

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



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

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

**“MAGÍSTER EN ESTADÍSTICA CON MENCIÓN EN GESTIÓN DE
LA PRODUCTIVIDAD Y CALIDAD”**

TEMA:

DESARROLLO DE UN MODELO DE PUNTUACIÓN PARA
CRÉDITOS EN UNA EMPRESA COMERCIAL.

AUTOR:

SINDY VICTORIA MACIAS CABRERA

Guayaquil - Ecuador

2020

RESUMEN

Este trabajo desarrolla el proceso para la elaboración de un modelo de perfilamiento o puntuación para créditos en una empresa comercial. Para ello se utilizan tres técnicas, las mismas que se buscan comparar para reconocer la potencia de predictibilidad y ajuste. Las técnicas que se usan son Árboles de Decisión, Análisis Discriminante y Regresión Logística, de estas se obtienen los indicadores de bondad de ajuste, logrando las puntuaciones y segmentación de los prospectos basados en la característica de ser o no ser un cliente moroso. Se realiza la comparación y validación de cada uno de los modelos obtenidos por estas tres técnicas, basados principalmente en el estadístico de Kolmogórov-Smirnov, se realiza la gráfica de Curvas COR entre los resultados del Análisis Discriminante y Regresión Logística. Posterior a la clasificación o segmentación se determinan grupos de riesgo usando la Calificación Buró de Crédito y con esto se efectúa el proceso de determinación de una línea de crédito, basado principalmente en factores de riesgo e ingresos.

Palabras Clave: Análisis Discriminante, Árboles de Decisión, Grupos de Riesgo, Línea de Crédito, Regresión Logística.

ABSTRACT

This paper develops the process for the development of a profiling or scoring model for credits in a commercial company. For this, three techniques are used, the same ones that are sought to compare to recognize the power of predictability and adjustment. The techniques used are Decision Trees, Discriminant Analysis and Logistic Regression, from these the indicators of goodness of fit are obtained, achieving the scores and segmentation of the prospects based on the characteristic of being or not being a delinquent client. The comparison and validation of each of the models obtained by these three techniques is carried out, based mainly on the Kolmogórov-Smirnov statistic, the COR Curves graph is made between the results of the Discriminant Analysis and Logistic Regression. After classification or segmentation, risk groups are determined using the Credit Bureau Rating and with this the process of determining a credit line is carried out, based mainly on risk factors and income.

Keywords: Credit Line, Decision Trees, Discriminant Analysis, Logistic Regression, Risk Groups.

DEDICATORIA

Dedico este trabajo a todos aquellos que busquen orientación en cuanto a modelación, segmentación o perfilamiento de prospectos de clientes.

AGRADECIMIENTO

Agradezco a mis padres que han sido pilar fundamental en mi desarrollo personal y académico.

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación, me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo, corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

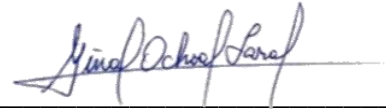


Sindy Victoria Macías Cabrera

TRIBUNAL DE GRADUACIÓN



Omar Ruiz Barzola, Ph.D.
PRESIDENTE



Mgtr. Gina Ochoa Jara
DIRECTOR



Mgtr. Francisco Moreira Villegas
VOCAL 1



Johny Pambabay Calero, Ph.D.
VOCAL 2

TABLA DE CONTENIDO

LISTADO DE FIGURAS	X
LISTADO DE TABLAS	XI
CAPÍTULO 1.....	1
1. INTRODUCCIÓN.....	1
1.1. Descripción del problema.....	2
1.2. Objetivos	5
1.2.1. Objetivos específicos	5
1.3. Alcance.....	5
CAPÍTULO 2.....	6
2. MARCO TEORICO.....	6
2.1. Introducción	6
2.2. Proceso de captación de clientes	7
2.2.1. Prospección de futuros clientes.....	7
2.2.2. Aprobación o Negación.....	7
2.3. Modelos scoring.....	8
2.4. Riesgo empresarial	8
CAPÍTULO 3.....	12
3. METODOLOGÍAS	12
3.1. Análisis Discriminante.....	12
3.1.1. Función Discriminante.....	12
3.1.2. Estadístico Lambda de Wilks	13
3.2. Regresión Logística	14
3.2.1. Criterio de Máxima Verosimilitud.....	16
3.2.2. Estadístico de Wald.....	16
3.3. Árboles de Decisión	17
3.3.1. Estadístico K-S (Kolmogorov – Smirnov).....	19
3.4. Muestra y Selección de Variables	20
CAPÍTULO 4.....	22
4. ELABORACION DE MODELOS Y RESULTADOS	22
4.1. Elaboración del Modelo por Árboles de Decisión	22
4.2. Elaboración del Modelo por Análisis Discriminante	25
4.3. Elaboración del Modelo por Regresión Logística.....	28
4.4. Comparación de los modelos	31
4.5. Curvas COR	32
4.6. Esquema de Matriz de Riesgo de los nuevos prospectos.....	33

CAPÍTULO 5.....	35
5. CONCLUSIONES Y RECOMENDACIONES.....	35
6. REFERENCIAS	36

LISTADO DE FIGURAS

Figura 3.1 Gráfica de la función Logística	15
Figura 3.3.1 Árbol de decisión “Riesgo en la concesión del Crédito”	18
Figura 3.3.2 Ejemplo de Estadístico K-S.....	19
Figura 4.1.1 Participación por perfiles – Árboles de Decisión.....	23
Figura 4.1.2 Porcentaje de Malos por segmentos de Riesgo	24
Figura 4.1.3 Distribuciones Acumuladas – Estadístico K-S.....	24
Figura 4.2.1 Porcentaje de Distribución de la muestra – Análisis Discriminante	26
Figura 4.2.2 Porcentaje de Malos por segmento de riesgo – Análisis Discriminante	27
Figura 4.2.3 Distribuciones Acumuladas por Análisis Discriminante – Estadístico K-S	27
Figura 4.3.1 Porcentaje de Distribución de la muestra – Regresión Logística ..	30
Figura 4.3.2 Porcentaje de Malos por segmento de riesgo – Regresión Logística	30
.....	30
Figura 4.3.3 Distribuciones Acumuladas por Análisis Discriminante – Estadístico K-S	31
Figura 4.5.1 Gráfica de Curva COR	32

LISTADO DE TABLAS

Tabla 3.4.1 Descripción de Variables.....	21
Tabla 4.1.1 Segmentación mediante Árboles de Decisión	23
Tabla 4.2.1 Coeficientes del Análisis Discriminante	25
Tabla 4.2.2 Agrupación de los Puntajes del Análisis Discriminante.....	25
Tabla 4.3.1 Indicadores de Ajuste del Modelo.....	28
Tabla 4.3.2 Prueba de Máxima Verosimilitud	28
Tabla 4.3.3 Pruebas de significancia individual.....	29
Tabla 4.3.4 Agrupación de las probabilidades de Regresión Logística.	29
Tabla 4.4.1 Estadístico KS por cada técnica	31
Tabla 4.6.1 Tipos de Riesgo	33
Tabla 4.6.2 Estructura del Gasto Corriente de consumo monetario Mensual ...	34
Tabla 4.6.3 Tabla de factores por segmento de Riesgo	34

CAPÍTULO 1

1. INTRODUCCIÓN

Las compañías buscan en la actualidad el incremento de sus ventas, para esto han optado por otorgar facilidades de consumo a sus clientes, incursionado así en el otorgamiento de “Crédito Directo”, con el cual el cliente puede generar una deuda a corto, mediano y largo plazo, por lo general hasta 24 meses. Los créditos generan ingresos adicionales importantes tales como: intereses por financiamiento, gastos de cobro, emisión de estados de cuenta, etc., de los cuales unos dependen del monto de la deuda adquirida y otros simplemente por tener el crédito.

En general las compañías comerciales en Ecuador se basan en algunos criterios sociodemográficos y comportamiento crediticio para generar un crédito al Solicitante. La compañía analiza estos criterios de forma subjetiva y así otorga dicho crédito, basado en un modelo muy conocido llamado “Las 5 C del Crédito”, las mismas que detallan lo siguiente:

CAPACIDAD

El solicitante debe tener la “capacidad” de pago suficiente para poder estar al día en las obligaciones contraídas, esto se refiere a los ingresos del mismo.

CAPITAL

El “capital” del solicitante son todos los bienes que registre.

COLATERAL

Se refiere a las garantías personales por parte de personas con solvencia moral y material.

CARÁCTER

Esta característica depende mucho del historial de crédito del solicitante, de las referencias bancarias y comerciales.

CONVENIENCIA

Se refiere a que tanto el prestatario, como el solicitante, deben estar en posibilidades de obtener un rendimiento adecuado de los créditos otorgados.

Los analistas de crédito, consideran estas características y la experiencia para otorgar o no un “Crédito Directo” de forma subjetiva, este proceso, ¿en qué medida repercute en el riesgo?, ¿se está otorgando el monto de crédito adecuado?, ¿se está clasificando adecuadamente a nuestros clientes?, estas son algunas de las interrogantes que le surge a la compañía.

De acuerdo con (Rayo, Lara, & Camino, 2010) una forma de aclarar estas preguntas es disponer de un “modelo de puntuación crediticia” el cual permita medir la probabilidad de que un prospecto llegue a ser un cliente de alto o bajo riesgo.

La compañía busca desarrollar un sistema con base científica que le permita identificar de forma objetiva a los clientes de mejor propensión de pago, minimizando el riesgo y ahorrando tiempo en los procesos de concesión de los créditos.

1.1. Descripción del problema

El objetivo principal de la compañía es la venta de artículos de ropa, hogar y tecnología, es por esto que para incrementar el flujo de las ventas ha otorgado créditos y así los clientes puedan diferir sus compras.

Los analistas otorgan el crédito directo a los prospectos de clientes, basándose en políticas de comportamiento crediticio (Buró de Crédito), y otras condicionantes las cuales son evaluadas por un área específica, dicha

aprobación o negación se realiza basada en la experiencia y de forma empírica, para la compañía el nivel de morosidad de los clientes es uno de los indicadores de mayor importancia que se miden diario y mensualmente.

Se ha notado que en los últimos meses el indicador de morosidad que registran los clientes nuevos a partir del cuarto mes de ser aprobado el crédito se ha incrementado, por esto y por motivos de implementación de técnicas científicas en la toma de decisiones se cree necesario el desarrollo de un nuevo modelo que profile o discrimine a los mejores prospectos (clientes que provoquen menor morosidad) de aquellos considerados de alto riesgo.

Para el desarrollo del modelo que permita a la compañía discriminar a los solicitantes del “Crédito Directo” entre los de mejor propensión de pago versus los de menor y con base científica, existen varias técnicas estadísticas, entre ellas Árboles de Decisión, Regresión Logística, Análisis Discriminantes y algunas otras.

Quizás unas de las técnicas de mayor uso para este tipo de segmentación son los Árboles de Decisión, (Cardona Hernández, 2004) en su artículo realiza un estudio en el que busca determinar la importancia de la aplicación de árboles de decisión para el cálculo de probabilidades de incumplimiento en crédito que enfrenta una institución financiera, con esto obtuvo un modelo confiable de alta potencia utilizando medidas como la de bondad de ajuste $K - S$ y otras, la importancia de tener un modelo de alta capacidad de discriminación radica en que esta impacta en el cálculo de las provisiones afectando directamente al balance y las utilidades que podría tener la compañía.

Sin embargo se pueden usar otras metodologías, es así (Ruiz Lopez, 2017) realiza un estudio para el diseño de un modelo matemático para la calificación de clientes morosos en una entidad comercial utilizando Árboles de Decisión y otras metodologías como Regresión Logística y Análisis

Discriminante, de este obtuvo que estas tres metodologías resultan estables; sin embargo, Árboles de Decisión tuvo menor potencia, lo favorable de éste es que describe claramente los perfiles de clientes en cada grupo de segmentación a diferencia de las otras dos metodologías que sugieren probabilidades.

Otro de los artículos que se toma como referencia bibliográfica corresponde al de los autores Rayo, Lara y Camino (2010) en el que proponen diseñar un modelo credit scoring para una institución especializada en créditos y analizar el riesgo del incumplimiento del pago de los microcréditos, utilizando regresión logística obtuvieron un modelo que pudo predecir correctamente el 78% de los créditos de dicha empresa y además sirve como herramienta de análisis para los analistas de crédito que deciden a quien otorgar los mismos.

Andrea Valencia en su proyecto de tesis realizado en 2017 tiene como objetivo usar la metodología de Regresión Logística y entender a través de la construcción de un modelo Scoring, cuáles son las posibles causas más comunes de que una pyme pague oportunamente o no, lo cual conllevaría a un equilibrio entre la búsqueda de maximización de utilidad de una entidad financiera, que en su mayoría es a través de la colocación de recursos y el nivel de riesgo que se desea asumir. El resultado de la investigación concluye que el tamaño de una pyme medida a través de las ventas y el activo resultan ser variables indispensables y con una relación directa al comportamiento de pago de una pyme.

Los autores Barrientos, Cruz y Acosta (2009) realizaron un estudio cuyo objetivo fue evaluar el desempeño de tres tipos de algoritmos para construir árboles de decisión sobre evaluación del diagnóstico de cáncer de mama en dos tipos de información, una que refiere a los diagnósticos de un “solo especialista” y otra que contiene diagnósticos de “varios médicos”, en la investigación plantean un modelo de árboles de decisión usando los algoritmos ID3, J48 y Naive Bayes, teniendo como resultado que este último

algoritmo resulta mayormente eficaz para ambos tipos de información; sin embargo, se identificó además que los árboles de decisión clasifican mejor si la información sobre las sintomatologías corresponden a un único especialista.

1.2. Objetivos

El objetivo general de este proyecto es elaborar un modelo que permita la separación de clientes que presentan mayor propensión de pago, identificando las características comunes.

1.2.1. Objetivos específicos

Los objetivos específicos que se busca desarrollar para lograr un entendimiento completo son:

- Desarrollar una matriz de riesgo que diferencie grupos de mayor riesgo en los nuevos prospectos que apliquen al crédito.
- Proponer directrices para la asignación del monto del crédito a los nuevos prospectos.
- Comparar la potencia entre las técnicas estadísticas de árboles de decisión, análisis discriminante y regresión logística.

1.3. Alcance

El estudio abarcará el análisis de los clientes que tienen aprobado el crédito durante el período (Jul – Dic 2017), que fueron captados en las ciudades de Guayaquil y Quito.

CAPÍTULO 2

2. MARCO TEORICO

2.1. Introducción

Los negocios comerciales medianos o grandes buscan vender sus productos al mejor precio de la mejor manera, por lo general con la venta en efectivo o con tarjetas de crédito, en cualquiera de los dos casos la empresa no incurría en riesgo alguno. Sin embargo, los medianos y grandes negocios vieron la oportunidad de generar otro tipo de ingreso a través del “Crédito”.

El poder otorgar crédito a los clientes conlleva no solo a la oportunidad de generar el ingreso de los intereses del financiamiento por la venta, sino más bien el riesgo que debe asumir la empresa en el caso del “no pago” de la deuda adquirida por el cliente. Desde siempre se ha considerado el juicio del “evaluador u otorgador” del crédito en la decisión de quien merece una línea de crédito, basados en herramientas como el comportamiento de cliente prospecto en el “Buró de Crédito”.

Mas sin embargo este tipo de evaluaciones están sujetas a un comportamiento externo de este prospecto en otro tipo de empresas, que si bien es cierto es una variable muy fuerte y válida en el momento de la evaluación de su solicitud de crédito, no está completamente ajustada al propio negocio.

La actividad principal de la compañía comercial es la venta al por menor de productos de vestir, hogar y tecnología, utilizando todos los medios de pagos (Efectivo, Cheques, Tarjetas de Créditos), ya desde hace algunos años la

compañía otorga créditos a los clientes para que realicen sus compras dentro de los locales.

Para otorgar un crédito existe un proceso actual en el mismo que se aprueba o se niega las aplicaciones de los interesados, el cual se basa principalmente en lo siguiente:

- Puntuación favorable en el Buró de Crédito
- Tipo de Vivienda (Propia, Con Parientes, Arrendada)
- Posee Tarjeta de Crédito
- Edad del cliente
- Sector donde se domicilia
- Estabilidad laboral

Estas son las principales características que se consideran al momento de evaluar a un futuro cliente (prospecto) para proporcionar el crédito.

2.2. Proceso de captación de clientes

Se detallan los conceptos y procedimientos necesarios para comprender el proceso que abarca la obtención de un crédito

2.2.1. Prospección de futuros clientes

Un área específica de la compañía se dedica a buscar futuros clientes, llenando un formulario con:

- Datos Personales
- Actividad económica
- Referencias personales y bancarias
- Evaluación del comportamiento crediticio y su respectivo consentimiento

2.2.2. Aprobación o Negación

Posterior a la recepción del formulario de prospección, este ingresa al sistema para que sea analizado por el área específica de verificación, la

misma que se dedica a analizar y corroborar la información ingresada y con base a las políticas.

2.3. Modelos scoring

La empresa en estudio tiene controlado y ampliamente comprendido como asignar un crédito a un cliente, sin embargo, considera que debe desarrollarse un modelo estadístico para la selección adecuada de clientes basada en fundamentos técnicos y no empíricos, como ofrece actualmente.

Un concepto de los que son Modelos Scoring lo detalla (Valencia, 2017) como “una de las herramientas más utilizadas en los últimos años para la medición y control del riesgo. Los modelos de Credit Scoring, también denominados scored-cards o clasifiers, son “métodos estadísticos utilizados para clasificar a los solicitantes de crédito, o incluso a quienes ya son clientes de la entidad evaluadora, entre las clases de riesgo bueno o malo”.

Existen algunas técnicas estadísticas y econométricas para el desarrollo de los Modelos Scoring, entre estas están Árboles de decisión, Redes Neuronales, Análisis Discriminante, Regresión Logística, algunas de estas técnicas nos otorgan puntajes o probabilidades y otras categorías, para clasificar e identificar si el prospecto es bueno o malo para otorgarle crédito.

Los Modelos Scoring nos brindan un esquema basado en técnica estadística para la clasificación o categorización de los clientes identificando cuales de estos tendrán mayor riesgo para la empresa, cuidando así el valor de la línea del crédito que se pueda otorgar.

2.4. Riesgo empresarial

Todas las empresas siempre generan un riesgo en el tipo de negocio que gestionen. El riesgo empresarial o de un negocio es la posibilidad de que los flujos en efectivo de una empresa sean insuficientes para cubrir los gastos de operación. Los gastos de operación son los que una empresa

ocupa al realizar sus operaciones normales. Estos incluyen los salarios, el alquiler, las reparaciones, los impuestos, el transporte y la venta, los gastos administrativos y los gastos generales. Sin el flujo de caja suficiente para pagar estos gastos, las empresas se vuelven más propensas a fallar. El riesgo empresarial se refiere a la probabilidad de que esto ocurra y se divide en dos tipos: el riesgo sistemático y el riesgo no sistemático. (Guzman, 2020)

Riesgo Sistemático

El riesgo sistemático, también conocido como "riesgo de mercado" o "riesgo no diversificable", engloba al conjunto de factores económicos, monetarios, políticos y sociales que provocan las variaciones de la rentabilidad de un activo.

Riesgo No Sistemático

El riesgo no sistemático, también conocido como "riesgo diversificable", engloba al conjunto de factores propios de una empresa o industria, y que afectan solo a la rentabilidad de su acción o bono.

En otras palabras, el riesgo no sistemático surge de la incertidumbre que rodea a una empresa por el desarrollo de su negocio, ya sea por las propias circunstancias de la empresa o por las del sector al que pertenezca. (Teruel, 2016)

Dentro del riesgo No Sistemático tenemos el Riesgo Crediticio el mismo que se define como el riesgo.

Riesgo Crediticio

El riesgo de crédito es la posibilidad de sufrir una pérdida como consecuencia de un impago por parte de nuestra contrapartida en una operación financiera, es decir, el riesgo de que no nos pague. (Ucha, 2020).

Las empresas soportan el riesgo de crédito cuando, tras la venta a plazo de un producto, el cliente termina impagado. Para evitar este tipo de situaciones, las empresas lo que hacen es contratar servicios externos que estudian el riesgo de crédito de los clientes que solicitan una compra-venta a plazo. (MyTripleA, 2020)

En vista de este tipo de riesgo se decide desarrollar Modelos Scoring para la mitigación o disminución de dicho riesgo con una mejor clasificación de los clientes de acuerdo con el comportamiento de pago en la empresa.

Algunas de las técnicas estadísticas que se utilizarán son Análisis Discriminante, Regresión Logística y Árboles de Decisión, esta última será la técnica principal que se desarrollará en este proyecto. Estas metodologías de análisis son usadas principalmente para la clasificación o identificación de comportamientos ya sea esto con personas, productos entre otros.

El Análisis Discriminante es una técnica estadística inferencial multivariante de clasificación que permite identificar grupos de acuerdo a la característica de interés, compuesta de la variable dependiente (los grupos de estudio) y de las variables independientes (las características de los grupos). El objetivo de esta técnica es encontrar la mejor combinación lineal de las variables independientes que permitan discriminar a los grupos, una vez que se determinó esta combinación o ya la llamada función discriminante, con esto se podrá clasificar los nuevos casos.

Esta técnica al igual que otras técnicas inferenciales se las puede aplicar en diversas áreas, especialmente la médica, así como las financieras y otras más. En recursos humanos se utiliza para filtrar las hojas de vida previo a la entrevista final. En el sector financiero se lo utiliza para identificar riesgos crediticios y en el área de seguros para predecir la siniestralidad. (Natividad, 2018)

Otra de las técnicas que se revisará en este proyecto es Regresión Logística, metodología multivariante que busca identificar y clasificar las características de los grupos de estudio. Al igual que el Análisis Discriminante está basada en un esquema de regresión lineal generalizada.

La Regresión Logística es una de las técnicas estadístico-inferenciales más empleadas en la producción científica contemporánea. Surge en la década del 60, su generalización dependía de la solución que se diera al problema de la estimación de los coeficientes. El algoritmo de Walker-Duncan para la obtención de los estimadores de máxima verosimilitud vino a solucionar en parte este problema, pero era de naturaleza tal que el uso de computadoras era imprescindible. (Fernández, 2011).

La metodología de Árboles de Decisión es una técnica inferencial No paramétrica que permite clasificar la población en segmentos o perfiles. Esta técnica tiene algunas ventajas, al ser una técnica no paramétrica la población no necesita ajustarse a una distribución estadística. Sin embargo, una de sus desventajas es la rigidez en su estructura es decir que un cambio mínimo en comportamiento de la población de estudio requerirá la elaboración de un nuevo árbol o modelo.

CAPÍTULO 3

3. METODOLOGÍAS

3.1. Análisis Discriminante

Es una técnica estadística de análisis multivariante que permite determinar si existe diferencias significativas entre los grupos planteados o variable dependiente (en el caso de este análisis Bueno o Malo), utilizando variables independientes o discriminantes las mismas que deben ser numéricas o continuas.

El Análisis Discriminante es una regresión lineal, con lo que se busca encontrar combinaciones lineales de las variables continuas que discriminen en grupo los entes. Este análisis determina una regla de asignación o decisión que permita establecer la pertenencia de un nuevo ente a los grupos ya establecidos a priori bajo cierto riesgo; cabe indicar que es desconocida la pertenencia del nuevo ente. (Fernández, 2011)

3.1.1. Función Discriminante

La función discriminante es la combinación lineal de las variables independientes de mayor potencia y discriminación que permitirá asignar a los nuevos entes u objetos a cada grupo.

El número de funciones discriminantes a obtener depende del número de grupos definidos por la variable dependiente, ya que se obtienen tantas funciones como grupos menos uno.

Se plantea la siguiente expresión para la función discriminante:

$$f_{kn} = \lambda_0 + \lambda_1 x_{1kn} + \lambda_2 x_{2kn} + \dots + \lambda_p x_{pkn}$$

Donde f_{kn} es la puntuación en la función discriminante para el caso n en el grupo k ; x_{ikn} es el valor de la variable discriminante para el caso n en el grupo k y λ_i son los coeficientes de las variables discriminante x_i . (Marín Diazaraque, 2014)

Supuestos en el Análisis Discriminante:

- No debe haber combinación lineal entre las variables discriminantes.
- Las matrices de varianzas y covarianzas deben ser homogéneas dentro de los grupos.
- Las variables discriminantes o independientes deben seguir distribución Normal.
- Deben existir al menos dos grupos y dos o más casos para cada grupo.

3.1.2. Estadístico Lambda de Wilks

Indica en qué medida los valores tomados por una variable independiente son diferentes en cada uno de los grupos de la variable dependiente. Su rango de variación va de 0 a 1. Valores altos de Lambda (próximos a 1) indican que la medida de la variable independiente correspondiente es igual en cada grupo. En cambio, valores bajos de lambda (próximos a 0) indican que la media es diferente. (Fundacion para la Universidad Oberta de Cataluña)

Wilks (1932) basándose en el principio de la razón de verosimilitud generalizada planteo el estadístico Lambda como:

$$\Lambda = \frac{\text{Suma de Cuadrados intra - grupos}}{\text{Suma de Cuadrados total}} = \frac{|S|}{|T|}$$

Este estadístico se aproxima a una chi-cuadrado, se usa tanto el valor de lambda como el de la chi-cuadrada con $(p-f)(k-f-1)$ donde p es el número de variables independientes, k cantidad de grupos, f funciones discriminantes, cuando se analizan dos grupos $f = 0$, si los valores Lambda son cercanos a 1 y el valor de chi-cuadrado es de sig. o nivel crítico 0.00 o menor a 0.05, rechazamos la hipótesis nula que se plantea a continuación:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

3.2. Regresión Logística

Al igual que el Análisis discriminante una de las principales aplicaciones de la Regresión Logística es la de clasificación y especialmente la binaria, en el que las observaciones se clasifican en un grupo u otro dependiendo del valor que tome la variable empleada como predictor.

Definida la variable dependiente como la ocurrencia o no de un acontecimiento, el modelo de Regresión Logística la expresa en términos de probabilidad, utilizando la función logística para estimar la probabilidad de que ocurra el acontecimiento o de que un individuo elija la opción uno de la variable dependiente, dados determinados valores de las variables explicativas, mediante la siguiente formulación:

$$P(X) = \beta_0 + \beta_1 X$$

Llevando esta ecuación al termino logarítmico, tendremos lo siguiente:

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 X$$

$$\frac{p}{1-p} = e^{\beta_0 + \beta_1 X}$$

$$p = (1-p)e^{\beta_0 + \beta_1 X}$$

$$p = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

$$p + p e^{\beta_0 + \beta_1 X} = e^{\beta_0 + \beta_1 X}$$

$$p = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

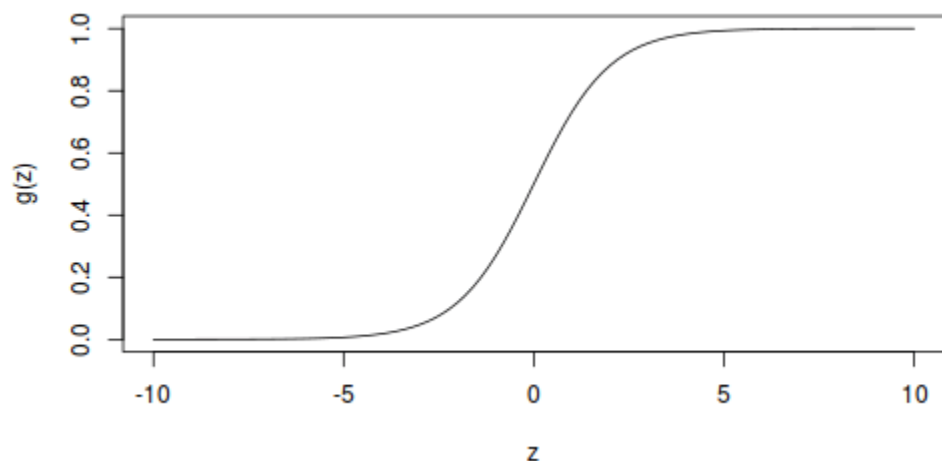
Puesto que el modelo anterior no es lineal respecto a las variables independientes, se considera la inversa de la función logística, que es el Logit o logaritmo de la odds de que un suceso ocurra, definiéndose ésta como el cociente entre la probabilidad de que ocurra un acontecimiento y la probabilidad de que no ocurra, que es su complementaria, como puede observarse en la siguiente expresión: (Mures Quintana, Garcia Gallego, & Vallejo Pascual, 2005)

$$\text{logit}(P(X)) = \ln p - \ln 1 - p = \beta_0 + \beta_1 X$$

$$\frac{p}{1 - p} = e^{\beta_0 + \beta_1 X}$$

La primera se la conoce como expresión Logit y la segunda como Odds – ratio o cociente de probabilidades.

Figura 3.1 Gráfica de la función Logística



Fuente: Research Gate - <https://www.researchgate.net/> (Charte, 2017)

3.2.1. Criterio de Máxima Verosimilitud

Al igual que en Regresión Múltiple, podemos usar los valores observados y predichos para evaluar el ajuste del modelo. La medida que usamos es log-likelihood (logaritmo de la razón de verosimilitud).

El estadístico log-likelihood es análogo a la suma de cuadrados residual en la Regresión Múltiple en el sentido de que es un indicador de cuánta información sin explicar queda en la variable respuesta tras haber ajustado el modelo. Grandes valores del log-likelihood indican un pobre ajuste del modelo, cuanto mayor sea este valor, más variabilidad sin explicar queda en el modelo. (Ferre Jaén, 2019)

3.2.2. Estadístico de Wald

Como en la regresión lineal, queremos saber no sólo la bondad de ajuste del modelo, sino también la contribución individual de cada una de las variables predictoras. Esta información la proporciona el estadístico de Wald (z-statistic) que sigue una distribución normal.

Al igual que el estadístico t en Regresión Lineal, el estadístico z nos dice si los coeficientes β_i para cada variable independiente o predictora son significativamente diferentes de cero. Si es distinto de cero asumimos que la variable predictora está haciendo una contribución significativa al modelo para predecir la respuesta.

Su valor para un coeficiente concreto viene dado por el cociente entre el valor del coeficiente y su correspondiente error estándar. La obtención de significación indica que dicho coeficiente es diferente de 0 y merece la pena su conservación en el modelo.

Sin embargo, en modelos con errores estándar grandes, el estadístico de Wald puede proporcionar falsos resultados. Si el coeficiente β_i del modelo es grande, el error estándar tiende a inflarse y esto incrementa la probabilidad de rechazar un predictor cuando en realidad está haciendo una contribución al modelo. (Ferre Jaén, 2019)

Se plantea el siguiente contraste de hipótesis, para este estadístico:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

3.3. Árboles de Decisión

Es una técnica estadística no paramétrica que busca clasificar o identificar patrones de comportamiento a cierto grupo de entes o individuos. En árboles de decisión al igual que en las dos técnicas detalladas previamente la variable dependiente es una variable categórica que define grupos de análisis en general binarios, por ejemplo, Cliente moroso o No moroso, Compra o No Compra, etc.

Ventajas de los Árboles de Decisión

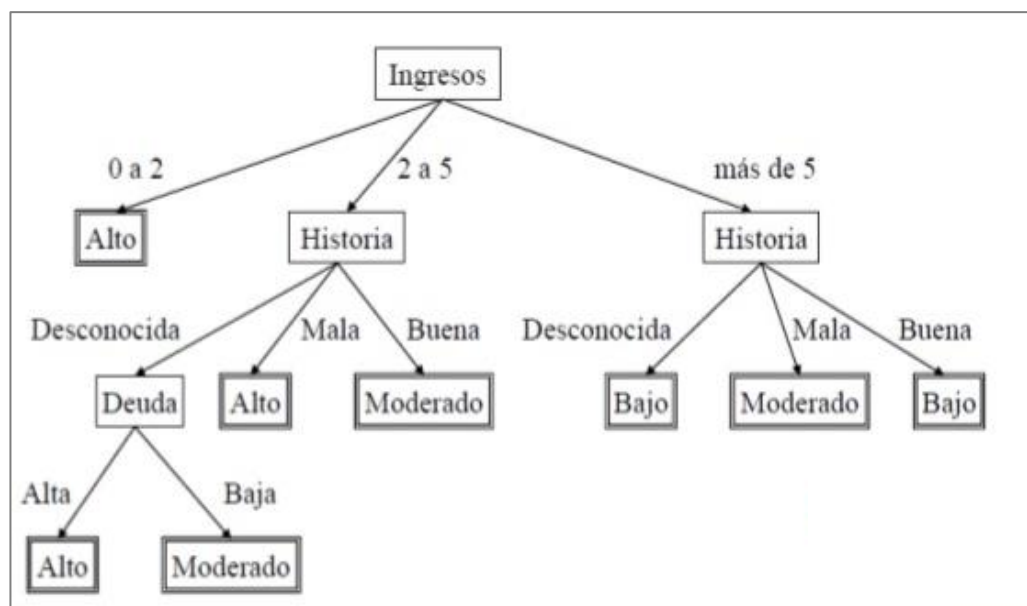
No es necesario que las variables independientes sean transformadas o que sean exclusivamente continuas.

Esta técnica al ser una técnica no paramétrica no es necesario que se ajuste a alguna distribución poblacional.

Los Árboles de Decisión en el desarrollo de un scoring son usados como herramienta para el cálculo de los momios de incumplimiento (odds, disparidad o razón de oportunidades) y representa un método efectivo para la estimación. Un mismo modelo permite diferentes usos, como mantenimiento de clientes considerados como buenos (probabilidades bajas de incumplimiento), cobranza proactiva y discriminada por nivel de riesgo para los clientes calificados como malos o con probabilidades altas de llegar a incumplimiento. Cuando en la administración de riesgos se busca qué perfil socio-demográfico pertenece a un nivel determinado de riesgo, se construyen una serie de tablas que permiten ver la asociación existente entre variables. (Espin-García & Rodríguez-Caballero, 2013).

En la Figura 3.3.1 se muestra un ejemplo de árboles de decisión en el cual se determinan las variables discriminatorias y sus valores de cortes, que clasifican de mejor forma si un nuevo prospecto tendrá riesgo “Alto” o “Moderado”.

Figura 3.3.1 Árbol de decisión “Riesgo en la concesión del Crédito”



Fuente: Aplicación de árboles de decisión en modelos de riesgo crediticio.

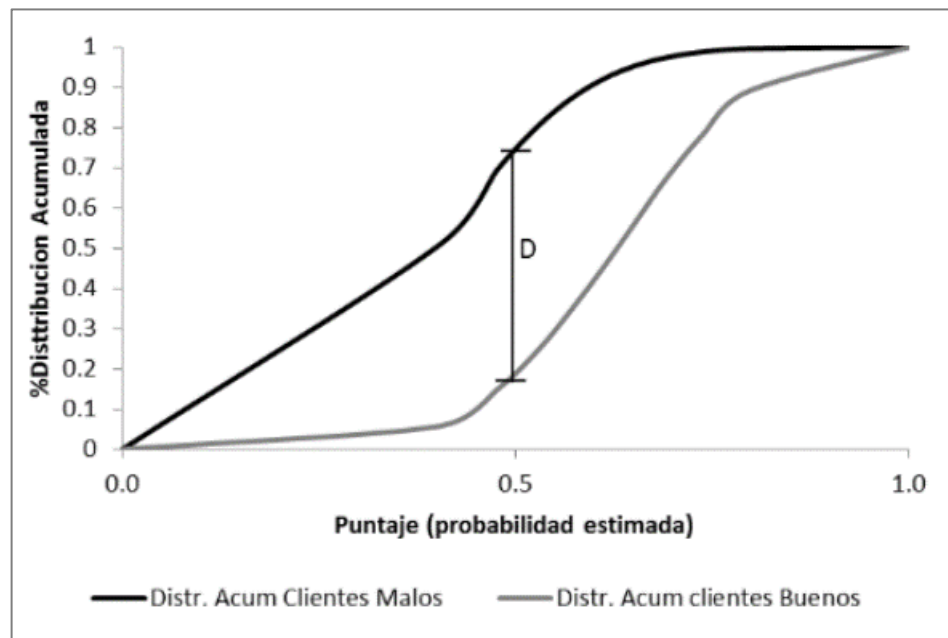
<https://es.slideshare.net>.

3.3.1. Estadístico K-S (Kolmogorov – Smirnov)

El estadístico también conocido como distancia Kolmogorov–Smirnov (K-S), se define como la distancia vertical máxima entre las funciones de distribución acumulada empíricas de dos muestras, o entre una función de distribución empírica y una función de distribución acumulada teórica de referencia. Para el caso de árboles de decisión es comprobar la distancia máxima entre las distribuciones acumuladas de los grupos tales como “No Moroso” y “Moroso”. (Capelo Vinza, 2012)

La ventaja principal de este estadístico es que es sensible a diferencias tanto en la localización como en la forma de la función de distribución acumulada.

Figura 3.3.2 Ejemplo de Estadístico K-S



Fuente: Comparación de Modelos de Riesgo de Crédito: Modelos Logísticos y Redes Neuronales. <https://docplayer.es>

La gráfica (Docplayer, s.f.) muestra la distribución acumulada de dos grupos, Clientes Buenos y Clientes Malos, la distancia D representa el valor del K-S para este ejemplo.

3.4. Muestra y Selección de Variables

En esta sección se detalla los datos, fuentes y variables que se utilizarán para la elaboración del modelo de segmentación de prospección de nuevos clientes, así como las metodologías a desarrollar, explicadas en el capítulo anterior.

Para la elaboración y construcción del modelo se toma una muestra para modelamiento y otra testing o prueba. Se seleccionan los créditos captados y aprobados entre Julio y Diciembre del 2017.

La muestra aleatoria es de 13000 créditos captados y aprobados, de estos se selecciona el 70% para la construcción del modelo y el otro 30% para el testing o prueba.

Definición de los Grupos

De acuerdo con el planteamiento inicial la compañía desea poder determinar qué clientes serán “Buenos” o “Malos” en su comportamiento de pagos.

Para esto se considera como variable dependiente (Comportamiento), la misma que se categoriza como “Bueno” o “Malo”; dependerá del comportamiento del cliente con el crédito en un periodo de 12 meses (ventana de tiempo) posterior a su aprobación.

Se define como **Buena**: aquel que en los 12 meses posterior a la aprobación del crédito no haya llegado a más de 30 días de mora.

Mala: sería el resto.

Selección de Variables

Para la construcción de este tipo de modelos, como se detalla en el capítulo 2, se utilizan variables sociodemográficas así como el Score o comportamiento crediticio regido por el “Buro de Crédito”, las mismas que se describen en la tabla 3.4.1.

Tabla 3.4.1 Descripción de Variables

Nombre corto	Descripción	Tipo
EDAD_CLTE	Edad del Cliente	Numérica
ESTADO_CIVIL	Estado Civil	Categoría - Polidicotómica
GENERO	Género	Categoría - Dicotómica
TPO_CASA	Tiempo Reside en Domicilio	Numérica
TPO_TRAB	Antigüedad Laboral	Numérica
CARGAS	Cargas Familiares	Numérica
Region_domicilio	Región de Domicilio	Categoría - Polidicotómica
Region_Trabajo	Región de Trabajo	Categoría - Polidicotómica
STS_CASA_PROPIA	Tipo de vivienda	Categoría - Polidicotómica
NIVEL_INST	Nivel de Instrucción	Categoría - Polidicotómica
TIENE_TC	Tiene Tarjeta de crédito	Categoría - Dicotómica
TIPO_TARJ	Tipo de tarjeta	Categoría - Dicotómica
SUELDO	Sueldo del cliente	Numérica
ZONA_DOMI	Zona de Domicilio	Categoría - Polidicotómica
SI_UNI	Es universitario	Categoría - Dicotómica
PUNTAJBR	Puntaje Buró de Crédito	Numérica

CAPÍTULO 4

4. ELABORACION DE MODELOS Y RESULTADOS

Para la elaboración de los modelos como se detalló en las secciones previas se realiza por las técnicas de Análisis Discriminante, Regresión Logística y Árboles de Decisión. Se recodifican las variables categóricas como variables dummy para que puedan ser incluidas dentro del análisis de Regresión y Análisis Discriminante.

Uno de los objetivos de este proyecto es la comparación de estas tres técnicas estadísticas, para esto se realiza un ajuste de las medidas que se obtengan de las técnicas: Análisis Discriminante y Regresión Logística a la de Árboles de Decisión.

Para las tres técnicas a utilizar se tendrá una proporción del 82.22% de Buenos y el restante 17.78% de Malos.

4.1. Elaboración del Modelo por Árboles de Decisión

La técnica de Árboles de decisión se basa principalmente en el estadístico KS (Kolmogorov – Smirnov). La Tabla 4.1.1 detalla el valor K-S, el cual resultó en 19.7% para la base de construcción del modelo y 19.2% para la base de Testing o prueba, la diferencia es mínima, lo que nos indicaría que es un modelo estable a nivel de ambas muestras.

Según el informe en el sitio (Flacso Andes, 2020), la evidencia empírica indica que, para un modelo de aprobación de nuevos créditos, los niveles mínimos permitidos de K-S van desde el 25% y los de Gini van desde el 35%. Esperar que las variables sociodemográficas hablen sobre el comportamiento esperado del cliente por encima de un 25% es muy optimista pero poco realista. Por este motivo los modelos de comportamiento (compras, pagos, etc.) tienen mejores resultados en los indicadores de rendimiento del modelo.

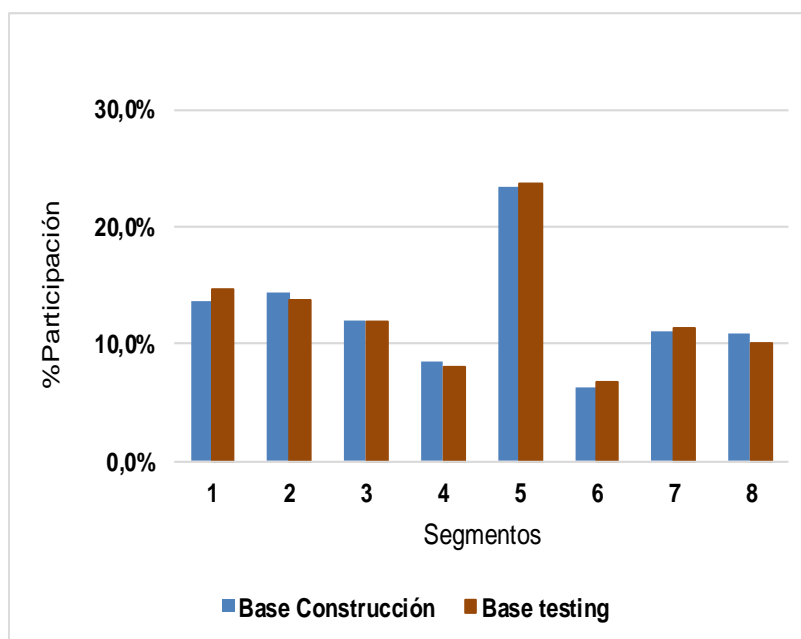
Tabla 4.1.1 Segmentación mediante Árboles de Decisión

Segmentos	Perfiles	Construcción		Testing	
		%Malos	%Partc.	%Malos	%Partc.
1	Puntaje Buró>0 y Puntaje Buró<=700	29,8%	13,7%	29,6%	14,6%
2	Puntaje Buró>700 y Puntaje Buró<=870 y Estado Civil=Soltero o Estado Civil=Viudo o Estado Civil=Union Libre	22,9%	14,3%	24,8%	13,7%
3	Puntaje Buró<=0 y Tiempo de Trabajo<=1,08	20,6%	11,9%	20,5%	11,9%
4	Puntaje Buró>700 y Puntaje Buró<=870 y Estado Civil=Casado o Estado Civil=Divorciado	17,7%	8,5%	20,1%	8,0%
5	Puntaje Buró>870 y Puntaje Buró<=933	15,0%	23,3%	15,9%	23,6%
6	Puntaje Buró<=0 y Tiempo de Trabajo>1,08	11,5%	6,2%	14,1%	6,7%
7	Puntaje Buró>933 y Puntaje Buró<=955	11,5%	11,1%	10,0%	11,3%
8	Puntaje Buró>955	6,2%	10,9%	7,9%	10,1%
KS TOTAL		19,7%		19,2%	

Las variables que discriminaron por el método de Árboles de Decisión fueron el Puntaje del Buró de Crédito (Calificación entre 0 y 1000), Estado Civil y Tiempo de trabajo. El mejor segmento indica que los clientes con Puntaje Buró de Crédito superior a los 955 pts tendrán un porcentaje de malos de 6.7%, en promedio entre las dos bases la de construcción y la de testing.

Es importante mencionar que la tendencia de los perfiles es a menor segmento mayor porcentaje de malos.

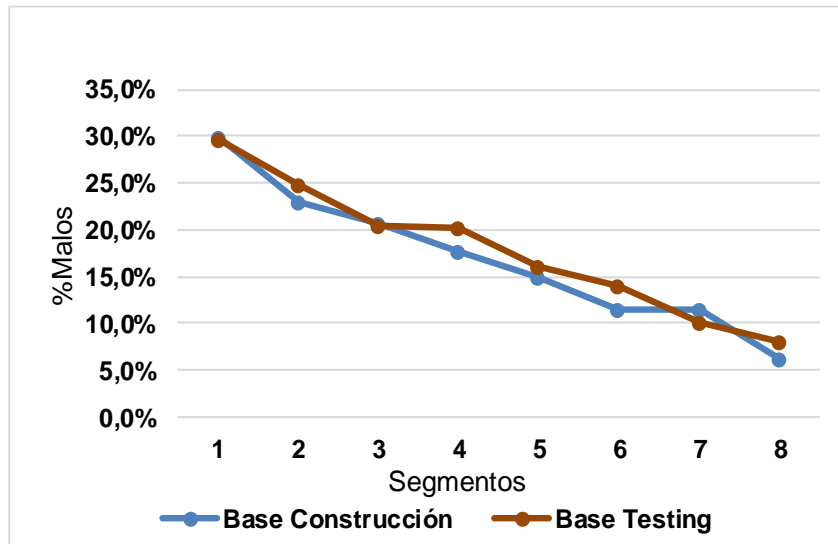
Figura 4.1.1 Participación por perfiles – Árboles de Decisión



En total el 48% de la muestra tanto para la de construcción como para la de testing se acumula en los segmentos del 1 – 4. (Ver figura 4.1.1)

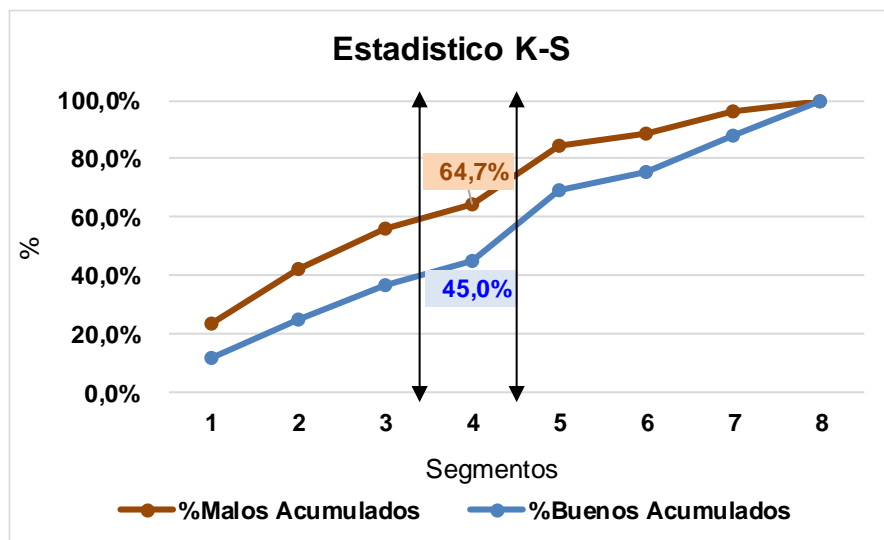
En la Figura 4.1.2 se detalla el porcentaje de malos por segmentos, como se observa el mejor segmento registra un 6.7% de malos (*valor promedio entre ambas muestras*), es decir 10 pts por debajo de la tasa de malos global previo al análisis detallada al inicio de esta sección (17.78%).

Figura 4.1.2 Porcentaje de Malos por segmentos de Riesgo



Se observa en la figura 4.1.3 la determinación del estadístico K-S, mediante las distribuciones acumuladas de los porcentajes de los Buenos y Malos, en la cual la máxima diferencia es de 19.7% siendo este el KS para la base de construcción.

Figura 4.1.3 Distribuciones Acumuladas – Estadístico K-S



Resulta más comprensible identificar las características de los mejores y peores clientes para la aprobación de un crédito, por lo cual la técnica Árboles de Decisión proporciona este tipo de detalle, sin embargo, se realizan las técnicas por Análisis discriminante y Regresión Logística para poder comprobar la efectividad de las mismas.

4.2. Elaboración del Modelo por Análisis Discriminante

Los coeficientes de la función discriminante estandarizados y no estandarizados, se detallan en la tabla 4.2.1, en la misma se observa la significancia y aporte obtenida de cada variable. Además de incluir las variables sociodemográficas, considera el Puntaje del cliente con el Buró de Crédito y el hecho de tener o no Tarjeta Externa (Tarjetas de Crédito o Comercial).

Tabla 4.2.1 Coeficientes del Análisis Discriminante

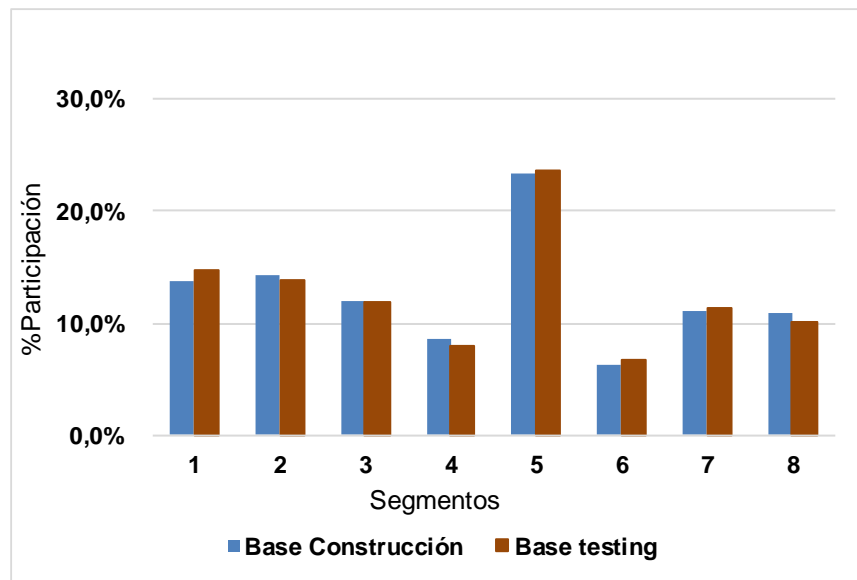
Variables Discriminantes	Coeficientes de la función discriminante canónica	Coeficientes de la función discriminante canónica (Estandarizados)	F		Descripción de las variables
			Estadístico	Sig.	
TPO_CASA	0,02	0,18	25,18	0,00	Tiempo que reside en casa
PUNTAJBR	0,00	0,31	28,86	0,00	Puntaje Buró de Crédito
EDAD_CLTE	0,03	0,40	61,19	0,00	Edad del Cliente
EC_3_DUM	-0,65	-0,21	22,46	0,00	Estado Civil - Unión Libre
RT_Q_DUM2	0,47	0,20	20,39	0,00	Región de Trabajo - Resto Costa
GENERO_DUM1	-0,61	-0,30	38,35	0,00	Género - Masculino
NIV_INST_DUM3	0,81	0,31	32,58	0,00	Instrucción - Superior
TIPO_TARJ_DUM2 (Constante)	-0,77 -1,33	-0,36	47,89	0,00	Sin Tarjeta Externa

Para poder ejecutar uno de los objetivos inicialmente planteados en este proyecto, que es la comparación de las técnicas estadísticas a utilizar, se decide agrupar las puntuaciones obtenidas del Análisis Discriminante, en rangos de ocho segmentos (*Número de segmentos obtenidos de la técnica de Árboles de decisión*) y realizar la comparación con Árboles de Decisión.

Tabla 4.2.2 Agrupación de los Puntajes del Análisis Discriminante

Segmento	Puntuación Discriminante
1	$\leq -2,62$
2	$> -1,10$ y $\leq -0,65$
3	$> -0,65$ y $\leq -0,32$
4	$> -0,32$ y $\leq -0,10$
5	$> -0,10$ y $\leq 0,59$
6	$> 0,59$ y $\leq 0,80$
7	$> 0,80$ y $\leq 1,30$
8	$> 1,30$

Figura 4.2.1 Porcentaje de Distribución de la muestra – Análisis Discriminante



La figura 4.2.1 muestra el porcentaje de malos por segmentos bajo la agrupación de rangos de coeficientes discriminante, en este caso el mejor segmento obtiene el 12% de malos a diferencia de Árboles de Decisión en el que se obtiene 6.2%.

Para este caso no se ajustan exactamente las muestras de construcción, sin embargo, mantiene la tendencia.

Figura 4.2.2 Porcentaje de Malos por segmento de riesgo – Análisis Discriminante

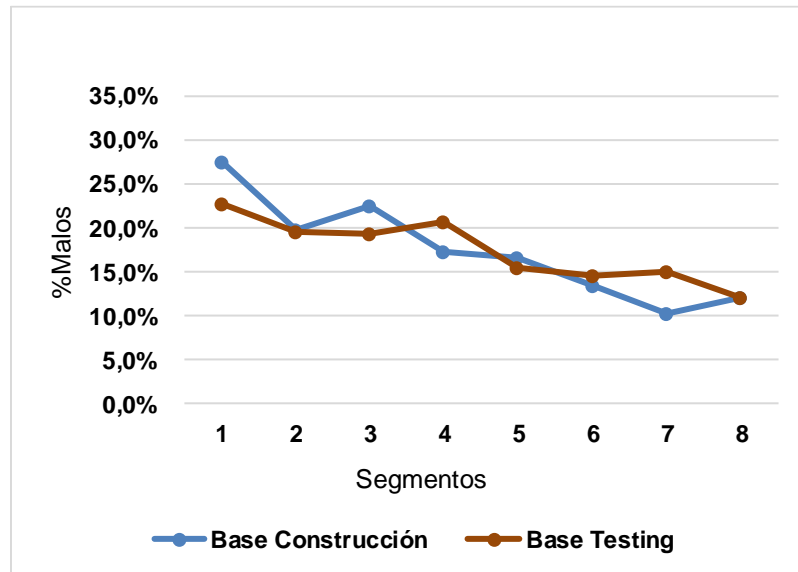
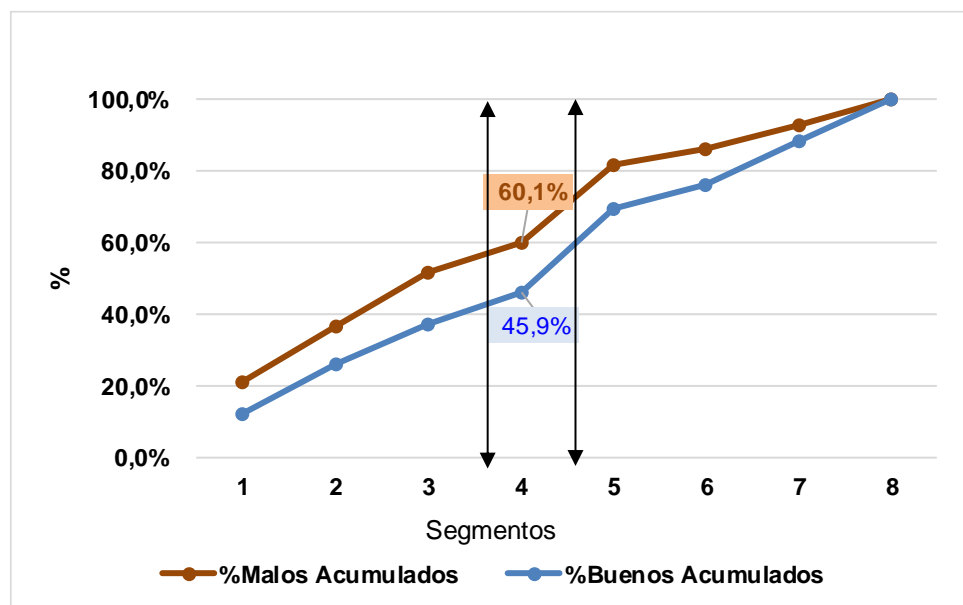


Figura 4.2.3 Distribuciones Acumuladas por Análisis Discriminante – Estadístico K-S



En cuanto al estadístico KS el obtenido por esta técnica fue de 14,5% 5 pts por debajo del obtenido por Árboles de decisión.

Podríamos decir que esta técnica no brinda una confiabilidad alta para la elaboración de este modelo en específico.

4.3. Elaboración del Modelo por Regresión Logística

Para el desarrollo del modelo de Regresión Logística se utiliza en SPSS el método de introducción de variable a variable por pasos hacia adelante. Los resultados obtenidos determinaron 8 pasos o el ingreso de 8 variables.

De acuerdo con la tabla 4.3.1 este modelo no tiene un buen ajuste ya que los valores R son muy bajos, sin embargo, en la tabla 4.3.2 obtenemos la estimación por máxima verosimilitud en la inclusión variable a variable, lo que nos indica que rechazamos la hipótesis nula de que todos los coeficientes del modelo son iguales a cero.

Tabla 4.3.1 Indicadores de Ajuste del Modelo

Resumen del modelo			
	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	8488,574	0,007	0,012
2	8453,757	0,011	0,018
3	8435,066	0,013	0,021
4	8419,416	0,015	0,024
5	8407,435	0,016	0,026
6	8399,725	0,017	0,027
7	8394,045	0,017	0,029
8	8387,764	0,018	0,030

Tabla 4.3.2 Prueba de Máxima Verosimilitud

Prueba de Hosmer y Lemeshow			
Paso	Chi-cuadrado	gl	Sig.
1	23,825	8	0,002
2	20,477	8	0,009
3	18,244	8	0,019
4	9,958	8	0,268
5	28,560	8	0,000
6	23,381	8	0,003
7	25,709	8	0,001
8	14,729	8	0,065

El aporte individual por cada variable que se incluye en el modelo de regresión logística, se detalla en la tabla 4.3.3. La significancia de cada variable bajo estadístico de Wald determina que se rechaza la hipótesis nula de que los coeficientes de cada variable son iguales a cero.

Tabla 4.3.3 Pruebas de significancia individual

Variables en la ecuación							Descripción de las variables
	B	Error estándar	Wald	gl	Sig.	Exp(B)	
EDAD_CLTE	0,013	0,003	22,454	1	0,000	1,013	Edad del cliente
TPO_CASA	0,007	0,003	5,416	1	0,020	1,007	Tiempo que reside en casa
PUNTAJBR	0,000	0,000	10,733	1	0,001	1,000	Puntaje Buró de Crédito
EC_3_DUM	-0,210	0,081	6,679	1	0,010	0,810	Estado Civil - Unión Libre
RT_Q_DUM2	0,178	0,072	6,147	1	0,013	1,194	Región de Trabajo - Resto Costa
GENERO_DUM1	-0,210	0,056	14,110	1	0,000	0,811	Género - Masculino
NIV_INST_DUM3	0,303	0,077	15,351	1	0,000	1,354	Instrucción - Superior
TIENE_TC_DUM1	0,294	0,069	18,234	1	0,000	1,342	Tiene Tajeta de Crédito
Constante	0,788	0,105	56,837	1	0,000	2,199	Constante

Al igual que en Análisis Discriminante, se agrupan las probabilidades obtenidas del modelo en rangos de 8 segmentos, con lo cual se determina el porcentaje de malos por segmento, así como el estadístico KS.

Tabla 4.3.4 Agrupación de las probabilidades de Regresión Logística.

Segmento	Probabilidades Agrupados
1	<=0,77
2	>0,77 y <=0,79
3	>0,79 y <=-0,82
4	>0,82 y <=-0,10
5	>-0,10 y <=0,59
6	>0,59 y <=0,80
7	>0,89 y <=0,86
8	>0,95

Las figuras 4.3.1 y 4.3.2 detallan la población de la muestra y porcentaje de malos por segmentos de las muestras, como se observa mantienen

la misma tendencia, similar que, en Árboles de Decisión y Análisis Discriminante, aunque no se ajustan a la perfección la muestra de la construcción del modelo versus la del testing.

Figura 4.3.1 Porcentaje de Distribución de la muestra – Regresión Logística

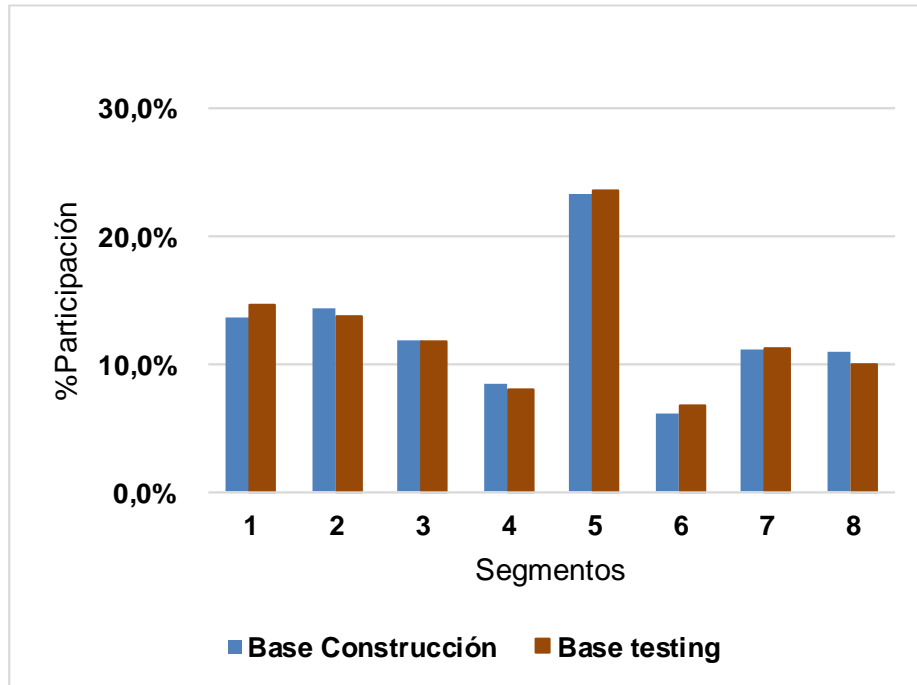


Figura 4.3.2 Porcentaje de Malos por segmento de riesgo – Regresión Logística

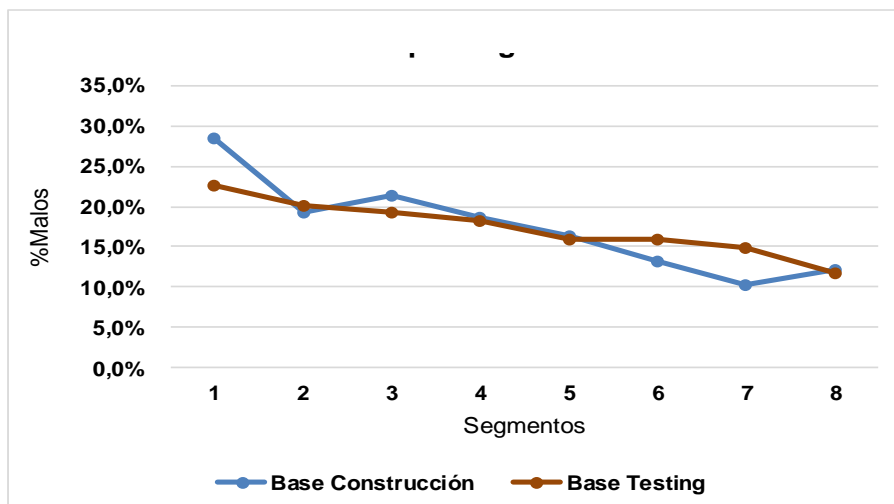
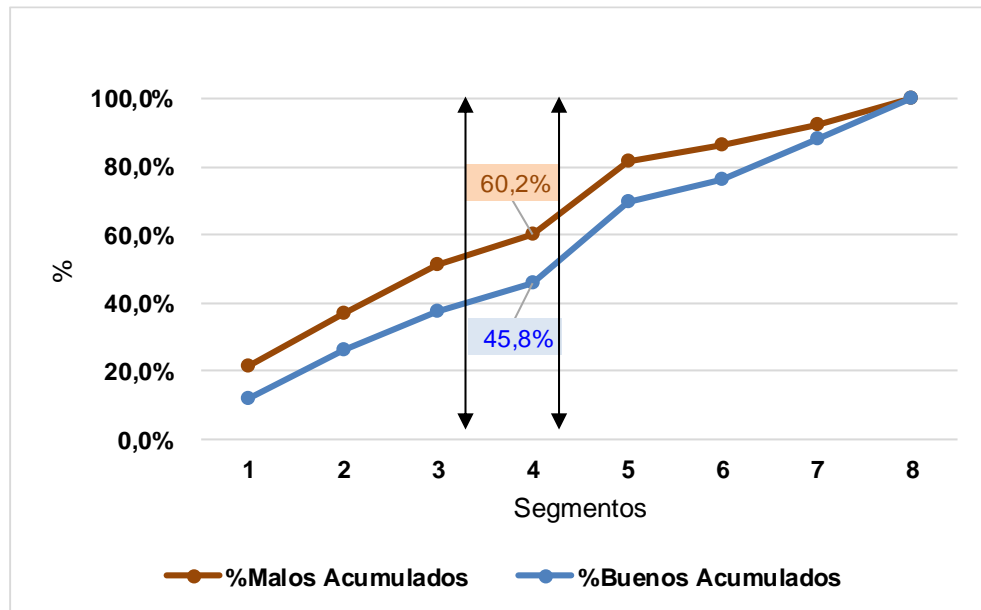


Figura 4.3.3 Distribuciones Acumuladas por Análisis Discriminante – Estadístico K-S



El KS para el modelo bajo Regresión Logística 14,3% similar al obtenido por Análisis Discriminante.

4.4. Comparación de los modelos

En la sección 4.3 se detallaron las tres técnicas utilizadas, así como el estadístico de KS obtenidos para cada uno de ellos, la tabla 4.4.1 recoge de estos valores, en la misma se puede observar que la técnica que se ajusta mejor es Árboles de Decisión ya que resulta tener un KS más elevado que el resto de técnicas.

Tabla 4.4.1 Estadístico KS por cada técnica

Técnica	Estadístico KS
Árboles de Decisión	19,7%
Análisis Discriminante	14,5%
Regresión Logística	14,3%

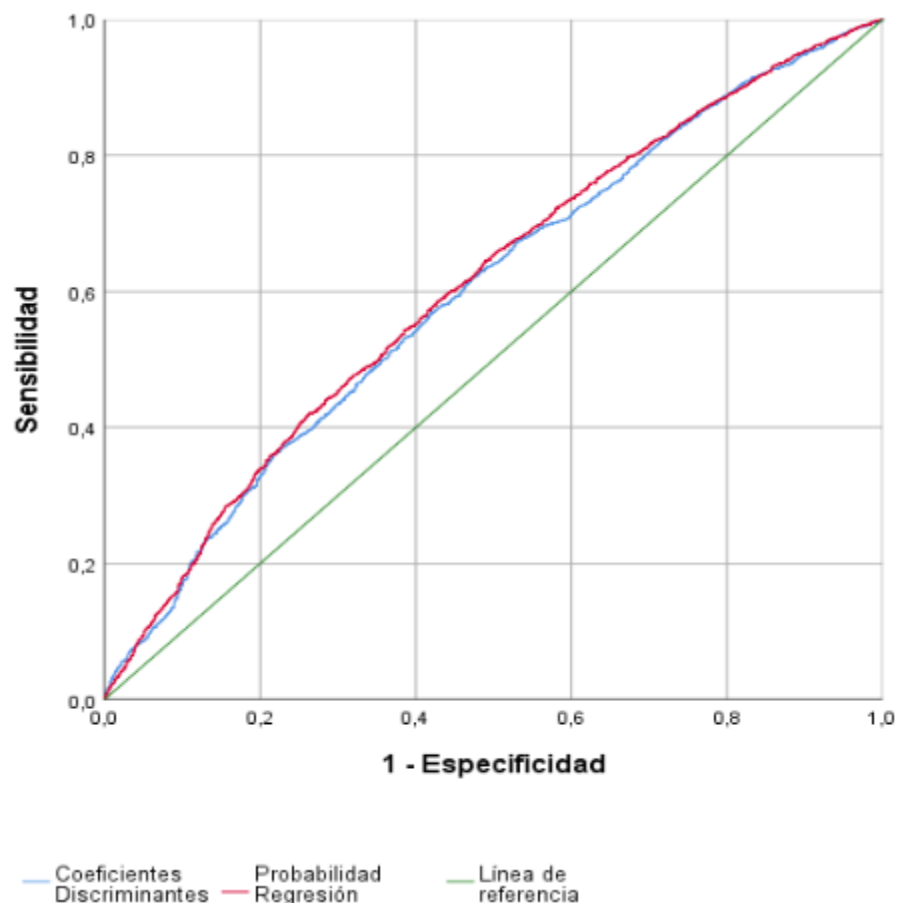
4.5. Curvas COR

Es un gráfico en el que se representa la sensibilidad en función de (1 – especificidad). Si vamos modificando los valores de algún valor de corte y representamos la sensibilidad (en ordenadas) frente a (1 – especificidad) (en abscisas) tenemos la curva COR.

Es una curva cóncava que conecta los puntos (0, 0) y (1, 1). Cuanto mayor sea el área bajo la curva, mejores serán las predicciones. La curva COR ofrece un mejor resumen de la capacidad predictiva que una tabla de clasificación, porque presenta la potencia predictiva para todos los posibles valores de referencia. (Marín Diazaraque, 2014)

En la Figura 4.4.1 se muestra la gráfica de curva COR entre los resultados de las probabilidades del modelo logístico versus los valores discriminantes. Los puntos de cortes de ambas curvas resulto en 0.6

Figura 4.5.1 Gráfica de Curva COR



4.6. Esquema de Matriz de Riesgo de los nuevos prospectos

Se considera el modelo obtenido por la técnica de Árboles de Decisión, sin embargo, esto es el 50% del proceso para la selección de nuevos clientes. La empresa busca identificar zonas de riesgo basados no únicamente en el modelo. Para esto una variable que no fue incluida dentro del modelamiento fue la Calificación Crediticia obtenida del Buró de crédito al ser una combinación de condicionamientos puntuales para empresas comerciales y del Puntaje Buró de Crédito.

Definición de Matriz de Riesgo

Se propone establecer la matriz de riesgo entre la calificación del buró y los segmentos de riesgo del modelo, la tabla 4.5.1 se muestra el porcentaje de malos para cada zona de riesgo, la cual se determina por la intersección entre los segmentos y calificación.¹

Tabla 4.6.1 Tipos de Riesgo

Segmento	Calif. Buró Crédito		%Malos
1	AAA		
2	AA	Riesgo Bajo	8%
3	A	Riesgo Medio Bajo	15%
4	Revisión	Riesgo Alto	22%
5	Rechazado	Riesgo Muy Alto	39%
6			
7			
8			

Para los tipos de Riesgo Alto y muy Alto se propone realizar una investigación exhaustiva de la información proporcionada por el cliente. Para el riesgo bajo, se puede solo realizar la confirmación de aceptación del crédito. Con las de riesgo medio bajo realizar ciertas verificaciones.

¹ Por petición de la empresa no se coloca detalladamente la Matriz de Riesgo.

Determinación de Línea de Crédito

Una de las formas que por lo general se realiza la asignación de una línea de crédito o cupo inicial tanto en tarjetas como en bancos es mediante capacidad de pago del prospecto, para ello se pretende asignar el cupo a este nuevo cliente, utilizando como referencia la participación del gasto a nivel nacional (por tipos de gastos de consumo), como se detalla en la tabla 4.5.2.

Tabla 4.6.2 Estructura del Gasto Corriente de consumo monetario Mensual

GRUPOS DE GASTO	GASTO (%)
GASTO DE CONSUMO	100%
01 ALIMENTOS Y BEBIDAS NO ALCOHÓLICAS	24,4%
02 BEBIDAS ALCOHÓLICAS, TABACO Y ESTUPEFACIENTES	0,7%
03 PRENDAS DE VESTIR Y CALZADO	7,9%
04 *ALOJAMIENTO, AGUA, ELECTRICIDAD, GAS y OTROS COMBUSTIBLES	7,4%
05 MUEBLES Y ENSERES	5,9%
06 SALUD	7,5%
07 TRANSPORTE	14,6%
08 COMUNICACIONES	5,0%
09 RECREACIÓN Y CULTURA	4,6%
10 EDUCACIÓN	4,4%
11 HOTELES Y RESTAURANTES	7,7%
12 BIENES Y SERVICIOS DIVERSOS	9,9%

Fuente: INEC 2011 - 2012

Se estima un factor de riesgo para cada segmento de riesgo, basados en los grupos de consumo de prendas de vestir y muebles y enseres. La Tabla 4.5.3 muestra el detalle del cálculo de estos factores. También se determinan por tipo de Calificación del Buró. Es decir, si un cliente resulta tener calificación AAA y Segmento 8 y tiene ingresos de \$1000, la línea de crédito será de \$1300.

Tabla 4.6.3 Tabla de factores por segmento de Riesgo

Segmento	INEC ^(*)	Factor	%Ingresos	Riesgo Buró	%Ingresos
Riesgo 1	13,8%	4,8%	48,3%	AAA	100%
Riesgo 2	13,8%	5,5%	55,2%	AA	90%
Riesgo 3	13,8%	6,9%	69,0%	A	78%
Riesgo 4	13,8%	8,3%	82,8%	Rechazado	30%
Riesgo 5	13,8%	9,7%	96,6%	Revisión	64%
Riesgo 6	13,8%	11,0%	110,4%		
Riesgo 7	13,8%	12,4%	124,2%		
Riesgo 8	13,8%	13,1%	131,1%		

^(*) Fuente INEC

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES

Entre las conclusiones resultantes de esta investigación se detallan a continuación:

- En los modelos realizados por las tres técnicas estadísticas la variable Puntaje de Buró estuvo presente.
- De todas las técnicas la de Árboles de Decisión fue la técnica con mejor predictibilidad, así como mejor ajuste en la determinación del porcentaje de clientes potencialmente Malos.
- Todas las técnicas mantuvieron tendencia en la distribución de porcentaje de malos por segmentos, es decir que a bajo segmento mayor tasa de malos.
- Los K-S para el Análisis Discriminante y Regresión Logística fueron de 14.5% mientras que para Árboles de Decisión fue de 19.5%, 5 pts superior.
- El porcentaje de malos por riesgo agrupado se ajusta perfectamente, los de bajo riesgo registran 6% de malos y los altos 22%, superior al 17% de malos que es la tasa general.
- Se consideran los segmentos mediante Árboles de Decisión para la determinación de la línea de crédito.

Como recomendaciones se establece lo siguiente:

- Establecer un tipo de corte para Buenos y Malos más conservador y determinar nuevamente los modelos por análisis Discriminante y Regresión Logística.
- Sería interesante crear submodelos que combinen las variables existentes y así se tendrían variables compuestas a modo de segmentos.

6. REFERENCIAS

- Barrientos, R. E., Cruz, N., Acosta, H., Rabatte, I., Gogeochea, M. d., Pavón, P., & Blázquez, S. (2009). Árboles de decisión como herramienta en el diagnóstico médico. *Medigraphic, Literatura Biomédica*, 9(2), 20-24.
- Bookdown. (s.f.). *Bookdown*. Obtenido de <https://bookdown.org/content/2031/arboles-de-decision-parte-i.html>
- Capelo Vinza, J. A. (2012). *Modelo de Aprobación de Tarjetas de Crédito en la población ecuatoriana bancarizada a través de una metodología analítica*. Quito: Escuela Politécnica Nacional.
- Cardona Hernández, P. A. (2004). Aplicación de árboles de decisión en modelos de riesgo crediticio. *Revista Colombiana de Estadística*, 139 - 151.
- Charte, D. (01 de Junio de 2017). *Research Gate*. Obtenido de https://www.researchgate.net/figure/Figura-74-Grafica-de-la-funcion-logistica_fig3_318888351
- Docplayer. (s.f.). Obtenido de <https://docplayer.es>: <https://docplayer.es/1516139-Comparacion-de-modelos-de-riesgo-de-credito-modelos-logisticos-y-redes-neuronales.html>
- Espin-García, O., & Rodríguez-Caballero, C. (2013). Metodología para un Scoring de Clientes sin Referencias Crediticias. *Cuadernos de Economía*, 139-165.
- Fernández, S. d. (2011). *Regresión Logística - Análisis Discriminante*. Madrid: Universidad Autónoma de Madrid.
- Ferre Jaén, M. E. (04 de Abril de 2019). *Universidad de Murcia*. Obtenido de <https://gauss.inf.um.es/>: <https://gauss.inf.um.es/feir/45/>
- Flacso Andes. (15 de Septiembre de 2020). *Flacso Andes*. Obtenido de <https://biblio.flacsoandes.edu.ec/catalog/resGet.php?resId=18022>
- Fundacion para la Universidad Oberta de Cataluña. (s.f.). Obtenido de http://cv.uoc.edu/moduls/UW03_84003_01131/web/nwin/m1/analisi_disc.pdf
- Guzman, O. (20 de Julio de 2020). *La Voz Houston*. Obtenido de <https://pyme.lavoztx.com/qu-es-el-riesgo-empresarial-4339.html>
- IBM. (s.f.). Obtenido de <https://www.ibm.com>: https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhel_p_client_ddita/clementine/nodes_treebuilding.html
- Marín Diazaraque, J. M. (8 de Abril de 2014). *Portal Universidad Carlos III de Madrid*. Obtenido de <http://portal.uc3m.es/>
- Mures Quintana, M. J., Garcia Gallego, A., & Vallejo Pascual, M. E. (2005). Aplicación del Análisis Discriminante y Regresión Logística en el estudio de la morosidad en las entidades financieras. Comparación de resultados. *Pecvnia*, 175-199.
- MyTripleA. (23 de Julio de 2020). *My Triple A*. Obtenido de <https://www.mytriplea.com/diccionario-financiero/riesgo-de-credito/>
- Natividad, R. A. (15 de Septiembre de 2018). *DOCPLAYER*. Obtenido de <https://docplayer.es/65781968-Capitulo-23-analisis-discriminante-el-procedimiento-discriminante-introduccion.html>
- Rayo, S., Lara, J., & Camino, D. (2010). Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II. *Journal of Economics, Finance and Administrative Science*, 90 - 124.
- Ruiz Lopez, H. (2017). Diseño de un modelo Matemático para la Calificación de Clientes Morosos en una Entidad Comercial mediante las metodologías de Árboles de

Decisión, Análisis Discriminante y Regresión Logística. *INNOVA Research Journal*, 176 - 188.

Sitio Big Data. (14 de Diciembre de 2019). *Sitio Big Data*. Obtenido de <https://sitiobigdata.com/2019/12/14/arbol-de-decision-en-machine-learning-parte-1/#>

Slideshare. (s.f.). Obtenido de <https://es.slideshare.net>:
<https://es.slideshare.net/franklinecano/aplicacin-de-rboles-de-decisin-en-modelos-de-riesgo-credificio>

Teruel, S. Y. (1 de Diciembre de 2016). *Gestión Pasiva*. Obtenido de <https://www.gestionpasiva.com/riesgo-sistemico-y-riesgo-no-sistemico/>

Ucha, A. P. (22 de Julio de 2020). *Economipedia*. Obtenido de <https://economipedia.com/definiciones/riesgo-de-credito.html>

Valencia, A. (2017). Modelo Scoring para el otorgamiento de crédito de las pymes. Medellín, Colombia.