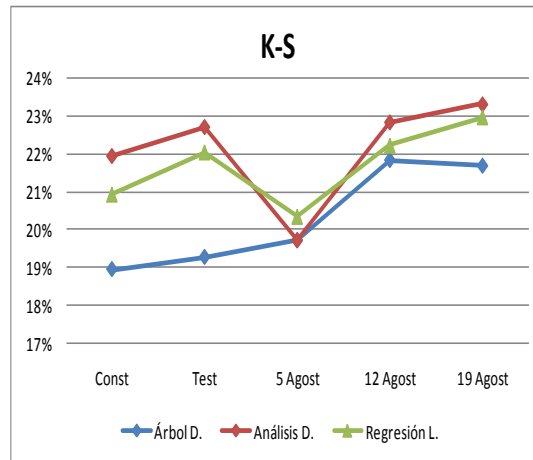


4.3. Comparación de modelos

Una vez presentado los resultados obtenidos tanto en la construcción como en la validación de los modelos, efectuamos a continuación un análisis comparativo de los mismos.

K-S.- El método de Árboles de Decisión tiene el K-S más bajo, mientras que el modelo realizado a través de Análisis Discriminante el más alto. El modelo bajo Regresión Logística tiene prácticamente la misma tendencia que el de Análisis Discriminante.

Figura 4.10 Comparativo K-S por Modelos



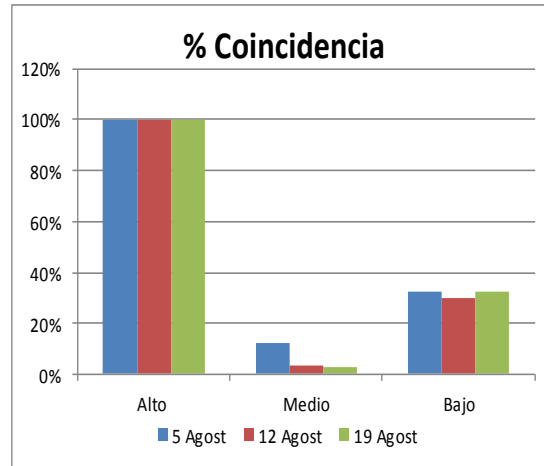
Comparativo entre segmentos.- Los modelos están compuesto por siete segmentos, estos segmentos se los clasificó en tres niveles de riesgo para una mejor comparación entre modelos, a continuación la clasificación y su respectivo gráfico de tasa de malos por riesgo:

Tabla 4.1 Clasificación del Riesgo por Segmentos

Nivel de Riesgo	Segmentos
Alto	1-3
Medio	4-5
Bajo	6-7

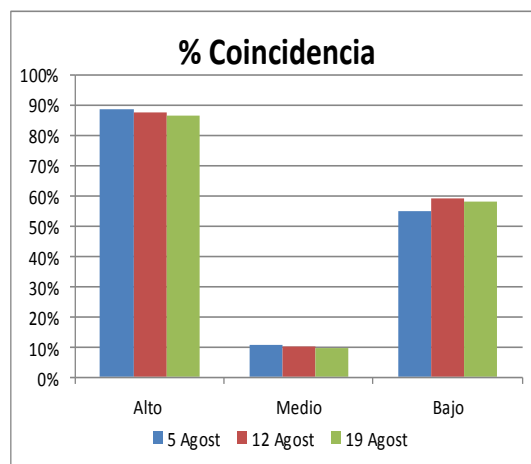
Análisis Discriminante – Regresión Logística.- Estos dos modelos coinciden el 99% en la clasificación de los clientes de riesgo alto, no así en los demás riesgos donde existen diferencias marcadas. En general tuvieron en promedio una coincidencia del 53%.

Figura 4.11 % Coincidencia Análisis Discriminante vs Regresión Logística



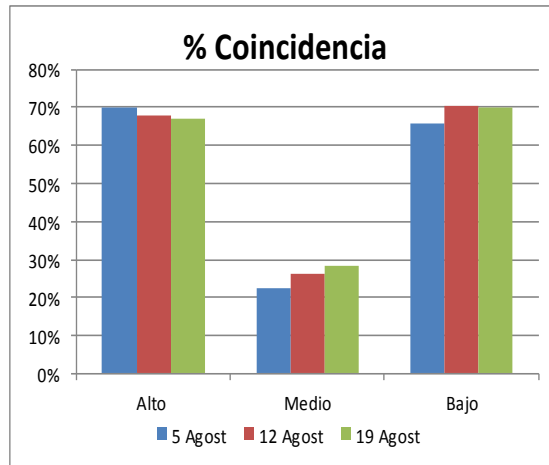
Análisis Discriminante – Árbol de Decisión.- Tienen una coincidencia inicial del 60% de la calificación de los clientes, los clientes de riesgo alto tienen un mayor porcentaje de coincidencia (87%), mientras que los clientes de riesgo medio tienen el menor porcentaje de coincidencia (10%).

Figura 4.12 % Coincidencia Análisis Discriminante vs Árbol de Decisión



Regresión Logística – Árbol de Decisión.- Tienen una coincidencia del 61%, los riesgos alto y bajo tienen una coincidencia del 68%, mientras que los clientes de riesgo medio tienen una coincidencia del 26% en promedio.

Figura 4.13 Coincidencia Regresión Logística vs Árbol de Decisión



4.4. Aplicación del modelo bajo programación matemática.

Se realizó la aplicación del modelo realizado bajo la modalidad de programación matemática solamente para la ciudad de Quito, a continuación presentamos los resultados:

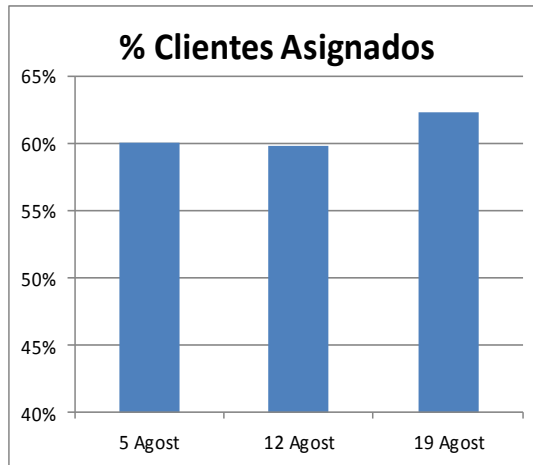
Tabla 4.2 Datos Iniciales Modelo Programación Matemática

Datos	5 Agust	12 Agust	19 Agust
Valor Vencido	\$ 1,304,490	\$ 1,323,023	\$ 1,285,823
Número de Gestores	28	28	28
Número de Cuentas	28,112	28,094	26,891

Resultados:

- Se Asignó él en promedio al 60% de los clientes.

Figura 4.14 % Clientes Asignados Modelo Programación Matemática



- El modelo asigno en mayor proporción a clientes de los segmentos con menor tasa de malos.

Tabla 4.3 Asignación de Clientes Modelo Programación Matemática

Segmento R.L	5 Agust	12 Agust	19 Agust
1	44.4%	43.3%	41.0%
2	51.6%	54.1%	53.5%
3	61.4%	64.3%	67.2%
4	71.3%	74.7%	77.1%
5	76.3%	81.0%	84.6%
6	84.0%	89.9%	92.4%
7	88.6%	97.0%	97.1%

- El promedio del valor vencido de los clientes asignados es \$ 66 dólares, mientras que de los no asignados es de \$ 17 dólares.

Tabla 4.4 Valor Vencido Promedio Asignado y no Asignado

Semana	Asignado	No Asignado
5 Agosto	\$ 67	\$ 16
12 Agosto	\$ 68	\$ 16
19 Agosto	\$ 67	\$ 16

- La recuperación esperada de acuerdo al modelo es del 23% del valor vencido total.

Tabla 4.5 Recuperación de Cartera Esperada Modelo Programación Matemática

Datos	5 Agust	12 Agust	19 Agust
Valor Vencido	\$ 1,304,490	\$ 1,323,023	\$ 1,285,823
Valor Esperado	\$ 309,106	\$ 296,782	\$ 317,808
% Recuperación Esperado	23.7%	22.4%	24.7%

Capítulo 5

Conclusiones y Recomendaciones

5.1. Conclusiones.

- El problema de planificación de cobranza puede ser modelado mediante un enfoque de programación matemática tal como se lo demuestra en el presente trabajo, ayudando a proyectar la recuperación y a planificar las estrategias de cobro.
- El modelo asigna automáticamente los clientes vencidos a los agentes de cobro, permitiendo de esta manera asignar metas para cada uno de ellos.
- Permite proyectar el incremento de gestores en función de los valores vencidos y el valor esperado de la recuperación.
- Los tres modelos estimados mediante árboles de decisión, análisis discriminante y regresión logística resultaron significativos y estables, tal como lo muestran los estadísticos calculados para su respectiva valoración.
- En términos del K-S el modelo de análisis discriminante resultó con mejores indicadores, el modelo de árboles de decisión mostró el menor K-S de los tres modelos.
- De acuerdo al análisis comparativo de los tres modelos el realizado por medio de regresión logística clasificó mejor a los clientes.
- Los clientes de riesgo alto son más fáciles de detectar en los tres modelos realizados.
- Los clientes con mayor tasa de malos son más fáciles de segmentar por los tres modelos.
- Las variables MMORMAX6 (mora máxima en los últimos seis meses) y total_np (número de pagos en los últimos seis meses) se repiten en los tres modelos.

5.2.Recomendaciones

- Se recomienda profundizar en la automatización del modelo desarrollando un sistema computacional que permita a los encargados de la cobranza utilizar la herramienta para planificación.
- Un tema interesante sería la modificación del modelo desarrollado para obtener el número óptimo de gestores que permita incrementar la recuperación esperada.
- Para futuras investigaciones y cambios del modelo de Programación Matemática, se deberá primero analizar la estabilidad del modelo de Regresión Logística que sirve como input principal del modelo.

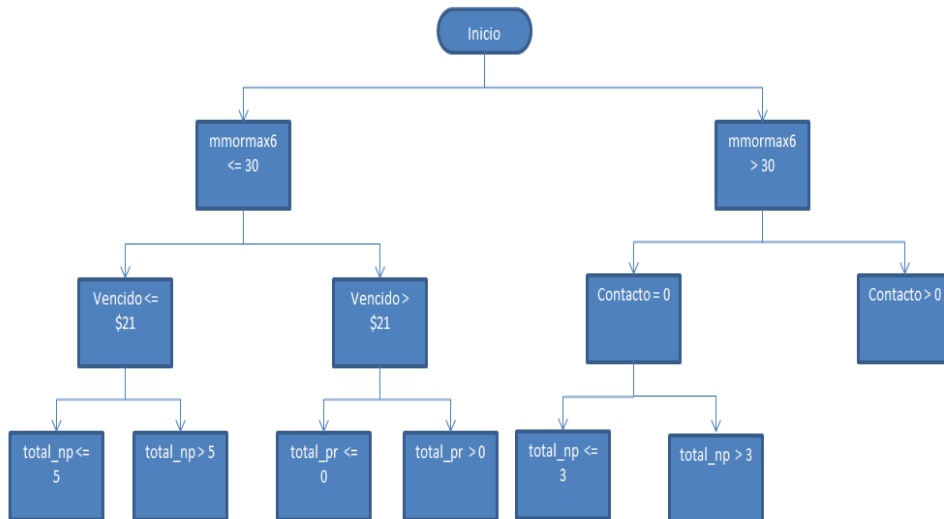
Bibliografía

- [1]. Jesús Mures, Ana García, Eva Vallejo, Aplicación del Análisis Discriminante y Regresión Logística en el Estudio de la Morosidad en las Entidades Financieras, Pecuaria 2005.
- [2]. Alfredo Añez, Paper de Riesgo Crediticio, 2004
- [3]. http://www.automind.cl/articulos/metodos_estadisticos/Metodos_estadisticos_con_arboles_de_decision.htm Empresa dedicada al desarrollo de modelos crediticios.
- [4]. Evelyn Muñoz, La técnica del Análisis Discriminante una Aplicación para el área Bancaria, Agosto
- [5]. <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema2dm.pdf>
- [6]. Enrique Castillo, Antonio J. Conejo, Pablo Pedregal, Ricardo Garcia y Natalia Alguacil, Formulación y Resolución de Modelos de Programación Matemática en Ingeniería y Ciencias, 20 Febrero 2002.
- [7]. Antoine Michele Sauré Valenzuela, Modelo de Programación Matemática para un Sistema de Cobranzas Caso: Tesorería General de la República, 2002.
- [8]. Jimmy Reyes, Carlos Escobar, Juan Duarte, Pedro Ramirez, Una Aplicación del Modelo de Regresión Logística en la Predicción del Rendimiento Estudiantil, Chile 2007.
- [9]. <http://www.aulafacil.com/spss-avanzado/curso/Temario.htm> Curso de SPS
- [10]. Irene Moral Peláez, Modelos de regresión: Lineal simple y regresión Logística, 20

Anexos

Anexo I

Modelo Árbol de Decisión



Anexo II

Script Variables de Comportamiento Crediticio

/*-----

/*Activacion

compute marca_saldo12=0.

compute marca_saldo06=0.

vector saldototal = Saldo306 to Saldo411.

loop #n = 1 to 12.

```
do if saldototal(#n)>0.  
  compute marca_saldo12=1.  
end if.  
end loop.
```

```
loop #i = 13 to 18.  
  do if saldototal(#i)>0.  
    compute marca_saldo06=1.  
  end if.  
end loop.
```

```
execute.
```

```
variable labels marca_saldo12 'Actividad 12 meses atras'.
```

```
variable labels marca_saldo06 'Actividad 6 meses adelante'.
```

```
value labels marca_saldo12
```

```
1 'Activo'
```

```
0 'Inactivo'.
```

```
value labels marca_saldo06
```

```
1 'Activo'
```

```
0 'Inactivo'.
```

```
/*-----
```

```
/* Clasificacion de malos ( mora max 120 dias) entre junio 04-nov 04
```

```
vector edad = edadv406 to edadv411.
```

```
compute marca_edad=0.
```

```
loop #i = 1 to 6.
```

```
    do if edad(#i)>=120.
```

```
        compute marca_edad=1.
```

```
    end if.
```

```
end loop.
```

```
execute.
```

```
variable labels marca_edad 'Mora max 120 '.
```

```
value labels marca_edad
```

```
0 'Bueno'
```

```
1 'Malo'.
```

```
/* -----Variables Indirectas-----  
--
```

```
/* Mora actual : mora en fecha de analisis (Mayo 04)
```

```
compute moract = edadv405.
```

```
execute.
```

```
/*Mora maxima en ultimos 3,6 y 12 meses , trimestre anterior , semestre anterior y mes  
en que ocurrio la mora maxima
```

```
vector M = edadv306 to edadv405.
```

```
compute Mormax3=0.
```

```
compute Mormax6=0.
```

```
compute Mormax12=0.
```



```
compute MormaxT=0.  
compute MormaxS=0.  
compute Lmax3=0.  
compute Lmax6=0.  
compute Lmax12=0.
```

```
/* mora max ultimos tres meses
```

```
Loop #n=10 to 12.  
do if M(#n)>Mormax3.  
compute Mormax3=M(#n).  
compute Lmax3=#n.  
end if.  
end loop.
```

```
/* mora max ultimos seis meses
```

```
Loop #n=7 to 12.  
do if M(#n)>Mormax6.  
compute Mormax6=M(#n).  
compute Lmax6=#n.  
end if.  
end loop.
```

```
/* mora max ultimos doce meses
```

```
Loop #n=1 to 12.  
do if M(#n)>Mormax12.  
compute Mormax12=M(#n).  
compute Lmax12=#n.
```

end if.

end loop.

/* mora max trimestre anterior

Loop #n=7 to 9.

do if $M(\#n) > M_{\text{ormaxT}}$.

compute $M_{\text{ormaxT}} = M(\#n)$.

end if.

end loop.

/* mora max semestre anterior

Loop #n=1 to 6.

do if $M(\#n) > M_{\text{ormaxS}}$.

compute $M_{\text{ormaxS}} = M(\#n)$.

end if.

end loop.

execute.

/* evolucion mora ultimo trimestre

compute $E_{\text{trimax}} = M_{\text{ormax3}} - M_{\text{ormaxT}}$.

/* evolucion de mora ultimo semestre

compute $E_{\text{semmax}} = M_{\text{ormax6}} - M_{\text{ormaxS}}$.

execute.

/* veces que supero la mora de 0, 30, 60 y 90 días en 3 meses

vector M = edadv306 to edadv405.

compute Nmo0_3 =0.

compute Nmo30_3=0.

compute Nmo60_3=0.

compute Nmo90_3=0.

Loop #n=10 to 12.

do if M(#n)>0.

compute Nmo0_3=Nmo0_3+1.

end if.

do if M(#n)>30.

compute Nmo30_3=Nmo30_3+1.

end if.

do if M(#n)>60.

compute Nmo60_3=Nmo60_3+1.

end if.

do if M(#n)>90.

compute Nmo90_3=Nmo90_3+1.

end if.

end loop.

execute.

/*veces que supero la mora 0, 30, 60 y 90 dias en 6 meses

vector M = edadv306 to edadv405.

```
compute Nmo0_6=0.
```

```
compute Nmo30_6=0.
```

```
compute Nmo60_6=0.
```

```
compute Nmo90_6=0.
```

```
loop #n=7 to 12.
```

```
do if m(#n)>0.
```

```
compute Nmo0_6=Nmo0_6+1.
```

```
end if.
```

```
do if m(#n)>30.
```

```
compute Nmo30_6=Nmo30_6+1.
```

```
end if.
```

```
do if m(#n)>60.
```

```
compute Nmo60_6=Nmo60_6+1.
```

```
end if.
```

```
do if m(#n)>90.
```

```
compute Nmo90_6=Nmo90_6+1.
```

```
end if.
```

```
end loop.
```

```
execute.
```

```
/*veces que supero la mora 0, 30, 60 y 90 dias en 12 meses
```

```
vector M = edadv306 to edadv405.
```

compute Nmo0_12=0.

compute Nmo30_12=0.

compute Nmo60_12=0.

compute Nmo90_12=0.

loop #n=1 to 12.

do if m(#n)>0.

compute Nmo0_12=Nmo0_12+1.

end if.

do if m(#n)>30.

compute Nmo30_12=Nmo30_12+1.

end if.

do if m(#n)>60.

compute Nmo60_12=Nmo60_12+1.

end if.

do if m(#n)>90.

compute Nmo90_12=Nmo90_12+1.

end if.

end loop.

execute.

/*-----Saldos-----*/

/* Saldo Promedio ultimos 3 meses y numero de meses con saldo

vector s = Saldo306 to Saldo405.

compute salpro3=0.

compute nsal3=0.

loop #n = 10 to 12.

do if s(#n)>0.

compute salpro3= salpro3 + s(#n).

compute nsal3 =nsal3 + 1.

end if.

end loop.

compute salpro3=salpro3/3.

execute.

/* Saldo Promedio ultimos 6 meses y numero de meses con saldo

vector s = Saldo306 to Saldo405.

compute salpro6=0.

compute nsal6=0.

loop #n = 7 to 12.

do if s(#n)>0.

compute salpro6= salpro6 + s(#n).

compute nsal6 =nsal6 + 1.

end if.

end loop.

compute salpro6=salpro6/6.

execute.

```
/* Saldo Promedio ultimos 12 meses y numero de meses con saldo
```

```
vector s = Saldo306 to Saldo405.
```

```
compute salpro12=0.
```

```
compute nsal12=0.
```

```
loop #n = 1 to 12.
```

```
do if s(#n)>0.
```

```
compute salpro12= salpro12 + s(#n).
```

```
compute nsal12 =nsal12 + 1.
```

```
end if.
```

```
end loop.
```

```
compute salpro12=salpro12/12.
```

```
execute.
```

```
/* Saldo Promedio trimestre anterior
```

```
vector s = Saldo306 to Saldo405.
```

```
compute SalproT=0.
```

```
loop #n = 7 to 9.
```

```
do if s(#n)>0.
```

```
compute salproT= salproT + s(#n).
```

```
end if.
```

```
end loop.
```

```
compute salproT=salproT/3.
```

```
Execute.
```

```
/* Saldo Maximo ultimos tres meses y mes con mayor saldo
```

```
vector s = Saldo306 to Saldo405.
```

```
compute Salmax3=0.
```

```
compute Lsmax3=0.
```

```
loop #n=10 to 12.
```

```
do if s(#n)>salmax3.
```

```
compute salmax3=s(#n).
```

```
compute lsmax3=#n.
```

```
end if.
```

```
end loop.
```

```
Execute.
```

```
/* Saldo Maximo ultimos seis meses y mes con mayor saldo
```

```
vector s = Saldo306 to Saldo405.
```

```
compute Salmax6=0.
```

```
compute Lsmax6=0.
```

```
loop #n=7 to 12.
```

```
do if s(#n)>salmax6.
```

```
compute salmax6=s(#n).
```

```
compute lsmax6=#n.
```

```
end if.
```

```
end loop.
```

```
execute.
```

```
/* Saldo Maximo ultimos doce meses y mes con mayor saldo
```

```
vector s = Saldo306 to Saldo405.
```



```

compute Salmax12=0.
compute Lsmax12=0.

loop #n=1 to 12.
do if s(#n)>salmax12.
compute salmax12=s(#n).
compute lsmax12=#n.
end if.
end loop.
execute.

/* Saldo Maximo trimestre anterior
vector s = Saldo306 to Saldo405.
compute SalmaxT=0.

loop #n=7 to 9.
do if s(#n)>salmaxT.
compute salmaxT=s(#n).
end if.
end loop.
execute.

/*evolucion de saldos-----

vector s = Saldo306 to Saldo405.
compute s12_11=0.
compute s12_sm12=0

```

```
do if s(11)>0.  
compute s12_11= trunc(100*s(12)/s(11)).  
end if.
```

```
do if (salmax12>0).  
compute s12_sm12= trunc(100*s(12)/salmax12).  
end if.  
execute.
```

```
/*-----relaciones cupo-----  
-----
```

```
compute salcup=0.  
compute salmcup=0.  
compute compcup=0.  
compute pagcup=0.
```

```
Do if cupc1405>0.  
compute salcup=trunc(100*saldo405/cupc1405).  
compute salmcup=trunc(100*salmax12/cupc1405).  
compute compcup=trunc(100*(comp0405+compd405)/cupc1405).  
compute pagcup=trunc(100*(pago0405+pagod405)/cupc1405).  
end if.  
execute.
```

```
/*-----Edad  
compute #a_nac = trunc(fecna8/10000) .
```

```
compute edad = trunc((12*(2004 - #a_nac) + 3 - (trunc(fecna8/100) - 100*#a_nac))/12)
```

```
.
```

```
Do if ((edad < 18) | (edad >85)) .
```

```
compute edad = -1.
```

```
end if.
```

```
Execute.
```

```
/******
```

```
/** Antigüedad de la cuenta
```

```
compute #a_aper = trunc(fecap8/10000) .
```

```
compute antcta_m = 12*(2004 - #a_aper) + 3 - (trunc(fecap8/100) - 100*#a_aper) .
```

```
compute antcta_a = trunc( antcta_m/12) .
```

```
execute.
```

```
/******
```

```
/** Ultimo cambio cupo
```

```
compute #a_ccup = trunc(cupam8/10000) .
```

```
compute ulcacu_m = 12*(2004 - #a_ccup) + 3 - (trunc(cupam8/100) - 100*#a_ccup) .
```

```
compute ulcacu_a = trunc(ulcacu_m/12) .
```

```
EXECUTE.
```

```
/*-----Compras -----
```

```
/* Resta comprot - compdif para aislar el efecto de cuotas cargadas.
```

```
COMPUTE Compra306 = comp0306-compd306 .
```

```
COMPUTE Compra307 = comp0307-compd307 .
```

```
COMPUTE Compra308 = comp0308-compd308 .
```

```
COMPUTE Compra309 = comp0309-compd309 .
COMPUTE Compra310 = comp0310-compd310 .
COMPUTE Compra311 = comp0311-compd311 .
COMPUTE Compra312 = comp0312-compd312 .
COMPUTE Compra401 = comp0401-compd401 .
COMPUTE Compra402 = comp0402-compd402 .
COMPUTE Compra403 = comp0403-compd403 .
COMPUTE Compra404 = comp0404-compd404 .
COMPUTE Compra405 = comp0405-compd405 .
```

execute.

```
/* Suma total compras ultimos 3 meses y numero de meses con compra
```

```
vector c = compra306 to compra405.
```

```
compute comsum3=0.
```

```
compute ncomp3=0.
```

```
compute lucom3=0.
```

```
loop #n = 10 to 12.
```

```
do if c(#n)>0.
```

```
compute comsum3= comsum3 + c(#n).
```

```
compute ncomp3 =ncomp3 + 1.
```

```
Compute lucom3 = #n .
```

```
end if.
```

```
end loop.
```

```
execute.
```

```
/* Suma total compras ultimos 6 meses y numero de meses con compra  
vector c = compra306 to compra405.
```

```
compute comsum6=0.
```

```
compute ncomp6=0.
```

```
compute lucom6=0.
```

```
loop #n = 7 to 12.
```

```
do if c(#n)>0.
```

```
compute comsum6= comsum6 + c(#n).
```

```
compute ncomp6 =ncomp6 + 1.
```

```
compute lucom6=#n.
```

```
end if.
```

```
end loop.
```

```
execute.
```

```
/* Suma total compras ultimos 12 meses y numero de meses con compra  
vector c = compra306 to compra405.
```

```
compute comsum12=0.
```

```
compute ncomp12=0.
```

```
compute lucom12=0.
```

```
loop #n = 1 to 12.
```

```
do if c(#n)>0.
```

```
compute comsum12= comsum12 + c(#n).
```

```
compute ncomp12 =ncomp12 + 1.
```

```
compute lucom12=#n.
```

```
end if.
```

```
end loop.
```

```
execute.
```

```
/* compra Maximo ultimos tres meses y mes con mayor compra
```

```
vector c = compra306 to compra405.
```

```
compute commax3=0.
```

```
compute Lcommax3=#n.
```

```
loop #n=10 to 12.
```

```
do if c(#n)>commax3.
```

```
compute commax3=c(#n).
```

```
compute lcommax3=#n.
```

```
end if.
```

```
end loop.
```

```
Execute.
```

```
/* Compra Maxima ultimos seis meses y mes con mayor compra
```

```
vector c = compra306 to compra405.
```

```
compute commax6=0.
```

```
compute Lcmax6=#n.
```

```
loop #n=7 to 12.
```

```
do if c(#n)>commax6.
```

```
compute commax6=c(#n).
```

```
compute lcmax6=#n.
```

end if.

end loop.

execute.

/* Compra Maxima ultimos doce meses y mes con mayor compra

vector c = compra306 to compra405.

compute commax12=0.

compute Lcmax12=0.

loop #n=1 to 12.

do if c(#n)>commax12.

compute commax12=c(#n).

compute lcmax12=#n.

end if.

end loop.

execute.

/*cupo-----

Do if (cupc1405 > 0) .

compute cmxcup3 = trunc(100*commax3 / cupc1405) .

compute smccup3 = trunc(100*comsum3 / cupc1405) .

compute cmxcup6 = trunc(100*commax6 / cupc1405) .

compute smccup6 = trunc(100*comsum6 / cupc1405) .

compute cmxcup12 = trunc(100*commax12 / cupc1405) .

compute smccup12 = trunc(100*comsum12 / cupc1405) .

end if.

```

/* Evolucion saldo

compute esal_cup=0.
do if (cupc1405>0).
compute esal_cup=trunc(100*(salpro3-salprot)/cupc1405).
End if.
execute.

```

Anexo III

Script Variables de Comportamiento en Compras

```

/*****total de compras Ultimos 12 meses*****/

```

```

vector CompraTotal = Co_0809 to Co_0908.

```

```

compute Compra12M=0.

```

```

loop #n = 1 to 12.

```

```

    do if CompraTotal(#n)>0.

```

```

        compute Compra12M=Compra12M+1.

```

```

    end if.

```

```

end loop.

```

```

execute.

```

```

/*****Compro mes actual*****/

```

```

vector CompraTotal = Co_0809 to Co_0908.

```

```

Compute Compra_Act = 0.

```



```

compute Compra_Ant = 0.
loop #n = 1 to 12.
    do if CompraTotal(12) >0.
        Compute Compra_Act = 1.
    end if.
    do if CompraTotal(11) >0.
        Compute Compra_Ant = 1.
    end if.
end loop.
execute.

```

/******total de compras Ultimos 6 meses*****

```

vector CompraTotal = Co_0809 to Co_0908.
compute Compra6M=0.
loop #n = 7 to 12.
    do if CompraTotal(#n)>0.
        compute Compra6M=Compra6M+1.
    end if.
end loop.

execute.

```

/******total de compras Ultimos 3 meses*****

```

vector CompraTotal = Co_0809 to Co_0908.
compute Compra3M=0.

```

```
loop #n = 10 to 12.  
    do if CompraTotal(#n)>0.  
        compute Compra3M=Compra3M+1.  
    end if.  
end loop.
```

```
execute.
```

```
/******Compra Anual*****
```

```
vector CompraTotal = Co_0809 to Co_0908.  
compute Compra_Anual=0.  
loop #n = 1 to 12.  
    do if CompraTotal(#n)>0.  
        compute Compra_Anual=Compra_Anual+CompraTotal(#n).  
    end if.  
end loop.
```

```
execute.
```

```
/******Compras último semestre*****
```

```
vector CompraTotal = Co_0809 to Co_0908.  
compute Compra_Semes=0.  
loop #n = 7 to 12.  
    do if CompraTotal(#n)>0.  
        compute Compra_Semes=Compra_Semes+CompraTotal(#n).  
    end if.  
end loop.
```

```
end if.  
end loop.
```

```
execute.
```

```
/******Compras último trimestre*****
```

```
vector CompraTotal = Co_0809 to Co_0908.
```

```
compute Compra_Trim=0.
```

```
loop #n = 10 to 12.
```

```
do if CompraTotal(#n)>0.
```

```
compute Compra_Trim = Compra_Trim+CompraTotal(#n).
```

```
end if.
```

```
end loop.
```

```
execute.
```

```
/******Compra Promedio Mensual*****
```

```
compute Compra_Pro_Men = Compra_Anual /12.
```

```
/******Compra anual para  
cupo*****
```

```
compute ComAn_Cupo = 0.
```

```
do if Cupo >0.
```

```
compute ComAn_Cupo = Compra_Anual / Cupo.
```

```
end if.
```

```
/******Ultima compra para el  
cupo*****
```

```
compute ComUlt_Cupo = 0.
```

```
do if Cupo >0.
```

```
compute ComUlt_Cupo = Co_0907/ Cupo.
```

```
end if.
```

```
/******Compra último trimestre para el cupo*****
```

```
compute ComUltTr_Cupo = 0.
```

```
do if Cupo >0.
```

```
compute ComUltTr_Cupo = Compra_Trim / Cupo.
```

```
end if.
```

```
/******Compra ultimo semestre para el cupo*****
```

```
compute ComUltSe_Cupo = 0.
```

```
do if Cupo >0.
```

```
compute ComUltSe_Cupo = Compra_Semes / Cupo.
```

```
end if.
```

```
execute.
```

```
VARIABLE LABELS ComAn_Cupo "Compra anual 'para' Cupo" .
```

VARIABLE LABELS ComUlt_Cupo "ultima-compra-para-cupo" .
VARIABLE LABELS ComUltSe_Cupo "Compra-ultimoSemes-Cupo" .
VARIABLE LABELS ComUltTr_Cupo "Compra-ultimoTrimes-Cupo" .
VARIABLE LABELS Compra_Pro_Men "Compra-prome-mensual" .
VARIABLE LABELS Compra12M "Numero-compras-12-meses" .
VARIABLE LABELS Compra_Act "Compro-mes-actual" .
VARIABLE LABELS Compra_Ant "Compro-mes-anterior" .
VARIABLE LABELS Compra6M "Numero-Compras-Ult-Semestre" .
VARIABLE LABELS Compra3M "Numero-Compras-Ult-trimestre" .
VARIABLE LABELS Compra_Anual "Monto-Compra-Anual" .
VARIABLE LABELS Compra_Semes "Monto-Compra-Ult-Semes" .
VARIABLE LABELS Compra_Trim "Monto-Compra-Ult-Trime" .

Anexo IV

Código Gams del Modelo

\$ title problema de asignación de gestores

option optcr=0

OPTION LIMCOL=0, LIMROW=0;

\$OFFLISTING;

OPTION SOLPRINT=OFF;

* primero se declaran y se definen los conjuntos

** el conjunto i recorre los origene

** el conjunto j recorre los destinos

Sets

i conjunto de cuentas vencidas /1*26891/

j numero de personal para asignar /j1*j28/

Parameters

S(i) saldo de las cuentas

/

\$include C:\Users\andres13d3\Desktop\Respaldo\parte 2\tesis\Programacion
Matemática\Dat\SaldoQ.txt

/

P(i) probabilidad de pago de cuenta en el mes

/

\$include C:\Users\andres13d3\Desktop\Respaldo\parte 2\tesis\Programacion
Matemática\Dat\ProbabilidadesQ.txt

/

T(i) Tiempo de demora de la gestion

/

\$include C:\Users\andres13d3\Desktop\Respaldo\parte 2\tesis\Programacion
Matemática\Dat\TiemposQ.txt

/

scalar HL total de minutos laborados por personal /1800/;

Variables

z valor de la funcion objetivo

x(i,j) si la cuenta i es asignada al gestor j;

Binary variables x(i,j);

equations

Saldo_Max funcion objetivo

Horas_Trab(j) horas máximas de trabajo

Aux(i) ecuacion de para no asignar una cuenta a dos gestores;

Saldo_Max.. z=e=Sum((i,j),x(i,j)*P(i)*S(i));

Horas_Trab(j).. Sum(i,x(i,j)*T(i))=l=HL;

Aux(i).. Sum(j,x(i,j))=l=1;

Model Cobranzas /Saldo_Max, Horas_Trab, Aux/;

solve Cobranzas using mip maximizing z;

Display x.l

Display z.l

File Res /"C:\Users\andres13d3\Desktop\Respaldo\parte 2\tesis\Programacion Matemática\reporte capacidad.txt"/

Put Res

"realizado por Hugo Ruiz"//

"Fecha: " System.date/

"Hora: " System.time/

Loop ((i,j) \$ (x.L(i,j)>0), Put "al agente agente: " j.tl:7 "debe asignarse a la cuenta " i.tl:7 /)

Anexo V

Tabla Correlaciones Variables Modelo Regresión Logística

		Correlaciones					
		gestion	total_pr	mmormax6	total_f	total_m	total_np
gestion	Correlación de Pearson	1	.036	.107	.155	.060	-.005
	Sig. (bilateral)		.000	.000	.000	.000	.597
	N	9540	9540	9540	9540	9540	9540
total_pr	Correlación de Pearson	.036	1	.046	.107	.034	.052
	Sig. (bilateral)	.000		.000	.000	.001	.000
	N	9540	9540	9540	9540	9540	9540
mmormax6	Correlación de Pearson	.107	.046	1	.531	.141	-.242
	Sig. (bilateral)	.000	.000		.000	.000	.000
	N	9540	9540	9540	9540	9540	9540
total_f	Correlación de Pearson	.155	.107	.531	1	.347	-.003
	Sig. (bilateral)	.000	.000	.000		.000	.762
	N	9540	9540	9540	9540	9540	9540
total_m	Correlación de Pearson	.060	.034	.141	.347	1	.023
	Sig. (bilateral)	.000	.001	.000	.000		.026
	N	9540	9540	9540	9540	9540	9540
total_np	Correlación de Pearson	-.005	.052	-.242	-.003	.023	1
	Sig. (bilateral)	.597	.000	.000	.762	.026	
	N	9540	9540	9540	9540	9540	9540