UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN

"PREDICCIÓN DEL RIESGO DE INCUMPLIMIENTO EN EL PAGO DE LOS CRÉDITOS DEL PORTAFOLIO DE UNA ENTIDAD FINANCIERA UTILIZANDO REGRESIÓN LOGÍSTICA"

TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR EL TÍTULO DE INGENIERA ESTADÍSTICA E INFORMÁTICA

ADRIANA MIRANDA PILCO

LIMA – PERÚ

2021

UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA
FACULTAD DE ECONOMÍA Y PLANIFICACIÓN

"PREDICCIÓN DEL RIESGO DE INCUMPLIMIENTO EN EL PAGO DE LOS CRÉDITOS DEL PORTAFOLIO DE UNA ENTIDAD FINANCIERA UTILIZANDO REGRESIÓN LOGÍSTICA"

Presentado por:

ADRIANA MIRANDA PILCO

TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR EL TÍTULO DE INGENIERA ESTADÍSTICA E INFORMÁTICA

SUSTENTADO Y APROBADO ANTE EL SIGUIENTE JURADO:

MA. Fernando René Rosas Villena	Dr. César Higinio Menacho Chiok
PRESIDENTE	**ASESOR**
Mg. Iván Dennys Soto Rodríguez	Mg. Diana Del Rocío Rebaza Fernández
MIEMBRO	**MIEMBRO**

LIMA – PERÚ

2021

DEDICATORIA

Dedico este trabajo con el mayor respeto y amor a:

Mis abuelos Rosa, Inocencia, Alfredo y Jorge, por ser mi ejemplo de perseverancia y constancia.

Mis padres y hermana Gabriela, quienes hacen de mí una mejor persona.

Mis sobrinas Vida y Hanan, porque son la alegría y el motor de toda mi familia.

AGRADECIMIENTOS

A Dios, por darme una familia maravillosa que me ha impulsado y apoyado para cumplir con las metas que me proponga.

A mi tía Rosa Chávez y familia, porque me acogieron en su hogar y apoyaron a lo largo de toda mi carrera universitaria.

A mi tío Carlos Miranda, quien con su experiencia y sabios consejos me impulsó a cumplir con esta meta.

Finalmente, al profesor Dr. César Menacho, por asesorarme y aconsejarme durante la elaboración de este trabajo.

¡Muchas gracias a cada uno por su apoyo incondicional!

ÍNDICE GENERAL

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

ÍNDICE DE ANEXOS

RESUMEN

El éxito de toda entidad financiera radica en la adecuada gestión de los riesgos a los que se encuentra expuesta, siendo uno de ellos el Riesgo de Crédito definido como la posibilidad de pérdida a consecuencia del incumplimiento de las obligaciones por parte del prestatario. Las herramientas analíticas usadas en la gestión de este tipo de riesgos han ido evolucionando a lo largo del tiempo e incluyendo a la estadística y la minería de datos como parte de estas. En esta memoria de Trabajo de Suficiencia Profesional se describe como la aplicación de la metodología de *Credit Scoring* conjuntamente con la metodología de minería de datos CRISP DM para la construcción de un modelo de riesgo comportamental en una entidad financiera, permitió obtener un indicador de gini de 64% y segmentar de mejor manera al portafolio de clientes de dicha entidad al incrementar en un 20% la participación de mejores clientes.

Palabras Claves: Sistema Financiero, Riesgo de Crédito, Modelos de Riesgo, *Credit Scoring,* Regresión Logística, Incumplimiento, Probabilidad de Incumplimiento.

ABSTRACT

The success of a financial institution lies in the proper management of the risks that it is exposed, being one of them a Credit Risk which is defined as the possibility of loss as a result of the borrower's failure to meet his obligations. The analytical tools used in the management of this type of risk have been evolving over time and include statistics and data mining as part of these tools. In this Professional Sufficiency Work report, it is described how the application of the Credit Scoring methodology together with the CRISP DM data mining methodology for the construction of a behavioral risk model in a financial institution, allowed to obtain a Gini coefficient of 64% and to better segment the client portfolio of that institution by increasing the participation of the best clients by 20%.

Keywords: Financial System, Credit Risk, Risk Models, Credit Scoring, Logistic Regression, Default, Probability of Default.

I. PRESENTACIÓN

La Banca es una de las industrias más antiguas del mundo, y su continua evolución ha ido de la mano con el avance de la tecnología y la digitalización, transformando y optimizando sus procesos y la forma de interactuar con sus clientes. Sin embargo, su modelo de negocio sigue siendo el mismo, Santos (2014) lo define como una matriz insumo – producto, ya que capta recursos vía depósitos, los que conjuntamente con su capital propio, son colocados en negocios viables mediante créditos bajo diversas modalidades.

Asimismo, en el Perú, uno de los principales aportes de la evolución de la banca ha sido y sigue siendo la dinamización de la economía a través de los créditos, impactando positivamente en el producto bruto interno (PBI), y en el bienestar de la población en general, ya que permite la satisfacción de sus necesidades. El crecimiento de la banca peruana, según los análisis en ASBANC (2015) se debe principalmente a su solidez, ya que se desenvuelve en un entorno de sólidos fundamentos macroeconómicos, y una responsable gestión del riesgo por parte de las diferentes entidades financieras.

La entidad financiera en estudio se encuentra dentro del top 10 de los bancos más grandes del país, al tener una participación aproximada del tres por ciento del total de colocaciones en el sistema financiero. Su éxito se debe a sus más de 40 años de experiencia en el rubro y el de estar presente en casi todo el territorio del país, lo que le ha conllevado a especializarse en diferentes tipos de productos crediticos según las necesidades de sus clientes, es así que su portafolio se encuentra conformado por personas naturales de la banca consumo, las micro, pequeñas y grandes empresas. Asimismo, como toda entidad financiera, uno de sus grandes pilares es la adecuada gestión de los riesgos a los que se encuentra expuesta, siendo uno de ellos el Riesgo de Crédito que es la posibilidad de pérdida a consecuencia del incumplimiento de las obligaciones por parte del prestatario.

Una de las áreas funcionales de la Gerencia de Riesgo de Crédito es el área de Modelos de Riesgos, cuya misión es la de ser el soporte analítico de riesgos de la entidad financiera, al proponer y desarrollar soluciones analíticas que permitan generar estrategias de mitigación del Riesgo de Crédito, siendo una de sus principales funciones el desarrollo, implementación y administración de modelos asociados a los siguientes objetivos: predicción del comportamiento de pago futuro del cliente, estimación del ingreso del cliente así como del nivel de sobreendeudamiento de este. En esta memoria se describirá el proceso de construcción de un modelo que estima la probabilidad de incumplimiento en el pago de los créditos del portafolio de la entidad financiera.

II. INTRODUCCIÓN

La importancia de los modelos de riesgos utilizados por la entidad financiera, radica en que estos forman parte de la toma de decisiones en las diferentes etapas del ciclo de vida de sus clientes, es decir forman parte de: (1) Prospección, ya que permiten estimar la probabilidad de impago de potenciales clientes,(2) Originación: permiten determinar los niveles de riesgo de clientes que soliciten directamente un crédito a la entidad; (3) Seguimiento, anticipan el deterioro de los créditos permitiendo generar estrategias de reducción y/o fidelización de los buenos clientes, (4) Cobranzas, permiten determinar los perfiles de los clientes morosos y generar estrategias de recobro.

Todo modelo de riesgo debe cumplir con los estándares de predictibilidad y calidad estipulados por la entidad financiera y la entidad reguladora (SBS). Es así, que surge la necesidad de la construcción de un modelo de comportamiento que estime la probabilidad de incumplimiento de aquellos clientes que tengan al menos una tarjeta de crédito con la entidad financiera, debido a la descalibración y baja capacidad predictiva del modelo en producción en este segmento de clientes, cuyas estimaciones son utilizadas en la generación de estrategias de incremento de líneas de crédito, reducción de tasas entre otros.

Para desarrollar la construcción del modelo predictivo, se utilizó la metodología de Credit Scoring, cuyos resultados permiten asignar un puntaje de riesgo a los clientes mediante modelos estadísticos. Asimismo, debido al avance de la tecnología, se cuenta con un gran volumen de información, que para ser explotada y analizada se debe hacer uso de las técnicas de minería de datos e inteligencia de negocios. Es así, que para la explotación de toda la información del presente estudio se utilizó la Metodología de Minería de Datos; CRISP DM (Cross Industry Standard Process for Data Mining), debido a que es una de las metodologías más utilizadas en los proyectos de análisis de datos ya que incluye el entorno del negocio en su desarrollo. Y, para la estimación del modelo predictivo se utilizó la técnica estadística de Regresión Logística, ya que a lo largo de la historia del Credit Scoring ha dado buenos

resultados, además de presentar la ventaja de medir la probabilidad de incumplimiento manteniéndola en un rango de variación entre cero y uno.

Como resultado del modelo propuesto se obtuvo un índice de gini de 64% y su ajuste se encuentra en los intervalos óptimos de calibración, lo que indica que el modelo tiene alta capacidad predictiva y un buen ajuste de la probabilidad estimada respectivamente. Lo que generó una considerable mejora al momento de identificar a los clientes de los segmentos de riesgos más bajos, incrementando su participación en más del 20%. Con estos resultados, las áreas usuarias podrán generar mejores estrategias de incremento de líneas de crédito así como de fidelización de estos clientes.

III. OBJETIVOS

3.1. Objetivo general

Proveer de una herramienta estadística a las áreas de negocio y riesgos de la entidad financiera, que permita segmentar a los clientes de su portafolio en función de su probabilidad de incumplimiento para la generación de estrategias de apalancamiento y/o reducción de líneas de crédito

3.2. Objetivos específicos

- Desarrollar un modelo estadístico predictivo del incumplimiento de los clientes del portafolio de la entidad financiera.
- Identificar las variables que influyen en el incumplimiento de los clientes.
- Asegurar la granularidad de la probabilidad de incumplimiento estimada.
- Transformar la probabilidad de incumplimiento estimada en una escala entendible, score, por las diferentes áreas usuarias.
- Determinar los cortes de score para la identificación de los segmentos de riesgo según el apetito de riesgo de la entidad financiera.

IV. CUERPO DEL TRABAJO

4.1. Descripción de las funciones desempeñadas:

A continuación, se describen las principales funciones desarrolladas en la entidad financiera:

- Desarrollo de las metodologías de validación y seguimiento de los modelos de riesgos. La adecuada gestión de los modelos de riesgos, hace que estos sean herramientas potentes en la gestión de los clientes del portafolio de las entidades financieras, Management Solutions (2014) indica que después de la modelación, dos tareas que deben implementarse son: La Validación de Modelos y El Seguimiento de Modelos, en los que se evalúa todo el procedimiento de construcción del modelo y el estado de salud de los modelos de manera periódica respectivamente. Es así que se construyeron los documentos metodológicos que detallan las fases a seguir para una adecuada validación y seguimiento de modelos.
- Automatización del seguimiento de los indicadores de predicción, estabilidad poblacional y calibración de los modelos en producción. En coordinación con la gerencia de riesgos, se estableció realizar los seguimientos de modelos de riesgos de manera mensual, para lo cual se automatizó los procesos de cálculo de los indicadores de evaluación de los modelos con el objetivo de optimizar el tiempo operativo y tomar acciones inmediatas dependiendo de los resultados obtenidos.
- Construcción de modelos de riesgos logísticos para las diferentes etapas del ciclo de crédito de los clientes: prospección (Modelo Buro), originación (Modelo Applicant) y seguimiento de los créditos (Modelo Comportamental). En todos los casos se siguió la metodología de Credit Scoring, obteniendo excelentes resultados.
- Segmentación de la cartera de clientes en función de la probabilidad estimada por los modelos. Una vez culminado el modelo, se debe generar los puntos de corte del score para la generación de los segmentos de riesgo.
- Automatización de reportes de análisis univariado y bivariado en el software R Project.

4.2. Puesta en práctica de lo aprendido

4.2.1. Revisión bibliográfica

En (Trejo, Ríos & Almagro, 2016) se detallan el proceso de actualización de un modelo de riesgo de crédito para la banca Revolvente de México. La necesidad de esta actualización surgió debido al aumento del porcentaje de créditos revolventes o de tarjetas de crédito en estado vencido, pasando de un 45% a un 49%. Para lo cual, construyeron un modelo de comportamiento logístico, cuya variable dependiente fue el incumplimiento de parte del cliente en el pago de un crédito en dos cuotas consecutivas y como variables predictivas a todas aquellas asociadas al comportamiento de pago del cliente con ese crédito (Historial de impagos, Antigüedad del crédito, Relación Pago Saldo, entre otros). Para la evaluación de la capacidad predictiva del modelo utilizaron la curva ROC y el indicador K-S obteniendo resultados satisfactorios. Asimismo, evaluaron el impacto en las provisiones cuyo calculo tiene como input la probabilidad estimada por el modelo, de este concluyeron que, si el modelo propuesto hubiera estado implementado en la Banca Revolvente se hubiera tenido un ahorro de MXM $6.2 millones y si la tendencia se siguiera en la Banca Consumo el ahorro habría sido de MXM $848 millones.

En (Pérez, 2017), se planteó un modelo de regresión logística para el otorgamiento de créditos financieros a las organizaciones de la Economía Social y Solidaria (OESS) de Ecuador, ya que es un sector productivo muy importante para el país al estar conformado por comunidades campesinas, grupos de artesanos y productores que necesitan ampararse en productos financieros adecuados a su contexto. Para el desarrollo del proyecto Pérez evaluó los factores atribuibles a la demanda de crédito, tales como historial del pago, así como los ingresos y ventas de la organización, garantías entre otros. El modelo resultante cumple con los supuestos de la técnica así como el de tener buena capacidad predictiva, al tener 70% de sensibilidad y 72% de especificidad. En sus conclusiones, Pérez rescata la importancia de la regresión logística como herramienta en el análisis de impago de las obligaciones crediticias de las OESS con las entidades financieras que en este caso serían las Cooperativas de Ahorro y Crédito (COAC) quiénes se han especializado en este nicho, permitiendo a éstas últimas optimizar y mejorar la Gestión del Riesgo de Crédito.

En (Rodríguez & Trespalacios, 2015), se plantean el uso de un modelo logístico que permita medir el riesgo de impago de la siguiente cuota de una cartera de crédito, dinamizado con simulaciones de Montecarlo para la identificación de la pérdida esperada de una entidad financiera de Colombia en el marco de lo estipulado en Basilea II. El modelo logístico que proponen cumple con los análisis de bondad de ajuste teniendo un 99.62% de sensibilidad y 94.18 de especificidad, por lo que, una vez obtenida la probabilidad de incumplimiento del cliente se calcularía por medio de simulaciones de Montecarlo (aproximadamente un millón de ensayos) la pérdida esperada del mismo cliente. En sus conclusiones, Rodríguez y Trespalacios, resaltan las ventajas e importancia de la regresión logística en los proyectos de Credit Score, y como con su interacción con las simulaciones de Montecarlo se genera una herramienta más dinámica de análisis financiero, no quedándose solo con la interpretación del modelo y la predicción del riesgo, sino con la evaluación de posibles escenarios con la estimación de la pérdida esperada en cualquier momento del tiempo.

4.2.2. Descripción de las técnicas estadísticas y/o informáticas utilizadas en la solución de la situación problemática en el ejercicio de su actividad laboral

En el presente estudio se utilizó la metodología de minería de datos CRISP – DM (Cross Industry Standard Process for Data Mining), que en base a los resultados obtenidos por KDnuggets. (2014) sigue siendo la metodología más utilizada para los proyectos de análisis de datos, con un 43% de participación en su última encuesta realizada el 2014. Según (Chapman, 2000) se describe en términos de un modelo de proceso jerárquico, que consiste en conjuntos de tareas descritas en cuatro niveles de abstracción (de general a específico): fase, tarea genérica, tarea especializada e instancia de proceso.

La descripción de fases y tareas como pasos discretos realizados en un orden específico representa una secuencia idealizada de eventos. En la práctica, muchas de las tareas se pueden realizar en un orden diferente, y a menudo será necesario retroceder reiteradamente a las tareas anteriores y repetir ciertas acciones. Lo descrito se ve reflejado en la Figura 1.

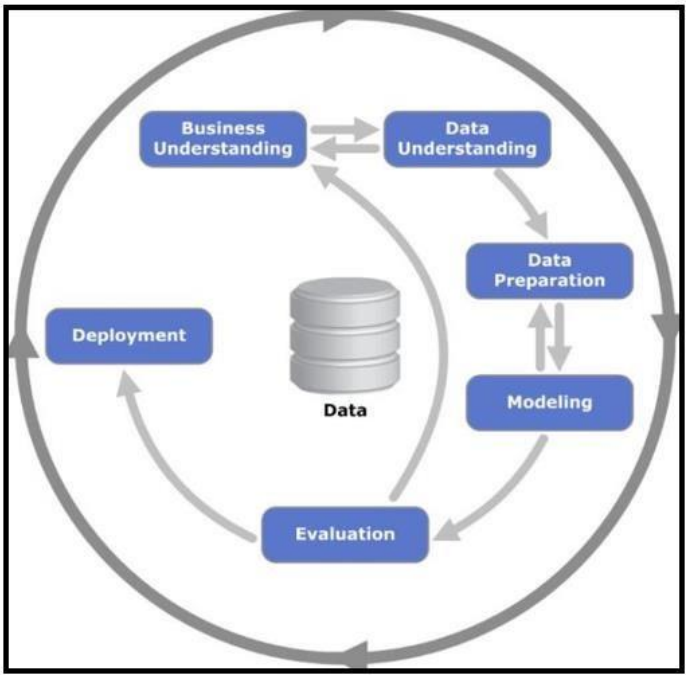


Figura 1: Fases del ciclo analítico

FUENTE: SPSS, 2000.

A continuación se describen las fases del ciclo analítico:

4.2.2.1. Comprensión del negocio

Es la primera fase del ciclo analítico donde se formulan los principales ejes del proyecto, en conjunto con las diferentes áreas involucradas en el proceso, existiendo cinco aspectos a definir:

- Objetivos del negocio.
- Estado actual del negocio.
- Objetivos del proyecto.
- Definición del Alcance y Diseño.
- Gantt del proyecto.

Para el desarrollo de esta fase, se debe coordinar reuniones con los especialistas de las áreas usuarias, donde serán expuestos los puntos de vista de cada equipo y las expectativas que se tiene respecto a los resultados del modelo. Así mismo se seleccionaran a los skateholders y los requerimientos necesarios para el desarrollo del proyecto.

Una de las metodologías para el adecuado levantamiento de la información descrita previamente, es el Design Thinking que según Brown (2008):

> *"Es una disciplina que usa la sensibilidad y métodos de los diseñadores para hacer coincidir las necesidades de las personas con lo que es tecnológicamente factible y con lo que una estrategia viable de negocios puede convertir en valor para la clientela, así como en una gran oportunidad para el mercado".*

Y, una de sus técnicas más completas y utilizadas es el Business Model Canvas (Lienzo de Modelo de Negocio), que permite diseñar e implementar sistemáticamente ideas visionarias que beneficie el desarrollo del proyecto, a las áreas interesadas y a la entidad en general.

> *"El punto de partida para cualquier buena discusión, reunión o taller sobre innovación del Business Model debe ser una comprensión compartida de lo que es realmente un Business Model. Se necesita un concepto de Business Model que todos entiendan: uno que facilite la descripción y la discusión. Es necesario comenzar desde el mismo punto y hablar sobre lo mismo. El desafío es que el concepto debe ser simple, relevante e intuitivamente comprensible, sin simplificar demasiado las complejidades de cómo funcionan las empresas".*

Osterwalder & Pigneur (2010) definieron nueve bloques de construcción que identifican las necesidades y la forma de como un proyecto puede ser exitoso. A continuación se describe cada uno de estos:

a. **Segmento de clientes:**
   Se ha de definir los segmentos de clientes que la entidad busca alcanzar y ofrecer valor.
b. **Propuesta de valor:**
   Se ha de identificar y describir los productos y/o servicios que crearan valor para los segmentos de clientes definidos.
c. **Canales:**
   Se ha de describir como la entidad se comunicará y entregará su propuesta de valor

a los clientes.

d. **Relaciones con los clientes:**

   Se identificaran los tipos de relaciones que una entidad tiene con los segmentos específicos de clientes.

e. **Fuente de ingresos:**

   Este bloque representa la forma en cómo los clientes generaran ingresos monetarios a la entidad.

f. **Recursos clave:**

   Se ha de identificar y describir los más importantes elementos necesarios para que el proyecto funcione.

g. **Actividades clave:**

   Se identificaran y describirán las principales actividades que la entidad debe realizar para que el proyecto funcione.

h. **Socios clave:**

   Se ha de identificar y describir a la red de proveedores y socios más importantes.

i. **Estructura de costos:**

   Se identificarán y describirán todos los costos en los que se incurrirán para la operatividad del proyecto.

Una vez definidos los principales ejes del proyecto y el plan de ejecución del desarrollo del mismo, se ha de proceder con las siguientes fases, sin embargo, dependiendo de los hallazgos en los análisis realizados en éstas fases, si es necesario, se redefinirán.

4.2.2.2. Comprensión de los datos

La factibilidad, viabilidad y calidad del modelo dependerá de las fuentes de información a utilizar durante su construcción e implementación.

> IBM (2015) *Esta fase implica estudiar más de cerca los datos disponibles. Este paso es esencial para evitar problemas inesperados durante la siguiente fase (preparación de datos) que suele ser la fase más larga de un proyecto*.

Para el desarrollo de esta fase, como primera tarea se deben identificar las fuentes primarias de información, existiendo diversos orígenes como:

a. **Datos existentes:** Incluye a toda información interna que maneje la entidad, como datos transaccionales, solicitudes, encuestas, etc.
b. **Datos adquiridos:** Es toda información externa con la que cuenta la entidad y si no la tuviera se debe considerar la necesidad de adquirirla.
c. **Datos adicionales:** En el caso de que la información anteriormente descrita no es suficiente, se debe evaluar qué información complementaria se debe obtener.

Una vez identificadas las fuentes de información, se debe hacer el requerimiento correspondiente a las áreas de tecnología de la información para luego realizar los análisis descriptivos y exploratorios, que permitirán evaluar la calidad y cantidad de los datos, asimismo se debe validar la veracidad de estas fuentes.

El análisis descriptivo permite conocer cómo se encuentran estructurados los datos en cuanto a la cantidad de registros llenos, los tipos de valores almacenados y los esquemas de codificación para ciertas variables.

En cuanto a la calidad de los datos, se debe evaluar la representatividad en el conjunto de datos de los siguientes problemas.

- **Datos perdidos:** Identificar los valores que se almacenaron como datos perdidos.
- **Los errores de data:** Suele ser errores tipográficos cometidos al momento de introducir los datos.
- **Los errores de mediciones:** Datos basados en esquema de mediciones incorrectos.
- **Incoherencias de codificación.**
- **Metadatos erróneos:** Errores entre el significado aparente de un campo incluido en un nombre o definición del campo.

4.2.2.3. Preparación de los datos

Es la fase que demora más tiempo, ya que en esta se procede a la preparación y transformación de los datos dependiendo de la técnica de modelado que se ha de utilizar posteriormente. Dependiendo de las necesidades del proyecto las tareas a realizar son:

a. Estructuración de los datos:

En esta etapa se ha de realizar la integración de datos y su respectivo formateo, que consiste en el consolidado de la información de forma estructurada en los sistemas de gestión de datos y la realización de transformaciones sintácticas sin modificar el significado de los datos.

El objetivo de esta tarea, es facilitar el acceso y análisis de la información permitiendo que esta sea lo más exacta y que refleje la realidad.

IBM (2015) *"Dedicar los esfuerzos adecuados a las primeras fases de comprensión comercial y comprensión de datos puede reducir al mínimo los gastos indirectos relacionados, pero aún deberá dedicar una buena cantidad de esfuerzo para preparar y empaquetar los datos para el proyecto".*

b. Definición de la población:

Una vez consolidada la información, se debe definir con qué población se ha de trabajar, las consideraciones y la elección del segmento de clientes responderán a lo requerido en la fase de comprensión del negocio. Sin embargo, en los proyectos de *credit scoring* se debe evaluar la estabilidad en cuanto a número de clientes y saldo adeudado que tiene el cliente con la entidad para la elección de los meses a analizar.

Asimismo, se ha de evaluar los perfiles de clientes que no deberían ser tomados en cuenta en la etapa de modelamiento, estas exclusiones serán las identificadas en la fase de comprensión del negocio, así como aquellos perfiles que el modelador considere que podrían generar algún tipo de sesgo al momento de construir el modelo.

Todas las decisiones finales en cuanto a la definición de la población deberán ser consultadas y aprobadas por las áreas usuarias y los comités correspondientes.

c. Definición de la variable dependiente:

Una vez que se haya definido la población con la que se modelará, habrá de definirse el target, para los proyectos de *credit scoring* se ha de evaluar todas las definiciones de incumplimiento (default) identificadas en la fase de comprensión del negocio.

Para una adecuada definición se deben considerar los siguientes aspectos:

- **Número de días de atraso:** La definición regulatoria especificada en Basilea II. (2001), estipula que un cliente habrá hecho default cuando haya incumplido en el pago de alguno de sus créditos por más de 90 días. Sin embargo, las entidades financieras pueden definir otro nivel de incumplimiento, siempre y cuando este sea menor a la definición regulatoria, por lo que es necesario realizar un análisis de Roll-Rate que según Siddiqi (2006) consiste en comparar los tramos de morosidad en un punto determinado respecto a su evolución a lo largo de un periodo de tiempo, con el fin de determinar el número máximo de días de atraso en los que el cliente podría revertir su estado, mientras que superándolos ya no existiría tal reversión, es decir el cliente no se curaría.

- **Ventana de desempeño:** Una vez definido el default, debe evaluarse el plazo de tiempo en el que se acumule la mayor cantidad de casos que hicieron default.

- **Estabilidad del ratio de default:** El *ratio de default* (RD) se define como la proporción de clientes que hicieron *default* en al menos uno de los meses de la ventana de desempeño, es así que se debe verificar la estabilidad de este en los diferentes meses de estudio.

En el caso de que se identificaran meses en los que el ratio de *default* fuera atípico, se deberá realizar los análisis correspondientes para detectar qué factores en la población impactaron en este.

- **Segmentación de la población:**

  Si bien se definió la población objetivo, es necesario realizar los análisis correspondientes de segmentación, ya que su importancia radica en la identificación de subpoblaciones de clientes que comparten rasgos comunes de predicción o patrones únicos de comportamiento Siddiqi (2006), define dos formas de segmentación:

  - Generación de segmentos en base a la experiencia y el conocimiento del negocio, para luego ser validados en el análisis.
  - Generación de segmentos en base a técnicas estadísticas tales como los árboles de decisión y clusterización.

  Una vez que se hayan definido las variables de segmentación, se deben evaluar los siguientes aspectos:

- **Materialidad:** Se debe verificar que la proporción en cuanto a número de clientes y saldo de las subpoblaciones respecto a la población sean representativas. Asimismo, se debe evaluar la estabilidad de la materialidad en los meses de estudio.
- **Heterogeneidad del ratio de *default*:** Aún si las subpoblaciones definidas por las variables de segmentación cumplieran con los requisitos de materialidad, es importante evaluar que estas tengan la capacidad de diferenciarse por los valores de sus ratios de *default*.
- **Partición de la base:**

  Según Vicente, Gonzáles, Parra y Beltrán (2019), en el proceso de estimación del modelo se debe utilizar una estrategia para que la técnica de clasificación se optimice, una de ellas es la práctica del muestreo para particionar la base de modelamiento en dos conjuntos de datos. La base de entrenamiento con la que se estima los parámetros del modelo y la base de test que se utiliza para ajustar y/o seleccionar el mejor modelo.

  Siddiqi (2006) describe diferentes maneras de realizar la división, sin embargo, normalmente el 70% hasta el 80% de la base es utilizada como la muestra de entrenamiento y/o desarrollo del modelo, y el 30% restante al 20% se utiliza como la muestra de testeo con la que se probará y/o validará de forma independiente el modelo.

- **Análisis de variables:**

  Antes de construir el modelo se debe explorar las variables de la muestra de desarrollo, para luego, en base a estos análisis seleccionar aquellas variables que puedan ser incluidas en el modelo. Consta de tres etapas:

1. **Análisis univariado de variables:**

   Este análisis se enfoca en el diagnóstico de la variable, respecto a su distribución, estadísticas de resumen y variabilidad, así como la presencia de valores perdidos y/o valores extremos con el objeto de detectar variables débiles o ilógicas. Los aspectos a evaluar en esta etapa son:

- **Concentración de registros:**

  Tanto para el caso de variables continúas como variables categóricas, se debe verificar que un intervalo o categoría respectivamente no concentre una cantidad considerable de registros, el máximo aconsejable es de 95%, sin embargo las variables que presenten esta condición deberán ser evaluadas cuidadosamente en la etapa de análisis bivariado.

- **Valores perdidos:** La mayoría de los datos de la industria financiera contienen valores que no tienen sentido para una característica en particular, éstos pueden ser campos que no fueron capturados, no estaban disponibles o no fueron completados por los solicitantes. Como en el caso anterior, el máximo aconsejable es de 95% de valores perdidos, sin embargo las variables que presentan esta condición deberán ser evaluadas cuidadosamente en la etapa de análisis bivariado.

- **Valores atípicos:** Son valores que están fuera del rango normal de valor para una determinada variable. Estos valores si no son analizados pueden afectar negativamente al modelo, por lo general son excluidos del análisis, sin embargo, en algunos casos estos podrían ser reemplazados por alguna cota cuando la variable es continua mientras que, cuando la variable es categórica estos valores podrían formar parte de alguna de las categorías. En todos los casos estos valores deben ser investigados previamente.

  En el caso de que se decida fijar cotas, esta se deberá hacer mediante un análisis visual de la distribución de la variable por percentiles y el nivel de riesgo para cada

corte, la cota será el valor del percentil que incluya los valores atípicos y no genere sesgo en los niveles de riesgo.

2. **Análisis bivariado de variables:**

   Este análisis se enfoca en el diagnóstico de la relación existente entre las variables predictivas y la definición de incumplimiento. Para lo cual se debe realizar las siguientes tareas:

- **Agrupación de Variables:** En su literatura Anderson (2007) recomienda agrupar y discretizar las variables categóricas y cuantitativas en función al ratio de *default* respectivamente, ya que ofrece las siguientes ventajas:
  - Ofrece una manera más fácil de lidiar y tratar los valores atípicos y valores perdidos.
  - Facilita la comprensión de las relaciones entre las variables predictoras y la variable respuesta.
  - Permite al usuario desarrollar información sobre el comportamiento de las variables predictivas de riesgo y aumentar el conocimiento de la cartera, que puede ayudar en el desarrollo de mejores estrategias para la gestión de esta.
- **Evaluación de Variables:** Según Naeem Sidiqi (2006), una vez agrupada las variables se deben evaluar los siguientes criterios:
  - **Poder Predictivo de cada Categoría:** Para este propósito se utiliza el estadístico WOE (Peso de la Categoría) que mide la fuerza de cada atributo en la separación de los clientes que hicieron *default* y los que no. Se calcula mediante la siguiente fórmula:

$$WOE = ln\left(\frac{PropNoDefault}{PropDefault}\right)$$

    Los valores negativos implican que el atributo en particular está aislando una mayor proporción de clientes que hicieron *default* de los que no.

  - Poder Predictivo de la Variable: La principal estadística que mide la capacidad predictiva de la variable es el *Information Value* (IV): Conocido como "Poder Total de la Variable", su forma de cálculo incluye al estadístico WOE.

$$\sum_{i=1}^{n}(PropNoDefault - PropDefault) * ln\left(\frac{PropNoDefault}{PropDefault}\right)$$

En base a este estadístico, Siddiqi (2006) clasifica a las variables:

Tabla 1: Clasificación de variables por nivel de predicción

Intervalo de IV	Nivel de Predicción
< 2%	No Predictora
[2%-10%>	Débil
[10%-30%>	Medio
[30%-50%>	Fuerte
>= 50%	Posible Sobre Predictora

FUENTE: Elaboración propia

- Tendencia y Sentido de Negocio: El poder predictivo de las variables no es el único factor de elección de éstas, ya que también se debe analizar la monotonía del ratio de *default* en función del orden lógico de la variable predictora. Asimismo, se debe evaluar que el sentido de la variable sea lógico con el negocio.

- **Transformación de Variables:** Según Anderson (2007), una vez realizada la selección de variables candidatas, estas deben ser transformadas, existiendo dos formas:

  - **Variables Duumy:** También llamadas variables ficticias, consiste en la creación de variables binarias para todos menos un atributo de cada variable.
  - **Variables Sustitutas de Riesgo:** El otro enfoque consiste en la creación de una sola variable transformada para cada una de las variables originales, con el fin de capturar el riesgo representado por cada una de las categorías. De esta forma se crea una relación lineal entre la variable predictora y la definición de incumplimiento. Existen diferentes formas de crear estas variables: la primera es

usar los WOEs como variable sustituta, la segunda es usar los porcentajes de clientes que no hicieron *default* por cada atributo y la tercera es seleccionar una transformación numérica de la variable (lineal, exponencial, etc.).

3. Análisis de correlación de variables:

En muchos casos, algunas variables estarán altamente correlacionadas entre sí, especialmente aquellas calculadas con los mismos insumos base o similares, o variables calculadas para diferentes periodos de tiempo usando la misma variable subyacente. Esto da lugar a una potencial multicolinealidad, que puede llevar a un rendimiento deficiente del modelo.

Para tal efecto, se realizó el análisis de correlación calculando los coeficientes de correlación de Pearson para todas las combinaciones de variables WOE.

$$r_{xy} = \frac{\sum x_i y_i - n\bar{x}y}{(n - s_y 1)s_x}$$

Posteriormente al cálculo de los coeficientes de correlación, las variables serán agrupadas en familias, en función de los estadísticos de predictibilidad y correlación.

d. Modelado

En esta fase se seleccionan las técnicas de modelado más apropiadas para el proyecto, en función de los siguientes criterios:

- Ser apropiada al problema.
- Disponer de los datos adecuados.
- Cumplir con los requisitos del problema.
- Tiempo adecuado para la construcción del modelo.

Para la selección de la técnica se debe considerar el objetivo principal del proyecto y la relación con la información disponible.

Para los proyectos de *credit scoring* la técnica supervisada de clasificación más utilizada es la Regresión Logística (RL), que Hosmer & Lemeshow (2000) la describe como una técnica que en base a un conjunto de variables predice la probabilidad de ocurrencia de un evento específico. Su notación simplificada está definida por:

$$\pi(x) = E(Y/x)$$

Mientras que su notación específica es:

$$(x) = \frac{e^{B_0+B_{1x}}}{1 + e^{B_0+B_{1x}}}$$

El procedimiento que se emplea para su estimación es el de máxima verosimilitud, un método de carácter iterativo que da solución tras varios pasos, asimismo para la estimación de los parámetros es necesario hacer uso de la función de enlace *logit*, cuya finalidad es el de linealizar la probabilidad y limitar el resultado de ésta a un intervalo de cero a uno.

La ecuación transformada es:

$$(x) = ln\,[\frac{\pi(x)}{1 - \pi(x)}]$$

$$(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

Donde:

p : Probabilidad de que un evento suceda.

x : Variables predictoras.

β_0 : Intercepto.

β_0: Parámetros.

Las principales ventajas de la Regresión Logística.

- **Plataformas de implementación:** Debido a la metodología de estimación de parámetros de la regresión logística es posible la construcción de una tabla de puntuación, lo que permite la implementación del modelo en el negocio.
- **Interpretabilidad de los resultados:** Los parámetros estimados por la Regresión Logística son de fácil interpretación por los colaboradores que no se encuentren especializados en la metodología y/o técnica.
- **Capacidad para realizar el seguimiento y diagnóstico del rendimiento del modelo:** Debido a la metodología de estimación del modelo es factible la réplica del modelo y el cálculo de los indicadores de poder predictivo para los seguimientos pertinentes del rendimiento del modelo.

e. Evaluación

En esta fase se debe evaluar los principales supuestos que la técnica estadística de modelado debe cumplir para ser implementada. Para el caso de la Regresión Logística se debe realizar las siguientes pruebas de bondad de ajuste que permiten comprobar cómo de bueno es el ajuste de los valores predichos por el modelo a los valores reales.

- **Evaluación de la Significancia del Modelo:**

En esta etapa se debe evaluar tanto la significancia del modelo global como la de los coeficientes de cada variable que compone el modelo:

- **Test de la Razón de Verosimilitud:** Este test tiene como objetivo comparar cuán mejor es un modelo logístico respecto a otro modelo logístico, en este caso como se desea evaluar la significancia del modelo global se debe comparar éste frente al modelo reducido que sería un modelo aleatorio sin la inclusión de variable alguna.

  La hipótesis nula establece que los coeficientes del modelo global son cero, mientras que el coeficiente del modelo reducido es diferente a cero:

$$Ho: B_{i\ldots n} = 0$$
$$H1: Al\ menos\ un\ B_{i\ldots t} \neq 0 \qquad para\ i = 1 \ldots hasta\ t$$

Donde t es el número de variables que componen el modelo global.
El estadístico a evaluar se calcula mediante la siguiente fórmula:

$$G = -2ln\left[\frac{Verosimilitud\ del\ modelo\ reducido\ (L_0)}{Verosimilitud\ del\ modelo\ global\ (L_1)}\right]$$

Donde G tiene una distribución Chi-Cuadrada con n − 1 grados de libertad.

- **Test de Wald:** Esta prueba tiene como objetivo evaluar la significancia de cada una de las variables que compone el modelo, para tal efecto se plantea la siguiente hipótesis por cada coeficiente:

$$Ho: B_i = 0$$
$$Ho: B_i \neq 0 \qquad para\ i = 1\ hasta\ t$$

Para lo cual se debe calcular el estadístico de Wald mediante la siguiente formula:

$$W = \frac{\hat{B}_i}{SB_i}$$

Donde B_i y SB_i son las estimaciones máximo verosímiles de los coeficientes de las variables y de su correspondiente desviación estándar.

Por lo que:

$$W = \frac{\hat{B}_i}{SB_i} \sim (0,1)$$

Lo que es equivalente a:

$$\left(\frac{i}{SB_i}\right)^2 \sim x^2$$

Esta distribución sirve para rechazar o aceptar la hipótesis nula.

- **Evaluación del Ajuste del Modelo:**

  Para determinar si la probabilidad estimada por el modelo tiene buen ajuste respecto al *ratio de default* se puede usar la prueba binomial de Vasicek, que es un contraste estadístico entre la PD y RD, cuya hipótesis nula es:

$$H_o: PD = RD$$

  Esta prueba asume independencia de eventos, es decir no existe correlación en el evento de *default* de diferentes individuos. Vasicek se basa en la prueba binomial y el *default* de Merton para construir su prueba estadística. Merton asume un modelo de un periodo y define como Rial rendimiento al final del periodo de un activo que un deudor posee. Si este rendimiento es menor a un umbral establecido (por ejemplo la cuota del crédito), el deudor entraría en *default*. Asimismo supone que Ri depende de un factor sistémico X (factor común a todos los deudores) y de un factor específico de cada deudor ei. Se tienen los siguientes supuestos adicionales:

  - $Ri \sim N(0,1)$
  - $X \sim N(0,1)$
  - $ei \sim N(0,1)$
  - $cov(X, ei) = 0$
  - $cov(ei, ej) = 0$, $i \neq j$
  - $cor(Ri, Rj) = \rho$
  - El tamaño de muestra es infinito.
  - El umbral establecido es el mismo para todos los deudores: (γ)
  - La correlación del evento de *default* entre todos los pares de deudores es la misma.

Bajo estos supuestos, el rendimiento de un activo puede expresarse como:

$$R_i = \sqrt{1-\rho e_i} + \sqrt{\rho X}$$

Un deudor entraría en *default* sí: $R_i < \gamma = \Phi^{-1}(PD)$
Donde Φ−1 denota la inversa de la función normal estándar acumulada. Incluyendo estos supuestos, la prueba binomial de Vasicek es:

$$(PD \leq z) = \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(z) - \Phi^{-1}(PD)}{\sqrt{\rho}}\right)$$

En base a esta prueba se pueden definir intervalos de aceptación según lo requerido por la entidad financiera en base a la probabilidad de incumplimiento, y evaluar si el ratio de *default* se encuentra dentro de los límites óptimos.

- **Evaluación de la Bondad de Ajuste del Modelo por Eficacia Predictiva:**
  En esta etapa se debe evaluar la capacidad discriminatoria del modelo, para lo cual se deben evaluar los siguientes indicadores que miden el poder predictivo:

  – **Indicador de Kolmogorov – Smirnov:** El objetivo de este test es verificar si dos muestras tienen la misma distribución, se define como la máxima diferencia absoluta entre las funciones de distribución acumuladas de ambas muestras. En este caso, será la distribución de los clientes que han hecho *default* versus la distribución de los clientes que no han hecho *default*.

$$F_{de}(a) = \frac{1}{n}\sum_{i=1}^{n} I(si \leq a \text{ A } Default = 1) \forall a \epsilon [L, H]$$

$$F_{node}(a) = \frac{1}{m}\sum_{i=1}^{n} I(si \leq a \text{ A } Default = 0) \forall a \epsilon [L, H]$$

Donde n es el total de clientes que han hecho *default* y m el total de clientes que no han hecho *default*, así como L y H son los valores mínimos y máximos de probabilidad respectivamente que estima el modelo.

El estadístico de Kolmogorov-Smirnov se estima:

$$KS = \max |F_{de}(a) - F_{nodef}(a)|$$

Según Mays (2004), los valores de KS van del 20%, por debajo del cual se debe cuestionar la capacidad predictiva del modelo, al 70%, por encima del cual podría existir sobreajuste. En tal sentido el valor mínimo establecido por la entidad financiera es de 40%.

- **Índice de Gini:** De igual forma que el estadístico de Kolmogorov-Smirnov, el objetivo de esta prueba es medir la desigualdad entre dos poblaciones, en este caso será la población de clientes que hicieron *default* versus la población de clientes que no hicieron *default*.

  Gráficamente, si las distribuciones son diferentes, la curva de Lorenz (rojo) estará más alejada de la diagonal, caso contrario la curva se aproximará a la diagonal.

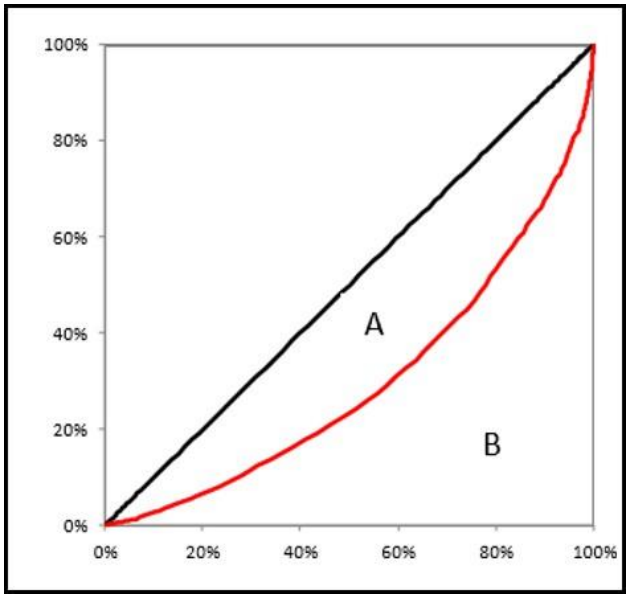


Figura 2*:* Índice de Gini

FUENTE: Elaboración propia

Por lo que el índice de Gini:

$$Gini = \frac{A}{A+B} = 2A$$

Con los resultados de las muestras, se calcula:

$$Gini = 1 - \sum_{i=1}^{n} (F_{de}(a) - F_{def}(a-1))\,(F_{def}(a) - F_{def}(a-1))$$

Para Siddiqi (2006) un índice de Gini igual o mayor a 50% es más que satisfactorio, mientras que un valor menor a 30% es posiblemente inaceptable.

En tal sentido el valor mínimo establecido por la entidad financiera para un modelo es de 50%.

– **Curva ROC (Receiver Operating Characteristics):** Es un gráfico en el que se representa la tasa de clasificación correcta (sensibilidad) en función de la tasa de error (1-especificidad).

La curva ROC ofrece un adecuado resumen de la capacidad predictiva del modelo, ya que presenta la potencia predictiva para todos los posibles valores de la probabilidad estimada.

Para determinar si el modelo a evaluar es adecuado, la curva de ROC correspondiente debe ubicarse próximo a la esquina izquierda superior, mientras que si se encuentra debajo de la diagonal el modelo sería deficiente.

En la Figura 3 se observa que el modelo evaluado tendría una buena capacidad predictiva.

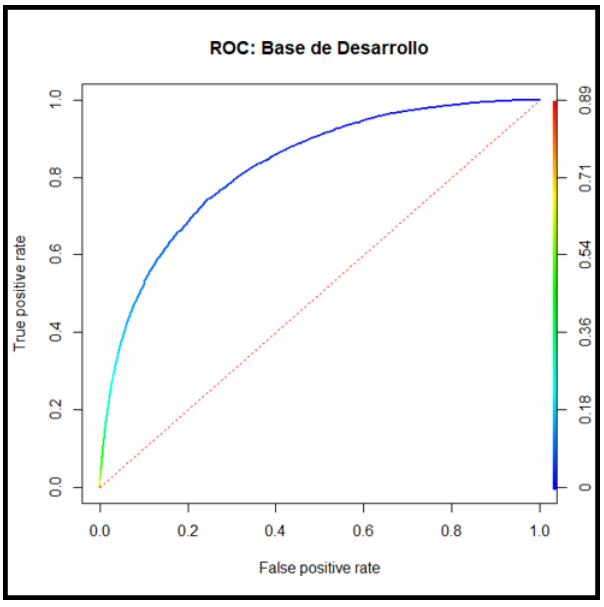


Figura 3: Curva ROC

FUENTE: Elaboración Propia

El área debajo de la curva definida como AUC ROC representa el grado de discriminación del modelo, cuánto mayor sea su valor mejor será el modelo para predecir el *default* del cliente.

Por lo que el valor mínimo permitido en la entidad financiera de AUC ROC es de 75%.

f. Implementación:

Una vez definido el modelo a usar, este debe ser implementado en los diferentes sistemas de la entidad. Por lo que, al ser un modelo de *credit scoring* se debe generar la regla de transformación de la probabilidad estimada a un *score*. Este proceso no afecta el poder predictivo del modelo y sus principales ventajas son:

- Permite replicar el modelo en plataformas diferentes a las usadas en el proceso de estimación de parámetros.
- Su formato es sencillo de interpretar por gestores de riesgo sin conocimientos avanzados de estadística o minería de datos
- Su proceso de generación es ampliamente entendido al no ser una caja negra, por lo tanto, cumple con cualquier normativa respecto a la transparencia del método.

Existen diferentes métodos de generar esta transformación, según Siddiqi (2006) la más utilizada es el uso de una escala logarítmica, que se basa en la transformación lineal del *odds*, ventaja de que el cliente sea bueno respecto a que sea malo, también definido como la relación de la cantidad de clientes que no han *default* respecto a la cantidad de clientes que han hecho *default*.

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

Donde el *score* es calculado usando un *odds* específico definido previamente, la cantidad de puntos necesarios para que el *odds* se duplique (pdo) y el *score* a obtener. El factor y el offset pueden ser calculados usando las siguientes fórmulas simultáneamente:

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

$$Score + pdo = Offset + Factor * \ln(odds)$$

Donde:

$Factor = pdo/\ln(2)$

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

Haciendo uso del Peso de Evidencia (WOE), la función para el *score* es:

$$score - (\sum_{j,i}^{k,n} WOE_j * \beta + \frac{\beta_0}{n}) * factor + offset$$

Donde:

WOE: Peso de evidencia para cada atributo.

β_i: Parámetro de cada variable.

β_0: Intercepto del modelo.

n: Número de variables.

k: Número de atributos por cada variable.

Luego de haber generado la regla de transformación, se debe definir los puntos de cortes para obtener las segmentaciones de riesgo, ya que para Rahal & Mungai (2015) la selección de un puntaje de corte es probablemente la decisión más importante que debe tomar una entidad financiera que desea implementar un sistema de calificación crediticia recientemente desarrollado.

4.2.3. Propuesta de alternativa de solución a la situación problemática

A continuación se ha de presentar los resultados de los análisis correspondientes a las diferentes fases de la metodología CRISP DM.

4.2.3.1. Comprensión del negocio

Se realizaron las reuniones correspondientes con las diferentes áreas de la entidad financiera, aplicando la metodología canvas, definiendo los principales ejes del proyecto:

a. Objetivo del negocio:

La entidad financiera requiere identificar cuáles de sus clientes tienen un buen comportamiento crediticio tanto con la entidad como con otras entidades del sistema financiero, ya que las áreas de productos y negocios están replanteando las estrategias comerciales de venta de productos crediticios y fidelización de sus mejores clientes, dando mayor esfuerzo a la oferta de incremento de líneas de tarjeta de crédito, disposición de efectivo por tarjeta de crédito, tarjetas adicionales, disminución de tasas entre otros.

b. Estado actual del negocio:

La entidad financiera cuenta con un modelo de comportamiento construido en el 2014 que es utilizado para la segmentación de riesgo de toda la cartera de la entidad financiera. Sin embargo cuando éste es evaluado solo en los clientes que cuentan con tarjeta de crédito, tanto los indicadores de poder predictivo y calibración son muy bajos, no alcanzando los umbrales mínimos establecidos por el gobierno de modelos de la entidad financiera. Asimismo, con el transcurso del tiempo, las políticas crediticias de otorgamiento de tarjetas de crédito han sufrido modificaciones, esto hace presumir que la población utilizada para la construcción del modelo que se encuentra en producción ha podido variar consideradamente.

c. Objetivo del proyecto:

Como se indicó en los puntos descritos anteriormente, al necesitar identificar los mejores clientes de la entidad financiera para aplicar estrategias que solo se aplicarían a aquellos clientes con tarjeta de crédito, se definió el objetivo del proyecto, siendo: Estimar un modelo de comportamiento que mida el nivel de riesgo de los clientes que cuenten con al menos una tarjeta de crédito en la entidad financiera.

d. Definición del alcance y diseño:

Se determinó en la reuniones con las áreas de riesgos y de negocio, que el modelo resultante sería utilizado para segmentar a aquellos clientes que tengan una o más tarjetas de crédito en los tres meses previos al mes de análisis. Asimismo, se ha de excluir a aquellos clientes que teniendo tarjeta de crédito no tengan saldo adeudado con la entidad financiera en el mes de estudio.

Con respecto al diseño del proyecto, este ha de seguir los pasos descritos en el punto 4.2.1, que fue aprobado juntamente con el Gantt de trabajo por la gerencia de riesgos.

4.2.3.2. Comprensión de los datos:

Como primer paso para esta etapa, se identificaron las fuentes primarías de información:

a. Información financiera:

Información extraída de los Reportes Crediticios Consolidados (RCC) otorgados por la Superintendencia de Banca y Seguros, que muestra, de manera mensual, la relación de los créditos que el cliente ha contratado con las empresas del sistema financiero, así como la calificación asignada por dichas entidades en base a su comportamiento de pago.

b. Información interna:

Fue extraída del datawarehouse de la entidad financiera, esta información engloba los datos referentes al comportamiento transaccional y de pago del cliente con la entidad financiera.

c. Información demográfica:

Información recolectada durante la apertura del crédito, y actualizada por los asesores de negocio en la apertura de otro crédito.

Respecto a la extracción de la información se realizaron los requerimientos correspondientes al área de base de datos y se realizaron los siguientes controles de calidad:

- **Datos perdidos:** Se analizó el porqué de los datos perdidos y se procedió a la codificación correspondiente:

Tabla 2: Codificación de datos perdidos

Tipo de Valor	Codificación
Missing por sistema	999999
Valor erróneo 0/0	111111
Valor erróneo a/0, (a numérico)	222222
Valor erróneo a/missing	333333
Valor erróneo missing/a	444444

FUENTE: Elaboración propia

- **Metadatos erróneos:** Se verificó que los valores almacenados tengan una relación lógica con el nombre del campo al que pertenecen, mediante el siguiente análisis:

Tabla 3: Control de metadatos erróneos

Tipo de Variable	Control
Categórica	Tamaño del campo, Descriptivos
Continua	Análisis de Estadísticas de Posición
Discreta	Descriptivos, verificación de valores negativos

FUENTE: Elaboración propia

4.2.3.3. Preparación de los datos:

- **Estructuración de los datos:**

  Los datos fueron almacenados de forma estructurada en el sistema de gestión de datos SQL Server, para lo cual se generaron los modelos de entidad relación. En la Figura 4 se puede observar de manera simplificada la integración de las fuentes de información:

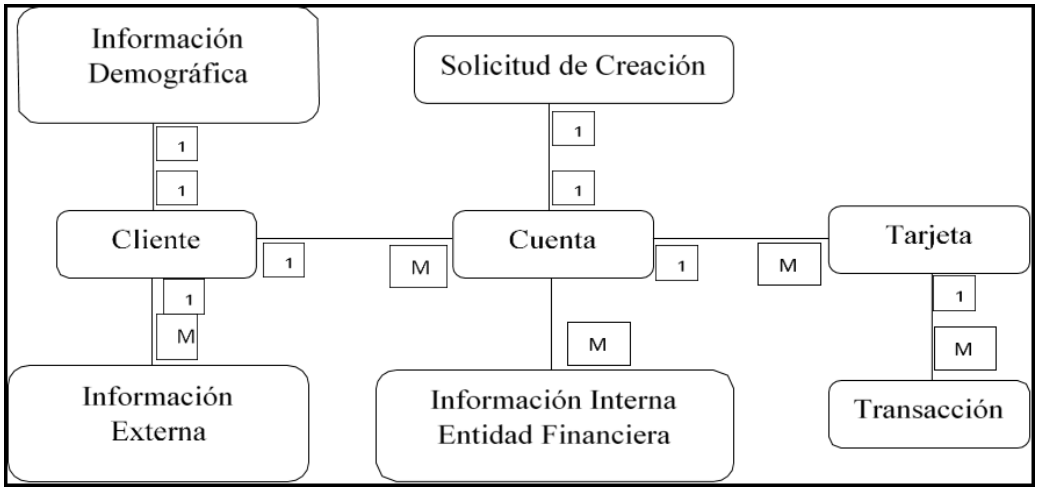


Figura 4: Estructura de los Datos

FUENTE: Elaboración propia

- **Definición de la Población:**

  Como se especificó en la definición del alcance y diseño del proyecto, la población a estudiar corresponde a todos los clientes que cuenten con al menos una tarjeta de crédito de la entidad financiera, sin embargo para la elección de los meses de estudio se analizó la estabilidad del total de clientes y saldo adeudado.

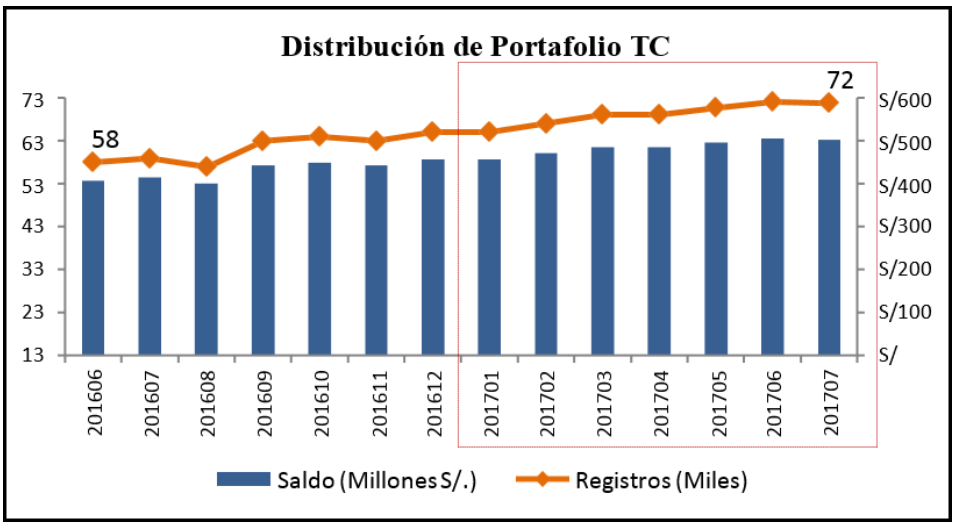


Figura 5: Distribución del Portafolio TC

FUENTE: Elaboración propia

En la Figura 5 se observa el incremento del portafolio de clientes con tarjeta de crédito, tanto en número de clientes como en saldo endeudado por este concepto. Por lo que se decidió trabajar con los periodos correspondientes al 2017.

Asimismo, se realizaron las siguientes exclusiones:

Tabla 4: Exclusiones de la población

Exclusiones	Descripción
Calidad	Número de Tarjetas inválidas, cronogramas errados
Antigüedad	Se excluyen a clientes con menos de 2 meses con saldo en la entidad financiera
Saldo	Clientes con tarjeta de crédito sin utilizarla en los últimos 12 meses
Estado	Clientes Castigados o Refinanciados tanto en la entidad financiera como en todo el sistema financiero

FUENTE: Elaboración propia

La población resultante por periodo:

Tabla 5: Cantidad de clientes a evaluar por periodo

Periodo	Clientes
201701	62,047
201702	62,149
201703	62,156
201704	62,097
201705	62,650
201706	63,099
201707	63,742

FUENTE: Elaboración propia.

- **Definición de la Variable Dependiente:**

  Para la definición de incumplimiento se evaluaron las siguientes propuestas de *default*:

  – Cliente con más de 30 días de atraso en el pago de al menos uno de sus créditos con la entidad financiera.

  – Cliente con más de 60 días de atraso en el pago de al menos uno de sus créditos con la entidad financiera.

– Cliente con más de 90 días de atraso en el pago de al menos uno de sus créditos con la entidad financiera.

Para la elección de algunas de estas definiciones se realizaron los análisis de las curvas acumulativas de *default* y análisis de Roll Rate. Para lo cual, se tomó la base de clientes correspondientes a los meses desde el 2015 hasta julio 2017 y se analizó el mes en el que el cliente hiso el primer *default* (para los tres casos).

En la Figura 6 se muestra la curva acumulativa de *default*, en donde se puede observar que aproximadamente el 80% de los clientes que cayeron en alguna de las definiciones de *default* lo hicieron durante los primeros 12 meses posteriores al mes de análisis.

Por lo tanto, la ventana de desempeño será de 12 meses.

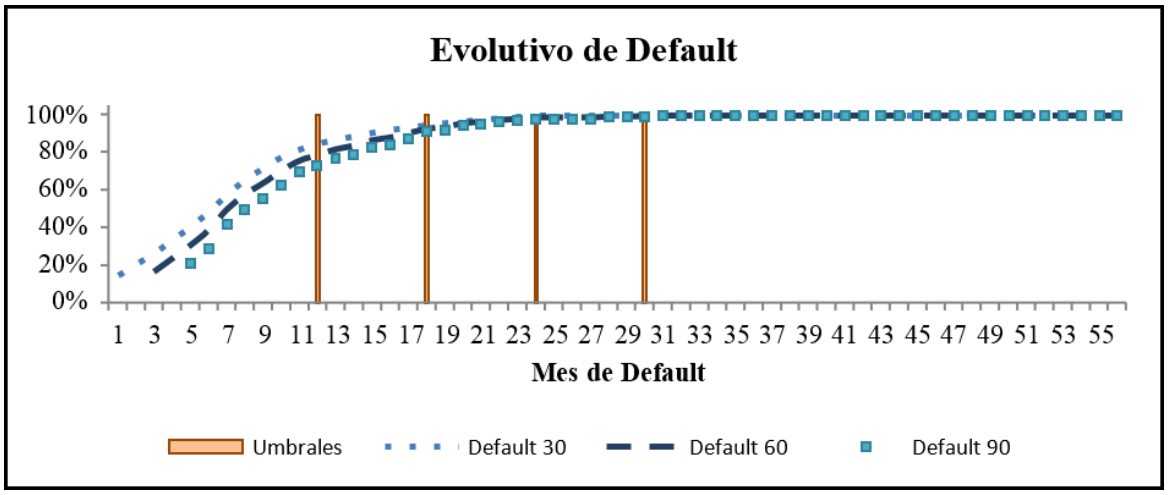


Figura 6: Evolutivo de *Default*

FUENTE: Elaboración propia.

Asimismo, para la construcción de las matrices de *Roll Rate* se consideraron las dos primeras propuestas de *default* y se evaluaron sus migraciones a una definición de *default* mayor a 120 días, ya que según la clasificación regulatoria si algún crédito del cliente cae en este *default* es considerado como un crédito perdido.

Tabla 6: Matriz de Roll Rate

Tramos de Mora en T0	Tramos de Mora a 12 Meses						Total
	0.[0]	0.[1-8]	1.[9-30]	2.[31-60]	3.[61-120]	4.>120	
0.[0]	67%	7%	13%	4%	4%	6%	100%
0.[1-8]	12%	27%	19%	12%	10%	20%	100%
1.[9-30]	19%	2%	27%	18%	15%	18%	100%
2.[31-60]	13%	4%	7%	11%	27%	39%	100%
3.[61-120]	7%	1%	8%	2%	12%	71%	100%
4.>120	4%	0%	0%	0%	0%	95%	100%

FUENTE: Elaboración propia

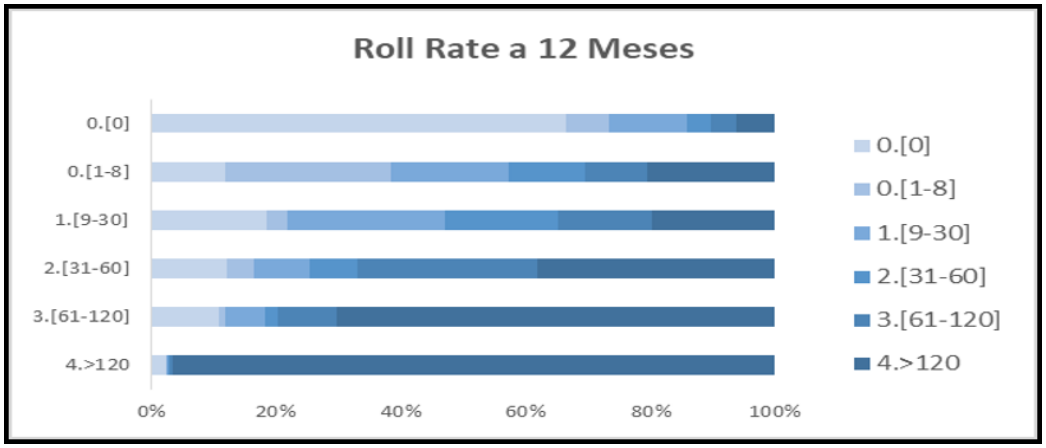


Figura 7: Roll Rate a 12 meses.

FUENTE: Elaboración propia.

Tanto en la Tabla 6 como en la Figura 7 se puede observar que más del 60% de los clientes que tenían inicialmente más de 60 días de mora evolucionaron al tramo de pérdida (más de 120 días de atraso), mientras que solo el 6% de los clientes que se encontraban al día en el periodo de estudio evolucionaron a pérdida.

En base a estos análisis se definió como variable respuesta, que un cliente habrá hecho *default* cuando incumpla en el pago de alguno de sus créditos en más de 60 días en alguno de los 12 meses posteriores al mes de referencia. Asimismo, se excluyó a aquellos clientes que meses anteriores al periodo de análisis hicieron *Roll Rate*. En la Figura 8, se observa la estabilidad del ratio de *default* en los meses de análisis.

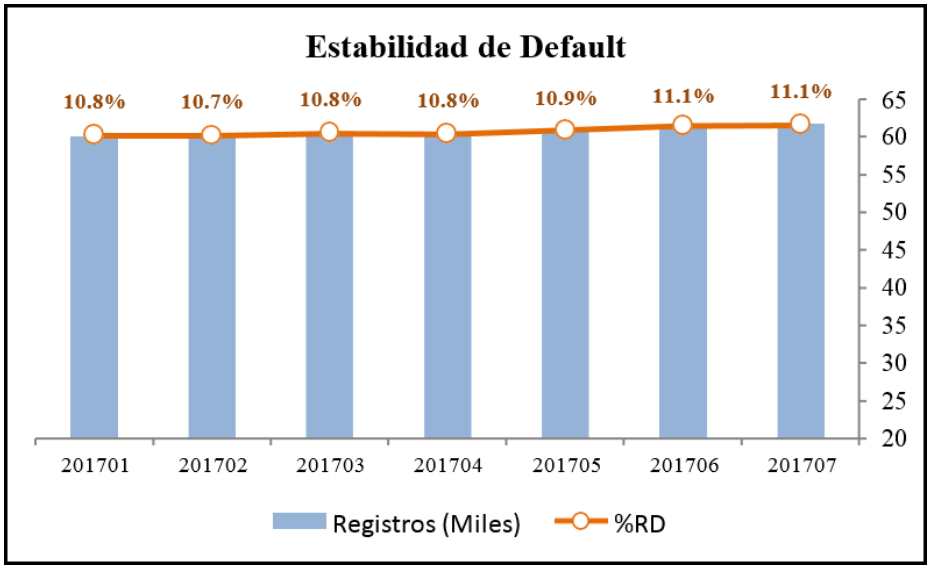


Figura 8: Estabilidad de *default*

FUENTE: Elaboración propia

- **Segmentación de la Población:**

  Para segmentar la población se evaluó el siguiente eje de segmentación:

  Nivel de atraso: El objetivo de este eje fue el de categorizar al cliente en Cliente Sin Atraso y Cliente Con Atraso, para lo cual se construyó la variable máximo de días de atraso en los último 12 meses, y se evaluó las diferentes casuísticas de segmentación, por ejemplo el cliente será considerado como "Sin Atraso" si tiene cero días de mora y en caso contrario se le considerará como "Con Atraso", de la misma forma la relaciones de un día de mora a más y así sucesivamente.

  En base a estos análisis y en conjunto con la gerencia de riesgo se decidió segmentar a la población en función a su nivel de atraso, teniendo como subpoblaciones:

  – Clientes sin atraso: A este segmento pertenecen los clientes que en los 12 meses anteriores al mes de análisis han tenido como máximo ocho días de atraso en alguno de sus créditos.
  – Clientes con atraso: A este segmento pertenecen los clientes que en al menos uno de los 12 meses anteriores al mes de análisis han tenido más de ocho días de atraso en alguno de sus créditos.

En la Figura 9 se observa la estabilidad de la participación de cada segmento:

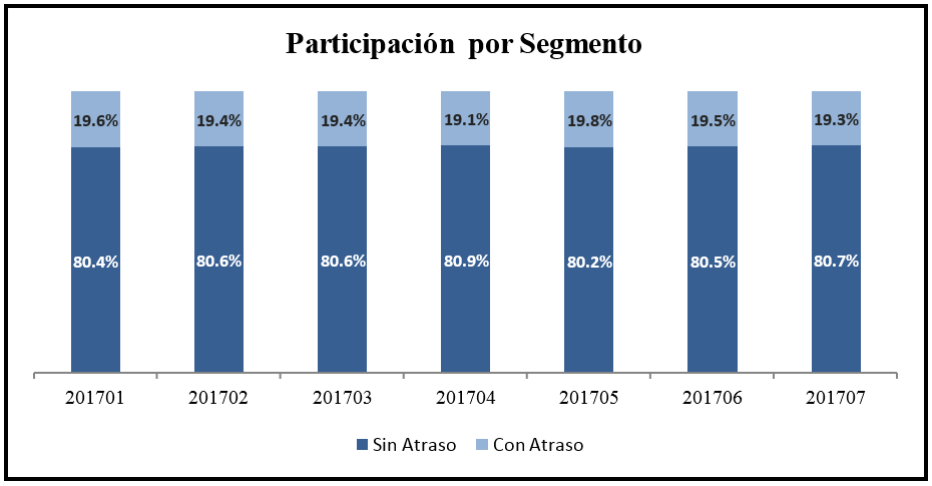


Figura 9: Participación por segmento

FUENTE: Elaboración propia.

Asimismo, en la Figura 10 se verifica la estabilidad del ratio de *default* por segmento:

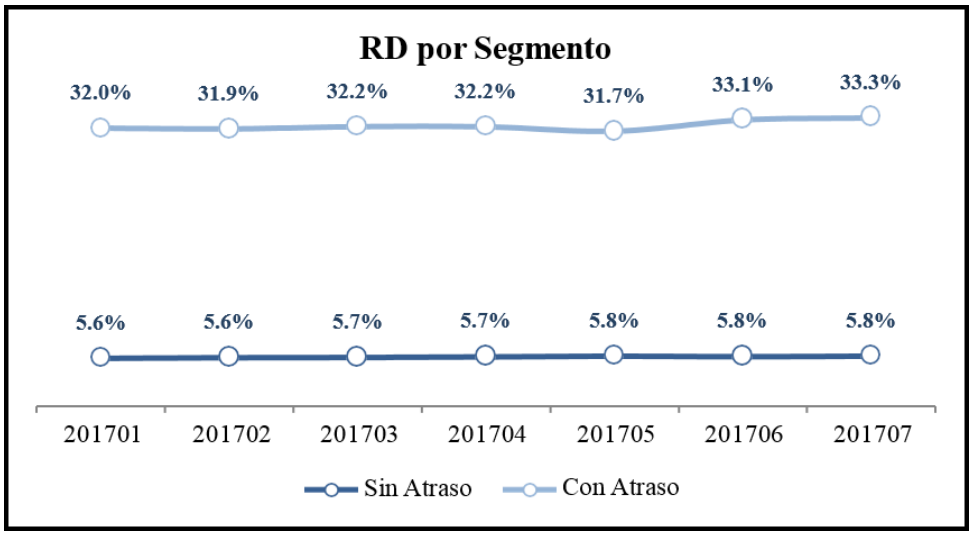


Figura 10: Ratio de *Default* por Segmento

FUENTE: Elaboración propia.

En la presente memoria se desarrollará la construcción del modelo correspondiente al segmento Sin Atraso, ya que los clientes pertenecientes al segmento Con Atraso serán gestionados por políticas diferentes de la gestión del riesgo.

- **Partición de la base:**

  Se dividió la base en: Base de desarrollo con el 75% de clientes del total de la población y base de testeo con el 25% restante. En la Tabla 7 se observa la estabilidad del ratio de *default* en ambas muestras:

Tabla 7: Partición de la base de modelamiento.

Base	Total	*Default*	%RD
Desarrollo	256,145	14,643	5.7%
Test	85,381	4,881	5.7%
Total	341,526	19,524	5.7%

FUENTE: Elaboración propia

- **Análisis de variables:**

  En la Tabla 8 se describen las familias de variables identificadas:

Tabla 8: Análisis de variables

Familia	Descripción
Solicitud	Variables obtenidas al momento de otorgar la Tarjeta de Crédito, tales como tipo de empleo, ingresos, entre otras.
Demográfica	Describen las características demográficas del cliente tales como Edad, Género y Ubigeo.
Endeudamiento	Miden el nivel endeudamiento del cliente en función al saldo.
Exposición	Miden el nivel de exposición del cliente, tales como las Líneas de Tarjeta de Crédito y Productos crediticios con los que cuenta elcliente.
Comportamiento	Describen el comportamiento de pago e incrementos de deuda del cliente.
Morosidad	Miden el nivel de morosidad del cliente.
Transaccionales	Describen las transacciones realizadas con la tarjeta de crédito, tales como medios de pago, establecimientos utilizados, entre otras

FUENTE: Elaboración propia.

1. **Análisis univariado de variables:**

   En esta etapa se aplicaron los criterios descritos en el punto 4.2.1 para el análisis univariado de variables. En la Tabla 9 se muestran la estadísticas de tendentica central, dispersión y de posición de las tres variables continuas que quedaron en el modelo final, así como de la variable discreta máximo atraso en los últimos seis meses.

Tabla 9: Análisis univariado de variables

Estadístico	Promedio de Utilización de Línea TC	Máximo Atraso	Proporción de Deuda en el último mes	Proporción de Deuda Revolvente
Observaciones	256,145	256,145	256,145	256,145
Missing Values	0	0	0	0
Mínimo	0	0	0	0
Máximo	1	8	1	1
Media	0.6	1.5	0.8	0.7
Desviación	0.25	2.12	0.19	0.32
P1	0.09	0	0.24	0.04
P2	0.1	0	0.31	0.06
P3	0.12	0	0.37	0.08
P4	0.15	0	0.4	0.11
P95	1	6	1	1
P96	1	7	1	1
P97	1	7	1	1
P98	1	8	1	1
P99	1	8	1	1
P100	1	8	1	1

FUENTE: Elaboración propia.

Para una mejor interpretación de los resultados en la Tabla 9, a continuación se muestran los gráficos descriptivos de estas variables:

- **Promedio utilización de línea TC en los últimos 12 meses:** Esta variable mide el uso de la tarjeta de crédito en el último año, se observa que más del 50% de clientes ha utilizado más del 60% de su tarjeta en este lapso de tiempo, asimismo se verifica la ausencia de valores atípicos.

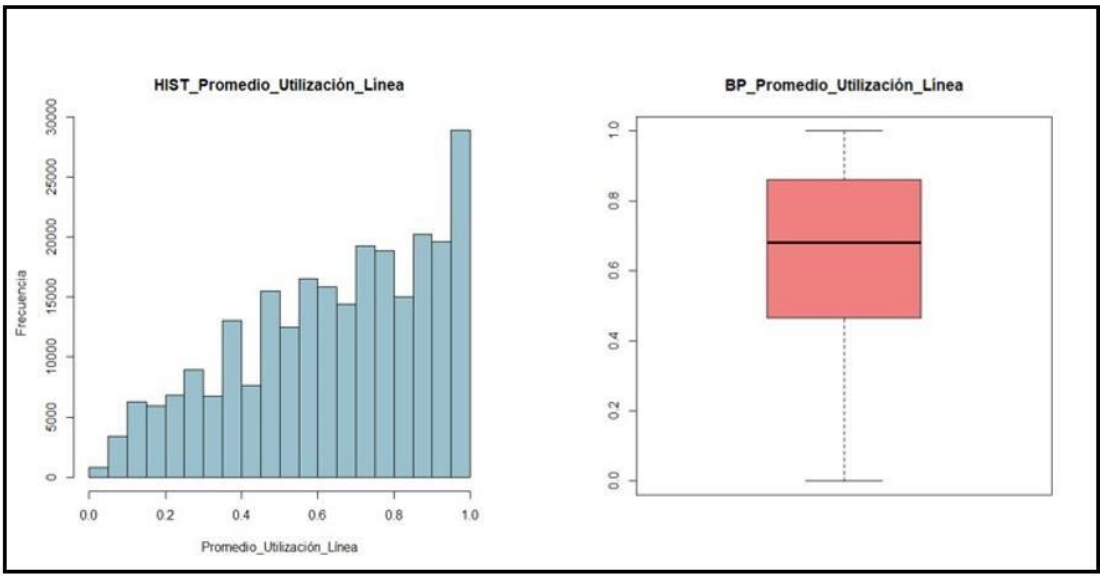


Figura 11: Univariado Promedio utilización de línea TC en los últimos 12 meses

FUENTE: Elaboración propia.

- **Máximo atraso en los últimos 6 meses:** Esta variable mide el número de días máximo que el cliente estuvo atrasado en al menos uno de sus créditos con la entidad financiera en los últimos 6 meses. En los Figuras se observa que la mayoría de los clientes no ha presentado atraso y que en promedio los clientes se atrasan no más a 2 días. Asimismo, se observa la presencia de valores atípicos, sin embargo se evaluará su tratamiento en el análisis bivariado.

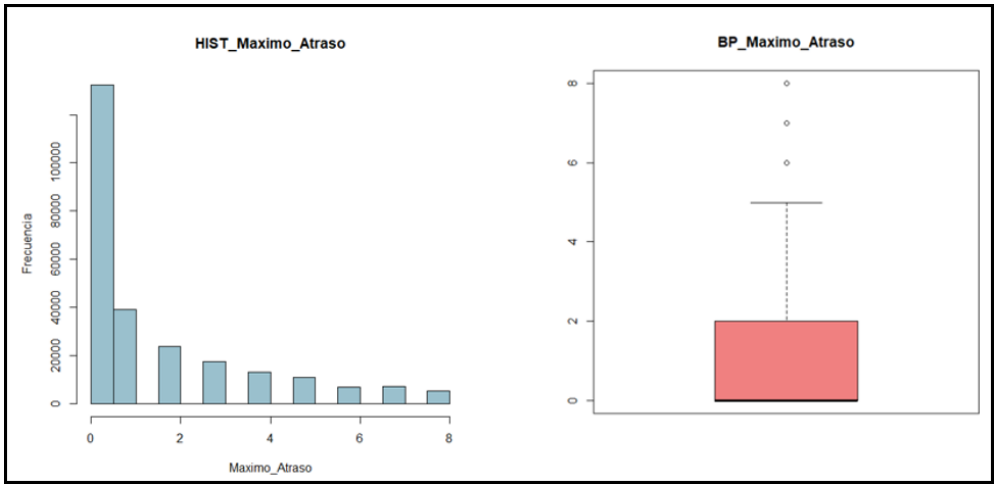


Figura 12: Univariado Máximo atraso.

FUENTE: Elaboración propia.

- **Deuda del último mes respecto al máximo de deuda de los últimos 12 meses:** Esta variable mide la proporción que representa en saldo total en el último mes respecto al máximo saldo en el último año, en las siguientes graficas se observa que la deuda de la mayoría de clientes representa más del 60% del saldo máximo del último año. Asimismo, se verifica la presencia de valores outliers, por lo que se ha fijado una cota de 0.4, siendo este valor el percentil cuatro.

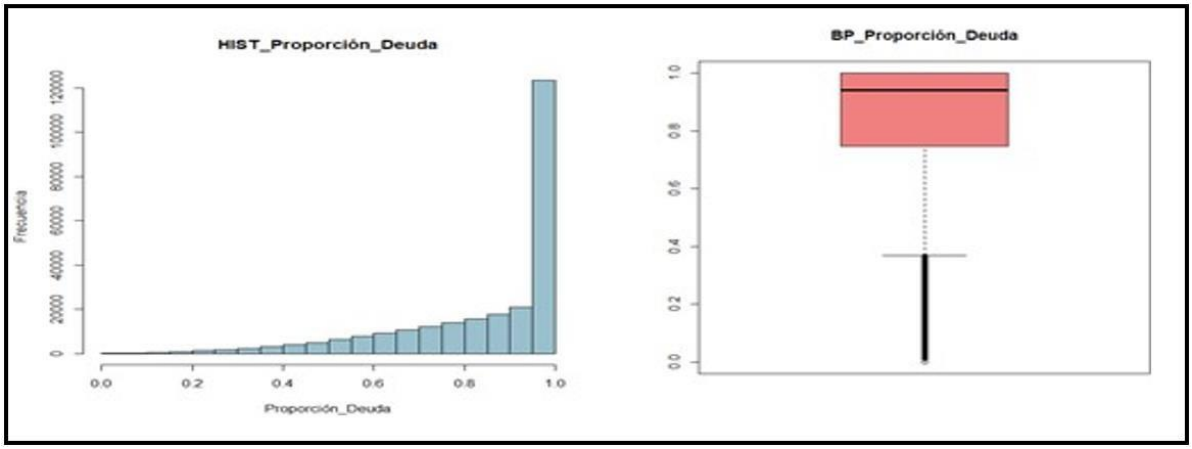


Figura 13: Deuda del último mes respecto al máximo de deuda de los últimos 12 meses

FUENTE: Elaboración propia.

- **Proporción de deuda revolvente respecto al total de deuda en los últimos 3 meses:** Esta variable mide la proporción que representa el saldo de consumo

revolvente (créditos cuyas cuotas no son fijas, existiendo el concepto de cuota mínima) en el último mes respecto al saldo total en el últimos tres meses.

En las siguientes graficas se observa que el saldo revolvente de la mayoría de clientes representa su saldo endeudado total. Asimismo, se verifica la ausencia de outliers.

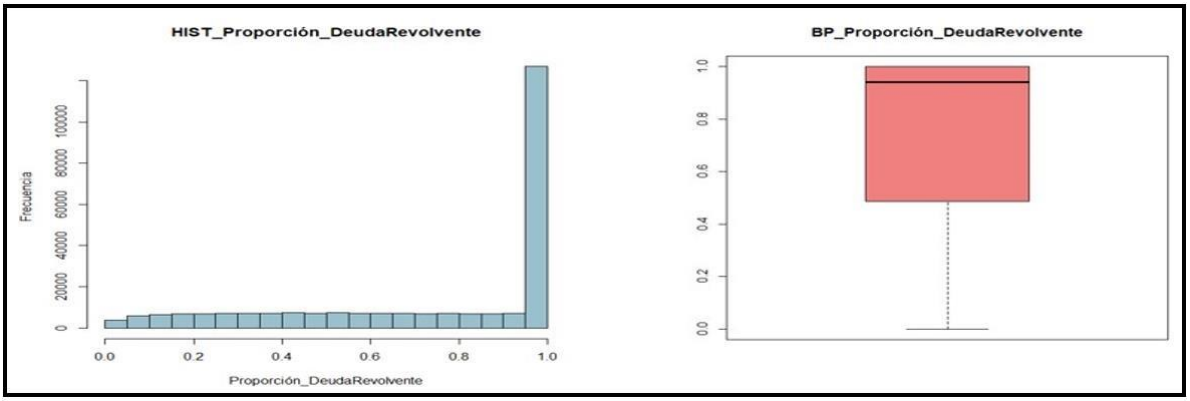


Figura 14: Proporción de Deuda Revolvente respecto al total de Deuda en los últimos 3 meses

FUENTE: Elaboración propia.

Con respecto a las variables categóricas y discretas, estas fueron tratadas en el análisis bivariado.

2. Análisis bivariado de variables:

En este proyecto, la agrupación de las variables dependiendo de sus características se trabajó de la siguiente forma:

- Las variables cuantitativas (continuas y discretas) se utilizó la técnica de Intervalos óptimos ofrecida por la herramienta estadística SPSS.
- Los valores perdidos se trataron como una categoría y fueron reagrupados con las categorías con un *default* similar.

A continuación se ha de describir los hallazgos de las variables tratadas y transformadas que componen el modelo final. En la Tabla 10 se detallan los indicadores de discriminación de las variables finales del modelo, siendo sus valores los adecuados.

Tabla 10: Análisis de variables

Variable	IV	KS	GINI
Promedio de Utilización de Línea TC	17%	16%	20%
Máximo Atraso en los últimos 6 meses	28%	22%	29%
Deuda del último mes respecto al Máximo de Deuda 12 Meses	41%	21%	27%
Número de Decrementos de Deuda	28%	23%	28%
Tipo de Ingresos	53%	33%	33%
Número Incrementos Consecutivos de Disposición de Efectivo	12%	15%	17%
Máxima Calificación	29%	17%	17%
Proporción de Deuda Revolvente respecto al Total	15%	18%	18%

FUENTE: Elaboración propia.

Para la inclusión de las variables al modelo, cada categoría fue reemplazada por el valor de WOE correspondiente. En las siguientes figuras se verifica la tendencia y monotonía de la relación de las variables categorizadas con el ratio de *default*, así como de la variable transformada.

- **Promedio de utilización de línea TC en los últimos 12 meses:**

  Como se observa en la Figura 15, esta variable presenta una relación positiva con el ratio de *default*, lo que indica que a mayor uso de la Línea de la Tarjeta de Crédito aumenta la probabilidad de *default*. Mientras que en la Figura 16 se observa que la variable transformada presenta una relación negativa con el ratio de *default*.

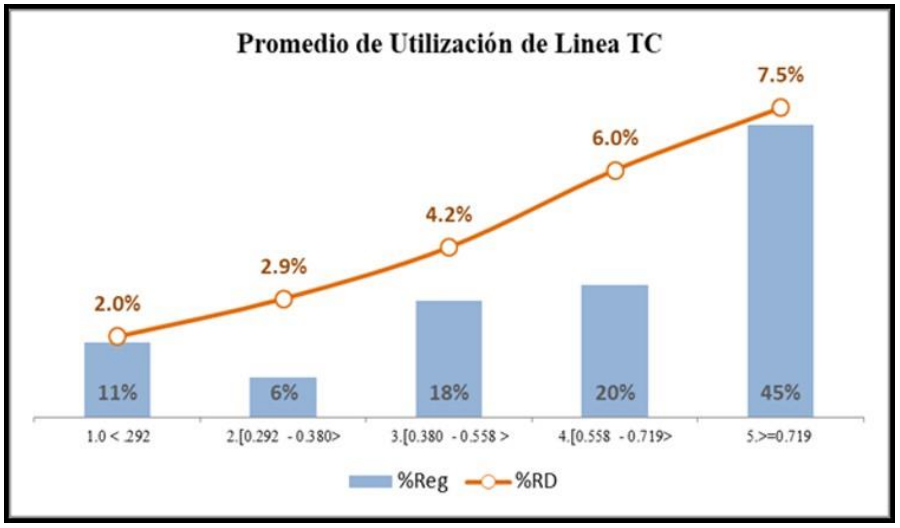


Figura 15: Tendencia del Promedio de Utilización de Línea TC en los últimos 12 meses

FUENTE: Elaboración Propia.

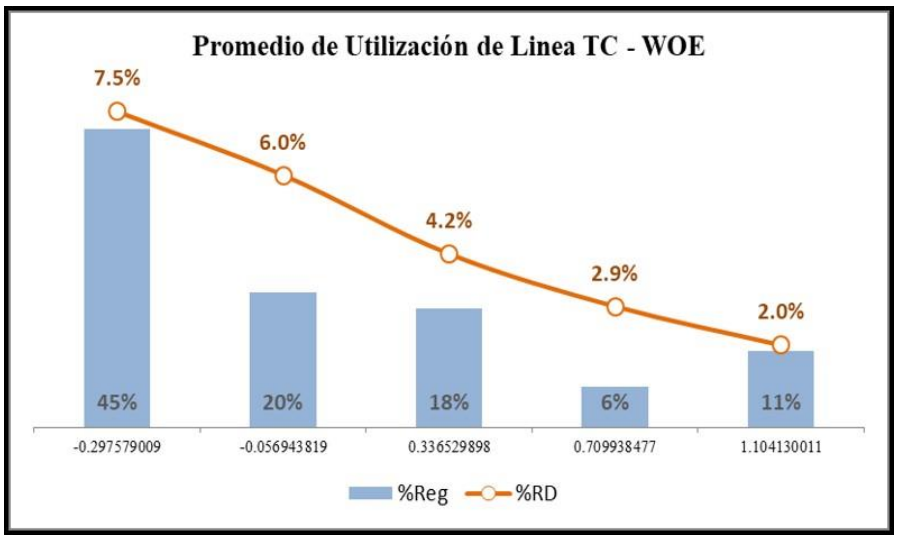


Figura 16: Tendencia del Promedio de Utilización de Línea TC en los últimos 12 meses WOE

FUENTE: Elaboración propia.

- **Máximo atraso en los últimos 6 meses:**

  Como se puede observar en la Figura 17, el atraso en los últimos seis meses presenta una relación positiva con el ratio de *default*, lo que indica que, cuando más días de atraso tenga el cliente mayor será su probabilidad de *default*. Mientras que la variable transformada presenta una relación negativa con el ratio de *default* evidenciada en la Figura 18.

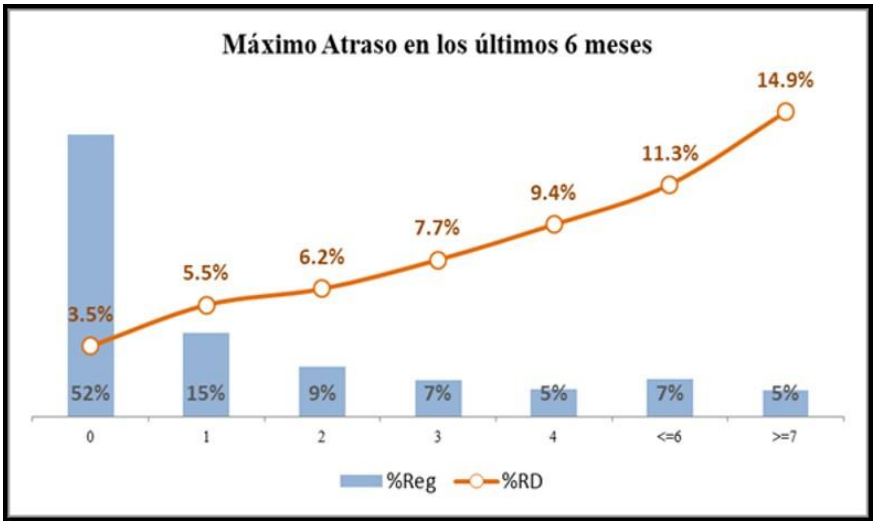


Figura 17: Tendencia del Máximo atraso en los últimos 6 meses

FUENTE: Elaboración propia.

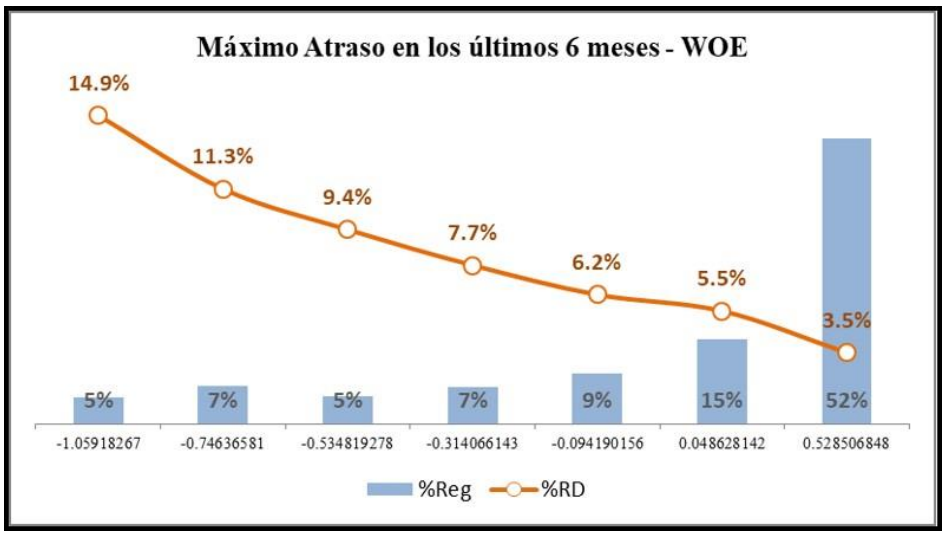


Figura 18: Tendencia del Máximo atraso en los últimos 6 meses WOE

FUENTE: Elaboración propia.

- **Deuda del último mes respecto al máximo de deuda de los últimos 12 meses:** Como se observa en la Figura 19 esta variable presenta una relación positiva con el ratio de *default*, lo que indica que cuan más cercano el valor del saldo en el último mes respecto al máximo saldo de los últimos 12 meses, la probabilidad de *default* es mayor. Mientras que la variable transformada presenta una relación negativa con el ratio de *default* mostrada en la Figura 20.

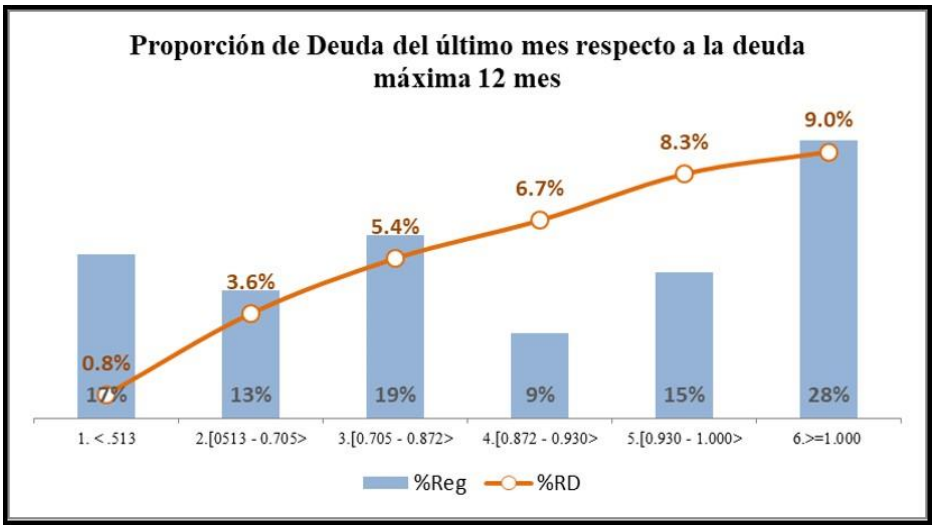


Figura 19: Tendencia de la Deuda del último mes respecto al Máximo de Deuda de los últimos 12 meses

FUENTE: Elaboración Propia

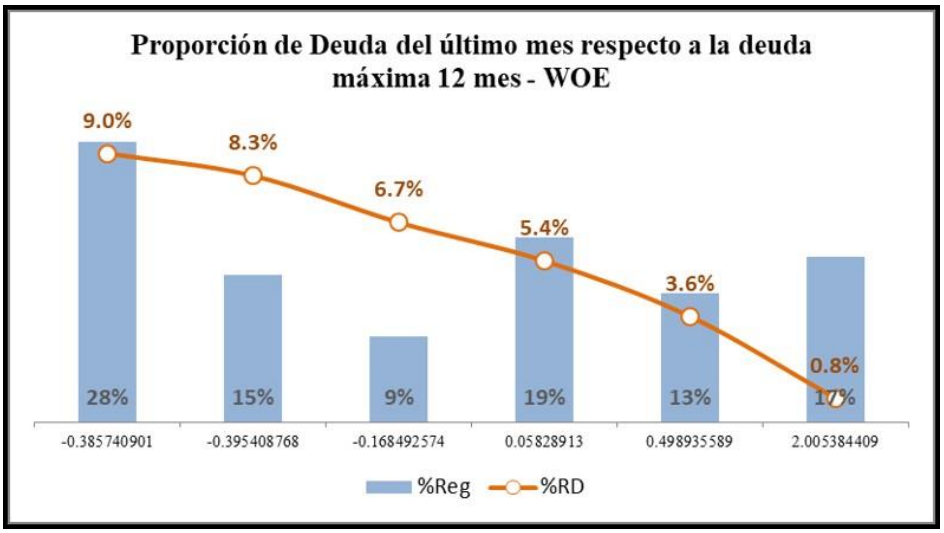


Figura 20: Tendencia de la Deuda del último mes respecto al Máximo de Deuda de los últimos 12 meses WOE

FUENTE: Elaboración Propia

- **Número de decrementos de deuda:**

  Como se observa en la Figura 21, esta variable presenta una relación negativa con el ratio de *default*, lo que indica que mientras más veces el cliente disminuya su saldo respecto al mes anterior su probabilidad de *default* será menor, de la misma forma sucede con la variable transformada mostrada en la Figura 22.

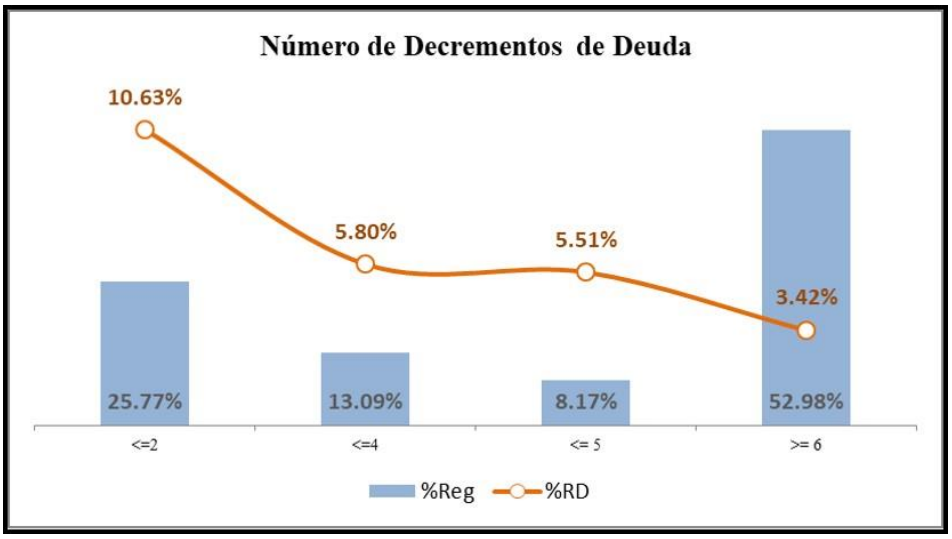


Figura 21: Tendencia Número de decrementos de Deuda.

FUENTE: Elaboración propia.

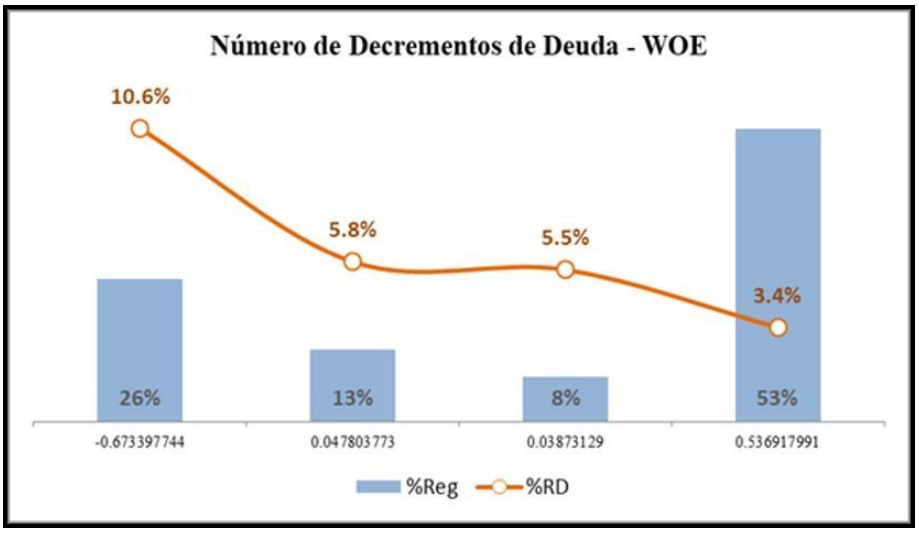


Figura 22: Tendencia Número de decremento de deuda WOE

FUENTE: Elaboración propia

- **Tipo de ingresos:**

  Como se observa en la Figura 23 los clientes que se encuentran inscritos en planilla tienen una probabilidad menor de incumplimiento en comparación de aquellos clientes que no se encuentren en planilla mostrada en la Figura 24.

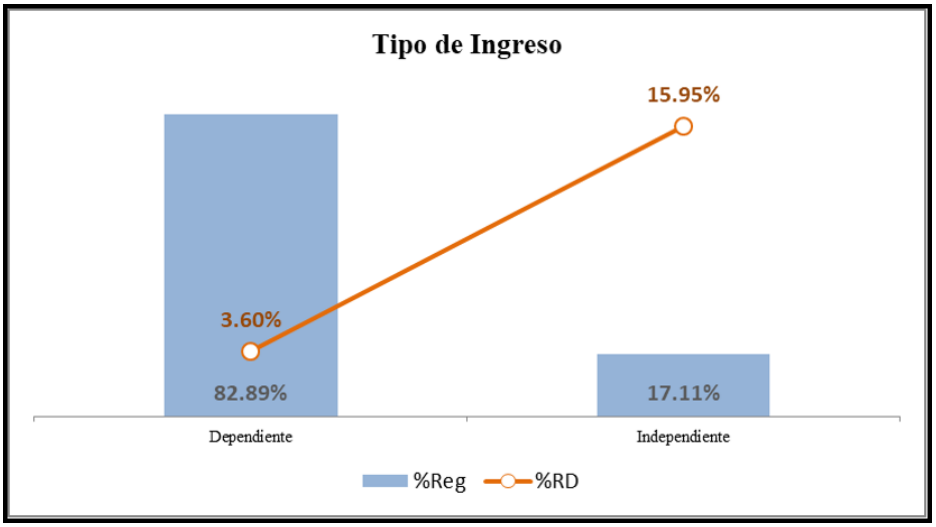


Figura 23: Tendencia Tipo de ingreso.

FUENTE: Elaboración propia.

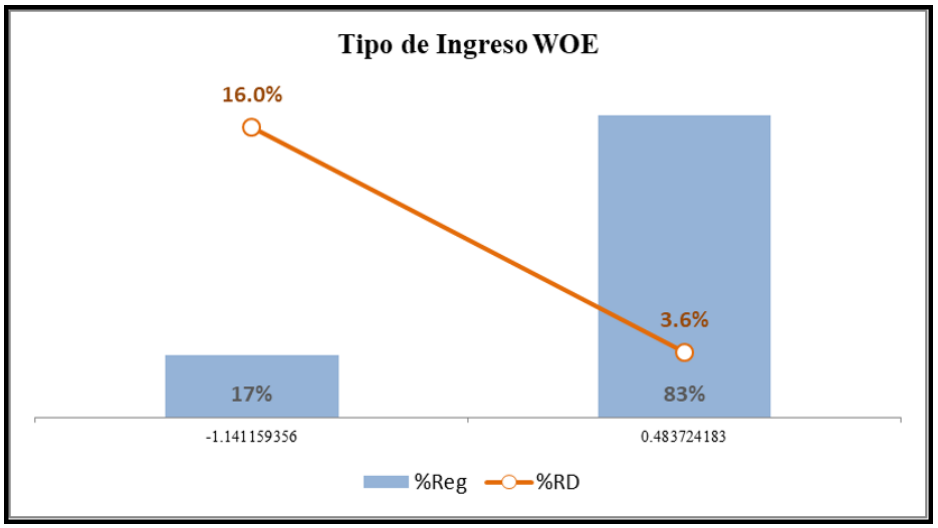


Figura 24: Tendencia Tipo de ingreso WOE

FUENTE: Elaboración propia.

- **Número de incrementos consecutivos de disposición de efectivo:**

  Como se observa en la Figura 25, esta variable tiene una relación positiva con el ratio de *default*, lo que indica que, los clientes que incrementan de manera consecutiva y recurrente su saldo por concepto de disposición de efectivo tienen mayor probabilidad de incumplimiento. Mientras que la variable transformada presenta una relación negativa con el ratio de *default* mostrada en la Figura 26.

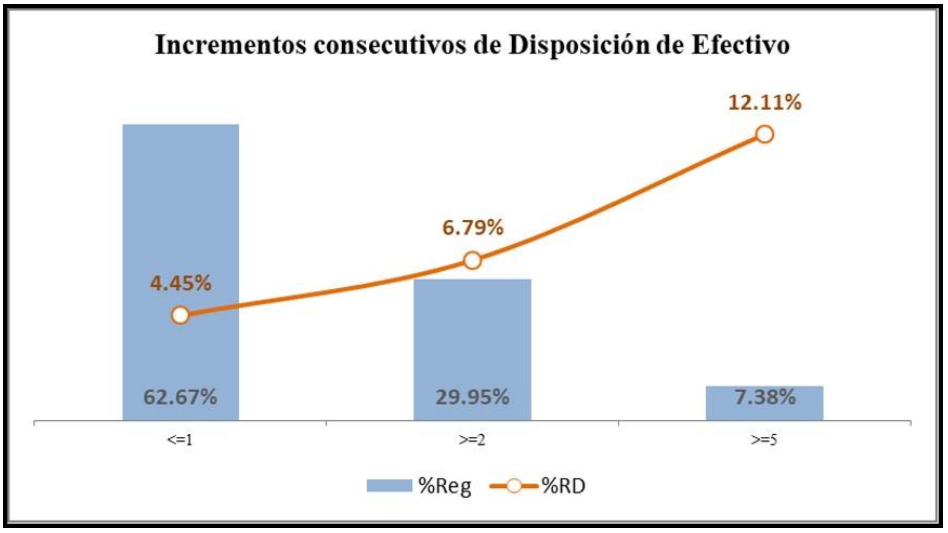


Figura 25: Tendencia de los Incrementos consecutivos de Disposición de Efectivo

FUENTE: Elaboración propia.

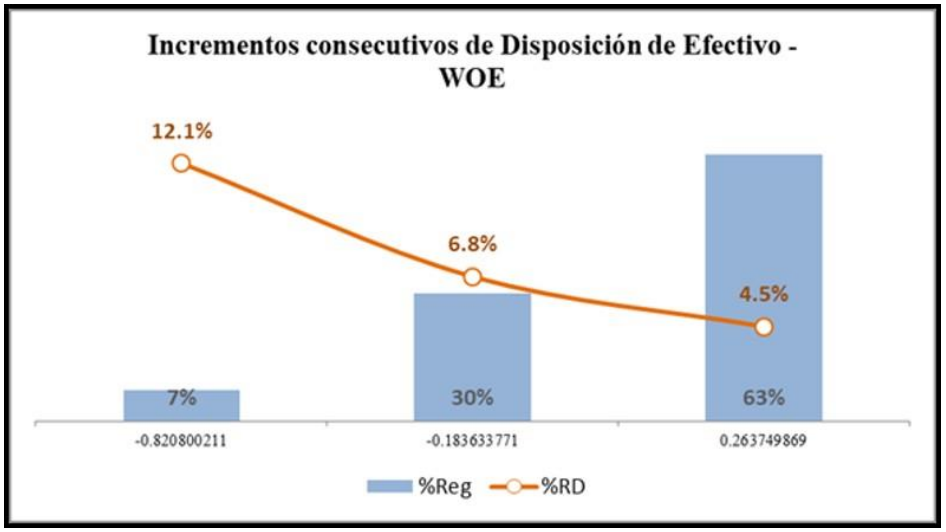


Figura 26: Tendencia de los Incrementos consecutivos de Disposición de Efectivo WOE

FUENTE: Elaboración propia

- **Máxima calificación en el sistema financiero:**

  Como se observa en la Figura 27, los clientes que tienen una calificación en el sistema financiero igual a Normal tienen menor probabilidad de *default* que los clientes que presentan una calificación diferente mostrada en la Figura 28.

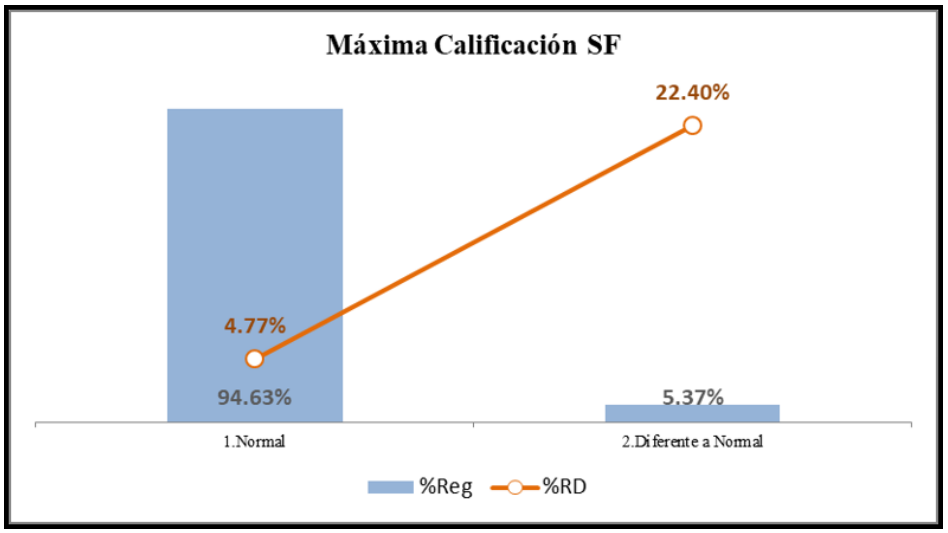


Figura 27: Tendencia Máxima calificación SF

FUENTE: Elaboración propia.

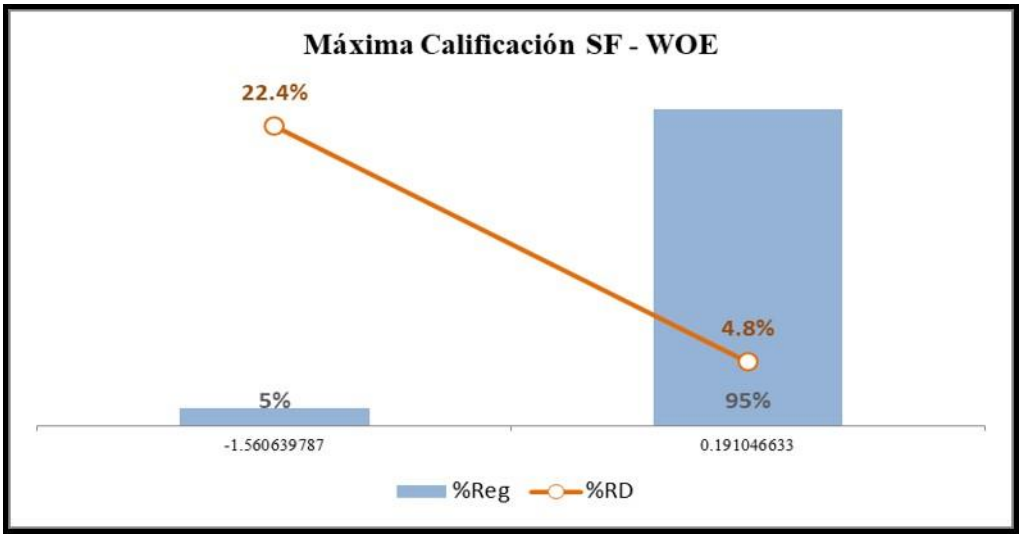


Figura 28: Tendencia Máxima calificación SF WOE

FUENTE: Elaboración propia.

- **Proporción de Deuda Revolvente respecto al total de Deuda en los últimos 3 meses:** Como se observa en la Figura 29, esta variable a diferencia de las variables anteriores presenta una relación cóncava con el ratio de *default*. Lo que indica que, los clientes que presentan una participación de saldo revolvente menor a 10% tienen menor riesgo que aquellos clientes cuya participación se encuentra entre el 10% y 80%, mientras que aquellos clientes que tienen una participación mayor al 80% su probabilidad de incumplimiento es menor.

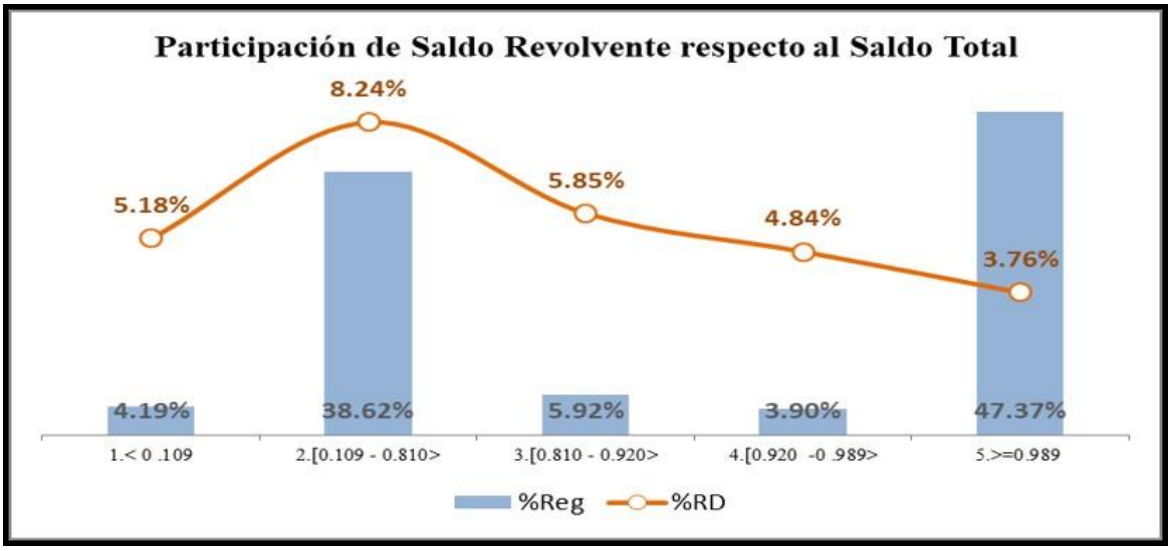


Figura 29: Tendencia de la Participación de saldo revolvente respecto al saldo total

FUENTE: Elaboración propia.

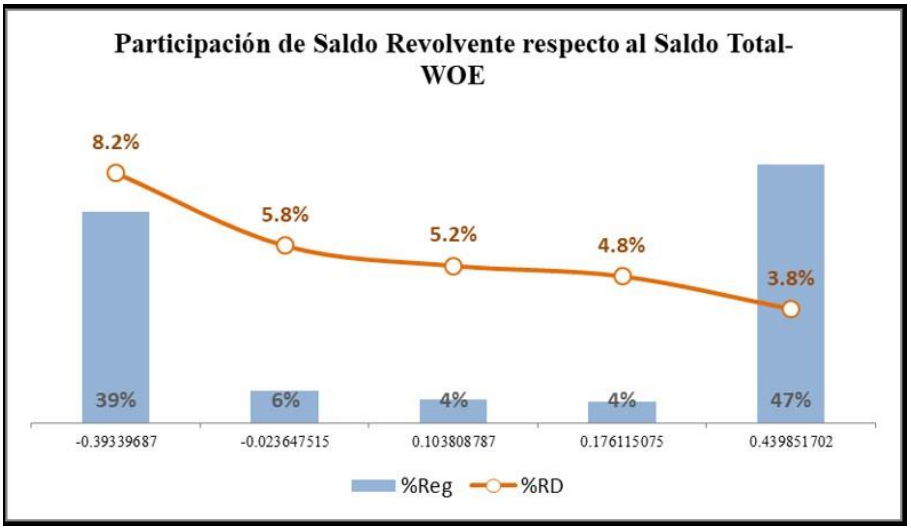


Figura 30: Tendencia de la Participación de saldo revolvente respecto al saldo total WOE

FUENTE: Elaboración propia.

3. Análisis de correlación de variables:

Para este estudio, se construyó la matriz de correlación de las variables transformadas y en función a esta se formaron los grupos de variables que se utilizaron en la etapa de modelado.

A continuación se describe el proceso de agrupamiento de variables:

1. Se elige la variable con mayor correlación con el target.
2. Se construye la familia identificando las variables con correlaciones desde 0.30 con la variable obtenida en el punto 1.
3. Repetir el paso 1 y 2 excluyendo a las variables obtenidas en los puntos anteriores.

A continuación se muestra los resultados de correlación de las variables finales del modelo, donde se verifica que ningún coeficiente es mayor a 0.6, por lo tanto no existe correlación entre las variables.

Tabla 11: Resultados de correlación de las variables finales del modelo

	Default	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**
Default	1	-0.21	-0.2	-0.17	-0.13	-0.13	-0.11	-0.09	-0.09
V1	-0.21	1	0.43	0.05	0.21	0.1	0.05	0.12	-0.06
V2	-0.2	0.23	1	0.04	0.17	0.21	0.06	0.21	-0.02
V3	-0.17	0.05	0.04	1	0.14	0.03	0.04	0.03	-0.01
V4	-0.13	0.21	0.17	0.14	1	0.01	0.01	-0.18	-0.02
V5	-0.13	0.1	0.21	0.03	0.01	1	0.03	0.08	-0.06
V6	-0.11	0.05	0.06	0.04	0.01	0.03	1	0.09	0.06
V7	-0.09	0.12	0.21	0.03	-0.18	0.08	0.09	1	0.03
V8	-0.09	-0.06	-0.02	-0.01	-0.02	-0.06	0.06	0.03	1

FUENTE: Elaboración propia

4.2.3.4. Modelado

Se construyeron aproximadamente 1000 modelos con la base de desarrollo, los criterios de selección del mejor modelo fueron:

- Capacidad Predictiva del Modelo
- Ajuste del Modelo
- Granularidad del Modelo
- Adecuada Distribución de los Pesos de las Variables

Tabla 12: Resultados del modelo

Fuente	Familia	Variable	Coeficiente	Peso
		Intercepto	-2.81541	
Interna	Exposición	Promedio de Utilización de Línea TC	-1.35429	19.00%
Interna	Morosidad	Máximo Atraso en los últimos 6 meses	-0.82082	17.40%
Externa	Endeudamiento	Deuda del último mes respecto al Máximo de Deuda 12 Meses	-0.86807	14.20%
Externa	Comportamiento	Número de Decrementos de Deuda	-0.80865	13.70%
Solicitud	Solicitud	Tipo de Ingresos	-0.53522	12.00%
Interna	Comportamiento	Número Incrementos Consecutivos de Disposición de Efectivo	-0.86218	8.80%
Externa	Morosidad	Máxima Calificación	-0.50645	8.70%
Externa	Endeudamiento	Proporción de Deuda Revolvente respecto al Total	-0.73488	6.10%

FUENTE: Elaboración propia

Como se observa en la Tabla 12, todos los coeficientes son negativos, esto debido a que se construyó el modelo con las variables transformadas a WOE, que tienen una relación monotónica decreciente con la definición de incumplimiento.

Asimismo, se verifica una adecuada distribución de los pesos de las variables, siendo el Promedio de la Utilización de Línea, el Máximo Atraso en los últimos seis meses y la Proporción de Deuda en el último mes respecto a los últimos 12 meses las tres variables más influyentes del modelo. De igual forma, se observa en las Tablas 13 y 14, la adecuada distribución de pesos de las familias de variables y fuentes de información.

Tabla 13: Pesos de las familias de variables

Familia	Peso
Morosidad	26%
Comportamiento	23%
Endeudamiento	20%
Exposición	19%
Solicitud	12%

FUENTE: Elaboración propia

Tabla 14: Fuentes de información

Fuente	Peso
Interna	45%
Externa	43%
Solicitud	12%

FUENTE: Elaboración propia.

4.2.3.5. Evaluación:

En la Tabla 15 se detalla los coeficientes estimados del modelo, y la significancia de los mismos.

Tabla 15: Coeficientes estimados del modelo y su significancia

Variable	Coeficiente	P-Value	Wald	Peso
Intercepto	-2.81541	0.000	73285.3	
Promedio de Utilización de Línea TC	-1.35429	0.000	2574.7	19.0%
Máximo Atraso en los últimos 6 meses	-0.82082	0.000	2353.4	17.4%
Deuda del último mes respecto al Máximo de Deuda 12 Meses	-0.86807	0.000	1922.1	14.2%
Número de Decrementos de Deuda	-0.80865	0.000	1856.1	13.7%
Tipo de Ingresos	-0.53522	0.000	1629.2	12.0%
Número Incrementos Consecutivos de Disposición de Efectivo	-0.86218	0.000	1189.4	8.8%
Máxima Calificación	-0.50645	0.000	1178.9	8.7%
Proporción de Deuda Revolvente respecto al Total	-0.73488	0.000	827.9	6.1%

FUENTE: Elaboración propia.

- **Evaluación de la significancia del modelo:**

– **Significancia global de los coeficientes estimados:**

Para lo cual se ha planteado la siguiente hipótesis:

$$Ho: B_{i...n} = 0$$
$$Ho: Al\ menos\ un\ B_{i...n} \neq 0 \quad para\ i = 1; n = 9$$

Tabla 16: Significancia global de los coeficientes estimados

Test	Valor	P-Valor
Razón de Verosimilitud	22130.24	0.00

FUENTE: Elaboración propia

Con un p-valor igual a 0% se verifica estadísticamente que al menos uno de los coeficientes estimados es diferente a cero.

– **Significancia individual de los coeficientes estimados:**

Para lo cual se ha planteado 8 hipótesis, por cada variable:

$$Ho: B_i = 0$$
$$Ho: B_i \neq 0 \quad para\ i = 1; n = 8$$

En la Tabla 15 se detalla el estadístico de Wald y el p-valor asociado a cada variable, ya que todos los p-valores son iguales a cero, se concluye estadísticamente que todos los coeficientes del modelo son diferentes a cero.

– **Evaluación del Ajuste del Modelo:**

Para la evaluación del ajuste de la probabilidad de incumplimiento con el ratio de *default*, se calcularon los intervalos de confianza, en base a los cuales se determinará el estado de calibración del modelo.

Tabla 17: Intervalos de probabilidad de default

Resultado	Límite Inferior	Límite Superior
Bueno	5.12%	6.20%
Aceptable	[4.85%-5.12%]	[6.20%-6.52%]
Crítico	<4.85%	> 6.52%

FUENTE: Elaboración propia.

En la Tabla 17, se observan los Intervalos de Probabilidad de *Default* para la muestra de test, cuyo ratio de *default* es igual a 5.7%, el cuál se encuentra comprendido en el primer intervalo definido como Bueno, por lo tanto el modelo se encuentra calibrado.

En las siguientes Figuras se observa el adecuado ajuste entre la Probabilidad de *Default* Estimada y el Ratio de *Default* por Ventiles de Probabilidad tanto para la base de desarrollo como la base de test. Asimismo, se verifica la granularidad de la probabilidad, ya que se observa claramente el ordenamiento y la monotonía del ratio de *default*, donde solo el 0.2% de los mejores clientes han hecho *default* mientras que casi el 35% de los peores clientes han hecho *default*.

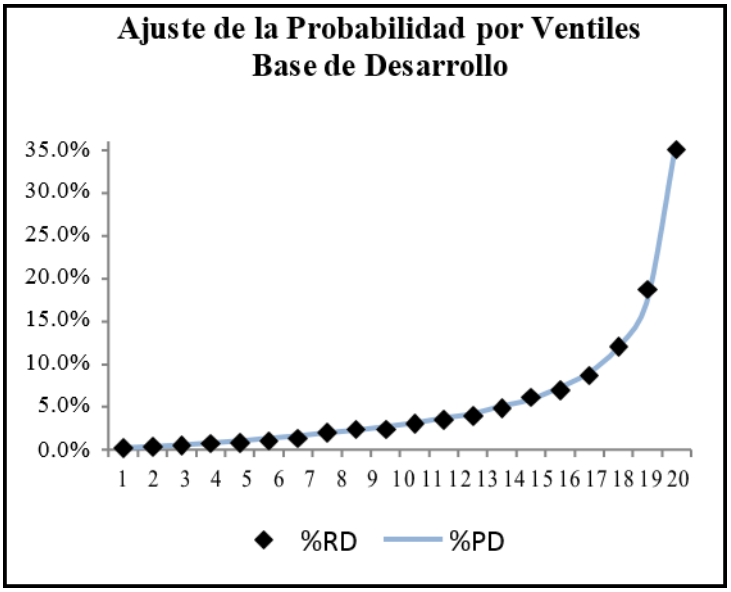


Figura 31: Ajuste de probabilidad base de desarrollo

FUENTE: Elaboración propia

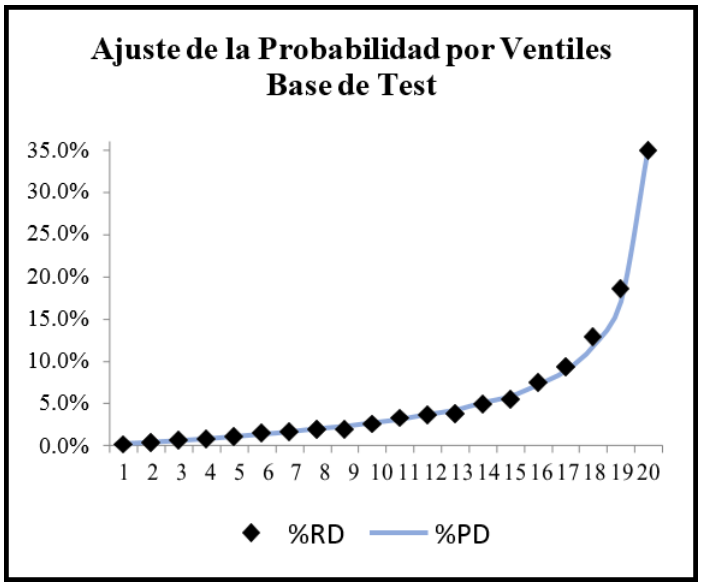


Figura 32: Ajuste de probabilidad por ventiles base de test

FUENTE: Elaboración propia.

- **Evaluación de la bondad de ajuste del modelo: Eficacia predictiva**

  Indicadores de Discriminación: En la Tabla 18 se verifica que, los indicadores de capacidad predictiva del modelo tanto en la base de desarrollo como en la de test se encuentran en los intervalos óptimos de discriminación, así como en los estipulados por la Entidad Financiera.

Tabla 18: Indicadores de discriminación del Modelo

Base	ROC	Gini	KS
Desarrollo	82.7	65.4	49.7
Test	82.8	65.6	50.0

FUENTE: Elaboración propia.

Lo expuesto líneas arriba se ratifica con las siguientes figuras, donde se observa que las curvas ROC son adecuadas.

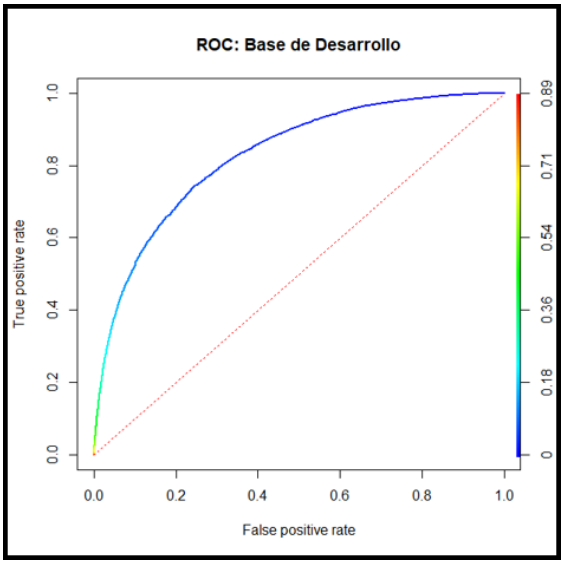


Figura 33: Curva ROC base de desarrollo

FUENTE: Elaboración propia

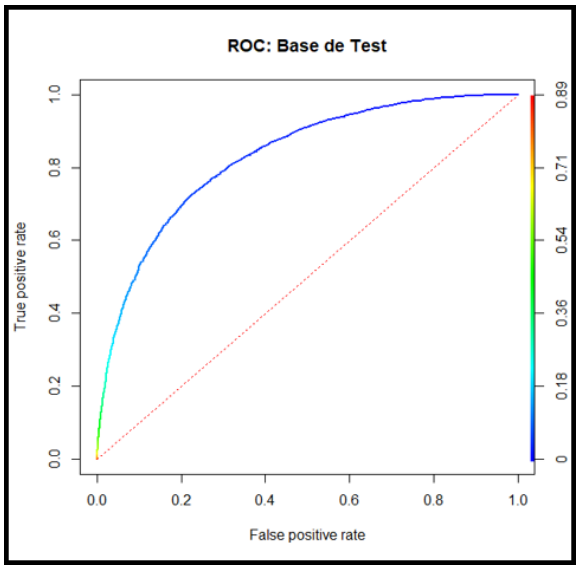


Figura 34: Curva ROC base de test

FUENTE: Elaboración propia

En base a los resultados de las tres evaluaciones; el modelo resultante es adecuado para la estimación del riesgo de los clientes de la entidad financiera que se encuentren en la subpoblación Sin Atraso.

4.2.3.6. Implementación:

En esta etapa se generó el "Plan de Implementación" con las diferentes áreas usuarias, estructurándose en las siguientes etapas.

- **Creación del *Score*:**

  Para la implementación del modelo en los diferentes sistemas de la entidad financiera, se generó la ecuación que transforma la probabilidad de *default* a una escala continua, usando la metodología descrita en la fase de implementación.

  Los criterios utilizados: como odds de 32 a 1, un pdo igual a 80 y un *Score* base igual a 1000.

$$score = -(\sum_{j,i}^{k,11} WOE_j * \beta_i + \frac{-1.893}{11}) * \frac{80}{ln(2)} + 600$$

  Quedando la fórmula:

$$score = 619.8619761 - 115.4156033xb$$

  Donde *xb* es el factor lineal obtenido por el modelo, en este caso:

$$xb = B_0 + B_1 xwoe_1 + B_2 xwoe_2 + \cdots + B_8 xwoe_8$$

  Por lo que, cuanto mayor sea el valor del *score* del cliente su riesgo será menor, mientras que para valores pequeños de *score* el riesgo asociado al cliente será mayor.

- **Definición de los cortes de *Score*:**

  Se definieron los cortes de *score* para la propuesta de segmentación de riesgo de los clientes, manteniendo el apetito de riesgo por segmento especificado en las políticas crediticias de la entidad financiera.

En la Tabla 19 se detalla los intervalos de *score* y la distribución de los clientes en función a la segmentación propuesta.

Tabla 19: Intervalos de score y distribución de clientes en función a la segmentación propuesta

Segmento	Intervalo de *Score*	Registros	%Registros	*Default*	%RD
Excepcional	1166.81 a Más	68,388	20%	296	0.43%
Muy Bueno	[1027.91 - 1166.81>	102,313	30%	1,667	1.63%
Bueno	[949.89 - 1027.91>	68,360	20%	2,612	3.82%
Moderado	[905.18 - 949.89>	34,159	10%	2,201	6.44%
Alto	[832.94 - 905.18>	34,155	10%	3,602	10.55%
Crítico	Menos de 832.94	34,151	10%	9,146	26.78%
Total		341,526	100%	19,524	5.70%

FUENTE: Elaboración propia.

En la Figura 35 se observa una adecuada distribución de los clientes, ya que aproximadamente el 70% de la población se encuentra en los mejores segmentos con un ratio de *default* no mayor a 3.8%, asimismo el segmento crítico se encuentra diferenciado respecto a los demás puesto que más del 26% del total de sus registros ha hecho *default*.

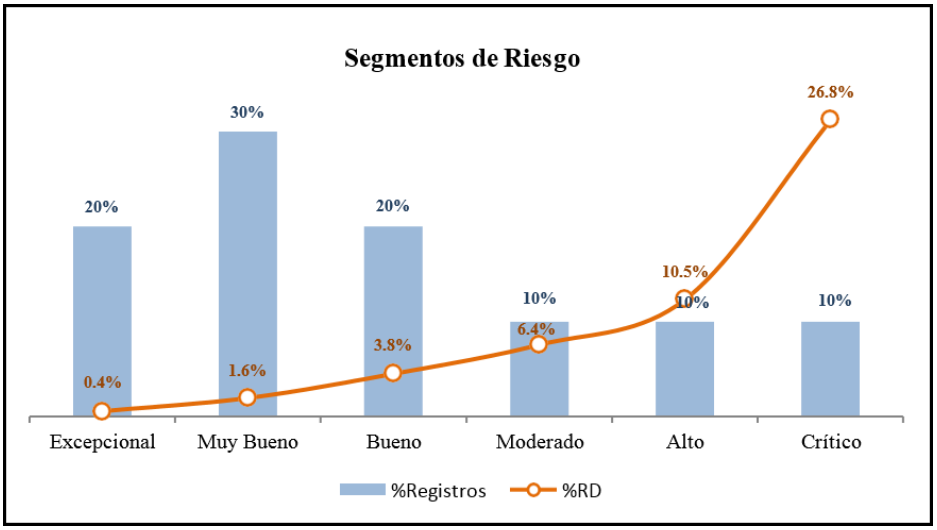


Figura 35: Segmentos de riesgo

FUENTE: Elaboración propia.

Asimismo, se evaluó el perfil de los grupos de riesgo en función de las variables finalistas del modelo, en la Figura 36 se observa la matriz de distribución por variable por cada segmento de riesgo:

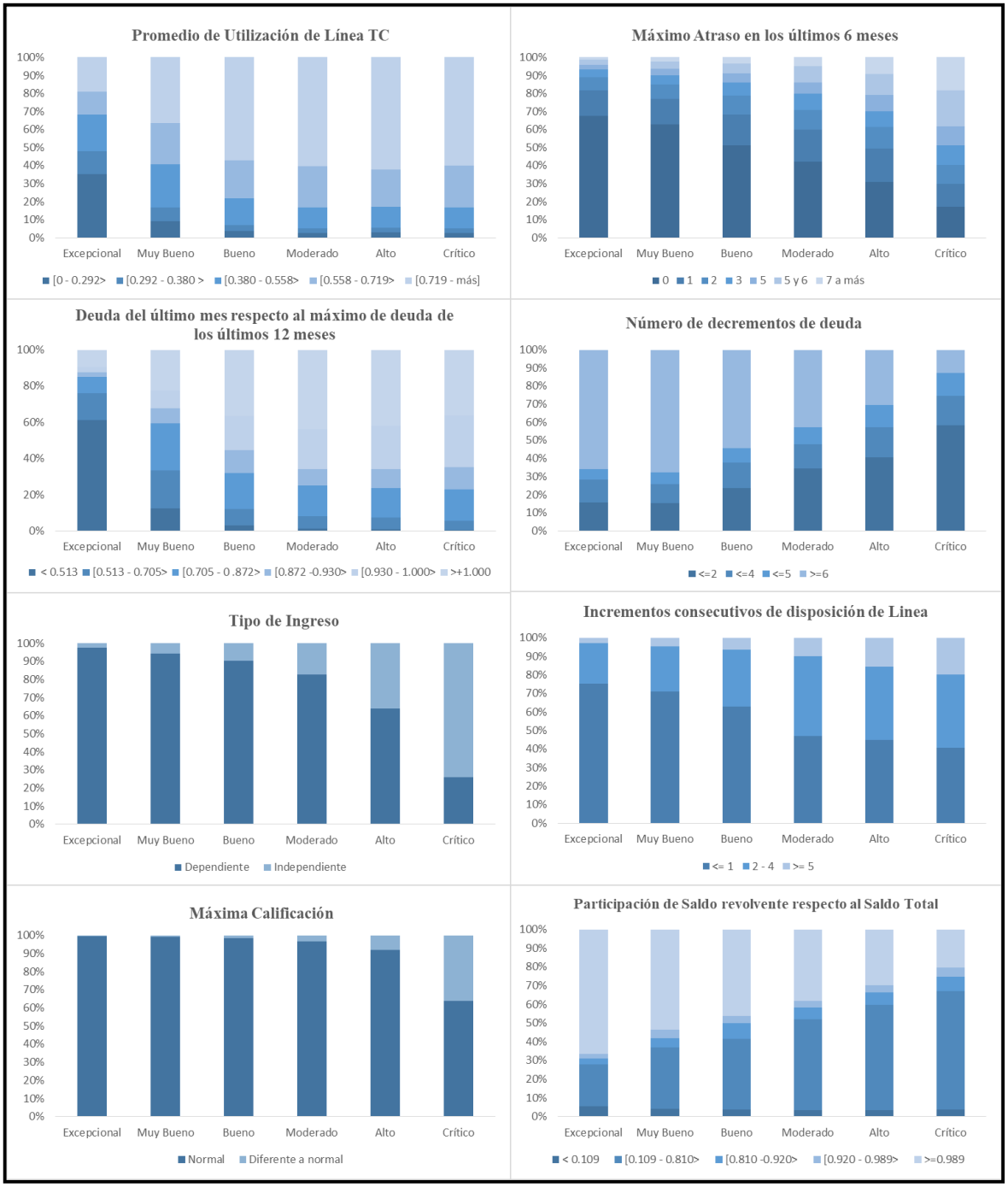


Figura 36: Distribución de Variables por Segmento de Riesgo

FUENTE: Elaboración propia.

En referencia a la figura anterior se puede describir los perfiles de los clientes que pertenecen a los segmentos de riesgo en base a tres aspectos: tipo de ingreso, tendencia a sobreendeudarse y cumplimiento de obligaciones crediticias.

De esta manera, un cliente excepcional es aquel que cuenta con un trabajo formal, que no tiende a sobreendeudarse al no usar más del 50% de su línea de crédito ni realizar contantemente disposiciones de efectivo de su tarjeta de crédito además de tener adeudado en el último mes un saldo no superior al 50% del saldo máximo que ha manejado en el último año, asimismo se encuentra cumpliendo con sus obligaciones crediticias al estar disminuyendo su deuda continuamente en el sistema financiero en los últimos 6 meses y al no presentar atrasos en el pago de sus cuotas manteniendo una calificación normal en el sistema.

Por otro lado, un cliente crítico, es aquel que no cuenta con un trabajo formal, que tiende a sobreendeudarse al usar de manera excesiva sus tarjetas de crédito y realizar disposiciones de efectivo continuamente además de tener adeudado en el último mes un saldo igual o mayor al saldo máximo manejado en el último año, asimismo es un cliente que no se encuentra al día con sus obligaciones crediticias al estar en una calificación peor a normal en el sistema.

Estos perfiles se encuentran en línea a lo esperado por el negocio y permiten evaluar a los clientes de manera diferenciada.

4.3. Contribución en la solución de situaciones problemáticas

Las áreas de riesgos y negocios necesitaban contar con un modelo probabilístico, que permita segmentar a los clientes que tienen al menos una tarjeta de crédito con la entidad financiera según su nivel de riesgo, ya que el modelo en producción no cumplía con los indicadores mínimos de predictibilidad. Es así que, el área de Modelos de Riesgos planteó la construcción de un nuevo modelo siguiendo los pasos de la metodología de *credit scoring*, donde se usó como técnica estadística de modelado a la Regresión Logística, ya que permite estimar la probabilidad de ocurrencia de un evento, en este caso que el cliente incumpla con sus obligaciones crediticias en una ventana de 12 meses posteriores al mes de análisis. Como resultado se obtuvo un modelo que consta de ocho variables predictivas y con una alta

capacidad de discriminación, permitiendo segmentar de una manera adecuada a los clientes de la entidad financiera.

4.4. Análisis de la contribución en términos de competencias y habilidades

El Plan de Estudios de la carrera de Estadística Informática permite el desarrollo de las destrezas técnicas (*hard skills*) que se requiere para el ejercicio de esta profesión en diferentes áreas de negocio e investigación.

Específicamente en las entidades financieras, el dominio de las técnicas de programación impartidas en los primeros ciclos de carrera y el manejo de base de datos son indispensables, ya que estas empresas al manejar grandes volúmenes de datos requieren de una adecuada administración y uso, permitiendo optimizar los recursos y dedicar más tiempo a los análisis de información. Asimismo, la variedad de programas estadísticos utilizados a lo largo de la carrera, tales como R *Project*, *Minitab,* SPSS, *Modeler* y *Python* permiten al profesional ser versátil en cuanto a los *softwares* que maneja así como la capacidad de adaptarse y aprender de nuevos *softwares* que tienen fines similares. Sin embargo, lo que distingue a un profesional de Estadística e Informática son sus competencias en cuánto a los fundamentos teóricos de las técnicas estadísticas que son más utilizadas en este rubro, destacando los cursos de Estadística Aplicada, Técnicas Multivariadas, Modelos Lineales, Inferencia Estadística, Modelos Lineales Generalizados y Técnicas de Computación.

Sin embargo, es necesaria la inclusión de cursos o talleres que desarrollen las habilidades blandas (*Soft Skills*) del profesional, tales como la comunicación, negociación, toma de decisiones, liderazgo, entre otras, ya que en el contexto actual del mundo empresarial, estas habilidades son tan valoradas como las destrezas técnicas al momento de seleccionar un nuevo colaborador y/o ascender a un colaborador, asimismo permite un crecimiento profesional más rápido y el manejo de situaciones críticas donde se deban tomar decisiones inmediatas.

4.5. Nivel de beneficio obtenido por el centro laboral

Al implementar los segmentos de riesgos obtenidos por el nuevo modelo, se incrementó en más de 20% la participación de los clientes pertenecientes a los segmentos más bajos, teniendo estos niveles más bajos de riesgo.

En la Tabla 20 se observa que el segmento Excepcional aumentó su participación en un 28% así como su ratio de *default* mejoró pasando de 2.1% a 0.4%, de la misma forma ocurre con los segmentos Muy Bueno y Bueno, aumentando su participación en más del 20% y disminuyendo su ratio de *default* en un 36% y 24% respectivamente. Esto permitirá a las diferentes áreas usuarias a generar estrategias para un número mayor de clientes que tienen menores niveles de probabilidad de incumplimiento. Asimismo, con el nuevo modelo se diferencia mejor a los clientes Buenos de los clientes Moderados, ya que con el anterior modelo sus niveles de riesgo eran muy similares, pudiendo traslaparse con el paso del tiempo. Y, finalmente la participación del segmento Crítico se mantuvo respecto al modelo anterior, sin embargo su diferenciación mejoró pasando de un ratio de *default* de 16.6% a 26.8%, lo que permitiría identificar de mejor manera a los clientes problemáticos y generar estrategias de cobranza temprana y/o realizar venta de cartera a otras entidades financieras.

Tabla 20: Comparativo de segmentos de riesgo

Segmentos	**Modelo En Producción**			**Modelo Propuesto**		
	Clientes	%Clientes	%RD	Clientes	%Clientes	%RD
Excepcional	53,218	16%	2.1%	68,291	20%	0.4%
Muy Bueno	82,169	24%	2.5%	102,410	30%	1.6%
Bueno	54,942	16%	5.0%	68,360	20%	3.8%
Moderado	55,900	16%	5.6%	34,159	10%	6.4%
Alto	55,846	16%	7.0%	34,150	10%	10.5%
Crítico	39,451	12%	16.6%	34,156	10%	26.8%
Total	341,526	100%	0.0%	341,526	100%	5.7%

FUENTE: Elaboración propia.

V. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

1. El Riesgo de incumplimiento de los clientes de la entidad financiera que tienen al menos una tarjeta de crédito que no hayan tenido más de ocho días de atraso en los 12 meses anteriores, es explicado por el modelo logístico propuesto, esto se evidenció con las pruebas de bondad de ajuste que exige la técnica estadística.
2. El modelo propuesto tiene alta capacidad predictiva, con un índice de Gini de 65.4% y ROC de 82.7 lo que indica que el modelo presenta menor tasa de falsos positivos y mayor tasa de verdaderos positivos. Esto lo convierte en una herramienta potente para la Gestión del Riesgo de Crédito, al poder discriminar a los clientes que harán *default* de los que que no.
3. La relación entre las variables que componen el modelo y el ratio de *default* tienen una tendencia lógica con el negocio, esto permite que el modelo tenga credibilidad y sea usado por las áreas usuarias.
4. Existe una adecuada distribución de pesos de las variables que conforman el modelo, esto asegura la estabilidad del modelo en el corto y mediano plazo, ya que si el comportamiento y distribución de alguna de las variables cambia, el impacto en la capacidad predictiva y calibración del modelo no sería muy alto, pudiéndose tomar acciones inmediatas tales como actualizar la transformación de la variable o retirarla del modelo.
5. Se identificó como una de las variables más influyentes en la probabilidad de incumplimiento al Promedio de Utilización de Línea TC en los últimos 12 años, ya que el 7.5% de los clientes que ha utilizado más del 70% de su Línea TC, ha superado los 60 días de atraso en alguno de los 12 meses posteriores a su mes de análisis.
6. Se evidenció que, los clientes que en los seis meses anteriores al mes de estudio han superado los tres días de atraso en alguno de sus créditos tienen mayor probabilidad de incumplimiento, llegando a tener un ratio de *default* mayor a 9%.

7. Se verificó la granularidad de la probabilidad de incumplimiento, lo que indica el adecuado ordenamiento del ratio de *default* por corte de probabilidad, esto asegura el no solapamiento del riesgo entre los segmentos de riesgo.
8. Se generó la regla de transformación de la probabilidad de incumplimiento a un *score*, que es manejado y entendido por áreas usuarias no especializadas con la construcción de modelos de riesgo.
9. Los segmentos de riesgo definidos por los cortes de *score* permiten una adecuada diferenciación de los clientes en cuanto a su probabilidad de *default*, mejorando la participación de clientes excepcional e identificando de una manera más óptima a los clientes críticos.
10. Los clientes que pertenecen al segmento de riesgo excepcional usan adecuadamente sus tarjetas de crédito, no tienden a sobreendeudarse y cumplen con sus obligaciones crediticias al no presentar atrasos. En contraparte, los clientes que se encuentran en el segmento crítico tienden a sobreendeudarse y presentan morosidad en los últimos meses, por lo cual deben existir estrategias diferenciadas por cada segmento de riesgo.

5.2. Recomendaciones

1. Para la inclusión de nuevas fuentes de información sin la necesidad de contratar el servicio de terceras empresas, se recomienda la implementación de técnicas de *Web Scraping*, cuyo objetivo es el de extraer información de sitios web.
2. Para la definición de incumplimiento, evaluar otras casuísticas de comportamiento de pago, diferentes a las evaluadas en este estudio, tales como número de meses que el cliente ha presentado atraso, máxima calificación en el sistema financiero, proporción de saldo adeudado en atraso, entre otras.
3. Para la segmentación de los clientes se recomienda la evaluación de otros ejes de segmentación, tales como cantidad de préstamos con la entidad financiera, participación del saldo en la entidad financiera respecto a la deuda total del cliente en todo el Sistema Financiero, entre otros.
4. Para la transformación de las variables se recomienda probar otros métodos de transformación, diferentes a lo utilizado en la entidad financiera, tales como transformaciones lineales, logarítmicas, exponenciales entre otros.

5. Evaluar los resultados de capacidad predictiva y calibración al usar otras técnicas predictivas, tales como Redes Neuronales, Random Forest, XGBoost entre otras.

VI. REFERENCIAS BIBLIOGRÁFICAS

Agresti A. (2007). *An Introduction To Categorical Data Analysis*. John Wiley & Sons. Asociación de Bancos del Perú. (2015). Solidez de la Banca Peruana. *Asbanc Semanal*, 1, 1-4.

Asociación de Supervisores Bancarios de las Américas. (2001). *El Nuevo Acuerdo de Capital de Basilea*. México DF. ASBA

Brown, T. (2008). *Design Thinking*. Estados Unidos. Harvard Business Review.

Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C. & Wirth R. (2000). *CRISP DM Step by Step data mining guide SPSS*. Estados Unidos: SPSS.

Hosmer, D. (2000). *Applied Logistic Regression*. New York. John Wiley & Sons, INC. IBM SPSS. (2015). *Manual de CRISP-DM de IBM SPSS Modeler*. Estados Unidos: IBM Corp.

Management Solutions. (2014). Aspectos cuantitativos y cualitativos de la gestión del riesgo de modelo. *Model Risk Management*, 1, 22-23.

Mays, E. (2001). *Credit scoring for risk managers*: *the handbook for lenders*. New York. AMACOM.

Nieto, S. (2010). *Crédito al Consumo*: *La estadística aplicada a un problema de riesgo crediticio*. (Tesis de Maestría). México. UAM. 96 pp.

Osterwalder A. & Pigneur Y. (2010). *Business Model Generation, A Handbook for Visionaries, Game Changers, and Challengers*. New Jersey, Estados Unidos: John Wiley & Sons, Inc.

Pérez J. (2017). *La regresión logística como modelo de predicción del riesgo crediticio en las organizaciones de la economía social y solidaria*, pp. 232-243, Universidad Católica de Colombia, Bogotá Colombia.

Piatetsky, G. (2014). *CRISP-DM, still the top methodology for analytics, data mining, or data scienceprojects*. de KDnuggets. Recuperado de https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html

Rahal, J.E. & Mungai, G. (2015).*Credit scoring for SME lending*. Nairobi, Kenya. FSD Kenya.

Raymond, A. (2007). *Theory and Practice for Retail Credit Risk Management and Decision Automation*. USA: Oxford University Press.

Rodríguez, D. & Carrasquilla, A. (2015). *Medición de Valor en Riesgo en Cartera de Clientes a Través de Modelos Logísticos y Simulación de Montecarlo*. Colombia.

Santos Jiménez, N. (2014). El negocio bancario. *Industrial Data*, 4(1), 025-033.

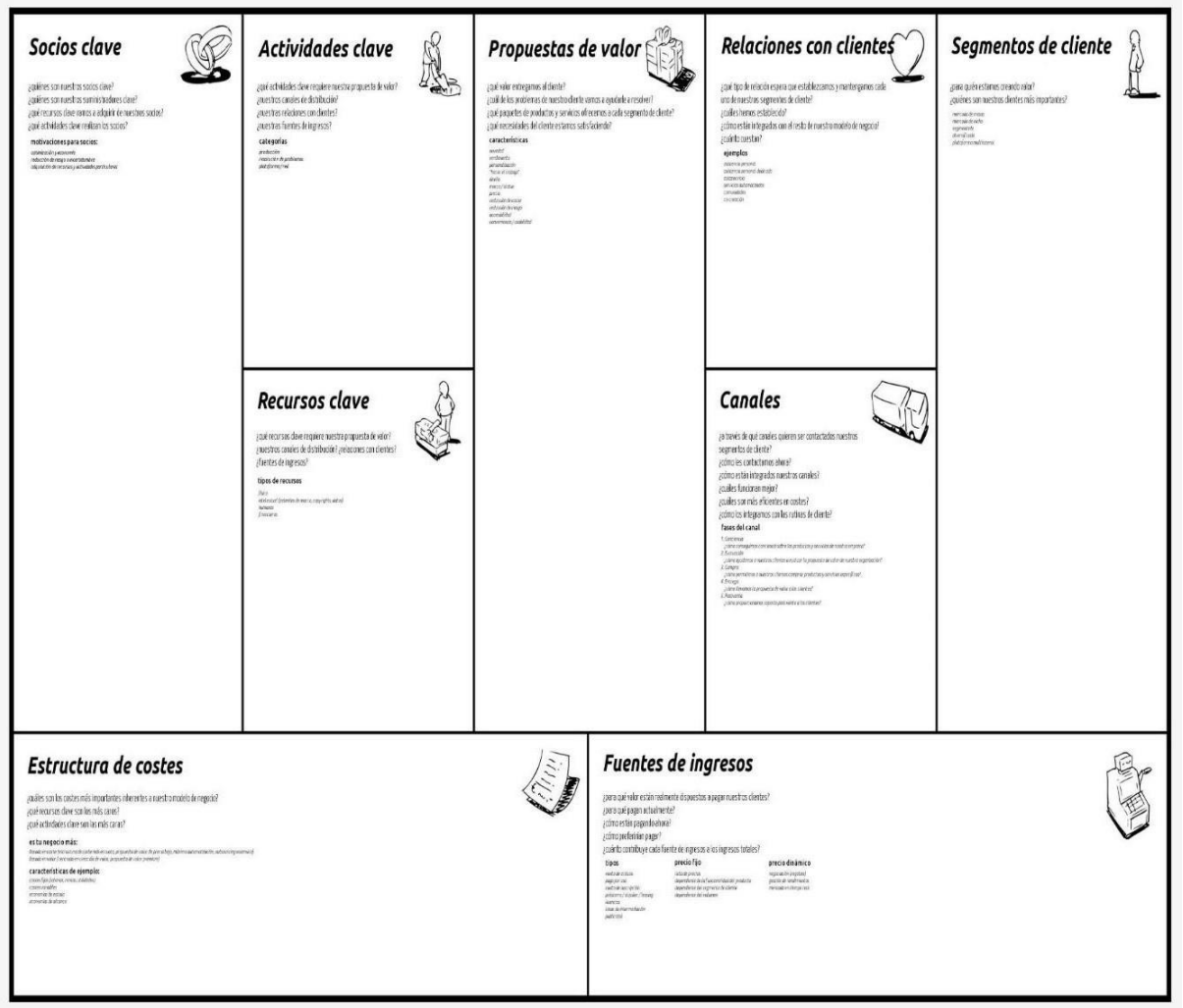
Siddiqi N. (2006). *Credit Risk Scorecards Developing and Implementing Intelligent Credit scoring*. New Jersey, Estados Unidos: John Wiley & Sons, Inc.

Trejo, J.; Ríos, H. & Almagro, F. (2016). Actualización del modelo de riesgo crediticio, una necesidad para la banca revolvente en México. *Revista Finanzas y Política Económica*, vol 8.núm 1, pp. 17-30 Universidad Católica de Colombia Bogotá, Colombia

Vicente, J.; Gonzáles, J.; Parra F. & Beltrán M. (2019). *Métodos de Data Science aplicados a la economía y a la dirección y administración de empresas*. Madrid. Universidad Nacional de Educación a Distancia Madrid.

VII. ANEXOS

Anexo 1: Modelo Canvas



Anexo 2: Descriptivos de variables finalistas por segmento de riesgo

– Variable: Promedio de Utilización de Línea TC

Segmento de Riesgo	Promedio de Utilización de Línea TC					Total
	[0 - 0.292>	[0.292 - 0.380 >	[0.380 - 0.558>	[0.558 - 0.719>	[0.719 - más]	
Número de Clientes						
Muy Bueno	9,535	7,664	24,562	23,183	37,369	102,313
Moderado	906	823	4,034	7,800	20,596	34,159
Excepcional	24,134	8,603	13,993	8,663	12,995	68,388
Crítico	894	847	4,024	7,859	20,527	34,151
Bueno	2,512	2,305	10,113	14,290	39,140	68,360
Alto	1,042	801	4,029	7,081	21,202	34,155
Total	**39,023**	**21,043**	**60,755**	**68,876**	**151,829**	**341,526**
Participación de Clientes por Segmento						
Excepcional	35%	13%	20%	13%	19%	100%
Muy Bueno	9%	7%	24%	23%	37%	100%
Bueno	4%	3%	15%	21%	57%	100%
Moderado	3%	2%	12%	23%	60%	100%
Alto	3%	2%	12%	21%	62%	100%
Crítico	3%	2%	12%	23%	60%	100%
Total	**11%**	**6%**	**18%**	**20%**	**44%**	**100%**
Ratio de Default						
Excepcional	0.48%	0.43%	0.36%	0.47%	0.40%	0.43%
Muy Bueno	2.07%	2.14%	1.42%	1.35%	1.72%	1.63%
Bueno	6.13%	4.99%	4.01%	3.65%	3.62%	3.82%
Moderado	9.49%	8.02%	7.24%	6.01%	6.25%	6.44%
Alto	10.56%	13.48%	11.96%	10.06%	10.33%	10.55%
Crítico	11.41%	17.47%	23.06%	26.30%	28.75%	26.78%
Total	**1.96%**	**3.03%**	**4.13%**	**5.99%**	**7.57%**	**5.72%**
Probabilidad Promedio de Default						
Excepcional	0.38%	0.42%	0.44%	0.49%	0.51%	0.44%
Muy Bueno	1.55%	1.61%	1.67%	1.89%	1.86%	1.77%
Bueno	3.81%	3.84%	3.85%	3.99%	4.00%	3.96%
Moderado	6.40%	6.41%	6.33%	6.40%	6.55%	6.48%
Alto	10.14%	10.26%	10.09%	10.26%	10.37%	10.31%
Crítico	22.22%	24.92%	26.44%	27.19%	25.96%	26.18%
Total	**1.79%**	**2.82%**	**4.26%**	**6.41%**	**7.38%**	**5.71%**

– Variable: Máximo Atraso en los últimos 6 meses

Segmento de Riesgo	Máximo atraso en los últimos 6 meses							Total
	0	1	2	3	5	5 y 6	7 a más	
Número de Clientes								
Excepcional	46,106	9,641	5,117	3,012	1,753	1,936	823	68,388
Muy Bueno	64,288	14,424	7,990	5,427	3,562	4,226	2,396	102,313
Bueno	35,077	11,531	7,243	4,879	3,458	3,761	2,411	68,360
Moderado	14,339	6,178	3,689	3,037	2,177	3,108	1,631	34,159
Alto	10,586	6,241	4,130	3,005	3,078	3,935	3,180	34,155
Crítico	5,883	4,242	3,587	3,744	3,566	6,863	6,266	34,151
Total	**176,279**	**52,257**	**31,756**	**23,104**	**17,594**	**23,829**	**16,707**	**341,526**
Participación de Clientes por Segmento								
Excepcional	67%	14%	7%	4%	3%	3%	1%	100%
Muy Bueno	63%	14%	8%	5%	3%	4%	2%	100%
Bueno	51%	17%	11%	7%	5%	6%	4%	100%
Moderado	42%	18%	11%	9%	6%	9%	5%	100%
Alto	31%	18%	12%	9%	9%	12%	9%	100%
Crítico	17%	12%	11%	11%	10%	20%	18%	100%
Total	**52%**	**15%**	**9%**	**7%**	**5%**	**7%**	**5%**	**100%**
Ratio de Default								
Excepcional	0.40%	0.39%	0.45%	0.33%	0.97%	0.83%	0.85%	0.43%
Muy Bueno	1.53%	1.74%	1.60%	1.95%	2.08%	1.89%	1.92%	1.63%
Bueno	3.46%	3.54%	4.16%	4.67%	3.99%	5.18%	5.31%	3.82%
Moderado	5.74%	5.89%	6.59%	7.21%	7.17%	7.82%	9.38%	6.44%
Alto	11.56%	9.39%	9.44%	10.38%	10.88%	10.37%	10.91%	10.55%
Crítico	28.71%	27.16%	25.73%	24.25%	25.69%	25.98%	28.33%	26.78%
Total	**3.47%**	**5.36%**	**6.32%**	**7.72%**	**9.30%**	**11.44%**	**14.70%**	**5.72%**
Probabilidad Promedio de Default								
Excepcional	0.42%	0.46%	0.48%	0.48%	0.49%	0.52%	0.50%	0.44%
Muy Bueno	1.76%	1.79%	1.78%	1.82%	1.74%	1.80%	1.78%	1.77%
Bueno	3.94%	3.87%	3.98%	4.12%	4.05%	4.01%	4.13%	3.96%
Moderado	6.51%	6.27%	6.39%	6.68%	6.56%	6.49%	6.68%	6.48%
Alto	10.39%	10.17%	10.22%	10.31%	10.41%	10.10%	10.55%	10.31%
Crítico	22.89%	24.45%	24.58%	24.54%	26.71%	27.35%	30.74%	26.18%
Total	**3.46%**	**5.37%**	**6.28%**	**7.56%**	**9.24%**	**11.39%**	**15.06%**	**5.71%**

– Variable: Deuda del último mes respecto al máximo de deuda de los últimos 12 meses

Segmento de Riesgo	Deuda del último mes respecto al máximo de deuda de los últimos 12 meses						Total
	< 0.513	[0.513 - 0.705>	[0.705 - 0.872>	[0.872 - 0.930>	[0.930 - 1.000>	>+1.000	
Número de Clientes							
Excepcional	41,721	10,358	6,041	1,804	1,866	6,598	68,388
Muy Bueno	12,724	21,518	26,513	8,353	10,272	22,933	102,313
Bueno	1,945	6,263	13,616	8,540	13,061	24,935	68,360
Moderado	451	2,365	5,713	3,108	7,567	14,955	34,159
Alto	293	2,156	5,634	3,507	8,143	14,422	34,155
Crítico	118	1,823	5,930	4,110	9,739	12,431	34,151
Total	**57,252**	**44,483**	**63,447**	**29,422**	**50,648**	**96,274**	**341,526**
Participación de Clientes por Segmento							
Excepcional	61%	15%	9%	3%	3%	10%	100%
Muy Bueno	12%	21%	26%	8%	10%	22%	100%
Bueno	3%	9%	20%	12%	19%	36%	100%
Moderado	1%	7%	17%	9%	22%	44%	100%
Alto	1%	6%	16%	10%	24%	42%	100%
Crítico	0%	5%	17%	12%	29%	36%	100%
Total	**17%**	**13%**	**19%**	**9%**	**15%**	**28%**	**100%**
Ratio de Default							
Excepcional	0.35%	0.54%	0.60%	0.39%	0.38%	0.67%	0.43%
Muy Bueno	1.29%	1.65%	1.68%	1.69%	1.48%	1.78%	1.63%
Bueno	4.22%	4.28%	3.77%	3.27%	3.29%	4.17%	3.82%
Moderado	6.65%	7.86%	7.09%	6.85%	5.29%	6.47%	6.44%
Alto	6.83%	13.03%	10.72%	9.81%	8.85%	11.32%	10.55%
Crítico	11.02%	23.04%	24.13%	23.38%	25.24%	31.08%	26.78%
Total	**0.79%**	**3.52%**	**5.41%**	**6.61%**	**8.23%**	**8.26%**	**5.72%**
Probabilidad Promedio de Default							
Excepcional	0.37%	0.49%	0.53%	0.58%	0.60%	0.61%	0.44%
Muy Bueno	1.47%	1.67%	1.86%	1.87%	1.94%	1.83%	1.77%
Bueno	3.74%	3.93%	3.96%	3.92%	3.91%	4.03%	3.96%
Moderado	6.38%	6.46%	6.39%	6.60%	6.31%	6.58%	6.48%
Alto	9.72%	10.02%	10.19%	10.23%	10.32%	10.42%	10.31%
Crítico	18.25%	22.38%	25.25%	26.16%	28.75%	25.24%	26.18%
Total	**0.86%**	**3.22%**	**5.52%**	**7.27%**	**9.55%**	**7.37%**	**5.71%**

– Variable: Número de Decrementos de Deuda

Segmento de Riesgo	Número de Decrementos de Deuda				Total
	<=2	<=4	<=5	>=6	
Número de Clientes					
Excepcional	10,616	8,628	4,175	44,969	68,388
Muy Bueno	15,494	10,840	6,531	69,448	102,313
Bueno	16,253	9,431	5,455	37,221	68,360
Moderado	11,754	4,545	3,247	14,613	34,159
Alto	13,834	5,661	4,197	10,463	34,155
Crítico	19,906	5,504	4,355	4,386	34,151
Total	**87,857**	**44,609**	**27,960**	**181,100**	**341,526**
Participación de Clientes por Segmento					
Excepcional	16%	13%	6%	66%	100%
Muy Bueno	15%	11%	6%	68%	100%
Bueno	24%	14%	8%	54%	100%
Moderado	34%	13%	10%	43%	100%
Alto	41%	17%	12%	31%	100%
Crítico	58%	16%	13%	13%	100%
Total	**26%**	**13%**	**8%**	**53%**	**100%**
Ratio de Default					
Excepcional	0.28%	0.49%	0.34%	0.47%	0.43%
Muy Bueno	1.52%	1.64%	1.50%	1.66%	1.63%
Bueno	3.93%	3.26%	2.80%	4.07%	3.82%
Moderado	5.84%	5.74%	5.17%	7.43%	6.44%
Alto	11.54%	9.61%	7.08%	11.12%	10.55%
Crítico	30.86%	20.11%	17.96%	25.40%	26.78%
Total	**10.62%**	**5.47%**	**5.41%**	**3.45%**	**5.72%**
Probabilidad Promedio de Default					
Excepcional	0.52%	0.43%	0.44%	0.42%	0.44%
Muy Bueno	1.75%	1.76%	1.78%	1.78%	1.77%
Bueno	4.32%	3.89%	3.94%	3.83%	3.96%
Moderado	6.83%	6.37%	6.34%	6.27%	6.48%
Alto	10.64%	10.14%	10.24%	9.99%	10.31%
Crítico	28.59%	24.24%	23.68%	20.12%	26.18%
Total	**10.24%**	**6.26%**	**7.21%**	**3.14%**	**5.71%**

– Variable: Tipo de Ingreso

Segmento de Riesgo	Tipo de Ingreso		Total
	Dependiente	Independiente	
Número de Clientes			
Excepcional	1,763	66,625	68,388
Muy Bueno	6,083	96,230	102,313
Bueno	6,744	61,616	68,360
Moderado	5,972	28,187	34,159
Alto	12,350	21,805	34,155
Crítico	25,336	8,815	34,151
Total	**58,248**	**283,278**	**341,526**
Participación de Clientes por Segmento			
Excepcional	97%	3%	100%
Muy Bueno	94%	6%	100%
Bueno	90%	10%	100%
Moderado	83%	17%	100%
Alto	64%	36%	100%
Crítico	26%	74%	100%
Total	**83%**	**17%**	**100%**
Ratio de Default			
Excepcional	0.41%	1.36%	0.43%
Muy Bueno	1.59%	2.29%	1.63%
Bueno	3.74%	4.54%	3.82%
Moderado	6.46%	6.36%	6.44%
Alto	10.19%	11.17%	10.55%
Crítico	23.36%	27.97%	26.78%
Total	**3.60%**	**15.99%**	**5.72%**
Probabilidad Promedio de Default			
Excepcional	0.44%	0.53%	0.44%
Muy Bueno	1.77%	1.82%	1.77%
Bueno	3.95%	4.07%	3.96%
Moderado	6.47%	6.51%	6.48%
Alto	10.19%	10.52%	10.31%
Crítico	19.62%	28.46%	26.18%
Total	**3.60%**	**15.95%**	**5.71%**

– Variable: Número de Incrementos Consecutivos de Disposición

Segmento de Riesgo	Número de Incrementos Consecutivos de Disposición			Total
	<= 1	2 - 4	>= 5	
Número de Clientes				
Excepcional	15,011	51,298	2,079	68,388
Muy Bueno	24,570	72,744	4,999	102,313
Bueno	20,822	43,000	4,538	68,360
Moderado	14,700	16,070	3,389	34,159
Alto	13,552	15,314	5,289	34,155
Crítico	13,546	13,819	6,786	34,151
Total	**102,201**	**212,245**	**27,080**	**341,526**
Participación de Clientes por Segmento				
Excepcional	75%	22%	3%	100%
Muy Bueno	71%	24%	5%	100%
Bueno	63%	30%	7%	100%
Moderado	47%	43%	10%	100%
Alto	45%	40%	15%	100%
Crítico	40%	40%	20%	100%
Total	**62%**	**30%**	**8%**	**100%**
Ratio de Default				
Excepcional	0.41%	0.47%	0.72%	0.43%
Muy Bueno	1.60%	1.61%	2.08%	1.63%
Bueno	3.82%	3.76%	4.12%	3.82%
Moderado	6.80%	5.78%	7.61%	6.44%
Alto	11.41%	9.81%	9.95%	10.55%
Crítico	25.62%	26.13%	30.45%	26.78%
Total	**4.43%**	**6.82%**	**11.65%**	**5.72%**
Probabilidad Promedio de Default				
Excepcional	0.42%	0.48%	0.51%	0.44%
Muy Bueno	1.78%	1.76%	1.80%	1.77%
Bueno	3.98%	3.89%	4.14%	3.96%
Moderado	6.41%	6.53%	6.57%	6.48%
Alto	10.18%	10.31%	10.69%	10.31%
Crítico	24.62%	26.36%	28.98%	26.18%
Total	**4.34%**	**7.09%**	**11.24%**	**5.71%**

– Variable: Máxima Calificación

Segmento de Riesgo	Máxima Calificación		Total
	Normal	Diferente a norma	
Número de Clientes			
Excepcional	68,228	160	68,388
Muy Bueno	101,627	686	102,313
Bueno	67,218	1,142	68,360
Moderado	33,045	1,114	34,159
Alto	31,405	2,750	34,155
Crítico	21,739	12,412	34,151
Total	**323,262**	**18,264**	**341,526**
Participación de Clientes por Segmento			
Excepcional	100%	0%	100%
Muy Bueno	99%	1%	100%
Bueno	98%	2%	100%
Moderado	97%	3%	100%
Alto	92%	8%	100%
Crítico	64%	36%	100%
Total	**95%**	**5%**	**100%**
Ratio de Default			
Excepcional	0.42%	6.88%	0.43%
Muy Bueno	1.61%	5.10%	1.63%
Bueno	3.75%	7.79%	3.82%
Moderado	6.21%	13.46%	6.44%
Alto	10.13%	15.35%	10.55%
Crítico	26.20%	27.80%	26.78%
Total	**4.75%**	**22.77%**	**5.72%**
Probabilidad Promedio de Default			
Excepcional	0.44%	0.55%	0.44%
Muy Bueno	1.77%	1.91%	1.77%
Bueno	3.96%	4.09%	3.96%
Moderado	6.48%	6.60%	6.48%
Alto	10.28%	10.64%	10.31%
Crítico	22.05%	33.40%	26.18%
Total	**4.62%**	**25.04%**	**5.71%**