

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA
CICLO OPTATIVO DE PROFESIONALIZACIÓN EN
MARKETING Y FINANZAS**



**“PROPUESTA DE UN MODELO DE REGRESIÓN LOGÍSTICA
PARA ANALIZAR EL RIESGO CREDITICIO EN LA EMPRESA
CARSA S.A.C.”**

Presentado por:

**EDUARDO ANGELO MARCOS SÁNCHEZ
ASESOR: MG. JESÚS SALINAS FLORES**

Trabajo de Titulación para Optar el Título de:

INGENIERO EN ESTADÍSTICA E INFORMÁTICA

Lima – Perú

2015

ÍNDICE

I.	INTRODUCCIÓN	1
1.1.	El Problema	2
1.1.1.	Problema Principal.....	2
1.1.2.	Problemas Secundarios	2
1.2.	Objetivos	3
1.2.1.	Objetivo general:	3
1.2.2.	Objetivos específicos:.....	3
II.	REVISIÓN DE LITERATURA.....	4
2.1.	El crédito	4
2.1.1.	Riesgo de Crédito	6
2.1.2.	<i>Credit Scoring</i>	9
2.1.3.	Construcción del <i>Scorecard</i>	11
2.1.4.	Determinación de clientes buenos y malos.....	13
2.1.5.	Ventana de muestra.....	14
2.1.6.	Proceso para determinar clientes buenos y clientes malos	16
2.2.	Empresa al minoreo.....	19
2.3.	Matrices de Transición	23
2.4.	Regresión logística	26
2.4.1.	Interpretación del modelo logístico	30
2.4.2.	Pruebas estadísticas del modelo Logit	32
2.4.3.	Criterios para la validación del modelo	34
2.5.	Marco conceptual	37
III.	MATERIALES Y MÉTODOS	41
3.1.	Tipo de investigación	41
3.2.	Diseño de la investigación.....	41
3.3.	Formulación de hipótesis	41
3.4.	Identificación de variables	42
3.5.	Población y muestra	44
3.6.	Procedimiento de análisis de datos.....	44

IV. RESULTADOS Y DISCUSIÓN	47
4.1. Análisis situacional comercial.....	47
4.1.1. Análisis de Procesos	47
4.1.2. Cadena de valor	48
4.1.3. Condiciones iniciales del proceso del otorgamiento de un crédito convencional	50
4.1.4. Análisis de las cinco fuerzas de Porter	53
4.1.5. Análisis de la matriz Boston consulting group	56
4.1.6. Análisis FODA	57
4.2. Análisis Estadístico	59
4.2.1. Identificación de Bueno y Malo	59
4.2.2. Análisis del modelo de regresión logística binaria	61
4.2.3. Pruebas Estadísticas del modelo Logístico.....	64
V. CONCLUSIONES	74
VI. RECOMENDACIONES	76
VII. REFERENCIAS BIBLIOGRÁFICAS.....	77
VIII. ANEXOS	81

ÍNDICE DE FIGURAS

Figura 1: El ciclo del Riesgo	6
Figura 2: Definición del rendimiento	14
Figura 3: Evolución de la tasa de malos.....	15
Figura 4: Supermercados PBI per cápita vs m ² per cápita	22
Figura 5: Flujo simplificado del proceso de la venta Crédito de Carsa S.A.C.....	48
Figura 6: Cadena de valor simplificado de Carsa S.A.C.....	49
Figura 7: Proceso simplificado de otorgamiento de un crédito convencional	50
Figura 8: Las cinco fuerzas competitivas de Porter de Carsa S.A.C.....	55
Figura 9: Matriz Boston Consulting Group de Carsa S.A.C.	57
Figura 10: Análisis FODA de Carsa S.A.C.....	58
Figura 11: Curva de Kolmogorov - Smirnov (K-S)	68
Figura 12: Curva de ROC.....	69

ÍNDICE DE CUADROS

Cuadro 1: Posibles estados al final de los 6 meses.....	17
Cuadro 2: Posibles estados al día en los primeros 6 meses.....	17
Cuadro 3: Ranking del sector al minoreo 2012.....	21
Cuadro 4: Matriz de Probabilidad de Transición	23
Cuadro 5: Consumidores y su preferencias del mes 1 al mes 12	25
Cuadro 6: Cuadro de criterios de clasificación	34
Cuadro 7: Regla de indicadores K-S	36
Cuadro 8: Regla de Indicadores Curva ROC	37
Cuadro 9: Variables independientes Seleccionadas	42
Cuadro 10: Indicador WOE por categoría e IV global por variable	45
Cuadro 11: Colocación Acumulada por Tipo de Línea.....	56
Cuadro 12: <i>Roll Rate</i> del total de créditos.....	60
Cuadro 13: Posibles estados de un cliente y su marca de clase asociado	60
Cuadro 14: Variables del Modelo	61
Cuadro 15: Pruebas ómnibus de coeficientes de modelo	64
Cuadro 16: Prueba de Hosmer y Lemeshow	65
Cuadro 17: Resumen del modelo	66
Cuadro 18: Tabla de clasificación	67
Cuadro 19: Tabla de Puntuaciones (<i>Score</i>)	72
Cuadro 20: Perdidas monetarias de solicitudes que debieron ser rechazadas	73

RESUMEN

El crecimiento de los créditos por consumo a nivel mundial, junto con la normativa internacional sobre requerimientos de capital (Basilea II), están impulsando a las instituciones de banca *retail* a una mayor competencia con las entidades bancarias por este segmento de negocio. Por lo tanto en la empresa Carsa S.A.C. tiende a utilizar estrategias de ventas a crédito para tener un mayor crecimiento en este mercado competitivo, sin embargo su falta de control e identificación de clientes buenos y malos hace que sus estrategias puedan causar grandes pérdidas.

Por ello, el objetivo de investigación del estudio es proponer un modelo de regresión logística, con el fin de analizar el riesgo crediticio en las ventas al minoreo en la empresa Carsa S. A. C. La investigación es de tipo descriptiva, explicativa transversal ya que comprende una población de los créditos generados en el periodo de enero y noviembre del año 2013 de electrodomésticos vendidos a nivel nacional. El modelo de regresión logística encontrado está compuesto por las variables: tipo de actividad, línea del producto, tipo de propiedad, estado civil, total de créditos, productos crediticios, plazo del crédito y mora máximo en el sistema financiero.

En consecuencia, la empresa Carsa S. A. C. debe incluir el modelo de *credit scoring* en su política de crédito como un filtro adicional al otorgamiento convencional del mismo, dicho modelo es el cual discrimina a los clientes buenos y malos basándose en el perfil del cliente previo al otorgamiento de crédito. En la actualidad, la empresa analizada tiene grandes pérdidas por clientes morosos o mal identificados, efecto causado por el otorgamiento del crédito por el método por experiencia. Las pérdidas monetarias de solicitudes que debieron ser rechazadas en el 2013 ascienden a S/. 658,000.00 nuevos soles.

Palabras Claves: riesgo de crédito, regresión logística, minoreo, definición de buenos/malos, *scorecard*.

I. INTRODUCCIÓN

En los últimos años ha habido un desarrollo importante en la economía peruana con un crecimiento de 2.68% en el mes de marzo 2015 en comparación con similar mes del año anterior (2014), siendo esta la tasa más alta de los últimos 11 meses (INEI, 2015)¹, esto se relaciona con que las empresas de banca de venta al minoreo sigan otorgando créditos de diferentes tipos como corporativo, grandes empresas, medianas empresas, hipotecarios, de consumo, y entre otros.

El presente trabajo de investigación se centró en el crédito por consumo y según el informe de evolución del sistema financiero emitido en abril del 2015, dicho crédito por consumo tuvo un incremento de 14.1% respecto el año anterior, y con una participación de 19.2% de la cartera de créditos (SBS, 2015)². A pesar del incremento en la colocación de créditos por consumo, la población no bancarizada, sigue teniendo la mayor participación con el 70% del mercado (América Economía, 2014)³, es por ello que se busca penetrar en este nicho de mercado, pero con el menor riesgo crediticio. Además la evidencia internacional reciente ha mostrado que cuando los ciclos económicos son muy favorables y el ritmo de las colocaciones se acelera, el riesgo crediticio que enfrentan las instituciones financieras tiende a aumentar significativamente. Esto último debido a que las instituciones financieras y los clientes se vuelven optimistas en periodos de expansión económica y, con ello, se tiende a subestimar el riesgo asociado al incumplimiento de pagos de los nuevos créditos. Como resultado, los créditos otorgados a deudores con reducida capacidad de pago se incrementan motivados por el crecimiento económico elevado. De esta manera, ante escenarios de crecimiento económico sostenido muy elevado, la regulación y supervisión

¹ www.inei.gob.pe/media/MenuRecursivo/noticias/nota-de-prensa-n064-2015-inei.pdf

² www.sbs.gob.pe/app/stats/EstadisticaBoletinEstadistico.asp?p=14#

³ www.americaeconomia.com/economia-mercados/finanzas/consultora-bancarizacion-del-peru-continuara-creciendo-y-se-situara-en-32

financiera deberían ser más rigurosas para evitar que el riesgo crediticio en el que incurren las instituciones financieras se materialice en una crisis al terminar el ciclo expansivo.

Bajo este escenario, se presenta como caso de investigación a la empresa Carsa S.A.C. quien es participante de este mercado de alto riesgo. Y se plantea la necesidad de estudiar y diseñar mecanismos que permitan analizar el incumplimiento de pago, haciendo con ello ofrecimientos y requerimientos más realistas a nuevos créditos. Evaluando si los actuales mecanismos de otorgamiento de crédito, presentan una medición acertada de la capacidad de pago, o pudieran ser sustituidos por otro medio, que tenga mejor capacidad de medición. Por ello, el problema principal y problemas secundarios en la empresa Carsa S.A.C. son:

1.1. El Problema

1.1.1. Problema Principal

¿Se puede utilizar un modelo de regresión logística para analizar el incumplimiento de pago en el riesgo crediticio, como una alternativa a la evaluación de expertos en la empresa Carsa S.A.C.?

1.1.2. Problemas Secundarios

¿Es el proceso del otorgamiento del crédito basado por experiencia, adecuado en la empresa Carsa S.A.C.?

¿Cuáles son las variables que el autor deberá considerar para plantear un modelo de regresión logística para el proceso del otorgamiento de crédito en la empresa Carsa S.A.C.?

¿En cuánto disminuye la pérdida de dinero basados en el tiempo utilizando un modelo de regresión logística versus un modelo por experiencia en la empresa Carsa S.A.C.?

1.2. Objetivos

En cuanto al objetivo general y objetivos específicos se encuentran:

1.2.1. Objetivo general:

Proponer un modelo de regresión logística para analizar el riesgo crediticio en las ventas al minoreo en la empresa Carsa S.A.C.

1.2.2. Objetivos específicos:

- a. Evaluar las condiciones iniciales del otorgamiento de un crédito convencional en la empresa Carsa S.A.C.
- b. Determinar las principales variables que influyen en el modelo de regresión logística crediticio en la empresa Carsa S.A.C., con el fin de clasificar con mayor eficiencia a los clientes morosos o no morosos.
- c. Medir el aporte del modelo de regresión logística basada en aspectos de dinero y tiempo en la empresa Carsa S.A.C.

El lugar de investigación fueron las tiendas Carsa S.A.C., que de igual forma viene en un crecimiento de más de 70% en sus ventas al minoreo, y la mayor concentración de sus ventas (70%) está en el nicho de mercado de los clientes no bancarizados, el cual actualmente el otorgamiento del crédito se trabaja por medio de gestión de expertos. La investigación es de tipo descriptiva, explicativa transversal. La población será de todos los créditos otorgados del año 2013 y para el análisis se utilizará el programa estadístico IBM SPSS Statistics.

II. REVISIÓN DE LITERATURA

En este capítulo se exponen los elementos teóricos generales y particulares, así como la explicación de los conceptos básicos en que se apoya la investigación, con el objetivo de comprender los aspectos de los procesos de la realidad que se está estudiando.

2.1. El crédito

Entenderemos como crédito al compromiso pactado entre una persona o institución que otorga capacidad de compra por adelantado al deudor. El crédito permite realizar ventas y satisfacer la necesidad de compra de los consumidores, conforme a su capacidad de pago. Las condiciones del convenio que permiten el acuerdo comercial pueden ser flexibles y negociables en cuanto a plazos, montos, tipo de interés, etc.; con la finalidad de concluir en buenos términos el compromiso de crédito adquirido.

Los clientes una vez que han sido aceptados deben dejar garantías reales o prendarias. Los pagos son convenidos a través de documentos por cobrar como facturas, letras, pagarés, etc. Estos documentos deben contener la tasa de interés pactada, monto de crédito, plazos y modalidad de pago.

Cuando los clientes no cumplen con los compromisos adquiridos, la institución crediticia los cataloga como clientes morosos o malos clientes, según sus políticas de cobro. Y ésta información es enviada al buró de crédito, institución que guarda la información del comportamiento de pago del cliente en todos sus créditos (Nieto, 2010).

Tipos de créditos:

Según Nieto (2010), los tipos de créditos se dividen en:

- **Crédito al consumo o Créditos comerciales.** Son aquellos créditos otorgados por empresas para la adquisición de bienes o servicios de uso personal en plazos determinados.

- **Créditos Empresariales.** Cuando una empresa requiere materia prima, insumos, servicios, etc. solicita el bien o servicio a otras empresas a crédito para continuar su actividad empresarial, realizando convenios para cubrir el adeudo en un futuro.
- **Créditos Bancarios.** Son los otorgados por entidades bancarias o empresas del sistema financiero a personas físicas o morales de bienes, servicios, pagar deudas, etc.

El crédito usual en el mercado al minoreo es el crédito al consumo o créditos comerciales, por lo que el crédito al consumo ha crecido en Perú en el último año en 14% (SBS, 2015).

Tipos de Riesgos:

Según Gómez y Partal (2010) los tipos de Riesgos Financieros son los siguientes:

- **Riesgo de Mercado.** Variación de los precios o tipos negociados en los mercados financieros, este riesgo se subdivide en riesgo de interés, de tipo de cambio, de renta variable, de volatilidad, etc.
- **Riesgo de crédito.** Incumplimiento del contrato por una variación en las condiciones o características de la contrapartida.
- **Riesgo de Liquidez.** Desequilibrio entre las exigibilidades a corto plazo y las disponibilidades de fondos en ese mismo periodo.
- **Riesgo Operativo.** Factores relaciones con el control interno de la empresa, como errores humanos o de medios de producción o gestión.
- **Riesgo Legal.** Incapacidad legal para ejercer derechos o se deriva del propio incumplimiento de la ley.

2.1.1. Riesgo de Crédito

Según Piñero (2010), el riesgo de crédito mide la posible pérdida que asume cualquier entidad, como consecuencia del incumplimiento de las obligaciones contractuales. En nuestro caso, nos preocupa el riesgo de crédito generado por la concesión del crédito comercial, sin entrar en detalle del riesgo crédito de las entidades financieras. El ciclo del riesgo se resume en la Figura 1.

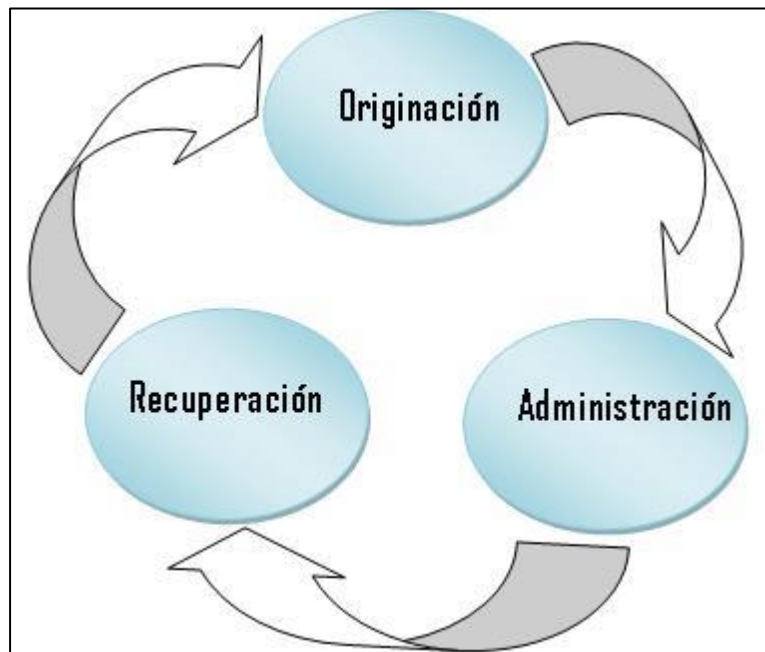


Figura 1: El ciclo del Riesgo

FUENTE: Piñero (2010)

- **Originación:** La intención en este punto es otorgar crédito a un cliente por primera vez en la institución.
- **Administración:** La intención en esta parte de ciclo es premiar a los clientes que se están “portando bien” (incrementos de límite de crédito) y castigar a los que se están “portando mal” (decrementos de límite de crédito); en ambos casos se

pueden combinar estrategias de incrementos o decrementos de tasas de interés, la reestructuración de deudas, etc. Aquí se busca la detección temprana de cuentas de alto riesgo y poder realizar acciones tempranas de corrección.

- **Recuperación:** En esta parte del ciclo de riesgo se pretende recuperar a todos aquellos clientes que dejaron de pagar. Se aplican actividades de recaudación a clientes con un alto puntaje de score según la empresa y determinar los clientes no recuperables para hacer el traspaso a una empresa recaudadora y así recuperar parte del capital perdido.

Según Gómez y Partal (2010), de entre los diferentes tipos de riesgos que afectan a la actividad de las empresas financieras, el riesgo de crédito es el típico, siendo el causante del mayor número de crisis financieras. Puede definirse como, “la pérdida potencial ocasionada por una variación en las condiciones y características de una contrapartida que altere la capacidad de ésta para cumplir con sus obligaciones contractuales”, riesgo estrechamente relacionado con solvencia.

Los motivos que pueden ocasionar una pérdida crediticia, se distingue entre riesgo de incumplimiento o impago (*default*) y, riesgo de pérdida de solvencia o de migración. El análisis de riesgo de crédito puede realizarse de forma individual y de cartera, el análisis individual o por posición crediticia, tiene en cuenta los siguientes factores de riesgo:

- **Probabilidad de incumplimiento:** hace referencia a la frecuencia relativa de ocurrencia del evento, donde el acreditado incumple con sus obligaciones contractuales. Esta probabilidad es diferente para cada acreditado al depender de las características particulares de cada uno.
- **Migración del crédito:** grado en que puede mejorar o deteriorarse la calidad crediticia o calificación del crédito.
- **Exposición crediticia:** hace referencia al tamaño previsto de la operación en el momento de incumplimiento de la contrapartida, considerando no recuperar ninguna parte del valor de la operación.

- **Tasa de pérdida o severidad:** porcentaje de la exposición perdido en caso de incumplimiento, es decir, porcentaje finalmente no recuperado, este importe equivale a uno menos la tasa de recuperación.

La pérdida crediticia esperada de una operación depende de las magnitudes siguientes:

- Probabilidad de que la contraparte incumpla sus obligaciones contractuales o probabilidad de incumplimiento (*probability of default – PD*).
- Exposición crediticia en el momento del incumplimiento (*exposure at default – EAD*).
- Porcentaje no recuperado del montante impagado o severidad (*loss given default – LGD*).

La pérdida esperada resulta del producto de estas magnitudes, expresada en valor absoluto:

$$\text{Pérdida crediticia esperada} = PD \times EAD \times LGD$$

Cada una de las variables que intervienen en el cálculo de la pérdida esperada están influidas por diferentes factores; la probabilidad de incumplimiento está estrechamente relacionada con la calidad crediticia (*scoring*) de la contrapartida; la exposición crediticia depende del tipo de operación; y la severidad, está condicionada fundamentalmente por las garantías aportadas.

2.1.2. *Credit Scoring*

El objetivo fundamental del *credit scoring*, es la maximización del beneficio derivado de la concesión de un volumen de crédito. Es por tanto un modelo de estimación de la probabilidad de incumplimiento de pago, a partir de las características del cliente que solicita el crédito, mediante la utilización del análisis de la base del comportamiento previo de otros individuos en similares condiciones (Piñeiro, 2010). En momentos como los actuales donde las pérdidas por morosidad que sufren todas las entidades comerciales están aumentando en consideración, se pone de manifiesto la importancia de una buena ejecución del planteamiento del *credit scoring*. El planteamiento que sigue cualquier modelo de *credit scoring* se puede resumir en los siguientes puntos:

- Problema dinámico, que involucra varios periodos dentro de un horizonte temporal.
- Método basado en la evaluación del valor presente descontado de los posibles beneficios y pérdidas derivados de la concesión a un individuo o sociedad.
- El candidato se le concederá el crédito si la esperanza del valor anterior es positivo.
- La máxima cuantía del crédito será aquella que iguale el valor anterior a cero.

Según Gómez y Partal (2010), los sistemas de *scoring* son utilizados en operaciones muy estandarizadas del segmento minorista (consumo, hipotecas, tarjetas de crédito, entre otras), cuando la intervención humana en la evaluación de la concesión, no reviste mayores ventajas frente al uso de puntuaciones fundamentadas en la información objetiva del cliente y la operación. Existen diferentes tipos de *scoring*:

- **Scoring reactivos**, herramientas de clasificación que determinan la posible concesión de nuevas operaciones a partir de información no relacionada con el comportamiento del cliente.

- **Scoring de comportamiento**, tienen en cuenta variables propias de la operación y del cliente disponibles internamente, en concreto, variables referentes al comportamiento de un producto en el pasado (retrasos en pago, mora, etc) y, el comportamiento del cliente con la entidad (saldo mantenidos, domiciliaciones, movimientos, cumplimiento de cuotas, etc.). Es utilizado para revisiones de límites y seguimiento del riesgo.
- **Scoring preactivos**, consideran las mismas variables que los *scoring* de comportamiento pero con una finalidad distinta, ofrecer nuevos productos al cliente, por ejemplo, realización de campañas de créditos preconcedidos a clientes particulares.

Según Arguedas et al. (2010) las ventajas de utilizar un *scoring*:

- Supone una reducción de costes y una mejora de la eficiencia en la evaluación de las operaciones por parte de los analistas. Con un sistema de *scoring*, dicha evaluación se hace de forma más objetiva en función de cuál sea el riesgo de la operación y no solo el importe de la misma. Por otro lado el sistema de *scoring* facilita el establecimiento de criterios mínimos para la concesión que asegure un nivel de riesgo aceptable dada la rentabilidad potencial, eliminándose costes en las solicitudes claramente rechazables y reduciendo tiempos de respuesta.
- Facilitan la medición del riesgo de crédito asumido por cada uno de los segmentos de la cartera, con el fin de garantizar que se obtenga un nivel aceptable de rentabilidad dado el coste potencial del riesgo. También sirven en la definición y gestión de políticas de crédito y establecimiento de límites de concentración
- Permiten hacer un seguimiento de la evolución de la cartera, al generar alertas sobre el posible deterioro de la calidad crediticia de los prestatarios. De esta forma

la entidad puede llevar a cabo las actuaciones necesarias con la mayor prontitud para minimizar las pérdidas en su cartera.

2.1.3. Construcción del *Scorecard*

Según Barberena y Berberana (2002) la metodología del proceso de modelación del *credit scoring* consta de las siguientes etapas fundamentales:

- **Preparación de la base de datos**, en esta fase inicial se procede a un estudio detallado de las funciones de distribución de cada una de las variables y a la limpieza de la base de datos, con el propósito de disminuir en lo posible la magnitud del problema.
- **Agrupamiento interactivo**, el objetivo de esta etapa es la clasificación, que es el proceso automático y/o interactivo de redistribución y agrupamiento de las variables de intervalos, ordinales y nominales con el propósito de:
 - a. Manejar el número de categorías por variable.
 - b. Mejorar el poder predictivo de la variable.
 - c. Seleccionar las mejores variables con nivel de poder de predicción.
 - d. Los atributos se forman tomando en cuenta su proporción de buenos y malos a través de una medida conocida como WOE (*weights of evidence*)

$$WOE_k = \ln \left(\frac{\text{Número de Buenos}}{\text{Número de Malos}} \right) \times 100$$

- **Determinar los clientes buenos y malos**, en este paso se utiliza la base de datos que registra el comportamiento de la moratoria mensual o pagos vencidos de los clientes. Se construye con estos datos una matriz de transición (*Roll Rate*) de un paso de seis meses. Basándose en esta matriz se determina cuáles de los clientes son buenos, malos e indeterminados en nuestra base de datos. En este punto se

forma una nueva base con los datos de los clientes que se clasificaron como clientes buenos y clientes malos (variable dependiente) con sus respectivas características (variables explicativas) extraídas de la base de solicitudes.

- **Determinar una función de clasificación**, una vez que se tiene la tabla de buenos y malos, con ella se procede a hacer una selección de las características que tienen mayor valor de predicción global. Para hacer esta selección se utiliza el Valor de Información (IV) que es una función de la proporción de buenos y malos clientes en los atributos de cada característica. Se forma una base que contiene buenos y malos, y las características con mayor poder de predicción. De esta base se genera una nueva base, de la siguiente manera:
 - a. Una columna para la variable dependiente con los valores 1, si el cliente fue clasificado como bueno y 0, si el cliente fue clasificado como malo. Una columna para cada característica, con los valores del WOE del atributo correspondiente para cada cliente.
 - b. Se obtiene una tabla que contiene únicamente los datos de los registros de los clientes buenos y malos de características con suficiente poder de predicción, reportando el valor del WOE correspondiente al atributo del cliente.
 - c. Finalmente con esta tabla se obtiene el modelo de predicción para los nuevos clientes, para ello se utiliza la regresión logística.

- **Elaboración del *scorecard***, aquí el WOE de cada atributo es multiplicado por el coeficiente de regresión de su característica para obtener el puntaje del *scorecard* del atributo. El puntaje total del solicitante es proporcional al logaritmo del *odds ratio* estimado.

- **Medir la eficiencia del *scorecard***, se utilizan métodos estadísticos como el índice de Gini, la prueba de Kolmogorov – Smirnov (K-S) y la curva de ROC para determinar que tan bien clasifica nuestro modelo de predicción.

- **Establecer el punto de corte**, se determina el punto de corte (*cut off*) que separará a las solicitudes nuevas en aceptados o rechazados.

2.1.4. Determinación de clientes buenos y malos

Según Nieto (2010), el riesgo crediticio puede dividir a los clientes en buenos o malos según la siguiente clasificación:

Se considera buen cliente a aquellos individuos que:

- Pagan el monto de su deuda en el periodo de gracia (entre los límites de pago o de corte).
- Cuando no cuenta con el capital para pagar la totalidad de la deuda pero pagan al menos el mínimo requerido por la empresa acreedora.
- Liquidan su adeudo en no más del tiempo determinado por la empresa. Generalmente como máximo a los tres meses.

Un cliente intermedio es aquel deudor que no se puede clasificar como bueno o malo. Debido a que tiene adeudos de capital e intereses acumulados y se requiere de más tiempo para ver la tendencia de su comportamiento.

Un mal cliente se refiere a aquel deudor que causa pérdidas económicas a la compañía. Este tipo de clientes no pagaron su cuenta, aún después de aplicarles técnicas de cobranzas. Comúnmente es considerado mal cliente después de 90 días de mora. Por experiencia se sabe que cuando ya está en esta etapa o estado es difícil que regrese al estado de buen cliente.

2.1.5. Ventana de muestra

Según Siddiqi (2006), los *scorecards* se desarrollan utilizando el supuesto de que "el rendimiento futuro reflejará los resultados anteriores". Basándose en esta suposición, el rendimiento de cuentas abiertas previamente se analiza con el fin de predecir el rendimiento de cuentas futuras. Para llevar a cabo este análisis, se necesita reunir datos para las cuentas abiertas durante un tiempo específico y luego monitorear su desempeño para otro período de tiempo específico para determinar si eran buenos o malos. Los datos recogidos (las variables), junto con la buena / mala clasificación (el objetivo) constituye la muestra de desarrollo a partir del cual se realiza el *credit scoring*.

La Figura 2 asume una nueva cuenta que es aprobada y se le concede crédito en un momento determinado. En algún momento en el futuro, es necesario determinar si esta cuenta había sido buena o mala (para asignar el rendimiento). "Ventana de Rendimiento" es la ventana de tiempo en el que el rendimiento de cuentas abiertas durante un período de tiempo particular (es decir, la ventana de muestra) se controla para asignar la clase (de destino). "Ventana de Muestra" se refiere al período de tiempo a partir del cual se seleccionarán los casos buenos y malos conocidos para la muestra de desarrollo.

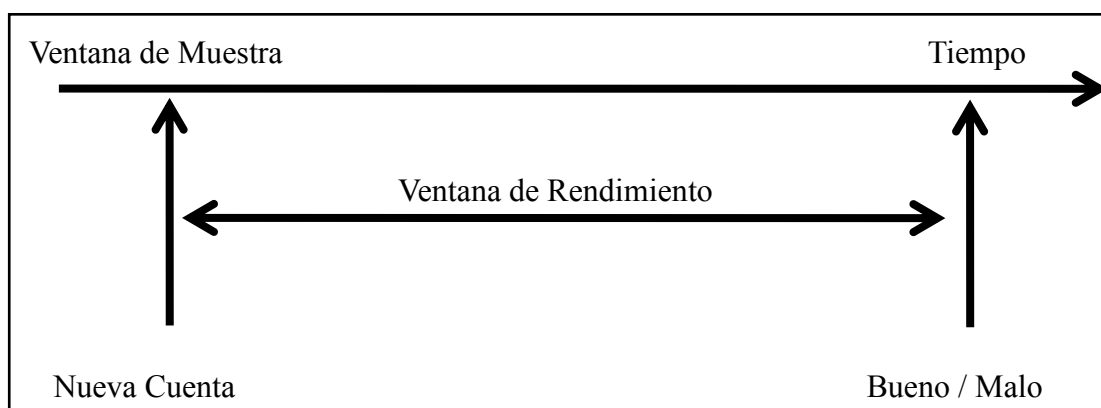


Figura 2: Definición del rendimiento

FUENTE: Siddiqi (2006)

La Figura 3 muestra un ejemplo de una cartera de tarjetas de crédito típico donde la tasa de malos clientes se ha graficado para las cuentas abiertas en un período de 14 meses. Esto demuestra que la tasa de malos clientes se desarrolla rápidamente en los primeros meses y luego se estabiliza cerca a los 12 meses.

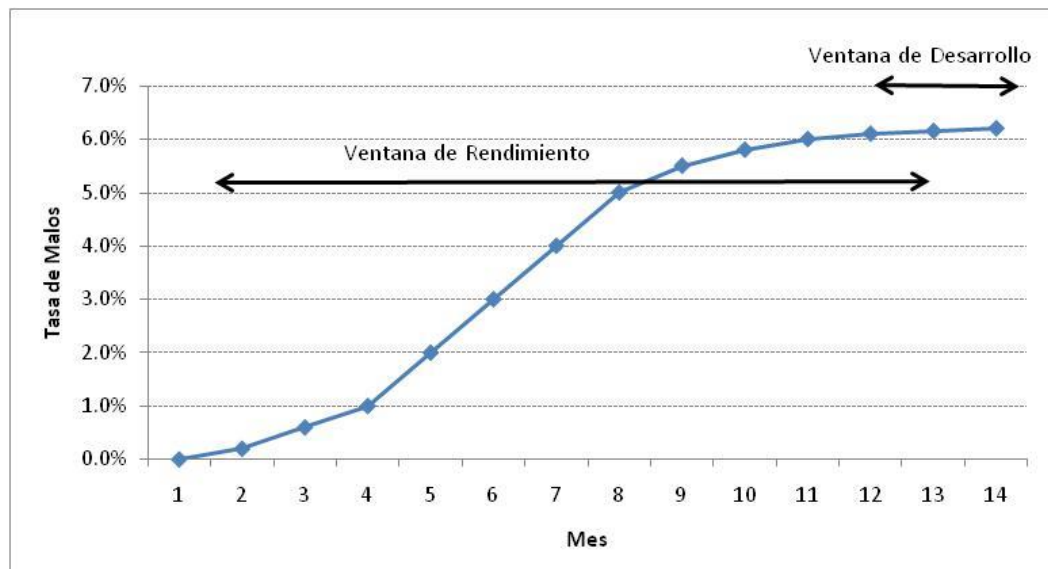


Figura 3: Evolución de la tasa de malos

FUENTE: Siddiqi (2006)

La ventana de desarrollo es elegida de un período de tiempo en que se considera que la tasa de malos clientes sea estable, o cuando se considere que la muestra de seguimiento ha madurado (es decir, donde la tasa de malos clientes comienza a estabilizarse). En el ejemplo anterior, una buena ventana de desarrollo podría estar en cualquier lugar entre 12 y 14 meses en el pasado.

Una vez que se conoce el tiempo que se requiere para llegar a la estabilización de las cuentas y la ventana de desarrollo se debe establecer que cuentas están contenidas en este periodo, para que formen parte de la muestra que se usará para construir los modelos

estadísticos, en consecuencia requiere determinar el periodo de tiempo donde estarán incluidas las fechas de alta de estas cuentas.

2.1.6. Proceso para determinar clientes buenos y clientes malos

Según Nieto (2010), para determinar a los clientes buenos y a los clientes malos se construye una matriz de transición para analizar la moratoria de los clientes. Esta matriz de transición en el contexto de *credit scoring* se le conoce como *Roll Rate*. Con esta matriz de transición veremos el comportamiento de las cuentas en la institución después de un periodo de tiempo, generalmente de seis meses. Aquí se identifican los estados que se utilizarán como marca para determinar si un cliente es bueno, indeterminado o malo.

Los estados de la matriz de transición (*Roll Rate*) se definen en función del número de pagos vencidos, y una vez que se tienen definidos los estados se estiman las probabilidades de transición, $P(X_2=j / X=i)$, con los datos de la base a trabajar. Independientemente del momento en que se apertura una cuenta el periodo de observación o ventana de muestra se contabiliza como primer mes, segundo mes, tercer mes, etc. a partir del momento de apertura, esto es, debemos “alinearse” todas las cuentas a un punto inicial igual a cero.

Para nuestro análisis el periodo de evaluación será de 12 meses.

Los estados deben de ser una partición de las cuentas; esto es, una cuenta debe pertenecer exclusivamente a un estado. Considerando la situación de la cuenta al final de los 6 meses, los estados se indican de acuerdo al Cuadro 1.

Cuadro 1: Posibles estados al final de los 6 meses

Estado	Descripción al final del periodo
PV0	Al día
PV1	1 - 30 días de mora
PV2	31 - 60 días de mora
PV3	61 - 90 días de mora
PV4	90 a más días de mora

FUENTE: Elaboración propia

Al incorporar la información de las cuentas durante los seis meses obtenemos un refinamiento de estos estados, por ejemplo, el estado PV 0 (al día) se particionan en cinco nuevos estados que se reportan en el Cuadro 2.

Cuadro 2: Posibles estados al día en los primeros 6 meses

Estado	Descripción al final del periodo
PV00	Al día y tiene máximo 0 días de atraso durante el periodo
PV01	Al día y tiene máximo entre 1-30 días de atraso durante el periodo
PV02	Al día y tiene máximo entre 31-60 días de atraso durante el periodo
PV03	Al día y tiene máximo entre 61-90 días de atraso durante el periodo
PV04	Al día y tiene máximo entre 90 a más días de atraso durante el periodo

FUENTE: Elaboración propia

Esta misma idea se utiliza para construir la siguiente partición para los demás estados. Así, el último estado será PV 44 como único elemento para 4 o más pagos vencidos en el sexto mes y como máximo 4 o más pagos vencidos en el mes 12. Así se construye una partición de 25 estados. El estado PV 44 es un comportamiento indeseado por las instituciones de crédito, los clientes que están en este estado se pueden identificar como malos "graves". El malo "grave" es típicamente definido basándose en castigo o *write off* o mora mayor a 90 días.

Ahora, se busca encontrar que estados en X_1 conducen al estado no deseado, PV44, en X_2 . Los elementos de la matriz de transición (*Roll Rate*) son las estimaciones de las probabilidades de estar en el estado j en los primeros seis meses a pasar al estado i en los seis meses siguientes:

$$\hat{P}(X_2 = i | X_1 = j) = \frac{\text{Numero de cuentas que estaban en el estado } j \text{ y pasaron al estado } i}{\text{Numero de cuentas que estaban en el estado } j}$$

Una vez que se tenga la matriz de transición (*Roll Rate*) correspondiente debe hacerse un análisis para determinar que estados tienen una alta probabilidad de pasar al estado PV 44 o en el siguiente periodo. La idea es estimar la probabilidad de caer en PV 44, esto es:

$$\hat{P}(X_2 = i | X_1 = j) \text{ donde } i = PV00, PV01, \dots, PV44$$

Y proceder de la siguiente manera:

$\hat{P}(X_2 = i | X_1 = j) < a$, las cuentas que están en el estado i se consideran buenas.
 $\hat{P}(X_2 = i | X_1 = j) > b$, las cuentas que están en el estado i se consideran malas.
 Si $a \leq \hat{P}(X_2 = i | X_1 = j) \leq b$, las cuentas que están en el estado i se consideran indeterminadas.

Los números a y b mencionados anteriormente satisfacen la relación $0 < a < b < 1$ y la institución crediticia es quien decide el valor de estos números.

2.2. Empresa al minoreo

La palabra “Al minoreo” se define como “Comercialización al por menor” o “Venta al detalle”. Aunque usualmente se utiliza para referirse al rubro de supermercados y tiendas por departamentos, en estricto rigor, los negocios tipo al minoreo abarcan desde el almacén de nuestro barrio o el quiosco de la esquina hasta las grandes multitiendas e hipermercados.

La industria al minoreo en Chile ha mostrado un gran dinamismo durante los últimos años. De hecho, las empresas nacionales más importantes se han instalado en otros países de la región, principalmente Argentina, Perú y Colombia (Duran y Kremerman, 2008).

Según Equilibrium Clasificadora de Riesgo (2013), la industria al minoreo en el Perú está conformada principalmente por la industria de supermercados las cuales son: Grupo de Supermercados Wong (Grupo Cencosud), Supermercados Peruanos (Grupo Interbank) e Hipermercados Tottus (Grupo Falabella), de los cuales una de las tres empresas es peruana. Adicionalmente con un formato de oferta diferente se encuentra el Grupo MAKRO que es una organización mayorista de productos alimenticios y no alimenticios con más de 50 años de experiencia en el mercado mundial.

Actualmente, se reconocen tres grandes actores en el negocio al minoreo a nivel Sudamericano: Cencosud, Falabella y Ripley (empresas Chilenas). Existen actores, como Cencosud y Falabella que integran completamente la operaciones del comercio al minoreo, siendo propietarios de supermercados, tiendas por departamento, tiendas para el mejoramiento del hogar, centros comerciales y participación activa en el minoreo financiero a través de tarjetas de créditos, banca y seguros (Duran y Kremerman, 2008).

Según Equilibrium Clasificadora de Riesgo (2013), de acuerdo a los distintos formatos o canales de venta que existen, el minoreo podría clasificarse en: i) Supermercados; ii)

Grandes Tiendas; iii) Ferreterías; iv) Farmacias y Perfumerías; v) Comercio Tradicional; vi) Consumo Local y vii) Otros formatos.

La industria al minoreo cubre un gran abanico de necesidades de las personas, tales como la alimentación, vestuario y calzado, equipamiento para el hogar (artículos electrónicos, electrodomésticos, menaje, ferretería, jardinería), remedios y artículos de cuidado personal e incluso accesorios básicos para automóviles.

En relación a los Artículos Electrónicos, Computación, Electrodomésticos y Línea Blanca, operan distintos distribuidores minoristas: tiendas grandes y medianas, supermercados y tiendas especializadas, las cuales tienen diversos formatos y utilizan distintos modelos de negocio (Duran y Kremerman, 2008).

Según Duran y Kremerman (2008), en el mercado al minoreo define en tres grandes grupos de tiendas, los cuales son:

- a. Tiendas por Departamento:** Corresponden a empresas que operan, en general, con gran cobertura. Su gran volumen de ventas les otorga un importante poder de negociación con los proveedores y, por lo tanto, obtienen una significativa ventaja en costos. Ofrecen amplias líneas de vestuario y productos para el hogar; poseen gran superficies de venta, presencia en *malls* y una extensa cartera de clientes.

- b. Multitiendas:** Operan con una menor gama de productos que las tiendas por departamento, se dirigen a segmentos socioeconómicos más bajos y no poseen presencia relevante en los centros comerciales más importantes. A pesar de poseer más de un local de ventas, se estima que el volumen de ventas de estas tiendas es menor.

- c. **Tiendas y Cadenas de Especialistas:** Las primeras están dirigidas a satisfacer pequeños nichos de mercado, que operan indistintamente en *malls* y sectores comerciales. Las segundas desarrollan su operación a nivel nacional y se especializan en una línea de productos, la que, sin embargo, puede ser muy amplia.

Cuadro 3: Ranking del sector al minoreo 2012

Puesto	País	Puntaje	Cambios respecto al 2012
1	Brazil	69.5	-
2	Chile	67.1	-
3	Uruguay	66.5	+ 1
4	China	66.1	-1
5	Emiratos Árabes Unidos	63.5	+ 2
6	Turquía	62.6	+ 7
7	Mongolia	62.5	+ 2
8	Georgia	61.4	- 2
9	Kuwait	58.4	+ 3
10	Armenia	58.2	N/A
11	Kazajistán	57.5	+ 8
12	Perú	56.5	- 2

FUENTE: Equilibrium Clasificadora de Riesgo (2013)⁴

Para el caso peruano, es de mencionar que ha retrocedido dos puestos en comparación al ranking del año 2012 en el sector al minoreo, encontrándose a la fecha como una economía “a considerar”. No obstante, el estudio destaca la favorable evolución de la economía en los últimos años, los alentadores índices de confianza del consumidor observados, y el

⁴ [http:// www. equilibrium. com. pe/sectorialsuperjun13.pdf](http://www.equilibrium.com.pe/sectorialsuperjun13.pdf)

crecimiento de la clase media producto de los mayores ingresos, las mayores tasas de empleo y el sólido acceso al crédito (Equilibrium Clasificadora de Riesgo, 2013)⁵. En comparación con mercados latinoamericanos más desarrollados como Chile o México, los cuales cuentan con empresas al minoreo con un promedio de 97 y 109 m² por cada 1,000 habitantes respectivamente. Perú tiene aún un amplio espacio para crecer, al mantener 17m² por cada 1,000 habitantes, no obstante el fuerte crecimiento observado en la industria nacional. De esta manera, es factible continuar esperando un incremento en la penetración de los supermercados y de la industria al minoreo en general, más aún cuando se espera que sea Perú el país latinoamericano que presente un crecimiento más dinámico del PBI en los próximos años.

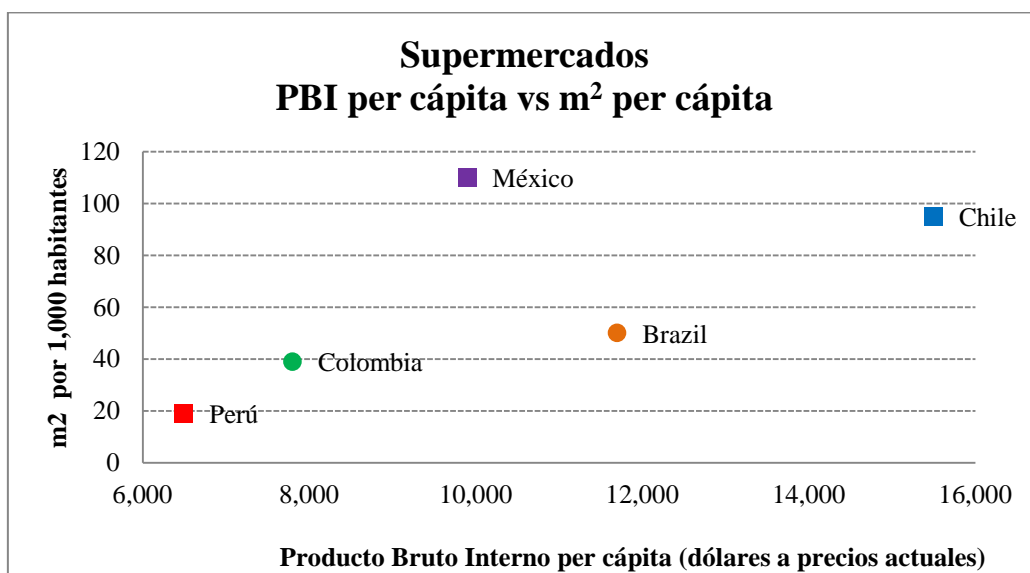


Figura 4: Supermercados PBI per cápita vs m² per cápita

FUENTE: Equilibrium Clasificadora de Riesgo (2013)⁶

⁵ [http:// www. equilibrium. com. pe/sectorialsuperjun13.pdf](http://www.equilibrium.com.pe/sectorialsuperjun13.pdf)

⁶ [http:// www. equilibrium. com. pe/sectorialsuperjun13.pdf](http://www.equilibrium.com.pe/sectorialsuperjun13.pdf)

2.3. Matrices de Transición

Una matriz de transición crediticia presenta una manera sucinta de describir la evolución de los créditos a lo largo de las clasificaciones crediticias, basada en un modelo de probabilidades de transición del tipo Markov. El modelo de probabilidades de transición de Markov empieza por definir los estados de la naturaleza. Para este caso es, clasificar a los clientes de acuerdo a su estatus moratorio, esto es de acuerdo al número de días que se encuentre en mora, se puede clasificar a los clientes en buenos, malos o indeterminados.

Se requiere hacer predicciones sobre el comportamiento futuro de los clientes y para este fin es conveniente conocer la probabilidad de que un cliente pase de un estado a otro a través del tiempo y para ello son útiles las matrices de transición.

La matriz de transición se define como las probabilidades de que los clientes con una calificación i migren a otra calificación j en un horizonte de tiempo dado, como se puede observar en el cuadro

Cuadro 4: Matriz de Probabilidad de Transición

		Categoría después de la transición				
		1	2	...	j-1	J
Categoría antes de la transición	1	P_{11}	P_{12}	P_{1j}
	2	P_{21}	P_{22}	P_{2j}

	i-1	$P_{(i-1)j}$
	i	P_{i1}	P_{i2}	...	$P_{(j-1)i}$	P_{ij}

Fuente: SBS (2013)⁷

⁷ http://www.sbs.gob.pe/repositorioaps/0/0/jer/ddt_ano2013/SBS-DT-001-2013.pdf

1. La primera columna de la representa la escala de calificaciones de inicio de período.
2. La primera fila superior representa la calificación en el período final (es decir, la calificación con que termina el cliente en el periodo analizado).
3. La intersección (diagonal), representa el porcentaje de calificaciones que se mantuvieron, que pueden aumentar o disminuir.
4. Las celdas por debajo de la diagonal representan las probabilidades de que las calificaciones mejoren.
5. Las celdas por encima de la diagonal representan las probabilidades de que las calificaciones empeoren.

Adicionalmente, la matriz de probabilidad de transición debe cumplir con las siguientes condiciones:

- Todos los elementos de la matriz deben ser positivos, es decir, $P_{ij} > 0$, para todo i, j .
- La suma de los elementos de cada fila debe ser igual a 1, es decir, $\sum_i P_{ij} = 1$, para todo i .

Donde P_{ij} representa la fracción de créditos con calificación de i después de un periodo tendrán calificación j .

$$P_{ij} = \frac{n_{ij}}{n_i}$$

Dónde:

- n_{ij} : Número de créditos que comenzaron al inicio del periodo en la calificación i y terminaron al finalizar el periodo en la calificación j .
- n_i : Número de créditos que estaban en la calificación i al comienzo del periodo.

Ejemplo 1: Suponga que el comportamiento de selección de compra de un operador de servicio de telefonía móvil de un grupo de clientes se puede modelar con una cadena de Markov. El Cuadro 5 muestra a 240 consumidores en un periodo de 12 meses, en el cual compiten tres operadores para el producto de servicio de telefonía móvil. La elección fue en el mes 1 y el cambio de marca fue en el mes 12.

Cuadro 5: Consumidores y su preferencias del mes 1 al mes 12

		Operador en el mes 12			Total
		Claro	Movistar	Entel	
Operador en el mes 1	Claro	86	20	8	114
	Movistar	14	56	7	77
	Entel	5	10	34	49
Total		105	86	49	240

Fuente: Elaboración propia

De los 240 consumidores que compraron el operador en el mes 1, 114 adquirieron a Claro, 77 a Movistar y 49 a Entel. Así mismo, de aquellos consumidores que compraron el servicio de Claro en el mes 1, visto en el mes 12, 86 se quedaron en Claro, mientras que 20 compraron Movistar y 8 compraron Entel. La matriz de transición se puede representar con una matriz de 3 estados, las probabilidades de transición se estiman de los datos del cuadro 5, de acuerdo a la formula $\hat{p}_{ij} = \frac{n_{ij}}{n_i}$, para todo i, j donde n_{ij} es el número de consumidores que escogieron el operador i en el mes 1 y cambiaron al operador j en el mes 12. El denominador n_i es el número de consumidores que escogieron el operador i . la matriz de transición estimada es:

$$\hat{P} = \begin{bmatrix} \hat{p}_{11} & \hat{p}_{12} & \hat{p}_{13} \\ \hat{p}_{21} & \hat{p}_{22} & \hat{p}_{23} \\ \hat{p}_{31} & \hat{p}_{32} & \hat{p}_{33} \end{bmatrix} = \begin{bmatrix} \frac{86}{114} & \frac{20}{114} & \frac{8}{114} \\ \frac{14}{77} & \frac{56}{77} & \frac{7}{77} \\ \frac{5}{49} & \frac{10}{49} & \frac{34}{49} \end{bmatrix}$$

$$\hat{P} = \begin{bmatrix} 0.75 & 0.18 & 0.07 \\ 0.18 & 0.73 & 0.09 \\ 0.10 & 0.20 & 0.70 \end{bmatrix}$$

La entrada $p_{12} = P(X_{12} = 2 | X_1 = 1) = 0.18$, la entenderemos como la probabilidad de que un cliente compra el servicio de telefonía móvil Claro, en el mes 1, prefiera comprar el servicio de telefonía móvil Movistar en el mes 12, es decir, se estima que el 18% haga esta transición.

2.4. Regresión logística

Según Agresti (2002), este es el modelo más importante para los datos de las respuestas categóricas. Se utiliza cada vez más en una amplia variedad de aplicaciones. Los primeros usos fueron en estudios biomédicos, pero los últimos 20 años también han visto mucho uso en la investigación de las ciencias sociales y el marketing.

Recientemente, la regresión logística se ha convertido en una herramienta popular en las aplicaciones empresariales. Algunas aplicaciones de *credit scoring* utilizan regresión logística para modelar la probabilidad de que un sujeto es digno de crédito. Por ejemplo, la probabilidad de que un sujeto pague una factura a tiempo puede utilizar predictores tales como el tamaño de la letra, el ingreso anual, la ocupación, las obligaciones hipotecarias y de deuda, porcentaje de facturas pagadas a tiempo en el pasado, y otros aspectos de un solicitante de historial de crédito. Así mismo una empresa que se basa en la venta por catálogo puede determinar si se debe enviar un catálogo a un cliente potencial al modelar la probabilidad de una venta en función de los índices de comportamiento de compra del pasado.

Según Nieto (2010), el modelo de Regresión Logística no requiere de los supuestos de la Regresión Lineal, como son el supuesto de normalidad de los errores, homocedasticidad y el supuesto que las variables involucradas sean continuas. En este sentido, la Regresión Logística, se aplica tanto a datos que se distribuyen como normal, como a datos que no lo son, y por lo tanto, el modelo de regresión logística es útil cuando la variable de respuesta “y” no está distribuida normalmente y tanto las variables predictoras como de respuesta tienen valores discretos, categóricos, ordinales o no ordinales.

La regresión lineal no es aplicable a este tipo de variables, dado que la variable respuesta “y” sólo presenta dos valores. La capacidad predictiva del modelo logístico se valora mediante la comparación entre el grupo de pertenencia observado y el pronosticado por el modelo. El modelo debe ser capaz de clasificar a los individuos en cada uno de los dos grupos: buenos o malos, basado en las variables que definen las características de los individuos.

Según Luque (2000), una excelente alternativa para garantizar que la respuesta prevista esté entre 0 y 1 es utilizar una función de enlace no lineal que se monótona, creciente y acotada entre dichos valores, de tal modo el modelo de Regresión Logística surge cuando se utiliza la función de distribución logística para modelizar la relación entre la probabilidad de $Y=1$, condicionada a un determinado valor de la variable(o variables) independientes, x_i :

$$\hat{p}_i = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \dots + \hat{\beta}_n x_{ni}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \dots + \hat{\beta}_n x_{ni}}}$$

Según Peña (2002), considerando el problema de la discriminación entre dos poblaciones, una forma de abordar el problema es definir una variable de clasificación, “y”, que tome el valor cero cuando el elemento pertenece a la primera población, P1, y uno cuando pertenece a la segunda, P2. Entonces, la muestra consistirá en “n” elementos del tipo (y_i, x_i) , donde y_i es el valor en ese elemento de la variable binaria de clasificación y x_i un

vector de variables explicativas. A continuación, se construye un modelo para prever el valor de la variable ficticia binaria en un nuevo elemento cuando se conocen las variables “x”. El primer enfoque simple es formular el modelo de regresión:

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

Si se estiman los parámetros por mínimos cuadrados este procedimiento es equivalente a la función lineal discriminante de *Fisher* y es óptimo para clasificar si la distribución conjunta de las variables explicativas es normal multivariante, con la misma matriz de covarianzas. Sin embargo, este modelo presenta problemas de interpretación. Tomando esperanzas para $x = x_i$

$$E[y|x_i] = \beta_0 + \beta_1'x_i$$

Denominando p_i a la probabilidad de que “y” tome el valor 1 (pertenezca a la población P_2) cuando $x = x_i$

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

La variable “y” es binomial y toma los valores posibles uno y cero con probabilidades p_i y $1-p_i$.

Su esperanza será:

$$E[y|x_i] = p_i \times 1 + (1 - p_i) \times 0 = p_i$$

De las ecuaciones anteriores, se concluye que:

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

Según Johnson y Wichern (2007), esta formulación tiene dos problemas principales:

Si se estima el modelo lineal, la predicción de la respuesta binaria “y” se podrían obtener probabilidades mayores que la unidad o negativas, porque la expresión lineal para su valor

esperado es ilimitada. Esto no es un problema para clasificar la observación, pero sí lo es para interpretar el resultado de la regla de clasificación.

En el análisis de regresión se supone que la varianza de “y” es constante a través de todos los valores de la variable predictora “x”. Se ha mostrado que esto no es el caso, para estimar los parámetros del modelo se debería utilizar mínimos cuadrados ponderados.

Según Peña (2002), a pesar de estos dos inconvenientes, este modelo simple estimado por mínimos cuadrados conduce a una buena regla de clasificación, ya que, según la interpretación de *Fisher*, maximiza la separación entre los grupos, sea cual sea la distribución de los datos. Sin embargo, cuando los datos no son normales, o no tienen la misma matriz de covarianzas, la clasificación mediante una ecuación de relación lineal no es necesariamente óptima.

Si se quiere que el modelo construido para discriminar nos proporcione directamente la probabilidad de pertenecer a cada población, debemos transformar la variable respuesta para garantizar que la respuesta prevista esté entre cero y uno. Escribiendo:

$$p_i = F(\beta_0 + \beta_1'x_i)$$

p_i estará entre cero y uno si escogemos F para que tenga esa propiedad. La clase de funciones no decrecientes acotadas entre cero y uno es la clase de las funciones de distribución, por lo que el problema se resuelve tomando como F cualquier función de distribución:

Tomar como F la función de distribución logística dada por:

$$1 - p_i = \frac{1}{1 + e^{\beta_0 + \beta_1'x_i}}$$

Resulta:

$$g_i = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1'x_i$$

Que es un modelo lineal en esta transformación que se denomina *Logit*. La variable *Logit*, g , representa en una escala logarítmica la diferencia entre las probabilidades de pertenecer a ambas poblaciones, y al ser una función lineal de las variables explicativas nos facilita la estimación y la interpretación del modelo.

$$g_i = \frac{p_i}{1-p_i} = \exp(\beta_0 + \beta_1'x_i)$$

Se obtiene:

$$p_i = \frac{\exp(\beta_0 + \beta_1'x_i)}{1 + \exp(\beta_0 + \beta_1'x_i)}$$

2.4.1. Interpretación del modelo logístico

Los parámetros del modelo son β_0 , la ordenada en el origen, y $\beta_1 = (\beta_1, \dots, \beta_p)$, las pendientes. A veces se utilizan también como parámetros $\exp(\beta_0)$ y $\exp(\beta_i)$, que se denominan los *odds ratios* o ratios de probabilidades, el cual hace referencia al incremento unitario en una de las variables explicativas del modelo.

$$odds = \frac{p_i}{1-p_i}$$

2.4.1.1. Estimación de Máxima Verosimilitud

Según Johnson y Wichern (2007), considere el modelo con varias variables de predicción. Sea $(x_{i1}, x_{i2}, \dots, x_{ir})$ los valores para los r predictores para la i -ésima observación.

Las estimaciones de β 's se pueden obtener por el método de máxima verosimilitud. La verosimilitud de L está dada por la distribución de probabilidad conjunta evaluado en los recuentos observados y_i . Por lo tanto:

$$L(b_0, b_1, \dots, b_r) = \prod_{i=1}^n p^{y_i(x_i)}(1 - p(x_i))^{1-y_i}$$

$$= \frac{\prod_{i=1}^n e^{y_i(b_0 + b_1 x_{i1} + \dots + b_r x_{ir})}}{\prod_{i=1}^n (1 + e^{b_0 + b_1 x_{i1} + \dots + b_r x_{ir}})}$$

Los valores de los parámetros que maximicen la probabilidad no se pueden expresar en una solución de forma cerrada agradable como en el caso de los modelos de la teoría lineales normales. En su lugar, se deben determinar numéricamente comenzando con una estimación inicial y la iteración hasta el máximo de la función de probabilidad. Técnicamente, este procedimiento se le llama método de mínimos cuadrados iterativos ponderados.

2.4.1.2. Intervalos de Confianza para los parámetros

Cuando el tamaño de la muestra es grande, $\hat{\beta}$ es aproximadamente normal con media β , y la matriz de covarianza aproximada:

$$\widehat{Cov}(\hat{\beta}) \approx \left[\sum_{i=1}^n \hat{p}(x_i)(1 - \hat{p}(x_i))x_i x_i' \right]^{-1}$$

Las raíces cuadradas de los elementos diagonales de esta matriz son las desviaciones estándar estimadas o errores estándar (SE) de los estimadores $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_r$ respectivamente. El intervalo de confianza para β_k con un 95% de confianza es:

$$\hat{\beta}_k \pm 1.96 SE(\hat{\beta}_k) \quad k = 0, 1, \dots, r$$

Los intervalos de confianza se pueden utilizar para evaluar la significación de los términos individuales en el modelo para el Logit. Intervalos de confianza de una muestra grande para el Logit y para la proporción de la población $p(x_i)$ también puede ser construido.

2.4.2. Pruebas estadísticas del modelo Logit

Según Johnson y Wichern (2007), una de las características deseables de los modelos utilizados es que sus estimadores tengan capacidad discriminadora. Para medir la capacidad discriminadora se aplican diferentes técnicas de prueba que a continuación veremos:

2.4.2.1. Devianza

Si la hipótesis nula es $H_0: \beta_k = 0$, los cálculos numéricos de nuevo dan la estimación de máxima verosimilitud del modelo reducido y, a su vez, el valor maximizado de la probabilidad.

$$L_{max,Reduced} = L(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{k-1}, \hat{\beta}_{k+1}, \dots, \hat{\beta}_r)$$

Al hacer la regresión logística, es común utilizar para probar la hipótesis nula (H_0) el valor menos el doble de la probabilidad log-ratio

$$-2 \ln \left(\frac{L_{max,Reduced}}{L_{max}} \right)$$

Que en este contexto, se llama *devianza* o desviación. Se distribuye aproximadamente como chi-cuadrado con 1 grado de libertad cuando el modelo reducido tiene un menor número de variables predictoras. La hipótesis nula ($H_0: \beta_k = 0$) es rechazada para un gran valor de la *devianza*.

2.4.2.2. Estadístico de *Wald*

Para determinar si una variable debe ser incluida en un modelo porque tiene un peso significativo se aplica la prueba de estadístico de *Wald*. La prueba resulta de contrastar la hipótesis nula

$$H_0: \beta_i = 0$$

Contra la alternativa

$$H_1: \beta_i \neq 0$$

Con un estadístico de prueba definido como

$$W = \left[\frac{\hat{\beta}}{SE(\hat{\beta}_i)} \right]^2$$

Sigue una distribución ji-cuadrado. Para variables categóricas, el número de grados de libertad será igual al número de categorías menos uno. La prueba de razón de verosimilitud es preferible a la prueba de *wald* ya que el nivel de esta prueba es generalmente más cercano de la nominal α .

2.4.2.3. Residuales de Pearson

Las pruebas de hipótesis que se contrastan son:

H_0 : El modelo es adecuado

H_1 : El modelo no es adecuado

Los residuos del modelo Logit se definen por:

$$\hat{e}_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}}$$

Si el modelo es adecuado \hat{e}_i tiene media cero y varianza uno, de aquí se construye el estadístico de prueba que se distribuye asintóticamente como:

$$\chi_c^2 = \sum_{i=1}^n \hat{e}_i^2$$

Que sigue una distribución χ^2 con $n - (p + 1)$ grados de libertad con p variables.

2.4.3. Criterios para la validación del modelo

2.4.3.1. Tabla de Clasificación

Para validar la eficacia del método de clasificación, se utilizan datos de clientes de la misma población los cuales se conoce a que población pertenece, pero diferentes a los utilizados para estimar el modelo, para poder determinar la tasa de error aparente expresada en porcentaje.

Cuadro 6: Cuadro de criterios de clasificación

	Clasificados		
Realidad	Buenos	Malos	Total
Buenos	n_{11}	n_{12}	n_1
Malos	n_{21}	n_{22}	n_2

FUENTE: Luque (2000)

n_{11} = número de clientes buenos clasificados como buenos.

n_{12} = número de clientes buenos clasificados como malos.

n_{21} = número de clientes malos clasificados como buenos.

n_{22} = número de clientes malos clasificados como malos.

Tasa de Error Aparente:

$$Error = \left(\frac{n_{12} + n_{21}}{n_1 + n_2} \right)$$

El error tipo I se origina al clasificar como “Mal” cliente a un cliente que en realidad era “Bueno” y, en caso contrario, el error tipo II se relaciona con clasificar a un cliente como “Bueno”, siendo que en realidad era un “Mal” cliente.

El error tipo I genera un costo de oportunidad al no recibir las utilidades generadas para la empresa por un crédito no cursado y, por su parte, el error de tipo II genera una pérdida para la empresa al tener que aumentar el riesgo total de la cartera aumentando la probabilidad de no recuperar el monto prestado.

2.4.3.2. Kolmogorov-Smirnov (K-S)

Según Mays (2001), el test mide la diferencia entre la distribución acumulada de buenos y la distribución acumulada de malos.

Esta prueba es una de las más utilizadas para validar un modelo de *scoring* ya que mide el poder de discriminación de la *Scorecard* entre buenos y malos y se ve en la máxima diferencia entre las dos distribuciones. Entre más grande sea este coeficiente mejor será el modelo, pero tener en cuenta que si el valor es mayor a 75% es sospechoso.

$$K - S = \max |F_{malo}(x) - F_{bueno}(x)|$$

$F_{malo}(x)$: Distribución Acumulativa de Malos Clientes

$F_{bueno}(x)$: Distribución Acumulativa de Buenos Clientes

Y la regla para identificar al mejor K-S es de acuerdo al siguiente cuadro

Cuadro 7: Regla de indicadores K-S

%	Resultado
K-S < 20	Malo
20 < K-S < 40	Regular
40 < K-S < 60	Bueno
61 < K-S < 75	Muy Bueno
K-S > 75	Sospechoso

FUENTE: Mays (2001)

2.4.3.3. Coeficiente Gini

Permite observar la discriminación entre buenos y malos por rango de Score. Mide la eficacia de la *scorecard* al comparar el porcentaje de buenos clientes frente al porcentaje de malos para los mismos puntajes. Si el porcentaje de malos clientes se traza contra el porcentaje de buenos clientes para una serie de bandas de puntajes el resultado es una curva.

$$Gini = 1 - \sum_{i=2}^n \left[\left(F_{malo(i)} - F_{malo(i-1)} \right) \times \left(F_{bueno(i)} + F_{bueno(i-1)} \right) \right]$$

2.4.3.4. Curva ROC

Según Hosmer y Lemeshow (2000), la curva de características operativas (*Receiver Operating Characteristic*) permite cuantificar la precisión discriminatoria de un modelo. Este análisis se realiza por medio de una gráfica denominada curva ROC, el cual es un análisis de sensibilidad, *sensitivity*, que es la habilidad para marcar valores positivos que están correctamente clasificados y especificidad, *specificity*, que es la habilidad de marcar valores negativos que están correctamente clasificados, por lo tanto la sensibilidad es la verdadera tasa de valores positivos y la especificidad la tasa de valores negativos. Cuando

esta curva tiende hacia la izquierda, entonces se dice que el modelo tiene un buen poder de predicción.

El rendimiento global de una prueba de diagnóstico se suele resumir el área bajo la curva ROC, el cual según:

Cuadro 8: Regla de Indicadores Curva ROC

%	Resultado
ROC = 50	Esto no sugiere discriminación
70 < ROC < 80	Esto se considera una discriminación aceptable
80 < ROC < 90	Esto se considera una discriminación excelente
ROC > 90	Esto se considera una discriminación sobresaliente

FUENTE: Hosmer y Lemeshow (2000)

2.5. Marco conceptual

- 1. Crédito:** Cantidad de dinero que presta una entidad bancaria a cambio de garantizar su devolución y de pagar un precio por disfrutarlo.(Thefreedictionary, 2015)⁸
- 2. Capacidad de pago o crédito:** Nivel de endeudamiento que puede soportar una persona física o jurídica con arreglo a su solvencia, sus ingresos y el ratio entre recursos ajenos y propios (La Caixa, 2015).⁹
- 3. Garantía:** Bien que se deja en prenda para garantizar el cumplimiento de un pago. (La Caixa, 2015).¹⁰

⁸ <http://es.thefreedictionary.com/crédito>

⁹ https://portal.lacaixa.es/docs/diccionario/C_es.html#CAPACIDAD-DE-CREDITO

4. **Prenda:** Bien que el deudor entrega al acreedor como garantía del cumplimiento de una obligación de pago. (La Caixa, 2015).¹¹
5. **Factura:** Documento que expide el vendedor y entrega al comprador en el que consta fecha de la operación, nombre de comprador y vendedor, importe, cantidad y descripción del producto o servicio, etc. (La Caixa, 2015).¹²
6. **Letra:** Denominación abreviada de la letra de cambio, título de crédito en el que el librado se obliga a pagar al librador o al beneficiario un importe en una fecha de vencimiento especificada. (La Caixa, 2015).¹³
7. **Pagaré:** Documento en el que figura el compromiso del emisor de abonar una cantidad determinada en una fecha prefijada. Se trata de deuda a corto plazo. (La Caixa, 2015)¹⁴
8. **Tasa:** Indicador básico que consiste en una relación entre dos variables o magnitudes. Tributo impuesto al disfrute de ciertos servicios o al ejercicio de determinadas actividades. (La Caixa, 2015).¹⁵
9. **Interés:** Remuneración que se paga o se percibe por el uso temporal de una cantidad de dinero, que se calcula como porcentaje de la cantidad tomada en préstamo o prestada. (La Caixa, 2015).¹⁶

¹⁰ https://portal.lacaixa.es/docs/diccionario/G_es.html#GARANTIA

¹¹ https://portal.lacaixa.es/docs/diccionario/P_es.html#PRENDA

¹² https://portal.lacaixa.es/docs/diccionario/F_es.html#FACTURA

¹³ https://portal.lacaixa.es/docs/diccionario/L_es.html#LETRA

¹⁴ https://portal.lacaixa.es/docs/diccionario/P_es.html#PAGARE

¹⁵ https://portal.lacaixa.es/docs/diccionario/T_es.html#TASA

- 10. Riesgo Crediticio:** Riesgo de que el prestamista no recupere el dinero prestado junto con los intereses generados, o al menos de que no recupere una parte. (La Caixa, 2015).¹⁷
- 11. Empresa al minoreo:** Comercio al por menor o al detalle. Dicho así suena fácil, pero es un poco más completo y conciso que esa breve explicación. Engloba el sector de negocios que va desde supermercados, pasando por tiendas de marca, grandes superficies, centros comerciales, hasta sucursales bancarias y en algunos casos restaurantes (ej.comida rápida) (AwanzoRetail, 2015)¹⁸
- 12. Regresión:** Técnicas estadísticas que permiten determinar la ecuación matemática que relaciona un efecto (variable dependiente o respuesta) con una o varias causas (variables independientes o predictoras). (Divestadística, 2015).¹⁹
- 13. Credit Scoring:** El *Credit Scoring* es un sistema de evaluación automática de solicitudes de operaciones de crédito, tales como préstamos al consumo, hipotecas o concesiones de tarjetas de crédito. (Definición ABC, 2015).²⁰
- 14. Probabilidad:** Medida adimensional que cuantifica la ocurrencia de los fenómenos producidos por azar. Sus valores son números comprendidos entre 0 y 1. Cuando la probabilidad es cero, el suceso no ocurrirá nunca. Si la probabilidad

¹⁶ https://portal.lacaixa.es/docs/diccionario/I_es.html#INTERES

¹⁷ https://portal.lacaixa.es/docs/diccionario/R_es.html#RIESGO-CREDITICIO

¹⁸ <http://www.retail.awanzo.com/2009/04/17/%C2%BFque-es-el-retail-largo-y-tendido/>

¹⁹ http://www.divestadistica.es/es/diccionario_estadistico.html#R

²⁰ <http://www.definicionabc.com/economia/scoring.php>

es uno, ocurrirá con total certeza. Los valores intermedios miden el gradiente de certidumbre. (Divestadística, 2015).²¹

15. Variable: Característica de cada individuo en una base de datos, el cual varía de individuo a individuo. Y por ello cada individuo tiene un valor para cada variable. (Divestadística, 2015).²²

16. Análisis multivariante: Tienen por objeto el estudio estadístico de varias variables medidas en elementos de una población, el cual pretende resumir el conjunto de variables en unas nuevas pocas variables, construidas como transformaciones de las originales, con la mínima pérdida de la información. (Peña, 2002).

²¹ http://www.divestadistica.es/es/diccionario_estadistico.html#R

²² http://www.divestadistica.es/es/diccionario_estadistico.html#R

III. MATERIALES Y MÉTODOS

A continuación se describen brevemente los procedimientos o métodos que se utilizarán para resolver el problema planteado, explicando cómo se realizará la investigación, las técnicas a utilizar, las muestras, entre otros.

3.1. Tipo de investigación

La presente investigación es de tipo descriptiva, explicativa transversal porque está orientada a la realidad tal como se presenta, a describir algunas características fundamentales de un fenómeno, hecho o situación determinada, y para este caso se propone un modelo de regresión logística que será aplicado al otorgamiento de un crédito convencional durante un periodo de tiempo (Hernández et al., 2006).

3.2. Diseño de la investigación

Según Hernández et al. (2006), el diseño de la presente investigación corresponde al de tipo transversal descriptiva, pues recolecta datos de un solo momento en un tiempo único. Su propósito es identificar las variables críticas en el estudio, y analizar su incidencia e interrelación en un periodo de tiempo proponiendo un modelo teórico susceptible a ser aplicado a voluntad.

3.3. Formulación de hipótesis

El modelo de regresión logística binaria permite discriminar a un cliente bueno de uno malo, para disminuir el riesgo de crédito en la empresa Carsa S.A.C.

3.4. Identificación de variables

Variable Dependiente:

- CLIENTE MALO: Aquel que en un periodo de 12 meses cae en mora mayor a 90 días.
- CLIENTE BUENO: Aquel que en un periodo de 12 meses no cae en atrasos.

Variables Independientes:

En el Anexo 4 se presentan el total de variables trabajadas, y a continuación se presentan las variables seleccionadas cuya categoría se presenta en el siguiente cuadro:

Cuadro 9: Variables independientes Seleccionadas

Variable	Descripción	Categorías
Perfil	Tipo de actividad económica del cliente	1. Dependiente formal
		2. Dependiente informal
		3. Independiente formal
		4. Independiente informal
Línea_Producto	Tipo de Línea del producto comprado	1. Audio
		2. Blanca
		3. Celulares
		4. Cómputo
		5. Entretenimiento
		6. Hogar
		7. Peds
		8. Video
Tipo_Propiedad	Tipo de propiedad de domicilio registrado	1. Alquiler
		2. Espacio cedido
		3. Familiar
		4. Propia

...continuación

Estado_Civil	Variable del estado civil del cliente.	1. Casado
		2. Conviviente
		3. Divorciado
		4. Soltero
		5. Viudo
Nro_Créditos_Total	Número de créditos que tiene el cliente en la empresa.	1. Nro. créditos=1
		2. Nro. créditos entre 2 y 3
		3. Nro. créditos >= 4
Producto_Crediticio	Tipo de Producto Crediticio de acuerdo a la evaluación de cada cliente.	1. Antecedentes crediticios
		2. Complementario
		3. Emprendedor
		4. Empresarial
		5. Otros
		6. Pre-aprobado vip
		7. Recurrente electro
		8. Tabla de negocios y oficios
		9. Facilito
Plazo	Plazo de Crédito	1. Plazo <= 12
		2. Plazo entre 13 y 18
		3. Plazo >= 19
Rango_Mora_Max	Máximo atraso del cliente de los últimos 12 meses en el sistema financiero	1. Atraso = null
		2. Atraso = [sin mora]
		3. Atraso = [1-30]
		4. Atraso = [31-60]
		5. Atraso = [61-90]
		6. Atraso = [mayor a 90]

FUENTE: Elaboración Propia

3.5. Población y muestra

- **Población Objetivo:** 17,979 créditos otorgados entre enero y noviembre del 2013 que fueron seleccionados entre buenos y malos.

Se ha considerado que el 80% de la población es la muestra de entrenamiento utilizada para desarrollar el modelo de regresión logística, y la diferencia (20%), fue utilizada para la validación del modelo.

3.6. Procedimiento de análisis de datos

- **Limpieza de base de datos**

Es recomendable para el análisis se realice la limpieza de base de datos antes del muestreo, ya que en el proceso de llenado de solicitudes se genera una serie de errores que deben de ser eliminados o corregidos. Se eliminaron datos inconsistentes como clientes menores a 18 años, campos vacíos o nulos, datos falsos. Los registros que son totalmente nulos son quitados de la base.

- **Selección de variables**

Según Siddiqi (2006) las variables con el valor de información (IV) menor a 0.02 es impredictivo, y mayor a 0.5 se considera sobrepredictivo.

$$IV = \sum_{k=1}^n (Distr\ Buenos_i - Distr\ Malos_i) \times \ln\left(\frac{Distr\ Buenos_i}{Distr\ Malos_i}\right)$$

$$WOE_k = \ln\left(\frac{Distr\ Buenos}{Distr\ Malos}\right) \times 100$$

Las variables seleccionadas de acuerdo al valor de información (IV) y Weight of Evidence (WOE) son las siguientes:

Cuadro 10: Indicador WOE por categoría e IV global por variable

Variable	Descripción	Categorías	WOE	IV
Perfil	Tipo de actividad económica del cliente	1. Dependiente formal	0.37	0.03
		2. Dependiente informal	-0.08	
		3. Independiente formal	0.19	
		4. Independiente informal	-0.07	
Línea_Producto	Tipo de Línea del producto comprado	1. Audio	-0.12	0.03
		2. Blanca	0.26	
		3. Celulares	-0.51	
		4. Cómputo	0.01	
		5. Entretenimiento	-0.45	
		6. Hogar	0.16	
		7. Peds	-0.25	
		8. Video	-0.13	
Tipo_Propiedad	Tipo de propiedad de domicilio registrado	1. Alquiler	-0.08	0.08
		2. Espacio cedido	1.36	
		3. Familiar	-0.17	
		4. Propia	0.51	
Estado_Civil	Variable del estado civil del cliente.	1. Casado	0.6	0.03
		2. Conviviente	0.09	
		3. Divorciado	0.28	
		4. Soltero	-0.07	
		5. Viudo	0.18	
Nro_Créditos_Total	Número de créditos que tiene el cliente en la empresa.	1. Nro. créditos=1	-0.23	0.2
		2. Nro. créditos entre 2 y 3	0.78	
		3. Nro. Créditos >= 4	1.41	

...continuación

Producto_Crediticio	Tipo de Producto Crediticio de acuerdo a la evaluación de cada cliente.	1. Antecedentes crediticios	0.07	0.1
		2. Complementario	-0.07	
		3. Emprendedor	-0.11	
		4. Empresarial	0.16	
		5. Otros	0.18	
		6. Pre-aprobado vip	0.18	
		7. Recurrente electro	0.02	
		8. Tabla de negocios y oficios	0.86	
		9. Facilito	-0.3	
Plazo	Plazo de Crédito	1. Plazo <= 12	0.36	0.05
		2. Plazo entre 13 y 18	0.14	
		3. Plazo >= 19	-0.19	
Rango_Mora_Max	Máximo atraso del cliente de los últimos 12 meses en el sistema financiero	1. Atraso = null	-0.18	0.86
		2. Atraso = [sin mora]	0	
		3. Atraso = [1-30]	0	
		4. Atraso = [31-60]	2.03	
		5. Atraso = [61-90]	-0.68	
		6. Atraso = [mayor a 90]	-2.69	

FUENTE: Elaboración Propia

Para el modelo, todas las variables mencionadas en el cuadro anterior tienen un valor de información (IV) mayor que 0.02, por lo tanto son variables que aportan a la discriminación del modelo. Los WOE negativos nos indican que la categoría presenta un elevado riesgo (número de casos malos), mientras que los valores positivos indican un bajo riesgo (casos buenos). Por ejemplo, si se observa la variable Rango_Mora_Max, en la categoría de [61-90] presenta un WOE de [-0.68] el cual indica que todas las cuentas que presenten esta característica tienen mayor riesgo de ser un cliente malo, en cambio la categoría [31-60], tienen un WOE de [2.03], el cual indica que tienen bajo riesgo de ser cliente malo, es decir, es cliente bueno.

IV. RESULTADOS Y DISCUSIÓN

En este capítulo se presentan los principales resultados y los comportamientos encontrados bajo la aplicación de las técnicas administrativas y estadísticas en la empresa Carsa S.A.C. A continuación se procede a explicar cada uno de los resultados encontrados.

4.1. Análisis situacional comercial

4.1.1. Análisis de Procesos

La empresa Carsa S.A.C. está a cargo de las operaciones comerciales de electrodomésticos (Figura 5), siendo la principal fortaleza: fuerza de ventas con experiencia en el mercado; y su principal defecto es la alta dependencia a los consumidores informales. En la industria de electrodomésticos sus principales competidores directos son: Hiraoka, Curacao y Efe, maneja proveedores con alto poder de negociación como Samsung, LG, Lenovo, Mabe, entre otros. En cuanto la cadena de valor de la empresa, su punto estratégico se encuentra en las actividades de operaciones (proceso de comercialización) y actividades de logística (cartera de proveedores y gestión despacho).

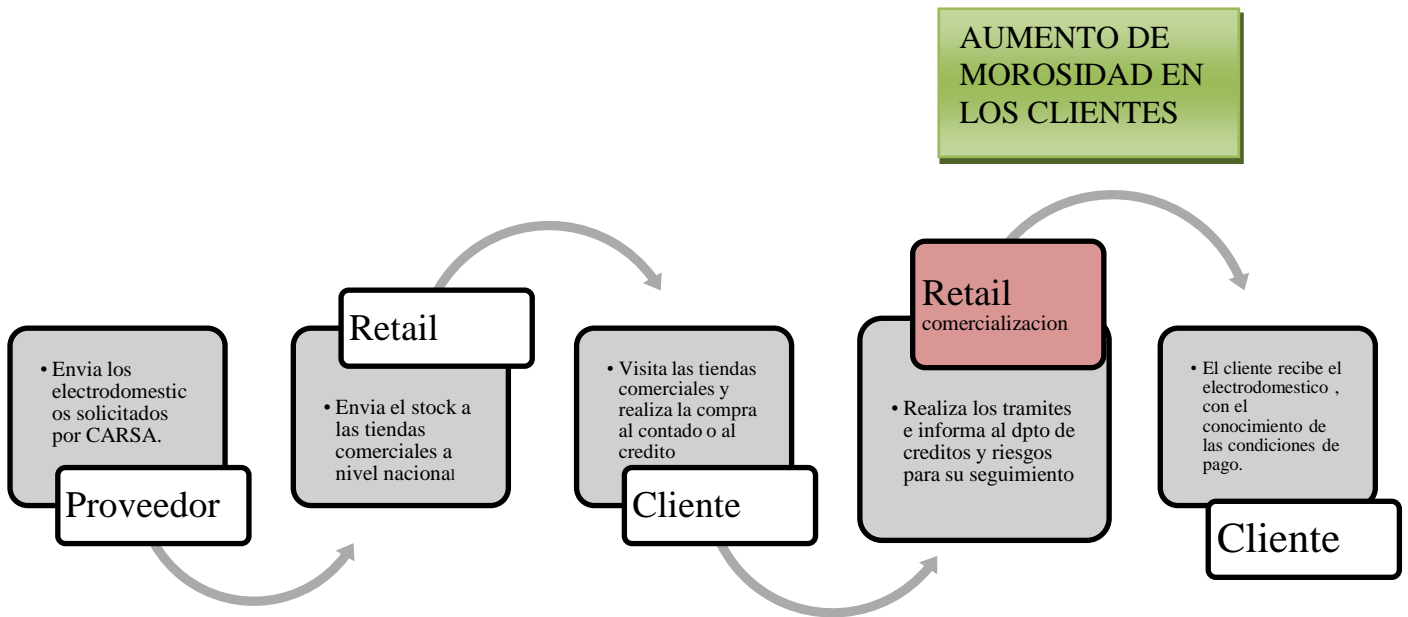


Figura 5: Flujo simplificado del proceso de la venta al Crédito de Carsa S.A.C.

FUENTE: Elaboración Propia

4.1.2. Cadena de valor

La cadena de valor de la empresa Carsa S.A.C. está conformada principalmente, por las actividades comerciales (actividad estratégica), a continuación en la figura 6 se detalla las actividades primarias y de apoyo.

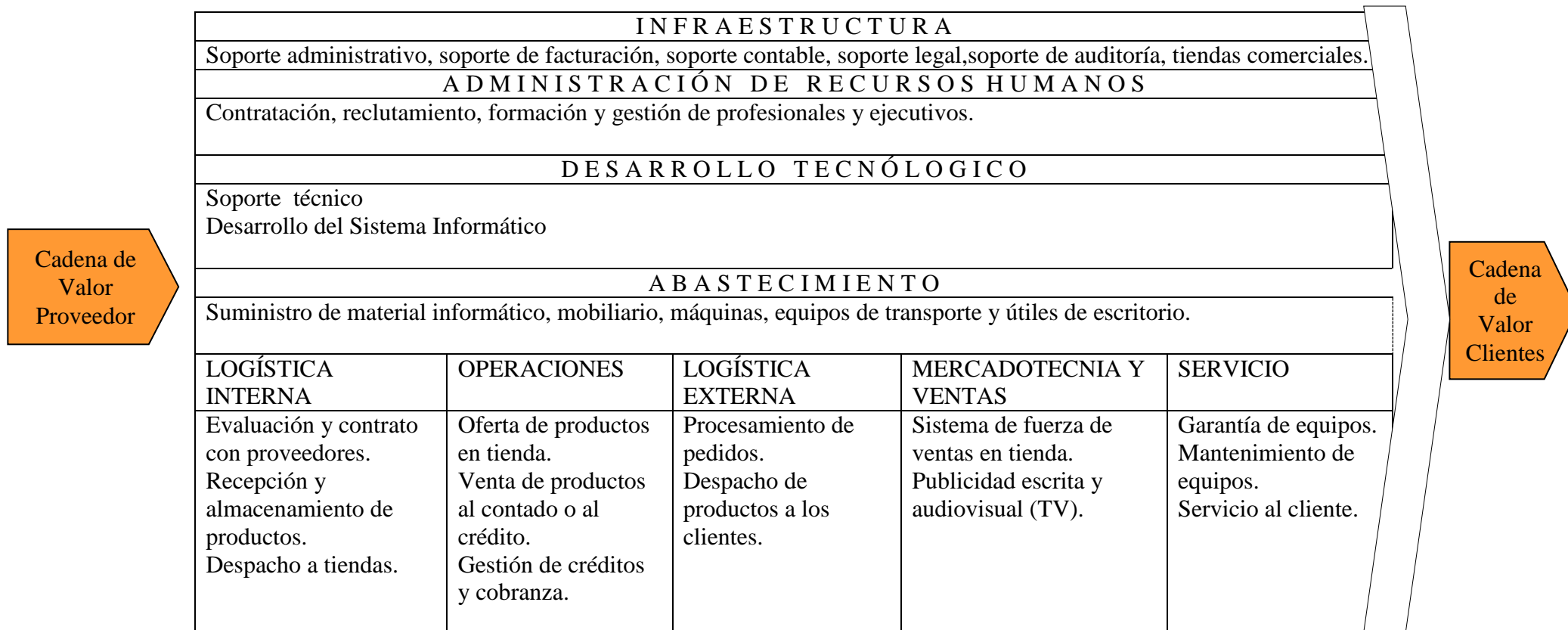


Figura 6: Cadena de valor simplificado de Carsa S.A.C.

FUENTE: Elaboración Propia

4.1.3. Condiciones iniciales del proceso del otorgamiento de un crédito convencional

La dinámica del proceso de otorgamiento de un crédito convencional de la empresa Carsa S.A.C. está representado principalmente por el vendedor (representante de la empresa ofertante), el analista de créditos, el verificador, el cajero y el usuario final, según se puede identificar el macro proceso en la Figura 7 y el detalle de cada proceso en anexos del 6 al 13.

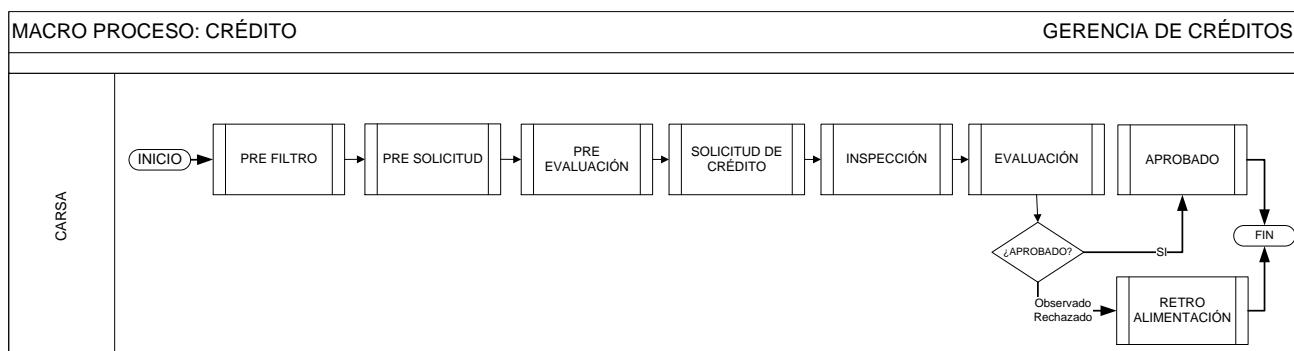


Figura 7: Proceso simplificado de otorgamiento de un crédito convencional

FUENTE: Elaboración Propia

Considerando que el segmento de mercado es un grupo de individuos con similares necesidades, Carsa S.A.C. atiende a diferentes segmentos concentrados del mercado, los cuales son los siguientes:

- Segmento 1: Todos las personas a nivel nacional del sector socioeconómico C.
- Segmento 2: Todos las personas a nivel nacional del sector socioeconómico D.
- Segmento 3: Todos las personas a nivel nacional del sector socioeconómico E.

Cada uno de los participantes realiza una función específica en el proceso de otorgamiento de crédito, a continuación se explica con más detalle:

a. Vendedor:

- Ofrece venta al crédito en Campo o Tienda
- Llena hoja de perfil de cliente
- Solicita filtro en las Centrales de Riesgo y Bases Externas
- Verifica en bases Internas
- Ofrece venta al contado

b. Analista de Créditos:

- Evaluar las solicitudes de crédito de clientes.
- Administrar los archivos de créditos generados en Tienda.
- Realizar el proceso de filtro para evaluar a los clientes.
- Realizar el proceso de retroalimentación.
- Aprobar solicitudes de crédito de acuerdo a su nivel de autonomía firmando y sellando las Solicitudes de Crédito aprobadas, de acuerdo a la normativa vigente.
- Dirigir e implementar la operatividad necesaria que garantice la ejecución de los procesos vigentes para la aprobación de créditos.
- Supervisar y controlar las operaciones de la Tienda a su cargo, relacionadas con la evaluación y aprobación de créditos generados en su plaza.

c. Verificador:

- Ejecutar las verificaciones e inspecciones domiciliarias y laborales en su respectiva plaza.

d. Cajero:

- Imprime y Emite la Boleta y/o Factura

Las condiciones iniciales en el otorgamiento de Carsa S.A.C. están determinadas por las siguientes políticas:

- En la evaluación de clientes, para el otorgamiento de crédito, se tendrá en cuenta la capacidad de pago del cliente en función a sus ingresos, el importe de sus obligaciones y las calificaciones asignadas por otras empresas del sistema financiero, un buen historial crediticio en Carsa S.A.C., en instituciones similares y en aquellas comprendidas dentro del sistema financiero.
- Para clientes que no cuentan con historial crediticio, la evaluación se fundamentará en sus ingresos y en la inspección domiciliaria y laboral del cliente, reflejadas en los respectivos formatos de verificación, los mismos que deberán estar debidamente sustentados.
- El Analista de Créditos, así como el Verificador, según corresponda, asumirán Responsabilidad Solidaria por cualquier beneficio económico y crediticio que reciba un cliente, basado en una indebida calificación y/o evaluación.
- No serán sujetos de crédito, aquellas personas naturales que:
 - Tengan créditos directos o indirectos vencidos, créditos calificados como pérdida, créditos castigados o en cobranza judicial con personas naturales, en el Sistema Financiero, ONGs, Cooperativas u otras instituciones, en los últimos 18 meses, a la fecha de solicitar el crédito
 - Registren protestos de letras o pagarés que, a la fecha de solicitar el crédito, no hayan sido regularizados y/o aclarados.

- No tengan domicilio fijo y verificable con recibos de agua, luz, contrato de alquiler, licencia municipal u otro documento probatorio. Excepcionalmente, el Informe de Visita del Inspector de Créditos podrá servir de sustento, en el cual se indique el arraigo del cliente en el domicilio y sustentado con referencias de, por lo menos, dos (02) vecinos identificados plenamente.
- Que domiciliados en zonas calificadas peligrosas o inaccesibles, salvo excepciones establecidas puntualmente en la Política de Créditos.
- Incumplan con otras restricciones establecidas en los “Requisitos mínimos Crediticios”, para el otorgamiento de crédito, de la Política de Créditos.

4.1.4. Análisis de las cinco fuerzas de Porter

El Análisis de las cinco fuerzas de Porter de la Empresa Carsa S.A.C. se representa por:

- El alto poder de negociación de los proveedores por ser marcas reconocidas por el mercado, mayormente son extranjeras y tienen más de 20 años de experiencia en el mercado. La empresa cuenta con 15 proveedores prestigiosos, los principales son:
 - Proveedor de Video: Samsung, Sony, Panasonic, LG, Miray, Philips, otras marcas.
 - Proveedor de Línea Blanca: Samsung, Sony, Mabe, Daewoo, Indurama, Electrolux, Coldex, Oster, Inresa, otras marcas.
 - Proveedor de Audio: Samsung, Sony, Panasonic, LG, Miray, Philips, otras marcas.

- Proveedor de PEDS: Samsung, Sony, Panasonic, LG, Miray, Philips, otras marcas.
- El poder de negociación de los compradores es alto por la representación de 37,265 compradores entre el sector C, D y E, incluyendo que dichos compradores cuentan con grandes ofertas de las diferentes empresas *retail*.
- La amenaza de ingreso de competidores potenciales es bajo por las altas barreras de ingreso, principalmente por las altas inversiones, los escasos canales de distribución y numerosos requisitos para ser distribuidores de los proveedores mencionados.
- El bajo ingreso de productos o servicios sustitutos, representado por las tiendas directas online (LINO, OLX, etc), pues aún el mercado peruano está en proceso de aprendizaje de este tipo de compra.
- La alta rivalidad de competidores existentes, ya que está representada por 17 empresas aproximadamente siendo sus principales competidores la Curacao (12.9% del mercado), Hiraoka (10.8% del mercado) y Efe (10.5% del mercado), por ser del sector especializado, mientras que Carsa tiene un 5.5% de participación. Los datos fueron recuperados de la encuesta “Tamaño de la Demanda 2013” realizada por Tendencias & Actitudes S.A.C.

La figura 8 muestra los principales factores de cada una de las fuerzas de Porter que presenta la empresa Carsa S.A.C.

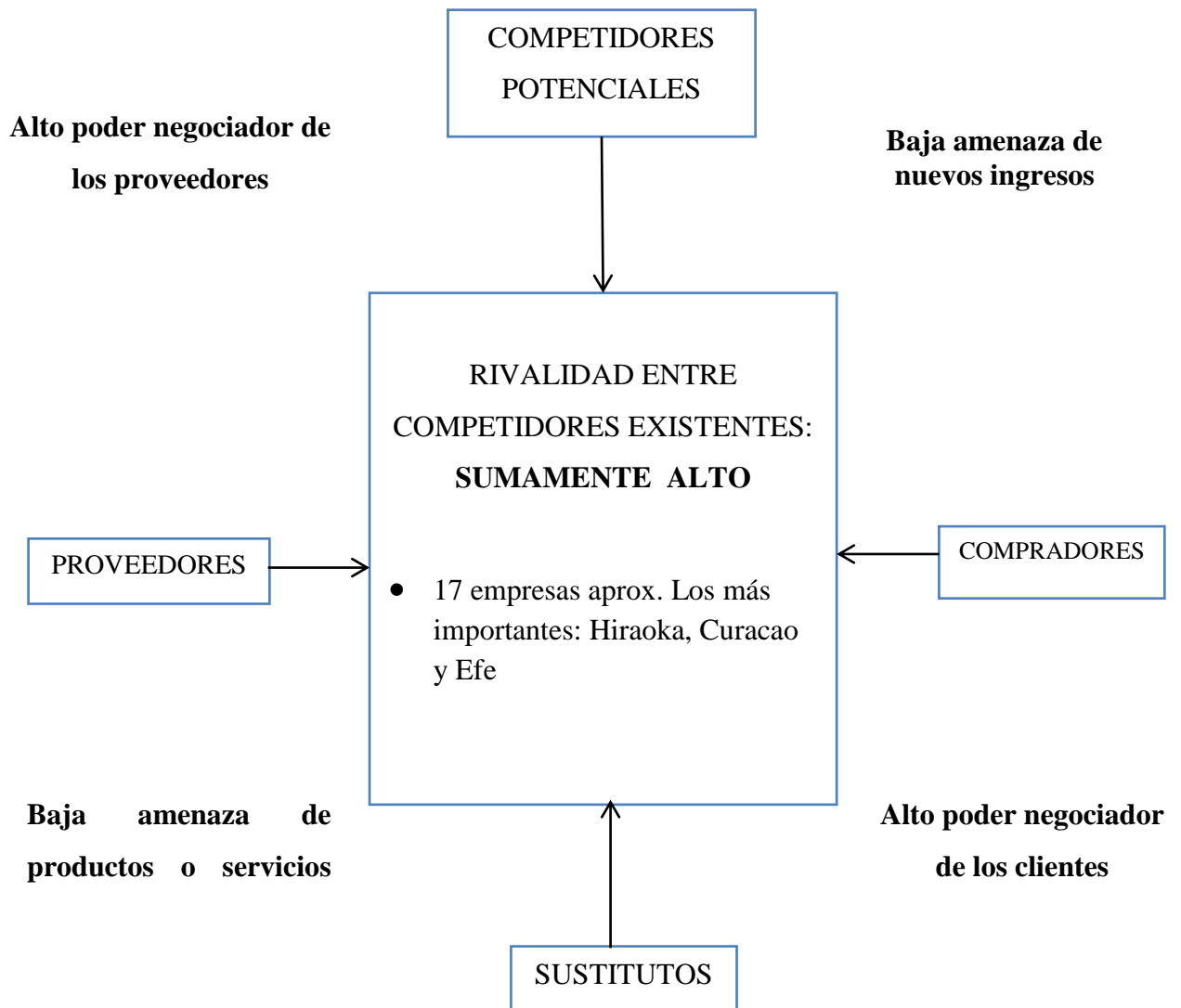


Figura 8: Las cinco fuerzas competitivas de Porter de Carsa S.A.C.

FUENTE: Elaboración propia

4.1.5. Análisis de la matriz Boston consulting group

La empresa Carsa S.A.C. comercializa electrodomésticos, sus productos se dividen en: Línea Blanca, Cómputo, Video, Audio, PEDS, Celulares, Entretenimiento y Hogar, según las ventas del 2013 por tipo de producto asciende a:

Cuadro 11: Colocación Acumulada por Tipo de Línea

Línea	Monto Facturado (Soles)	% Participación	Clasificación Matriz BCG
AUDIO	8,763,616	10.70	VACA
BLANCA	23,013,548	28.10	VACA
CELULARES	760,955	0.90	PERRO
COMPUTO	20,688,703	25.20	VACA
ENTRETENIMIENTO	373,226	0.50	INTERROGANTE
HOGAR	313,778	0.40	PERRO
PEDS	208,504	0.30	PERRO
VIDEO	27,915,260	34.00	VACA
Total	82,037,591	100.00	

FUENTE: Elaboración Propia



Figura 9: Matriz Boston Consulting Group de Carsa S.A.C.

FUENTE: Elaboración Propia.

4.1.6. Análisis FODA

En el ambiente interno de Carsa S.A.C., la principal fortaleza que presenta son los profesionales con experiencia en la cadena *retail* y su principal debilidad es su alta dependencia a los consumidores informales.

En el ambiente externo, la principal amenaza son las altas variaciones del tipo de cambio (importación de productos) y su principal oportunidad es el aumento del uso de tiendas *online*, lo cual disminuye la dependencia al alquiler de tiendas. Se presenta los principales factores del análisis FODA de la empresa en la Figura 10.

AMBIENTE INTERNO	
FORTALEZAS	DEBILIDADES
<p>F.1 Profesionales con experiencia en la cadena <i>retail</i>.</p> <p>F.2 Altos canales de distribución a nivel nacional (tiendas comerciales).</p> <p>F.3 Precios competitivos en el mercado</p> <p>F.4 Alta fidelidad de los clientes a nivel de provincia (20% aprox. del mercado).</p> <p>F.5 Alta cartera de productos de vanguardia.</p> <p>F.6 Gran número de proveedores.</p> <p>F.7 Alta cantidad de cobradores motorizados</p> <p>F.8 Fuerza de ventas eficiente.</p> <p>F.9 Toma de decisiones por directorio.</p>	<p>D.1 Alta dependencia a los consumidores informales</p> <p>D.2 Alta dependencia a los proveedores de marcas prestigiosas</p> <p>D.3 Personal desmotivado.</p> <p>D.4 Bajo control de las cobranzas en efectivo. (recibos a mano no cuadra con el dinero ingresado).</p> <p>D.5 Alta dependencia de la inversión por los accionistas del Banco Financiero</p> <p>D.6 Falta de mayor inversión en publicidad televisiva.</p>
AMBIENTE EXTERNO	
OPORTUNIDADES	AMENAZAS
<p>O.1 Aumento de uso de tiendas <i>online</i></p> <p>O.2 Disminuye el PBI en 4.8% (2014). Sin embargo se estima que para los años 2015 y 2016 se estima que el PBI crecerá en 6.0% y 6.3% respectivamente.</p> <p>O.3 Disminuye la tasa de desempleo del 2013 en comparación con el 2014 , actualmente se mantiene en 5.9%</p> <p>O.4 Aumento de la población para el 2017 en 31826 a nivel nacional</p> <p>O.5 El índice de precios del consumidor aumento a 3.08 para el año 2014 ,siendo evidencia de mayor disposición por la compra de bienes y servicios para el hogar</p>	<p>A.1 En cuanto al tipo de cambio, el documento elaborado por el Ministerio de Economía proyectó que el dólar cerrará en S/. 2,81 en 2014. La proyección del tipo de cambio del dólar para el 2015, 2016 es de S/. 3.20 y S/. 3.35 respectivamente.</p> <p>A.2 El Perú registrará una inflación de 3% este año, mayor a la estimación previa de 2,8%, en el límite del rango meta del Banco Central de Reserva (BCR) de entre 1% y 3%.</p> <p>A.3 Crecimiento lento de la demanda</p> <p>A.4 Alta rivalidad de la competencia directa.</p> <p>A.5 Variedad de productos de la competencia.</p>

Figura 10: Análisis FODA de Carsa S.A.C.

4.2. Análisis Estadístico

4.2.1. Identificación de Bueno y Malo

La matriz de transición en el contexto de *credit scoring* se le conoce como *Roll Rate*. De la población total de cuentas (37,265), se procede a generar la matriz *Roll Rate* como se muestra en el Cuadro 12 el cual identifica a aquellos clientes buenos, malos e indeterminados.

Para nuestro caso, las probabilidades para que se considere a un cliente bueno, malo o indeterminado es de acuerdo de la siguiente manera:

$\hat{P}(X_2 = PV44 | X_1 = i) < 0.2$, las cuentas que están en el estado i se consideran buenas.

$\hat{P}(X_2 = PV44 | X_1 = i) > 0.4$, las cuentas que están en el estado i se consideran malas.

Si $0.2 \leq \hat{P}(X_2 = PV44 | X_1 = i) \leq 0.4$, las cuentas que están en el estado i se consideran indeterminadas.

Cuadro 12: Roll Rate del total de créditos

En porcentaje (%)

PER12 PER6	PV00	PV01	PV02	PV03	PV04	PV11	PV12	PV13	PV14	PV22	PV23	PV24	PV33	PV34	PV44
PV00	50.6	35.4	2.6	0.2	0.1	3.6	2.2	0.3	0.0	1.5	0.5	0.1	1.4	0.2	1.4
PV01	0.0	64.9	6.9	0.9	0.2	6.2	6.6	0.8	0.1	3.8	1.6	0.2	3.3	0.9	3.6
PV02	0.0	0.0	45.6	2.3	0.9	0.0	16.2	2.4	0.5	6.3	5.4	0.9	7.1	2.9	9.5
PV03	0.0	0.0	0.0	37.4	1.1	0.0	0.0	15.4	0.5	0.0	9.9	0.5	9.9	3.8	21.4
PV04	0.0	0.0	0.0	0.0	20.0	0.0	0.0	0.0	5.0	0.0	0.0	15.0	0.0	15.0	45.0
PV11	0.0	8.8	14.5	2.7	1.0	3.3	14.3	2.8	0.8	5.9	6.8	1.9	8.8	6.0	22.4
PV12	0.0	0.0	12.3	2.7	0.9	0.0	15.7	4.1	0.5	7.2	10.3	1.6	11.1	9.6	24.0
PV13	0.0	0.0	0.0	13.2	2.7	0.0	0.0	11.4	3.6	0.0	10.9	3.6	9.1	10.9	34.5
PV14	0.0	0.0	0.0	0.0	14.3	0.0	0.0	0.0	9.5	0.0	0.0	0.0	0.0	23.8	52.4
PV22	0.0	0.0	4.2	4.9	2.0	0.0	2.4	4.2	1.7	0.8	7.4	2.1	6.7	10.4	53.2
PV23	0.0	0.0	0.0	5.2	1.8	0.0	0.0	3.8	1.6	0.0	12.6	2.9	5.9	14.3	51.8
PV24	0.0	0.0	0.0	0.0	15.9	0.0	0.0	0.0	4.5	0.0	0.0	11.4	0.0	22.7	45.5
PV33	0.0	0.0	0.0	1.8	2.1	0.0	0.0	1.5	1.6	0.0	1.8	1.7	2.3	8.0	79.3
PV34	0.0	0.0	0.0	0.0	1.9	0.0	0.0	0.0	1.7	0.0	0.0	4.2	0.0	16.1	76.2
PV44	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	99.4

FUENTE: Elaboración Propia

La clasificación de cada estado se encuentra en el siguiente cuadro:

Cuadro 13: Posibles estados de un cliente y su marca de clase asociado

Estado	Descripción al final del periodo	Marca
PV00	al día y tiene máximo 0 días de atraso durante el periodo	Bueno
PV01	al día y tiene máximo entre 1-30 días de atraso durante el periodo	Bueno
PV02	al día y tiene máximo entre 31-60 días de atraso durante el periodo	Bueno
PV03	al día y tiene máximo entre 61-90 días de atraso durante el periodo	Indeterminado
PV04	al día y tiene máximo 90 a más días de atraso durante el periodo	Malo
PV11	1-30 días de atraso y tiene máximo entre 1-30 días de atraso durante el periodo	Indeterminado
PV12	1-30 días de atraso y tiene máximo entre 31-60 días de atraso durante el periodo	Indeterminado
PV13	1-30 días de atraso y tiene máximo entre 61-90 días de atraso durante el periodo	Indeterminado
PV14	1-30 días de atraso y tiene máximo 90 a más días de atraso durante el periodo	Malo
PV22	31-60 días de atraso y tiene máximo entre 31-60 días de atraso durante el periodo	Malo
PV23	31-60 días de atraso y tiene máximo entre 61-90 días de atraso durante el periodo	Malo
PV24	31-60 días de atraso y tiene máximo 90 a más días de atraso durante el periodo	Malo
PV33	61-90 días de atraso y tiene máximo entre 61-90 días de atraso durante el periodo	Malo
PV34	61-90 días de atraso y tiene máximo 90 a más días de atraso durante el periodo	Malo
PV44	90 a más días de atraso y tiene máximo 90 a más días de atraso durante el periodo	Malo

FUENTE: Elaboración Propia

De acuerdo a los parámetros establecidos, en el cuadro 13 se identifica el tipo de cliente (bueno, malo e indeterminado), por ejemplo, si un cliente al final del periodo de análisis está al día, y tiene como máximo días de atraso entre [31-60] días, es un cliente bueno; si un cliente al final del periodo de análisis está entre [31-60] días de atraso y además tiene como máximo días de atraso entre [61-90] días, es considerando un mal cliente; si un cliente al final del periodo de análisis está entre [1-30] día de atraso y además tiene como máximo días de atraso entre [1-30] días, es considerado un cliente indeterminado, el cual aún no se puede identificar si es cliente bueno o malo. Es por ello que realizar el análisis del modelo de regresión logística se considera únicamente a aquellos clientes con estado Bueno y Malo, es decir, se retira a los clientes indeterminados, el cual nos queda una base de datos de 17,979 clientes.

4.2.2. Análisis del modelo de regresión logística binaria

El método de selección de variables fue por pasos hacia adelante, el cual comienza con el modelo que incluye solo el término constante y se van añadiendo al mismo modelo las variables independientes según su grado de relación con la variable dependiente y su significación estadística. La que a continuación se presenta:

Cuadro 14: Variables del Modelo

Variables en la ecuación							
		B	Error estándar	Wald	Gl	Sig.	Exp(B)
Paso 8 ^h	v_producto_creditos			51.224	8	0.000	
	v_producto_creditos(1)	1.407	0.791	3.165	1	0.075	4.084
	v_producto_creditos(2)	0.804	0.878	0.838	1	0.360	2.234
	v_producto_creditos(3)	0.154	0.057	7.286	1	0.007	1.166
	v_producto_creditos(4)	0.223	0.119	3.509	1	0.061	1.25
	v_producto_creditos(5)	1.768	0.798	4.906	1	0.027	5.858
	v_producto_creditos(6)	0.663	0.795	0.695	1	0.404	1.941
	v_producto_creditos(7)	0.245	1.191	0.042	1	0.837	1.278
	v_producto_creditos(8)	0.648	0.792	0.67	1	0.413	1.912

...continuación

V_LINEA			69.632	7	0.000	
V_LINEA(1)	-0.014	0.08	0.028	1	0.867	0.987
V_LINEA(2)	0.37	0.059	39.658	1	0.000	1.448
V_LINEA(3)	-0.558	0.185	9.138	1	0.003	0.572
V_LINEA(4)	0.155	0.069	5.125	1	0.024	1.168
V_LINEA(5)	-0.437	0.262	2.782	1	0.095	0.646
V_LINEA(6)	-0.161	0.274	0.345	1	0.557	0.851
V_LINEA(7)	-0.57	0.289	3.904	1	0.048	0.565
v_rango_maxima_mora			626.979	5	0.000	
v_rango_maxima_mora(1)	18.704	1,292.22	0	1	0.988	132,720,039
v_rango_maxima_mora(2)	18.748	755.648	0.001	1	0.980	138,734,854
v_rango_maxima_mora(3)	1.025	0.794	1.666	1	0.197	2.786
v_rango_maxima_mora(4)	-1.783	0.792	5.071	1	0.024	0.168
v_rango_maxima_mora(5)	-3.924	0.787	24.89	1	0.000	0.02
V_ESTADO_CIVIL			28.974	4	0.000	
V_ESTADO_CIVIL(1)	0.5	0.297	2.822	1	0.093	1.648
V_ESTADO_CIVIL(2)	0.174	0.291	0.36	1	0.549	1.19
V_ESTADO_CIVIL(3)	0.05	0.561	0.008	1	0.928	1.052
V_ESTADO_CIVIL(4)	0.008	0.285	0.001	1	0.978	1.008
V_TIPO_PROPIEDAD			90.392	3	0.000	
V_TIPO_PROPIEDAD(1)	-0.513	0.085	36.44	1	0.000	0.599
V_TIPO_PROPIEDAD(2)	1,370	1.153	1.411	1	0.235	3.937
V_TIPO_PROPIEDAD(3)	-0.558	0.06	86.548	1	0.000	0.572
V_TIPO_ACTIVIDAD			11.895	3	0.008	
V_TIPO_ACTIVIDAD(1)	0.302	0.111	7.342	1	0.007	1.352
V_TIPO_ACTIVIDAD(2)	0.167	0.067	6.176	1	0.013	1.182
V_TIPO_ACTIVIDAD(3)	0.112	0.165	0.465	1	0.495	1.119
v_rango_plazo			35.986	2	0.000	
v_rango_plazo(1)	0.423	0.072	34.809	1	0.000	1.527
v_rango_plazo(2)	0.148	0.054	7.51	1	0.006	1.16
v_rango_total_creditos			126.949	2	0.000	
v_rango_total_creditos(1)	-0.978	0.294	11.046	1	0.001	0.376
v_rango_total_creditos(2)	0.019	0.292	0.004	1	0.947	1.02
Constante	1.9	0.405	21.998	1	0.000	6.687

FUENTE: Elaboración propia

De acuerdo al cuadro detallado anteriormente, se puede observar que las variables en mención resultaron significativas. Según Aldana y Domínguez (2001) si alguna categoría resulta significativa la variable queda incluida en el modelo final, aun cuando el resto no lo sea.

Interpretaciones

- Dep Formal [$\exp(b)=1.352$], cuando el cliente es dependiente formal en lugar de otras actividades económicas, la ventaja de que el cliente sea bueno aumenta en 1.352 veces respecto a que sea malo, manteniendo las demás variables constantes.
- Linea Blanca [$\exp(b)=1.448$], cuando el cliente compra un producto de línea blanca en lugar de otros productos, la ventaja de que el cliente sea bueno aumenta en 1,448 veces respecto a que sea malo, manteniendo las demás variables constantes.
- Celular [$\exp(b)=0.572$], cuando el cliente compra un celular en lugar de otros productos, la ventaja de que el cliente sea bueno disminuye en 0.572 veces respecto a que sea malo, manteniendo las demás variables constantes.
- Casado [$\exp(b)=1.648$], cuando el cliente es una persona Casada, en lugar de otro estado, la ventaja de que el cliente sea bueno aumenta en 1,648 veces respecto a que sea malo, manteniendo las demás variables constantes.
- Antecedentes crediticios [$\exp(b)=4.084$], cuando el cliente se le evalúa como tipo de producto de Antecedentes Crediticios, en lugar de otro tipo de producto, la ventaja de que el cliente sea bueno aumenta en 4,084 veces respecto a que sea malo, manteniendo las demás variables constantes.

- Tipo de Vivienda Alquilada [$\exp(b)=0.599$], cuando el cliente vive en una casa alquilada, en lugar de otro tipo de vivienda, la ventaja de que el cliente sea bueno disminuye en 0.599 veces respecto a que sea malo, manteniendo las demás variables constantes.
- Máxima mora [$\exp(b)=0.168$], cuando el cliente tiene una mora máxima entre 60-90 días en los últimos 12 meses en otras entidades, en lugar de otros días de atraso, la ventaja de que el cliente sea bueno disminuye en 0.168 veces respecto a que sea malo, manteniendo las demás variables constantes.

4.2.3. Pruebas Estadísticas del modelo Logístico

a. Significatividad global del modelo de *credit scoring*.

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0, \quad i = 1,2,3 \dots 34$$

Cuadro 15: Pruebas ómnibus de coeficientes de modelo

		Chi-cuadrado	Gl	Sig.
Paso 8	Escalón	11,944	3	,008
	Bloque	4430,123	34	,000
	Modelo	4430,123	34	,000

FUENTE: Elaboración propia

Esta prueba se aplica para valorar la significación global del modelo, en el cual indica que todas las variables explicativas (Perfil, Línea_Producto, Tipo_Propiedad, Estado_Civil, Nro_Credito_Total, Producto_Crediticio, Plazo, Rango_Mora_Max) son significativas para la discriminación de que un cliente sea bueno.

b. Bondad de Ajuste

Según el cuadro 16, el estadístico de Hosmer – Lemeshow indica un mal ajuste si el valor de su significación es menor a 0.05, en este caso el modelo se adecua correctamente a los datos, porque valor de significación es de 0.385.

H₀: El modelo se adecua correctamente a los datos

H₁: El modelo no se adecua correctamente a los datos

Cuadro 16: Prueba de Hosmer y Lemeshow

Escalón	Chi-cuadrado	G1	Sig.
8	8,512	8	,385

FUENTE: Elaboración propia

c. Pseudo Estadísticas R²

Según Bernal (2014), el **R cuadrado de Cox y Snell** es un coeficiente de determinación generalizado que se utiliza para estimar la proporción de la varianza de la variable dependiente explicada por las variables independientes. Se basa en la comparación del logaritmo de la verosimilitud para el modelo respecto al logaritmo de la verosimilitud para un modelo de línea base. Los valores oscilan entre 0 y 1.

Para nuestro caso, según el cuadro muestra un valor de 0.266 que indica que el 26% de la variación de Cliente Bueno/Cliente Malo (variable dependiente) es explicadas por las variables incluidas en el modelo (variables independientes) los cuales son: Perfil, Linea_Producto, Tipo_Propiedad, Estado_Civil, Nro_Credito_Total, Producto_Crediticio, Plazo, Rango_Mora_Max.

El R cuadrado de Nagelkerke es una versión corregida de la R cuadrado de Cox y Snell, la R cuadrado de Cox y Snell tiene un valor máximo inferior a 1, incluso para un modelo “perfecto”. La R cuadrado de Nagelkerke corrige la escala del estadístico para cubrir el rango completo de 0 a 1.

Según el cuadro muestra un valor de 0.405, el cual indica que el modelo explica el comportamiento (Cliente Bueno/Cliente Malo) al 40.5%.

Cuadro 17: Resumen del modelo

Escalón	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
8	10914,689 ^a	,266	,405

FUENTE: Elaboración propia

d. Tabla de Clasificación

La tabla de clasificación muestra el porcentaje de clientes que ha sido correctamente clasificado. La base para generar el modelo de regresión logística, se obtiene un 82.3% de clasificación correcta, mientras que en la base de validación al procesar el modelo se obtiene un 80.8% de clasificación correcta.

Cuadro 18: Tabla de clasificación

Observado			Pronosticado					
			Base de Desarrollo			Base de Validación		
			Predictiva		Corrección de porcentaje	Predictiva		Corrección de porcentaje
			Malo	Bueno		Malo	Bueno	
Paso 8	PREDICTIVA	Malo	960	2295	29,5	245	632	27,9
		Bueno	238	10815	97,8	71	2723	97,5
		Porcentaje global			82,3			80,8

FUENTE: Elaboración propia

Se observa en el cuadro 18, que el 2.2% (resultado del complemento del 97.8%) de los clientes buenos ha sido clasificado como malo (error tipo II). Así mismo que el 70% (resultado del complemento del 29.5%) de clientes malos han sido clasificados como buenos (error tipo I).

El error tipo I genera un costo de oportunidad al no recibir las utilidades generadas para la empresa por un crédito no cursado y, por su parte, el error de tipo II genera una pérdida para la empresa al tener que aumentar el riesgo total de la cartera aumentando la probabilidad de no recuperar el monto vendido.

e. Discriminación del Modelo

- **Estadístico Kolmogorov-Smirnov (K-S):** El cual permite observar la máxima discriminación entre buenos y malos.

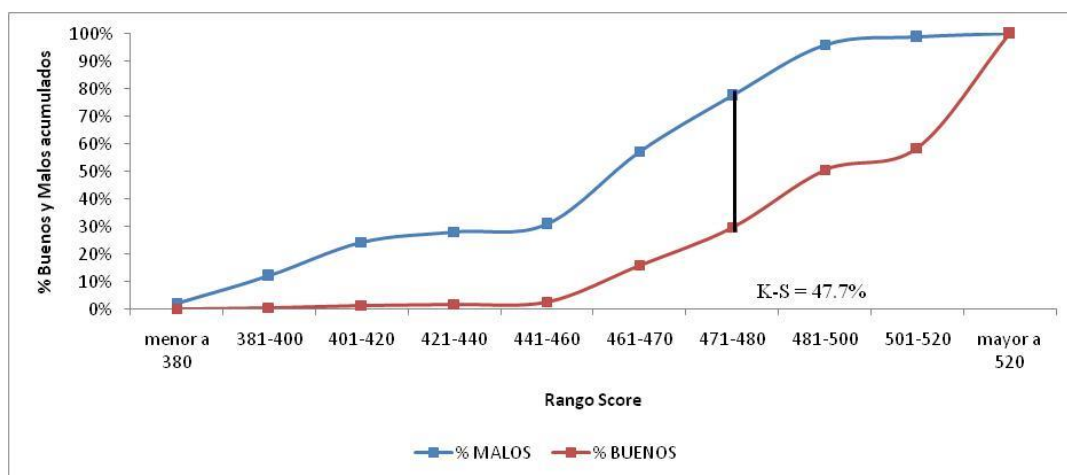


Figura 11: Curva de Kolmogorov - Smirnov (K-S)

FUENTE: Elaboración propia

El K-S es la máxima diferencia en valor absoluto entre la distribución acumulada de Buenos y Malos clientes obtenidos por los diferentes Score, el modelo presenta un K-S de 47.7%, el cual es considerado “Aceptable” (Anexo 2).

- **Coefficiente de Gini**

Permite observar la discriminación entre buenos y malos por rango de score. Mide la eficacia de la *scorecard* al comparar el porcentaje de buenos clientes frente al porcentaje de malos para los mismos puntajes. El modelo presenta un

Coefficiente de Gini de 65.3% (Anexo 2), el cual es considerado “aceptable” para un score de Originación (Simbaqueba, 2004)

- **Curva ROC**

La curva ROC permite cuantificar la precisión discriminatoria de un modelo, mediante el área bajo la curva. El área de bajo la curva para este modelo es de 0.834, el cual es la probabilidad de calcular un par de clientes como buenos y malos seleccionados al azar. (Ver figura 12).

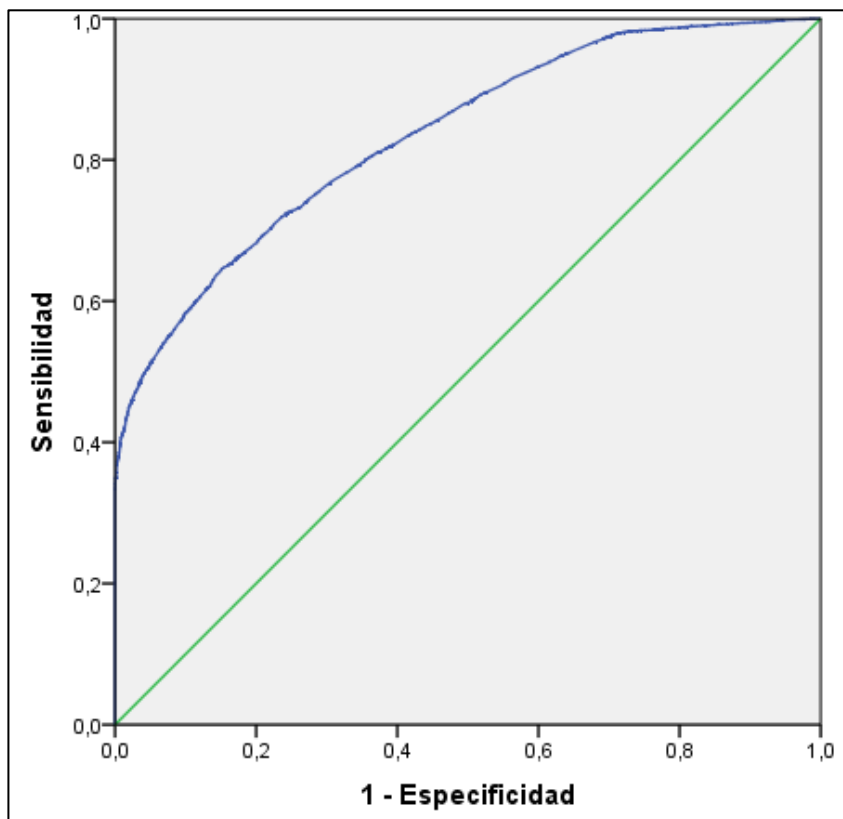


Figura 12: Curva de ROC

FUENTE: Elaboración propia

De acuerdo a todas las validaciones anteriormente mostradas, se concluye que el modelo se ajusta adecuadamente a los datos, y la precisión de discriminación del modelo es excelente.

f. Construcción del Scorecard

Según Siddiqi (2006), el *scorecard* se construye con los estimadores de los parámetros de la regresión logística. Para los puntajes del *score*, la relación entre las probabilidades y las puntuaciones se puede presentar como una transformación lineal, de acuerdo a la siguiente ecuación:

$$\text{Score} = \text{Offset} + \text{Factor} * \text{Ln} (\text{ODDS})$$

Donde *offset* es un término de traslación o compensación y **Factor** es un término de reescalamiento. *Offset* y **Factor** deben de satisfacer las condiciones impuestas por Carsa S.A.C. este procedimiento permite la estandarización del *score* para que diferentes *scorecards* puedan ser comparadas. Para obtener los valores de *Offset* y *Factor* se resuelve las ecuaciones siguientes:

$$\text{Score} = \text{Offset} + \text{Factor} * \ln (\text{odds})$$

$$\text{Score} + \text{pdo} = \text{Offset} + \text{Factor} * \ln (2 * \text{odds})$$

Resolviendo las ecuaciones anteriores para pdo, obtenemos:

$$\text{pdo} = \text{Factor} * \ln (2),$$

Por lo tanto

$$\text{Factor} = \text{pdo} / \ln (2);$$

$$\text{Offset} = \text{Score} - \{ \text{Factor} * \ln (\text{Odds}) \}$$

Por ejemplo, se construye un *scorecard* en el que el usuario quería probabilidades de 50:1 a 600 puntos y se quiere que las probabilidades se

dupliquen cada 20 puntos (es decir, pdo = 20), el factor y el desplazamiento son:

$$\text{Factor} = 20 / \ln(2) = 28.8539$$

$$\text{Offset} = 600 - \{28.8539 \ln(50)\} = 487.123$$

Y cada puntuación correspondiente a cada conjunto de probabilidades (o cada atributo) se puede calcular como:

$$\text{Score} = 487.123 + 28.8539 \ln(\text{odds})$$

Para la aplicación del modelo de *Credit Scoring* para la empresa Carsa S.A.C. los valores de *Offset* y *Factor* fueron calculados en función de los requerimientos de la empresa.

$$\text{Factor} = 20 / \ln(2) = 28.8539008$$

$$\text{Offset} = 50 - 28.8539008 * \ln(1) = 50$$

Con estos valores la función de *score* para el modelo de *credit scoring* es el siguiente:

$$\text{Score} = 50 + 28.8539008 \ln(\text{odds})$$

De acuerdo a la ecuación anterior el *score* para cada una de las variables que forman parte del modelo se encuentra en el Anexo 1.

El *score* que permite medir el riesgo de incumplimiento de pago para cada nuevo cliente que solicite un crédito sigue la siguiente ecuación:

$$\begin{aligned} \text{Score} = & 50*v_{\text{tipo_actividad}}+59*v_{\text{tipo_actividad}(1)}+55*v_{\text{tipo_actividad}(2)}+ \\ & 53*v_{\text{tipo_actividad}(3)}+50*v_{\text{linea}}+50*v_{\text{linea}(1)}+ 61*v_{\text{linea}(2)}+ 34*v_{\text{linea}(3)}+ \\ & 54*v_{\text{linea}(4)}+ 37*v_{\text{linea}(5)}+ 45*v_{\text{linea}(6)}+ 34*v_{\text{linea}(7)}+ 50*v_{\text{estado_civil}}+ \\ & 64*v_{\text{estado_civil}(1)}+ 55*v_{\text{estado_civil}(2)}+ 51*v_{\text{estado_civil}(3)}+ \end{aligned}$$

$50*v_estado_civil(4)+ 50*v_rango_total_creditos+ 22*v_rango_total_creditos(1)+$
 $51*v_rango_total_creditos(2)+50*v_productos_creditos+91*v_productos_creditos(1)+$
 $73*v_productos_creditos(2) + 54*v_productos_creditos(3)+$
 $56*v_productos_creditos(4)+ 101*v_productos_creditos(5)+$
 $69*v_productos_creditos(6)+ 57*v_productos_creditos(7)+$
 $69*v_productos_creditos(8)+50*v_rango_plazo+62*v_rango_plazo(1)+$
 $54*v_rango_plazo(2)+50*v_tipo_propiedad+35*v_tipo_propiedad(3)+$
 $90*v_tipo_propiedad(4)+$
 $34*v_tipo_propiedad(5)+50*v_rango_maxima_mora+590*v_rango_maxima_mora(1)+$
 $591*v_rango_maxima_mora(2)+ 80*v_rango_maxima_mora(3) -$
 $1*v_rango_maxima_mora(4) - 63*v_rango_maxima_mora(5)+105$

Tabla de niveles de Riesgo para la población objetivo

Se propone una tabla de asignación por niveles de riesgo, de acuerdo al puntaje obtenido, sin embargo el punto de corte para definir el nivel de riesgo que se está dispuesto a aceptar, corresponde al área de riesgos.

Cuadro 19: Tabla de Puntuaciones (Score)

SCORE	MALO	BUENO	Total Clientes	% Total Clientes	% MALOS	% BUENOS
menor a 380	95	13	108	1%	2%	0%
381-400	396	79	475	3%	10%	1%
401-420	503	132	635	4%	12%	1%
421-440	154	42	196	1%	4%	0%
441-460	125	129	254	1%	3%	1%
461-470	1107	1846	2953	16%	27%	13%
471-480	842	1876	2718	15%	20%	14%
481-500	742	2874	3616	20%	18%	21%
501-520	117	1099	1216	7%	3%	8%
mayor a 520	51	5757	5808	32%	1%	42%
Total	4,132	13,847	17,979	100%	100%	100%

Puntajes bajos, riesgo alto
 Puntajes altos, riesgo bajo

FUENTE: Elaboración propia

Analizando el cuadro anterior en la columna “% **Cientes**”, se puede observar que si se escoge el punto de corte en un *score* de 465, está en el rango de [461-470] el cual significaría un 26% de solicitudes rechazadas (suma acumulada desde el rango de score inicial [menor a 380] hasta el rango de score elegido [461-470]). Así mismo si se elige un punto de corte de 411 significaría un 8% de rechazos del total de solicitudes.

Como ya se había indicado anteriormente la elección del punto de corte depende del nivel de riesgo que asuma la empresa, el cual está dispuesto a asumir una pérdida por riesgo y rentabilidad deseada.

Por la cantidad de solicitudes que debieron ser rechazadas, Carsa S.A.C. obtuvo las siguientes pérdidas, según el cuadro 20:

Cuadro 20: Perdidas monetarias de solicitudes que debieron ser rechazadas

Detalle	Cantidad	Porcentaje de pérdida	Promedio de crédito unitario	Total (S/.)
120 días / analista de crédito	--			S/. 10,000.00
30% de créditos que se pierden al 100%	108	100%	2500	S/. 270,000.00
70% de créditos que se pierden el 60%	252	60%	2500	S/. 378,000.00
Total	360		--	S/. 658,000.00

Fuente: Elaboración propia.

V. CONCLUSIONES

Las conclusiones según el desarrollado en las secciones “Resultados” y “Discusión” son las siguientes:

1. Trabajando con la base de datos se obtuvo con el cual se puede identificar que el 8% de los solicitantes presentan una mayor probabilidad de incumplimiento de pago. Mientras que el 32% tiene una menor probabilidad de incumplimiento de pagos. Por lo tanto el modelo de regresión logística propuesto, disminuye el riesgo crediticio en las ventas al minoreo en la empresa Carsa S.A.C.
2. Las condiciones iniciales del otorgamiento de crédito convencional en la empresa Carsa S.A.C. presenta una deficiencia en la etapa de pre-evaluación a la etapa de solicitud de crédito, ya que no se encuentra con indicadores de discriminación para determinar si el cliente es bueno o malo, con el fin de iniciar el proceso de solicitud de crédito. Por lo tanto, el presente modelo de *credit scoring* es una herramienta adicional a la evaluación del analista de crédito, de acuerdo a las políticas de otorgamiento, el cual debe de incluirse en la etapa de pre-evaluación, que tiene como fin la mejora de la eficiencia del proceso de evaluación crediticia.
3. Para el modelo de regresión logística, se determinó que las variables o indicadores que discriminan de ser un cliente bueno y malo son: máximo atraso de los últimos 12 meses en el sistema financiero (sin información, sin mora, 1-30, etc.), tipo de línea del producto comprado (Audio, Blanca, Celular, etc.), tipo de propiedad del domicilio (Alquiler, Espacio cedido, Familiar, Propia), Tipo de actividad económica del cliente (Dependiente Formal, Dependiente Informal, etc.), Plazo de Crédito en meses (menor a 12, 13-18, mayor a 19), estado civil (Casado, conviviente, divorciado, soltero, viudo), número total de

créditos (1, entre 2-3 y mayor a 4) y producto de crédito (Antecedentes crediticios, complementario, facilito, entre otros).

4. El modelo de regresión logística propuesto, determinó un 8% de solicitudes que debió ser rechazada, el cual representa 360 créditos al mes aproximadamente. En aspectos de tiempo, los días/hombres perdidos son de 120 días que la compañía se ahorraría en dicha gestión y en aspectos monetarios la pérdida mensual asciende a S/. 658,000.

VI. RECOMENDACIONES

A partir del estudio realizado, se muestran las siguientes recomendaciones para completarlo o mejorarlo y estimular la ejecución de nuevos proyectos de aplicación de los métodos y/o resultados obtenidos.

1. Realizar una prueba *backtesting* para evaluar la calidad y estabilidad del modelo que se está empleando, en un periodo no menor a tres meses.
2. Para aquellos clientes que no tengan información en el sistema financiero, utilizar variables del perfil psicológico del cliente, con el fin de medir la honestidad de pago del cliente, el cual se denomina *scoring* psicométrico.
3. Se recomienda potenciar la comunicación entre el personal de Créditos y personal de cobranza con el fin de que exista una mayor sinergia, ya que existen grandes pérdidas mensuales. Así mismo se debería aplicar el modelo de *Scoring* de comportamiento para el área de cobranzas.
4. Se recomienda que la empresa Carsa S.A.C. tenga un comportamiento financiero conservador en las actividades de apoyo de la cadena de valor, para poder balancear las pérdidas generadas en el proceso crediticio.

VII. REFERENCIAS BIBLIOGRÁFICAS

1. AGRESTI, ALAN. 2002. *Categorical Data Analysis*. USA. Segunda edición. Editorial JhonWiley&Sons. 721 pp.
2. ALDANA, DEYSI Y DOMÍNGUEZ EMMA. 2001. *Regresión Logística. Un ejemplo de su uso en Endocrinología*. Revista Cubana Endocrinología 12(1): 58-64.
3. ARGUEDAS, RAQUEL; CASALS JOSÉ; PRA, INMACULADA Y RIOS, ANTONIO. 2010. *Gestión y Control del Riesgo de Crédito con Modelos Avanzados*. Madrid. Primera Edición. Editorial Ediciones Académicas. 475 p.p.
4. AMÉRICA ECONOMÍA. 2014. *Bancarización del Perú continuará creciendo y se situará en 32% al cierre de 2014* n°1. Consultado 25 julio 2014. Disponible en: <http://www.americaeconomia.com/economia-mercados/finanzas/consultora-bancarizacion-del-peru-continuara-creciendo-y-se-situara-en-32>.
5. HERNÁNDEZ, ROBERTO; FERNÁNDEZ, CARLOS Y BAPTISTA, PILAR. 2006. *Metodología de la Investigación*. Cuarta Edición. México. Editorial McGraw-Hill.848 pp.
6. BARBERENA, MANUEL Y BERBERANA, VITERBO. 2002. *La minería de datos en la industria financiera: Un nuevo enfoque de investigación de mercados*. AMAI. No. 31, Año 9
7. BERNAL, ENRIQUE. 2014. *Bioestadística básica para investigadores con SPSS*. Primera Edición. España. Editorial Bubok. 105 p.p.

8. BONAS, ARIANNA; LLANES, MARTA; USÓN, IDOYA Y VEIGA, NOA. 2007. *Riesgo de Crédito, Amenaza u Oportunidad*. Barcelona. Sexta Edición.
9. DURAN, GONZALO Y KREMERMAN, MARCO. 2008. *Fundación Sol: Caracterización del Sector Retail (una mirada general)* n° 7. Consultado 25 julio 2014. Disponible en <http://www.fundacionsol.cl/wp-content/uploads/2010/09/Cuaderno-7-Retail.pdf>
10. EQUILIBRIUM CLASIFICADORA DE RIESGO S.A. 2013. *Análisis del Sector Retail* n° 1. Consultado 23 mayo 2014. Disponible en <http://www.equilibrium.com.pe/sectorialsuperjun13.pdf>
11. GÓMEZ, PILAR Y PARTAL, ANTONIO. 2010. *Gestión y Control del Riesgo de Crédito en la Banca*. Madrid. Primera Edición. Editorial Delta. 284 p.p.
12. HOSMER, DAVID Y LEMESHOW, STANLEY. 2000. *Applied Logistic Regression*. USA. Segunda Edición. John Wiley&Sons, Inc. 375 p.p.
13. INEI. 2015. *Producción Nacional creció 2.68% en marzo del 2015*. Consultado el 12 junio 2015. Disponible en: www.inei.gob.pe/media/MenuRecursivo/noticias/nota-de-prensa-n064-2015-inei.pdf
14. JOHNSON, RICHARD Y WICHERN, DEAN. 2007. *Applied Multivariate Statistical Analysis*. USA. Sexta edición. Editorial Pearson Prentice Hall. 773 p.p.
15. LUQUE, TEODORO. 2000. *Técnicas de análisis de datos en investigación de mercado*. Madrid. Primera Edición. Ediciones Pirámide. 544 p.p.

16. MAYS, ELIZABETH. 2001. *Handbook of Credit Scoring*. USA. Primera Edición. The Glenlake Publishing Company, Ltd. 375 p.p.
17. NIETO, SORAIDA. 2010. *Crédito al Consumo: La estadística aplicada a un problema de riesgo crediticio*. Tesis Maestra. México. UAM. 96 pp.
18. PEÑA, DANIEL. 2002. *Análisis de datos multivariados*. Madrid. Primera edición. Editorial McGraw Hill. 544 pp.
19. PIÑEIRO SÁNCHEZ, CARLOS. 2010. *Dirección financiera: un enfoque centrado en valor y riesgo*. Madrid. Primera Edición. Editorial McGraw Hill. 375 p.p.
20. SIDDIQI, NAEM. 2006. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. New Jersey. Editorial John Wiley & Sons, Inc. 196 p.p.
21. SIMBAQUEBA, LILIAN. 2004. *¿Qué es el scoring? Una visión práctica de la gestión del riesgo de crédito*. Bogotá. Instituto del Riesgo Financiero
22. SUPERINTENDENCIA DE BANCA, SEGUROS Y AFP. 2013. *Indicadores alternativos de riesgo de crédito en el Perú: matrices de transición crediticia condicionadas en el ciclo económico*. N°1. Consultado 11 de agosto 2014. Disponible en: http://www.sbs.gob.pe/repositorioaps/0/0/jer/ddt_ano2013/SBS-DT-001-2013.pdf
23. SUPERINTENDENCIA DE BANCA, SEGUROS Y AFP. 2015. *Información del Sistema Financiero*. N°1. Consultado 11 de junio 2015. Disponible en: <http://www.sbs.gob.pe/app/stats/EstadisticaBoletinEstadistico.asp?p=14#>

24. TENDENCIA & ACTITUDES. 2014. *Tamaño de la demanda 2013: Peruana de Artefactos para el Hogar S.A.C.* N°1. Perú.

VIII. ANEXOS

Anexo 1: Modelo Final y Score

Variables – SPSS	Variables	Betas	Odds	Score
V_TIPO_ACTIVIDAD	Independiente Informal	0.00	1.00	50
V_TIPO_ACTIVIDAD(1)	Dependiente Formal	0.30	1.35	59
V_TIPO_ACTIVIDAD(2)	Dependiente Informal	0.17	1.18	55
V_TIPO_ACTIVIDAD(3)	Independiente Formal	0.11	1.12	53
V_LINEA	Video	0.00	1.00	50
V_LINEA(1)	Audio	-0.01	0.99	50
V_LINEA(2)	Blanca	0.37	1.45	61
V_LINEA(3)	Celulares	-0.56	0.57	34
V_LINEA(4)	Computo	0.16	1.17	54
V_LINEA(5)	Entretenimiento	-0.44	0.65	37
V_LINEA(6)	Hogar	-0.16	0.85	45
V_LINEA(7)	Peds	-0.57	0.57	34
V_ESTADO_CIVIL	Viudo	0.00	1.00	50
V_ESTADO_CIVIL(1)	Casado	0.50	1.65	64
V_ESTADO_CIVIL(2)	Conviviente	0.17	1.19	55
V_ESTADO_CIVIL(3)	Divorciado	0.05	1.05	51
V_ESTADO_CIVIL(4)	Soltero	0.01	1.01	50
v_rango_total_creditos	[4 a más]	0.00	1.00	50
v_rango_total_creditos(1)	[= 1]	-0.98	0.38	22
v_rango_total_creditos(2)	[2-3]	0.02	1.02	51
v_producto_creditos	TABLA DE NEGOCIOS Y OFICIOS	0.00	1.00	50
v_producto_creditos(1)	ANTECEDENTES CREDITICIOS	1.41	4.08	91
v_producto_creditos(2)	Complementario	0.80	2.23	73
v_producto_creditos(3)	Emprendedor	0.15	1.17	54
v_producto_creditos(4)	Empresarial	0.22	1.25	56
v_producto_creditos(5)	Facilito	1.77	5.86	101
v_producto_creditos(6)	Otros	0.66	1.94	69
v_producto_creditos(7)	Pre-aprobado vip	0.25	1.28	57
v_producto_creditos(8)	Recurrente Electro	0.65	1.91	69

...continuación

v_rango_plazo	>= 19	0.00	1.00	50
v_rango_plazo(1)	<= 12	0.42	1.53	62
v_rango_plazo(2)	13-18	0.15	1.16	54
V_TIPO_PROPIEDAD	Propia	0.00	1.00	50
V_TIPO_PROPIEDAD(1)	Alquiler	-0.51	0.60	35
V_TIPO_PROPIEDAD(2)	Espacio Cedido	1.37	3.94	90
V_TIPO_PROPIEDAD(3)	Familiar	-0.56	0.57	34
v_rango_maxima_mora	Null	0.00	1.00	50
v_rango_maxima_mora(1)	[sin mora]	18.70	132720038.80	590
v_rango_maxima_mora(2)	[1-30]	18.75	138734853.58	591
v_rango_maxima_mora(3)	[31-60]	1.03	2.79	80
v_rango_maxima_mora(4)	[61-90]	-1.78	0.17	-1
v_rango_maxima_mora(5)	[mayor a 90]	-3.92	0.02	-63
Constante		1.90	6.69	105

Anexo 2: Indicadores de Discriminación del Modelo para la Muestra

Score	Malo	Bueno	Total	% Malos	% Bueno	(%) F _{malo}	(%) F _{bueno}	(%) F _{malo} - F _{bueno}	(%) F _{malo(i)} - F _{malo (i-1)}	(%) F _{bueno(i)} + F _{bueno (i-1)}	(%) X _i
menor a 380	76	9	85	2.3	0.1	2.3	0.1	2.3	2.3	0.1	0.0
381-400	316	57	373	9.7	0.5	12.0	0.6	11.4	9.7	0.7	0.1
401-420	403	103	506	12.4	0.9	24.4	1.5	22.9	12.4	2.1	0.3
421-440	119	36	155	3.7	0.3	28.1	1.9	26.2	3.7	3.4	0.1
441-460	89	103	192	2.7	0.9	30.8	2.8	28.0	2.7	4.6	0.1
461-470	858	1465	2323	26.4	13.3	57.2	16.0	41.1	26.4	18.8	5.0
471-480	662	1520	2182	20.3	13.8	77.5	29.8	47.7	20.3	45.8	9.3
481-500	602	2295	2897	18.5	20.8	96.0	50.6	45.4	18.5	80.3	14.9
501-520	91	888	979	2.8	8.0	98.8	58.6	40.2	2.8	109.1	3.1
mayor a 520	39	4577	4616	1.2	41.4	100.0	100.0	0.0	1.2	158.6	1.9

Donde:

F_{malo}: Distribución acumulada de los clientes Malos.

F_{bueno}: Distribución acumulada de los clientes Buenos.

$$\text{Sea: } X_i = \left[(F_{malo(i)} - F_{malo(i-1)}) \times (F_{bueno(i)} + F_{bueno(i-1)}) \right]$$

$$K - S = \max |F_{malo}(x) - F_{bueno}(x)| = 47.7\%$$

$$Gini = 1 - \sum_{i=2}^n \left[(F_{malo(i)} - F_{malo(i-1)}) \times (F_{bueno(i)} + F_{bueno(i-1)}) \right] = 1 - \sum_{i=2}^n (X_i)$$

$$= 1 - 0.347 \approx 65.3\%$$

Anexo 3: Indicadores de Discriminación del Modelo para la Cartera

Score	Malo	Bueno	Total	% Malos	% Bueno	(%) F _{malo}	(%) F _{bueno}	(%) F _{malo} - F _{bueno}	(%) F _{malo(i)} - F _{malo(i-1)}	(%) F _{bueno(i)} + F _{bueno(i-1)}	(%) X _i
menor a 380	95	13	108	2.3	0.1	2.3	0.1	2.2	2.3	0.1	0.0
381-400	396	79	475	9.6	0.6	11.9	0.7	11.2	9.6	0.8	0.1
401-420	503	132	635	12.2	1.0	24.1	1.6	22.4	12.2	2.3	0.3
421-440	154	42	196	3.7	0.3	27.8	1.9	25.9	3.7	3.5	0.1
441-460	125	129	254	3.0	0.9	30.8	2.9	28.0	3.0	4.8	0.1
461-470	1107	1846	2953	26.8	13.3	57.6	16.2	41.4	26.8	19.0	5.1
471-480	842	1876	2718	20.4	13.5	78.0	29.7	48.2	20.4	45.9	9.4
481-500	742	2874	3616	18.0	20.8	95.9	50.5	45.4	18.0	80.2	14.4
501-520	117	1099	1216	2.8	7.9	98.8	58.4	40.3	2.8	108.9	3.1
mayor a 520	51	5757	5808	1.2	41.6	100.0	100.0	0.0	1.2	158.4	2.0

Dónde:

F_{malo}: Distribución acumulada de los clientes Malos.

F_{bueno}: Distribución acumulada de los clientes Buenos.

$$\text{Sea: } X_i = \left[\left(F_{malo(i)} - F_{malo(i-1)} \right) \times \left(F_{bueno(i)} + F_{bueno(i-1)} \right) \right]$$

$$K - S = \max |F_{malo}(x) - F_{bueno}(x)| = 48.2\%$$

$$\begin{aligned} Gini &= 1 - \sum_{i=2}^n \left[\left(F_{malo(i)} - F_{malo(i-1)} \right) \times \left(F_{bueno(i)} + F_{bueno(i-1)} \right) \right] = 1 - \sum_{i=2}^n (X_i) \\ &= 1 - 0.345 \approx 65.5\% \end{aligned}$$

Anexo 4: Total de variables Independientes

Variable	Descripción	Categorías
Perfil	Tipo de actividad económica del cliente	1. Dependiente formal
		2. Dependiente informal
		3. Independiente formal
		4. Independiente informal
Linea_Producto	Tipo de Línea del producto comprado	1. Audio
		2. Blanca
		3. Celulares
		4. Computo
		5. Entretenimiento
		6. Hogar
		7. Peds
		8. Video
Tipo_Propiedad	Tipo de propiedad de domicilio registrado	1. Alquiler
		2. Espacio cedido
		3. Familiar
		4. Propia
Estado_Civil	Variable del estado civil del cliente.	1. Casado
		2. Conviviente
		3. Divorciado
		4. Soltero
		5. Viudo
Nro_Creditos_Total	Número de créditos que tiene el cliente en la empresa.	1. Nro créditos=1
		2. Nro créditos entre 2 y 3
		3. Nro. Créditos \geq 4
Producto_Crediticio	Tipo de Producto Crediticio de acuerdo a la evaluación de cada cliente.	1. Antecedentes crediticios
		2. Complementario
		3. Emprendedor
		4. Empresarial
		5. Otros
		6. Pre-aprobado vip
		7. Recurrente electro
		8. Tabla de negocios y oficios
		9. Facilito
Plazo	Plazo de Crédito	1. Plazo \leq 12
		2. Plazo entre 13 y 18
		3. Plazo \geq 19

...continuación

Rango_Mora_Max	Máximo atraso del cliente de los últimos 12 meses en el sistema financiero	1. Atraso = null
		2. Atraso = [sin mora]
		3. Atraso = [1-30]
		4. Atraso = [31-60]
		5. Atraso = [61-90]
		6. Atraso = [mayor a 90]
Tipo_Cliente	Tipo de Cliente	1. Nuevo
		2. Recurrente
Genero	Genero del Cliente	1. Femenino
		2. Masculino
Grado_Instrucción	Grado de Instrucción del Cliente	1. Primaria
		2. Secundaria
		3. Técnica
		4. Universitaria
Tipo_Evaluación	Tipo de evaluación al origen del crédito	1. Bancarizados
		2. Otros
		3. Recurrentes
		4. Trámite Normal
Monto_Capital	Monto Capital Financiado del Crédito	1. Menor a 1500
		2. 1501-2000
		3. 2001-2500
		4. 2501-4000
		5. 4001 a más
Rango_Inicial	Valor de la inicial del crédito	1. Sin Inicial
		2. menor a 10%
		3. 11% - 20%
		4. mayor a 21%
Tiempo_Residencia	Tiempo de Residencia del cliente en el domicilio actual	1. [0-6 meses]
		2. [6-12 meses]
		3. [1-2 años]
		4. [2-3 años]
		5. [3 años a más]

...continuación

Material_Vivienda	Material de vivienda del cliente	1. Adobe
		2. Adobe con quincha
		3. Caña
		4. Estera
		5. M. Prensada
		6. Madera
		7. Noble
		8. Triplay
Zona_Domicilio	Zona del domicilio actual del cliente	1. Asentamiento Humano
		2. Otros
		3. Populosa
		4. Residencial
		5. Urbana
		6. Zona Comercial
		7. Zona Rural
Tipo_Teléfono	Tipo de Teléfono del cliente	1. Carece
		2. Celular
		3. Comunitario
		4. Domiciliario
Tipo_Pistas_Veredas	Tipo de Pistas y veredas de la zona de domicilio del cliente	1. Afirmadas
		2. Arenal
		3. Asfaltadas
		4. Carece
		5. Enripiado
		6. Trocha
Tipo_Ingresos	Tipo de Ingresos del cliente	1. Fijo
		2. Variable
Modalidad_Pago	Modalidad de pago que le otorgan al cliente	1. Honorarios
		2. Otros
		3. Planilla
		4. Recibo Simple
Frecuencia_Ingresos	Frecuencia de Ingresos del cliente	1. Diario
		2. Mensual
		3. Otros
		4. Por Obra
		5. Quincenal
		6. Semanal

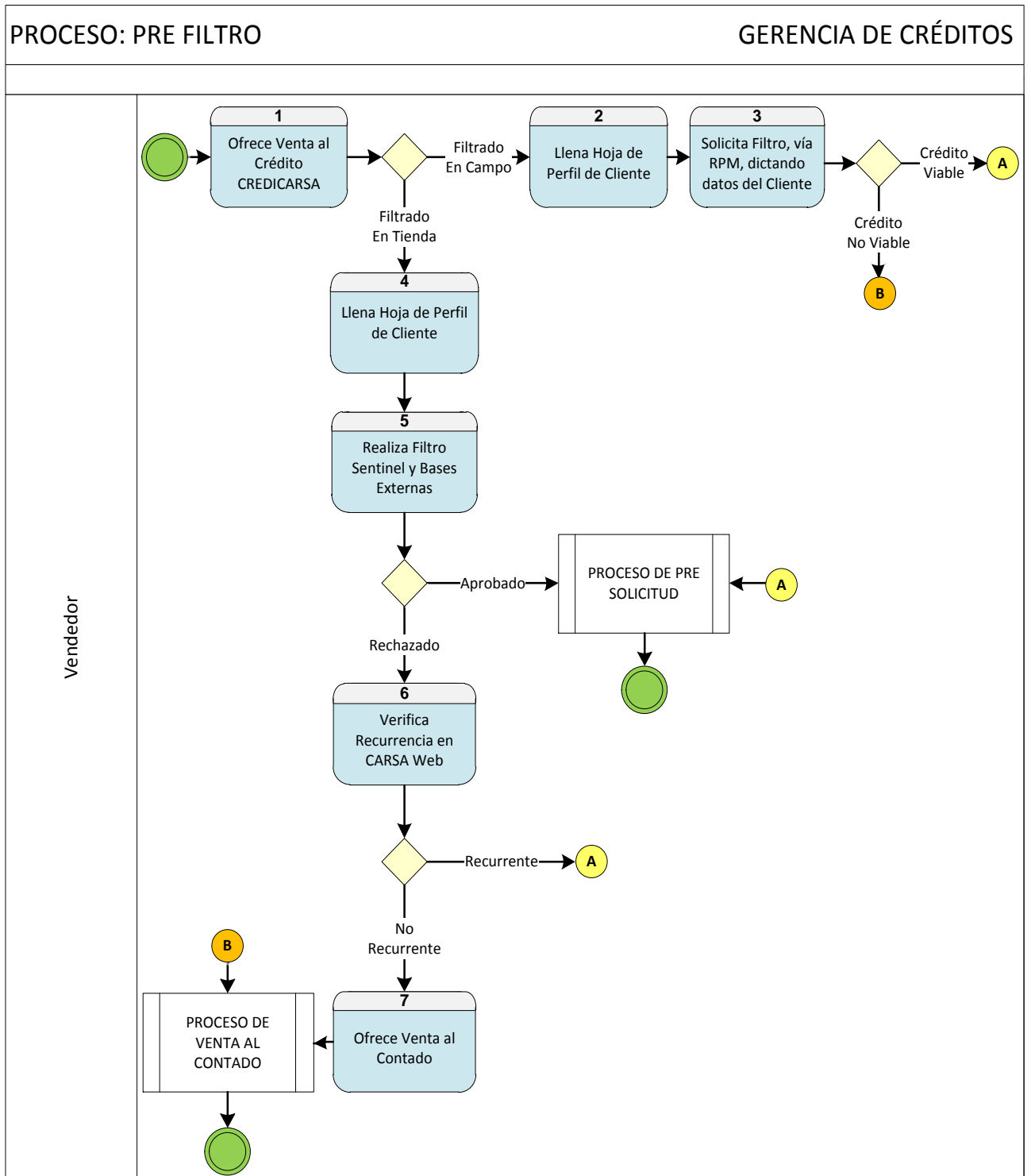
...continuación

Tipo_Puesto	Tipo de Puesto de trabajo del cliente	1. Fijo
		2. No Fijo
Tiempo de la empresa	Tiempo de creación de la empresa	1. [0-6 meses]
		2. [6-12 meses]
		3. [1-3 años]
		4. [3 años a más]
Tiempo laborando en la empresa	Tiempo laborando en la empresa	1. [0-6 meses]
		2. [6-12 meses]
		3. [1-3 años]
		4. [3 años a más]
Zona_Laboral	Zona del Centro Laboral actual del cliente	1. Asentamiento Humano
		2. Otros
		3. Populosa
		4. Residencial
		5. Urbana
		6. Zona Comercial
		7. Zona Rural
Formato_Puesto	Formato de Puesto del cliente	1. Ambulante
		2. Carretilla
		3. Kiosko
		4. Oficina
		5. Otros
		6. Puesto
		7. Taller
		8. Tienda
		9. Vehículo
		10. Vivienda
Propiedad_Negocio	Propiedad del negocio del cliente	1. No Propio
		2. Propio
Rango_Cuota	Rango de cuota del crédito del cliente	1. [menor a 100]
		2. [101-120]
		3. [121-150]
		4. [151-200]
		5. [201-400]
		6. [400 a más]

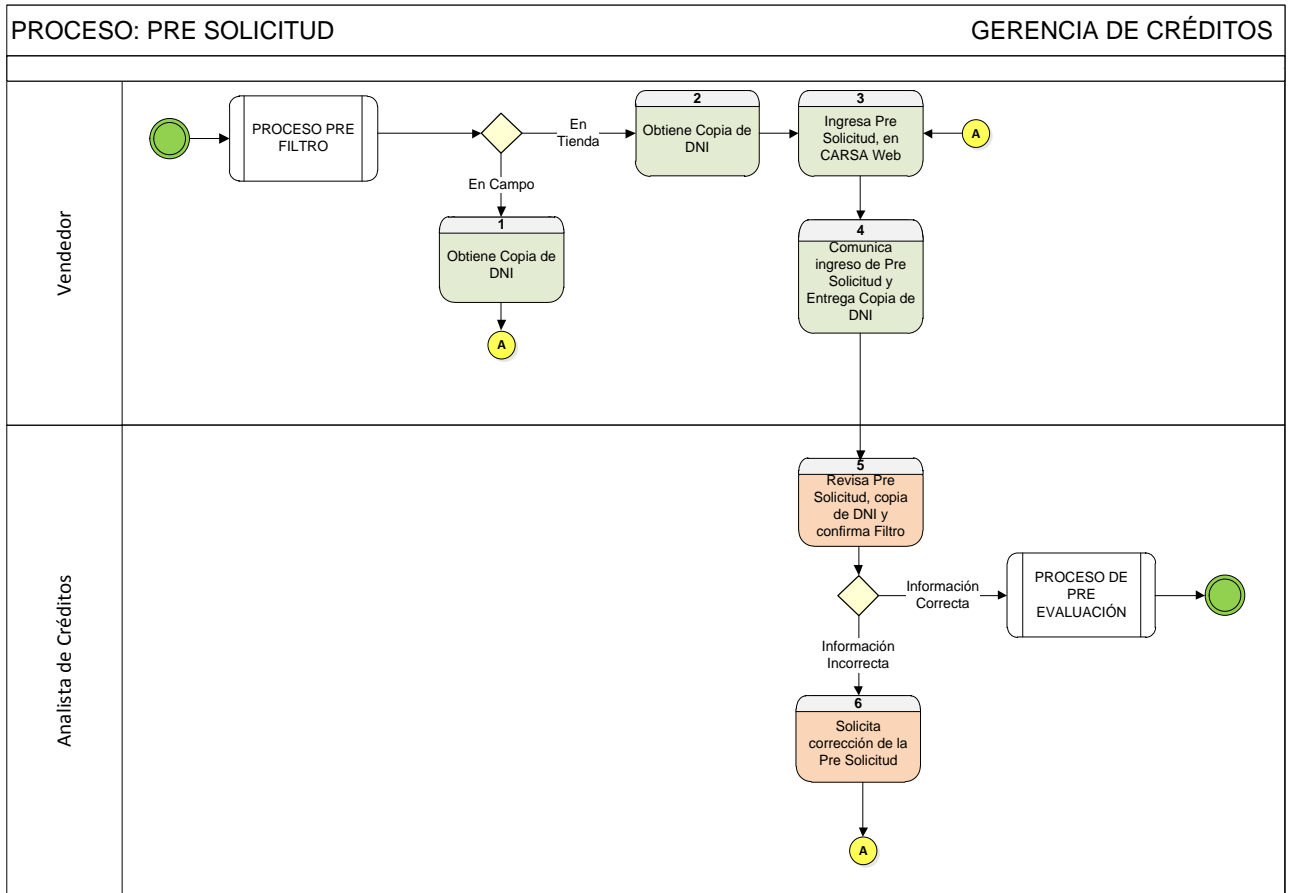
Anexo 5: Predicción de nuevos individuos

ID	TIPO_ACTIVIDAD	LINEA	ESTADO CIVIL	Total_Créditos	PRODUCTO DE CREDITO	TIPO PROPIEDAD	RANGO PLAZO	rango_maxima_mora	Score Cartera	Rango Score	Decision
1	INDEPENDIENTE INFORMAL	COMPUTO	SOLTERO	[= 1]	EMPRENDEDOR	FAMILIAR	<= 12	null	481	481-500	Aceptado
2	DEPENDIENTE FORMAL	COMPUTO	SOLTERO	[= 1]	TABLA DE NEGOCIOS Y OFICIOS	PROPIA	<= 12	null	502	501-520	Aceptado
3	DEPENDIENTE INFORMAL	COMPUTO	SOLTERO	[= 1]	RECURRENTE ELECTRO	FAMILIAR	<= 12	null	501	501-520	Aceptado
4	INDEPENDIENTE INFORMAL	BLANCA	SOLTERO	[= 1]	EMPRENDEDOR	Alquiler	13-18	[61-90]	430	421-440	Rechazado
5	INDEPENDIENTE INFORMAL	VIDEO	SOLTERO	[= 1]	TABLA DE NEGOCIOS Y OFICIOS	FAMILIAR	>= 19	[1-30]	1002	mayor a 520	Aceptado
6	INDEPENDIENTE INFORMAL	VIDEO	SOLTERO	[4 a más]	EMPRENDEDOR	Alquiler	<= 12	[31-60]	536	mayor a 520	Aceptado
7	DEPENDIENTE FORMAL	COMPUTO	SOLTERO	[= 1]	ANTECEDENTES CREDITICIOS	PROPIA	<= 12	[1-30]	1084	mayor a 520	Aceptado
8	INDEPENDIENTE INFORMAL	BLANCA	SOLTERO	[2-3]	EMPRENDEDOR	PROPIA	<= 12	[61-90]	482	481-500	Aceptado
9	DEPENDIENTE FORMAL	VIDEO	SOLTERO	[= 1]	ANTECEDENTES CREDITICIOS	PROPIA	<= 12	null	539	mayor a 520	Aceptado
10	INDEPENDIENTE INFORMAL	BLANCA	SOLTERO	[= 1]	TABLA DE NEGOCIOS Y OFICIOS	FAMILIAR	<= 12	[1-30]	1025	mayor a 520	Aceptado
11	INDEPENDIENTE INFORMAL	BLANCA	SOLTERO	[= 1]	FACILITO	FAMILIAR	>= 19	[1-30]	1064	mayor a 520	Aceptado
12	DEPENDIENTE FORMAL	VIDEO	SOLTERO	[= 1]	FACILITO	FAMILIAR	>= 19	[31-60]	551	mayor a 520	Aceptado
13	INDEPENDIENTE INFORMAL	VIDEO	SOLTERO	[= 1]	ANTECEDENTES CREDITICIOS	PROPIA	>= 19	[1-30]	1059	mayor a 520	Aceptado
14	DEPENDIENTE INFORMAL	COMPUTO	SOLTERO	[2-3]	FACILITO	PROPIA	>= 19	null	566	mayor a 520	Aceptado
15	INDEPENDIENTE INFORMAL	COMPUTO	SOLTERO	[= 1]	FACILITO	PROPIA	<= 12	[31-60]	574	mayor a 520	Aceptado
16	INDEPENDIENTE INFORMAL	VIDEO	SOLTERO	[= 1]	RECURRENTE ELECTRO	PROPIA	13-18	[31-60]	530	mayor a 520	Aceptado
17	INDEPENDIENTE INFORMAL	VIDEO	SOLTERO	[= 1]	EMPRENDEDOR	PROPIA	<= 12	null	493	481-500	Aceptado
18	INDEPENDIENTE INFORMAL	COMPUTO	CONVIVIENTE	[= 1]	RECURRENTE ELECTRO	PROPIA	>= 19	[61-90]	454	441-460	Rechazado
19	DEPENDIENTE FORMAL	COMPUTO	CASADO	[2-3]	TABLA DE NEGOCIOS Y OFICIOS	FAMILIAR	>= 19	null	517	501-520	Aceptado
20	DEPENDIENTE INFORMAL	AUDIO	CONVIVIENTE	[2-3]	FACILITO	FAMILIAR	>= 19	[1-30]	1092	mayor a 520	Aceptado
21	INDEPENDIENTE INFORMAL	VIDEO	CASADO	[= 1]	PRE-APROBADO VIP	PROPIA	>= 19	null	498	481-500	Aceptado
22	INDEPENDIENTE INFORMAL	BLANCA	CONVIVIENTE	[4 a más]	EMPRENDEDOR	Alquiler	13-18	[mayor a 90]	401	401-420	Rechazado
23	INDEPENDIENTE INFORMAL	COMPUTO	VIUDO	[2-3]	FACILITO	FAMILIAR	>= 19	null	545	mayor a 520	Aceptado
24	DEPENDIENTE FORMAL	COMPUTO	SOLTERO	[= 1]	RECURRENTE ELECTRO	PROPIA	<= 12	null	521	mayor a 520	Aceptado
25	INDEPENDIENTE INFORMAL	VIDEO	CONVIVIENTE	[= 1]	TABLA DE NEGOCIOS Y OFICIOS	FAMILIAR	13-18	null	470	461-470	Aceptado
26	DEPENDIENTE FORMAL	VIDEO	SOLTERO	[2-3]	FACILITO	FAMILIAR	<= 12	[31-60]	592	mayor a 520	Aceptado
27	INDEPENDIENTE INFORMAL	VIDEO	SOLTERO	[2-3]	FACILITO	FAMILIAR	<= 12	[61-90]	502	501-520	Aceptado
28	DEPENDIENTE INFORMAL	BLANCA	SOLTERO	[= 1]	EMPRENDEDOR	FAMILIAR	13-18	[1-30]	1026	mayor a 520	Aceptado
29	DEPENDIENTE FORMAL	VIDEO	CONVIVIENTE	[= 1]	EMPRENDEDOR	PROPIA	13-18	[1-30]	1040	mayor a 520	Aceptado
30	INDEPENDIENTE INFORMAL	VIDEO	VIUDO	[= 1]	EMPRENDEDOR	PROPIA	<= 12	null	493	481-500	Aceptado
31	DEPENDIENTE INFORMAL	COMPUTO	CASADO	[= 1]	EMPRENDEDOR	PROPIA	13-18	[61-90]	457	441-460	Rechazado
32	INDEPENDIENTE INFORMAL	BLANCA	CASADO	[2-3]	PRE-APROBADO VIP	PROPIA	13-18	[1-30]	1083	mayor a 520	Aceptado

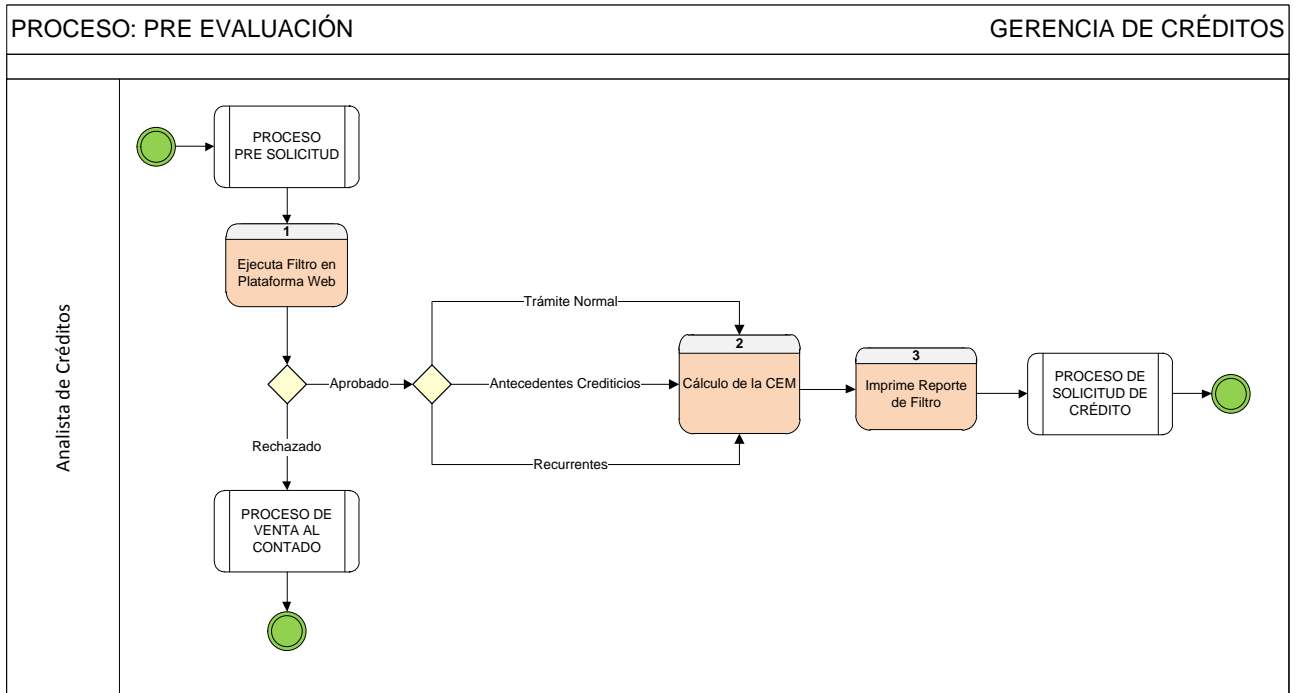
Anexo 6: Proceso de Pre-Filtro



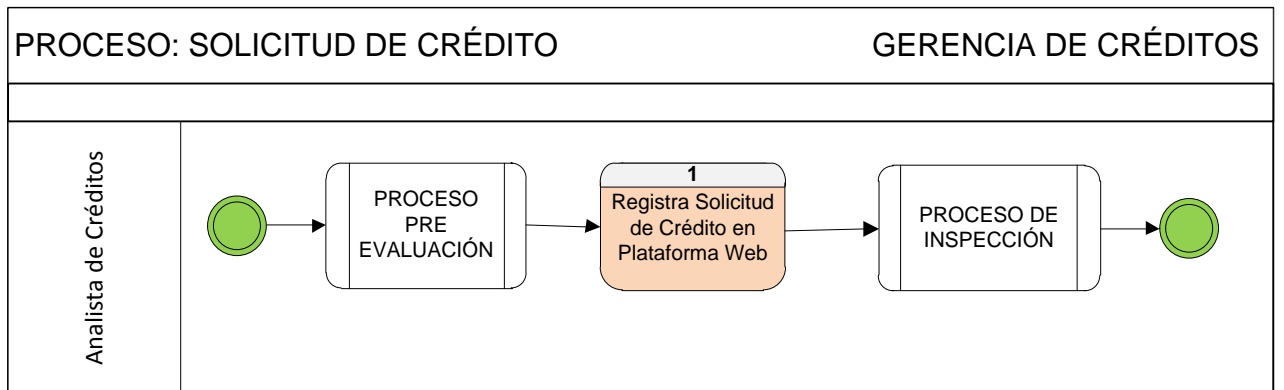
Anexo 7: Proceso de Pre-Solicitud



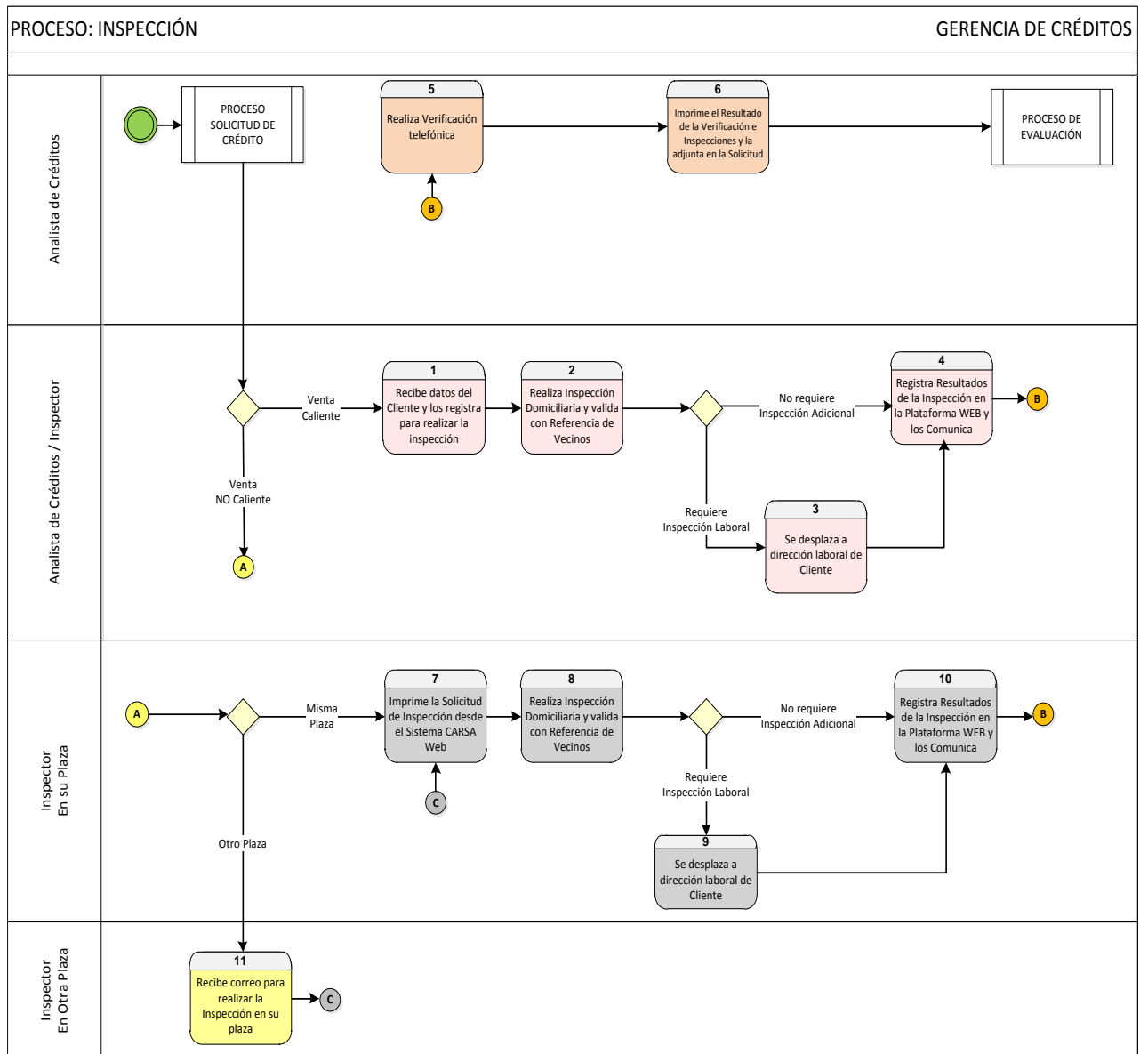
Anexo 8: Proceso de Pre-Evaluación



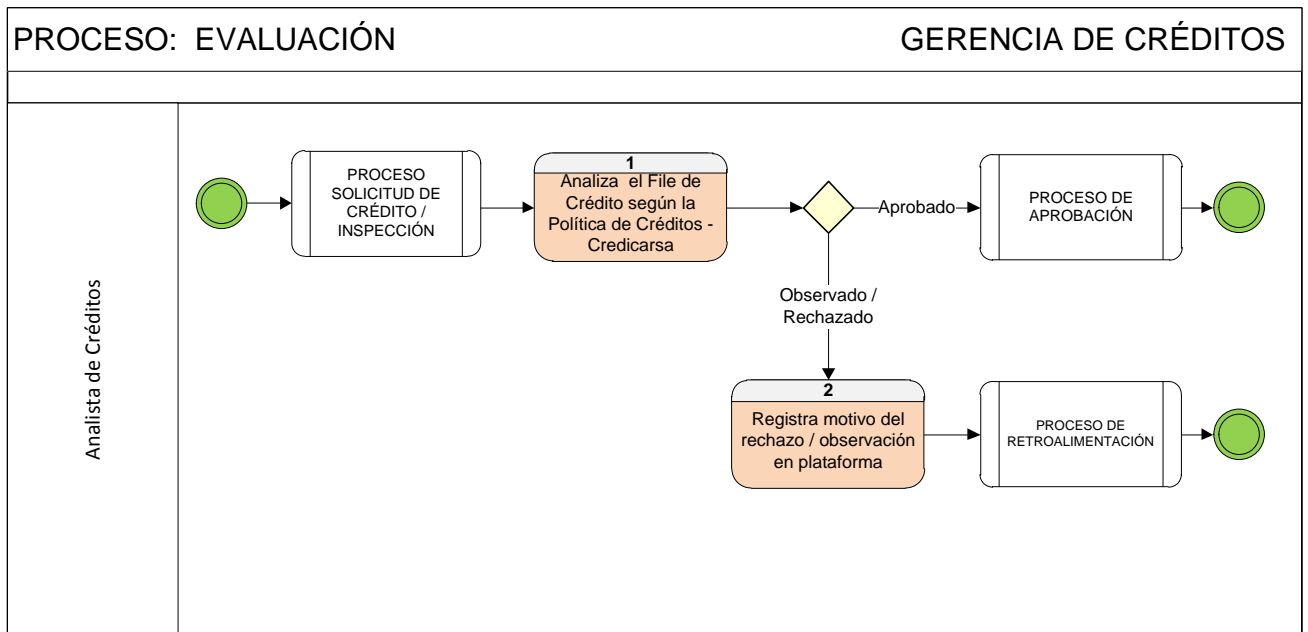
Anexo 9: Proceso de Solicitud de Crédito



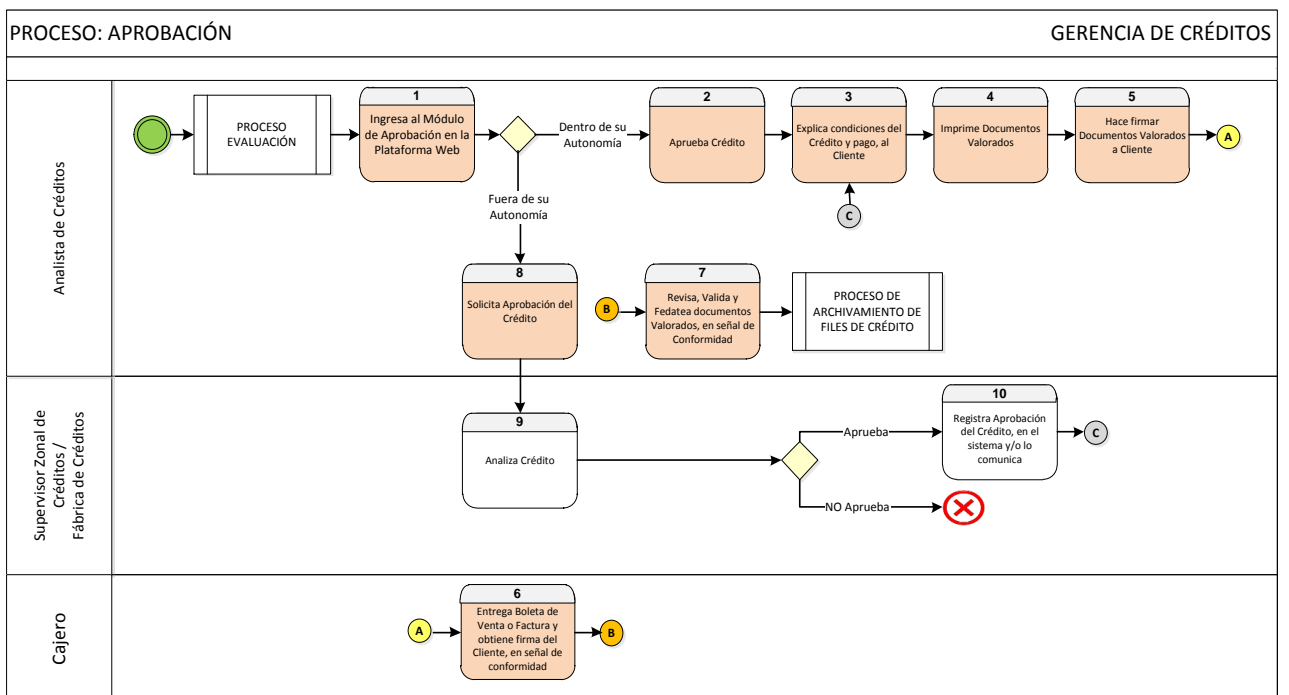
Anexo 10: Proceso de Inspección



Anexo 11: Proceso de Evaluación



Anexo 12: Proceso de Aprobación



Anexo 13: Proceso de Retroalimentación

