

UNIVERSIDAD NACIONAL DE INGENIERÍA

FACULTAD DE INGENIERÍA ECONÓMICA,
ESTADÍSTICA Y CIENCIAS SOCIALES



TRABAJO DE SUFICIENCIA PROFESIONAL

“Aplicación de árboles de inferencia condicional y regresión logística para el desarrollo de modelos de Score para Behaviour de créditos para clientes de persona jurídica”

PARA OBTENER EL TÍTULO PROFESIONAL DE INGENIERO ESTADÍSTICO


Elaborado por:

JUAN JUNIOR RODAS MENDOZA

 0009-0007-4448-6822

Asesor:

CARLOS ALVARO RISCO FRANCO

 0000-0003-1323-436X

**LIMA – PERÚ
2024**

Citar/How to cite	Rodas Mendoza[1]
Referencia/Reference	[1] J. Rodas Mendoza, “ <i>Aplicación de árboles de inferencia condicional y regresión logística para el desarrollo de modelos de Score para Behaviour de créditos para clientes de persona jurídica</i> ” [Trabajo de Suficiencia Profesional]. Lima (Perú): Universidad Nacional de Ingeniería, 2024.
Estilo/Style: IEEE (2020)	

Citar/How to cite	(Rodas, 2024)
Referencia/Reference	Rodas, C. (2024). <i>Aplicación de árboles de inferencia condicional y regresión logística para el desarrollo de modelos de Score para Behaviour de créditos para clientes de persona jurídica</i> . [Trabajo de Suficiencia Profesional, Universidad Nacional de Ingeniería]. Repositorio institucional Cybertesis UNI.
Estilo/Style: APA (7ma ed.)	

Dedicatoria

A mi amada familia, quienes han sido mi constante motivación, el motor que impulsa mis sueños y la fuente inagotable de inspiración en mi vida.

A mis queridos padres, les agradezco por su apoyo incondicional, por haberme guiado con amor y dedicación, y por haberme forjado en una vida de constante superación. Su ejemplo y aliento han sido la luz que ilumina mi camino hacia el éxito.

Este logro es también suyo, y esta dedicatoria es un modesto reconocimiento a la invaluable contribución que han tenido en mi viaje.

Resumen

En el último semestre del 2019 y primer semestre del 2020 la prima de riesgo de la cartera de clientes de personas jurídicas de una entidad financiera incrementó su prima de riesgo en un promedio del 50%. En ese periodo la cartera de clientes era administrada de acuerdo con la experiencia de los especialistas del producto. Una herramienta esencial para la gestión del riesgo de crédito es el comportamiento crediticio en el sistema financiero, una manera de resumir este comportamiento es mediante la generación de perfiles de riesgo creados en base a un modelo, en ese sentido, a fines del 2020 desarrollé modelos de score para Behaviour (comportamiento) en donde se aplicó árboles de inferencia condicional y regresión logística usando variables del Sistema Financiero, Estados Financieros e información financiera interna de los clientes. El objetivo del modelo fue identificar a los clientes, de persona jurídica, que puedan tener un riesgo de incumplimiento de pago de su deuda a partir de grupos de riesgo establecidos en base a rangos de puntaje, con la finalidad de tomar acciones para administrar la cartera de crédito del Banco. Al usar los modelos para Behaviour (comportamiento) se logró disminuir la prima de riesgo en un 30%.

Abstract

In the last half of 2019 and the first half of 2020, the risk premium of the client portfolio of legal entities of a financial institution increased its risk premium by an average of 50%. In that period, the portfolio was managed according to the experience of the product specialists. An essential tool for credit risk management is credit behavior in the financial system. One way to summarize this behavior is through the generation of risk profiles created based on a model. In this sense, at the end of 2020, I developed models of score for Behavior where conditional inference trees and logistic regression were applied using variables of the Financial System, Financial Statements and internal financial information of the clients. The objective of the model was to identify clients, legal entities, who may have a risk of non-payment of their debt based on risk groups established based on score ranges, in order to manage the credit portfolio. By using the models for Behaviour, it was possible to reduce the risk premium by 30%.

Indice

CAPITULO I: ANTECEDENTES	1
1.1. Datos de la Empresa o Entidad Financiera	1
1.2. Área donde se Desarrolló el Trabajo	1
1.3. Descripción del Portafolio	1
1.4. Formulación del Problema.....	2
1.4.1. Problema General	2
1.4.2. Problema Específico	2
1.5. Objetivos.....	3
1.5.1. Objetivo General.....	3
1.5.2. Objetivo Específico	3
CAPITULO II: DEFINICIÓN DE TÉCNICAS	4
2.1 Árboles de Decisión	4
2.2 Clúster de Variables.....	5
2.3 Análisis de Correlación	5
2.4 Indicador Pesos de Evidencia o Weight of Evidence (WOE)	5
2.5 Monotonicidad.....	6
2.6 Information Value (IV).....	6
2.7 Transformación de Variables.....	7
2.7.1 Truncamiento:.....	7
2.7.2 Reescalar por Rango:.....	7
2.8 Modelo Logit	7
2.9 Multicolinealidad.....	8
2.10 Principio de Parsimonia.....	9
2.11 Construcción del Scorecard (escalamiento)	9
2.12 Indicadores de Discriminancia	10
2.12.1 Coeficiente de Gini	10
2.12.2 Estadístico Kolmogorov-Smirnov (KS).....	11
2.12.3 Curva ROC.....	12
2.12.4 Divergencia	12
2.12.5 Estabilidad Poblacional.....	13
2.12.6 Bootstrap	13
CAPITULO III: CONSTRUCCIÓN DEL MODELO.....	15
3.1 Tratamiento de Información	15
3.1.1 Fuentes de Información.....	15
3.1.2 Descripción del Universo de Observaciones	16

3.1.3	Ventanas de Tiempo	16
3.1.4	Definición de Incumplimiento de Pago de Crédito (Default).....	16
3.1.5	Exclusiones del Modelo.....	18
3.1.6	Definición de Muestras de Trabajo.....	20
3.1.7	Definiciones	21
3.1.8	Construcción de la Base de Desarrollo	23
3.1.9	Construcción de Variables	23
3.1.10	Ventas_Soles.....	23
3.1.11	Grupo_Score_Empresa	26
3.2	Segmentación	29
3.3	Concentración y Penetración:.....	29
3.4	Criterio Information Value (IV) y Pesos de Evidencia (WOE).....	30
3.5	Criterio de Correlación	32
3.6	Significancia	33
3.7	Bootstrap.....	33
3.8	Aporte a Discriminancia.....	33
3.9	Elección del Modelo	33
CAPITULO IV: RESULTADOS.....		35
4.1	Segmentación	35
4.1.1	Patrón Externo	35
4.1.2	Patrón Interno.....	40
4.1.3	Nivel de Ventas (Ventas soles)	45
4.1.4	Segmentos Finales	48
4.2	Modelo para el Segmento 1 (Riesgo bajo)	49
4.2.1	Grupos de Riesgo.....	54
4.3	Modelo para el Segmento 2 (Riesgo medio)	55
4.3.1	Grupos de Riesgo.....	60
4.4	Modelo para el Segmento 3 (Riesgo alto)	61
4.4.1	Grupos de Riesgo.....	66
4.5	Modelo Total	67
4.5.1	Resumen de Variables.....	67
4.5.2	Grupos de Riesgo.....	68
CONCLUSIONES		69
RECOMENDACIONES		71
REFERENCIAS BIBLIOGRÁFICAS		72
ANEXOS		74

Índice de Tablas

Tabla 1	<i>Distribución del Portafolio - Enero 2021</i>	2
Tabla 2	<i>Criterios de Indicador K-S</i>	12
Tabla 3	<i>Criterios de Indicador de Estabilidad Poblacional.</i>	13
Tabla 4	<i>Matriz de Transición de Clasificación SBS de Deudores (Diciembre 2018 vs Diciembre 2019)</i>	18
Tabla 5	<i>Resumen de Caídas por Exclusiones - Segmento 1</i>	19
Tabla 6	<i>Resumen de Caídas por Exclusiones - Segmento 2</i>	19
Tabla 7	<i>Resumen de Caídas por Exclusiones - Segmento 3</i>	20
Tabla 8	<i>Resumen de Caídas por Exclusiones – Base Total</i>	20
Tabla 9	<i>Cantidad de Registros que no Cumplen el Principio Contable por año</i>	25
Tabla 10	<i>Distribución del Grupo de Riesgo - Junio 2019</i>	27
Tabla 11	<i>Rangos de PD por Grupo de Riesgo - Base Desarrollo Score Buro para Persona Natural - Abril 2018</i>	28
Tabla 12	<i>Rango del índice de Information Value (IV)</i>	30
Tabla 13	<i>Rango de Correlaciones (r)</i>	32
Tabla 14	<i>Resumen de la Distribución de la Segmentación del Patrón Externo con GINI y KS.</i>	39
Tabla 15	<i>Resumen de la Distribución de Variables en el Patrón Interno Según Nivel de Riesgo y Participación con GINI y KS de la Segmentación Patrón Interno.</i> ..	45
Tabla 16	<i>Cortes Según el Árbol para la Variable Ventas Soles</i>	46
Tabla 17	<i>Cortes Agrupados para la Variable Nivel Ventas (Ventas_Soles)</i>	46
Tabla 18	<i>Porcentaje de Distribución de los Niveles de Venta por Banca.</i>	47
Tabla 19	<i>Cortes Finales para la Variable Ventas Soles</i>	47
Tabla 20	<i>Porcentaje de Incumplimiento de Pago de Crédito (Default) de la Combinación de Variables que Forman los Segmentos Finales</i>	48
Tabla 21	<i>Análisis de Estimaciones de Máxima Verosimilitud.</i>	50
Tabla 22	<i>Indicadores de Discriminancia</i>	50
Tabla 23	<i>Aporte de las Variables al KS del Modelo en la Muestra de Análisis.</i>	51
Tabla 24	<i>Indicadores de discriminancia - Muestra de control (Fuera de la muestra).</i> ..	52
Tabla 25	<i>Grupos de Riesgo por Deciles en la Muestra de Análisis.</i>	54
Tabla 26	<i>Grupos de Riesgo por Deciles en la Muestra de Control</i>	55
Tabla 27	<i>Análisis de Estimaciones de Máxima Verosimilitud</i>	56

Tabla 28	<i>Indicadores de Discriminancia</i>	56
Tabla 29	<i>Aporte de las Variables al KS del Modelo</i>	57
Tabla 30	<i>Indicadores de Discriminancia - Muestra de Control (Fuera de la Muestra)</i>	58
Tabla 31	<i>Grupos de Riesgo por Deciles en la Muestra de Análisis.</i>	60
Tabla 32	<i>Grupos de Riesgo por Deciles en la Muestra de Control</i>	61
Tabla 33	<i>Análisis de Estimaciones de Máxima Verosimilitud</i>	62
Tabla 34	<i>Indicadores de Discriminancia</i>	62
Tabla 35	<i>Aporte de las Variables al KS del Modelo</i>	63
Tabla 36	<i>Indicadores de Discriminancia - Muestra de Control (Fuera de la muestra)</i>	64
Tabla 37	<i>Grupos de Riesgo por Deciles en la Muestra de Análisis.</i>	66
Tabla 38	<i>Grupos de Riesgo por Deciles en la Muestra de Control.</i>	67
Tabla 39	<i>Resumen de uso de Variables por Modelo</i>	67
Tabla 40	<i>Grupos de Riesgo por Deciles en la Muestra de Análisis.</i>	68
Tabla 41	<i>Grupos de Riesgo por Deciles en la Muestra de Control.</i>	68

Índice de Figuras

Figura 1	<i>Ejemplo de Árbol de Decisión</i>	4
Figura 2	<i>Ejemplo de Acumulación de Grupos en Base a la Curva de Lorenz</i>	10
Figura 3	<i>Ejemplo de Diferencia de Muestras como Indicador de Kolmogorov-Smirnov</i>	11
Figura 4	<i>Estado Financiero Auditados de la Superintendencia de Mercado de Valores (SMV)</i>	24
Figura 5	<i>Estado Financiero en el Sistema de la Entidad Financiera</i>	25
Figura 6	<i>Ejemplo de Gráfica WOE para Variable no Monotónica</i>	31
Figura 7	<i>Ejemplo Árbol de Decisión para Reescalar la Variable</i>	31
Figura 8	<i>Ejemplo de Gráfica WOE para Variable Transformada</i>	32
Figura 9	<i>Esquema de Variables que Forman el Patrón Externo</i>	35
Figura 10	<i>Árbol Inicial con las Variables que Forman el Patrón Externo</i>	36
Figura 11	<i>Árbol de Inferencia Condicional de la Variable P1</i>	36
Figura 12	<i>Árbol de Inferencia Condicional de la Variable P8_1</i>	37
Figura 13	<i>Árbol de Inferencia Condicional de la Variable P8_2</i>	37
Figura 14	<i>Árbol de Inferencia Condicional de la Variable P8_3</i>	38
Figura 15	<i>Distribución de Variables en el Patrón Externo Según Nivel de Riesgo y Participación</i>	39
Figura 16	<i>Esquema de Variables que Forman el Patrón Interno</i>	40
Figura 17	<i>Árbol Inicial con las Variables que Forman el Patrón Interno</i>	41
Figura 18	<i>Árbol de Inferencia Condicional de la Variable INT_P1</i>	41
Figura 19	<i>Árbol de Inferencia Condicional de la Variable INT_P8_1</i>	42
Figura 20	<i>Árbol de Inferencia Condicional de la Variable INT_P8_2</i>	42
Figura 21	<i>Árbol de Inferencia Condicional de la Variable INT_P8_3</i>	43
Figura 22	<i>Distribución de variables en el patrón interno según nivel de riesgo y participación</i>	44
Figura 23	<i>Árbol de Inferencia Condicional de la Variable Nivel de Ventas (Ventas_soles)</i>	45
Figura 24	<i>Participación de las Clases de la Variable Nivel de Ventas por Default</i> ...	47
Figura 25	<i>Participación e Incumplimiento de Pago de Crédito (Default) de los Segmentos Finales</i>	49
Figura 26	<i>Distribución de Clientes que no han Incumplido el Pago de su Crédito</i>	

	<i>(Buenos) y Clientes que han Incumplido el Pago de su Crédito (Malos) en la Muestra de Análisis (Desarrollo)</i>	51
Figura 27	<i>Distribución Acumulada de Buenos y Malos</i>	52
Figura 28	<i>Promedio Teórico del Incumplimiento de Pago de Crédito del Modelo (pd_prom) y Porcentaje de Clientes en Incumplimiento de Pago de Crédito Real (% malos).....</i>	52
Figura 29	<i>Distribución de Clientes que no han Incumplido el Pago de su Crédito (Buenos) y Clientes que han Incumplido el Pago de su Crédito (Malos) en la Muestra de Control.</i>	53
Figura 30	<i>Distribución acumulada de Buenos y Malos - Desarrollo vs Fuera de Muestra</i>	53
Figura 31	<i>Promedio Teórico del Incumplimiento de Pago de Crédito del Modelo (pd_prom) y Porcentaje de Clientes en Incumplimiento de Pago de Crédito Real (% malos) en la Muestra de Control</i>	54
Figura 32	<i>Distribución de Clientes que no han Incumplido el Pago de su Crédito (Buenos) y Clientes que han Incumplido el Pago de su Crédito (Malos) en la Muestra de Análisis(desarrollo).</i>	57
Figura 33	<i>Distribución acumulada de Buenos y Malos</i>	58
Figura 34	<i>Promedio Teórico del Incumplimiento de Pago de Crédito del Modelo (pd_prom) y Porcentaje de Clientes en Incumplimiento de Pago de Crédito Real (% malos) en la Muestra de Análisis (desarrollo)</i>	58
Figura 35	<i>Distribución de Clientes que no han Incumplido el Pago de su Crédito (Buenos) y Clientes que han Incumplido el Pago de su Crédito (Malos) en la Muestra de Control.</i>	59
Figura 36	<i>Distribución acumulada de Buenos y Malos - Desarrollo vs Fuera de Muestra</i>	59
Figura 37	<i>Promedio Teórico del Incumplimiento de Pago de Crédito del Modelo (pd_prom) y Porcentaje de Clientes en Incumplimiento de Pago de Crédito Real (% Malos) en la Muestra de Control.....</i>	60
Figura 38	<i>Distribución de Clientes que no han Incumplido el Pago de su Crédito (Buenos) y Clientes que han Incumplido el Pago de su Crédito (Malos) en la Muestra de Análisis (Desarrollo).</i>	63
Figura 39	<i>Distribución acumulada de Buenos y Malos</i>	64
Figura 40	<i>Promedio Teórico del Incumplimiento de Pago de Crédito del Modelo</i>	

	<i>(pd_prom)</i> y <i>Porcentaje de Clientes en Incumplimiento de Pago de Crédito Real (% Malos) en la Muestra de Análisis (Desarrollo)</i>	64
Figura 41	<i>Distribución de clientes que no han incumplido el pago de su crédito (Buenos) y clientes que han incumplido el pago de su crédito (Malos) en la muestra de control</i>	65
Figura 42	<i>Distribución Acumulada de Buenos y Malos - Desarrollo vs Fuera de Muestra</i>	65
Figura 43	<i>Promedio Teórico del Incumplimiento de Pago de Crédito del Modelo (pd_prom) y Porcentaje de Clientes en Incumplimiento de Pago de Crédito Real (% Malos) en la Muestra de Control</i>	66

INTRODUCCIÓN

La gestión efectiva de riesgos en el ámbito financiero se ha convertido en un pilar fundamental para garantizar la estabilidad y el crecimiento sostenible de las entidades. En este contexto, el presente trabajo surge como respuesta a una problemática específica dentro de una de las instituciones financieras más relevantes del Perú, un banco con más de dos décadas de trayectoria y un enfoque claro en su misión social: Generar un impacto trascendental en el desarrollo de la sociedad.

El problema general identificado en esta institución radica en la carencia de un modelo estadístico robusto para evaluar el comportamiento de personas jurídicas, lo cual ha conducido a un notable aumento de la prima de riesgo en el último año. Las decisiones, hasta el momento, se han basado principalmente en la experiencia de especialistas de productos, generando evaluaciones subjetivas y resultados inciertos. Esta situación es un desafío significativo que demanda la implementación de una herramienta efectiva para la gestión de riesgos, específicamente orientada al comportamiento de personas jurídicas.

El objetivo central de esta investigación es, por lo tanto, desarrollar un modelo predictivo que permita evaluar de manera precisa y oportuna el riesgo crediticio asociado a personas jurídicas. La ausencia de esta herramienta ha impactado negativamente en la prima de riesgo, un indicador clave en el sector financiero, que ha experimentado un aumento del 50% en el primer semestre del 2020. La necesidad de una solución cuantitativa y predictiva se convierte en imperativa para mitigar este incremento y fortalecer la posición financiera de la institución.

Metodológicamente, la construcción del modelo se ha guiado por un proceso estructurado que incluye el tratamiento de información, segmentación, aplicación de técnicas estadísticas avanzadas, y finalmente, la presentación de resultados. Se ha dado especial atención a la definición de incumplimiento de pago de crédito (default),

estableciendo criterios claros y excluyendo casos que no aportan a la precisión del modelo.

El documento se organiza de manera lógica y justificada. Inicia con antecedentes que contextualizan la problemática en el entorno de la entidad financiera, seguido de la definición de técnicas utilizadas en el proceso, detallando desde árboles de decisión hasta indicadores de discriminancia. La sección de construcción del modelo aborda el tratamiento de información, segmentación y criterios para la elección del modelo. Los resultados, presentados de manera detallada en la sección correspondiente, evidencian una disminución significativa de la prima de riesgo.

En conclusión, este trabajo aporta un enfoque estructurado y cuantitativo para la gestión de riesgos asociados a personas jurídicas en el sector financiero. Los submodelos desarrollados, respaldados por indicadores robustos, ofrecen a la institución una herramienta efectiva y diferenciada para la toma de decisiones estratégicas. Las recomendaciones finales, orientadas al monitoreo constante, la inclusión de variables macroeconómicas y la exploración de nuevas técnicas, apuntan a la sostenibilidad y mejora continua del modelo en el tiempo.

CAPITULO I: ANTECEDENTES

1.1. Datos de la Empresa o Entidad Financiera

La entidad financiera es uno de los 5 Bancos más grande del Perú. Fundado hace más de 20 años con sede en Lima, forma parte de un grupo empresarial de origen español. Su misión social es generar un impacto trascendental en el desarrollo de la sociedad, a través de acciones íntegras y transparente, su propósito es: Acompañamos sueños, generando desarrollo.

1.2. Área donde se Desarrolló el Trabajo

El trabajo se llevó a cabo en el área de Inteligencia de Riesgos, esta desempeña un papel fundamental en la creación de modelos predictivos destinados a ser empleados como herramientas estratégicas en la gestión de riesgos crediticios, con el propósito de minimizar el riesgo de crédito.

A lo largo de esta iniciativa, se logró la participación de los equipos de la Gerencia de Riesgos PJ, la Gerencia de Gestión de Portafolios de Personas Jurídicas y la Gerencia de Control y Modelos de Riesgos de Crédito.

1.3. Descripción del Portafolio

Para este portafolio se toma en consideración a todos los clientes que tienen un RUC que inicie en “20”, con saldo mayor a cero en deuda directa e indirecta y no se encuentren en un proceso de cobranza vigente. Se establecieron tres bancas importantes: Corporativa, Empresas y Negocios. La participación de estas bancas está ilustrada en la Tabla 1.

Tabla 1*Distribución del Portafolio - Enero 2021*

Banca	Operaciones	Clientes	Saldo Directo	Saldo indirecto
Corporativa	13,902	335	3,565	990
Empresa	5,761	585	1,568	391
Negocios	12,549	4,659	1,072	313
Otros	738	89	245	138

Nota. Los montos de saldo directo y saldo indirecto se encuentran en millones de soles.

1.4. Formulación del Problema

1.4.1. Problema General

¿Cómo se puede mejorar la gestión del riesgo crediticio para disminuir el valor de la prima de riesgo¹ en el contexto financiero, utilizando modelos de clasificación robustos como la regresión logística y los árboles de inferencia condicional?

1.4.2. Problema Específico

- ¿Existirán árboles de inferencia condicional que segmenten a los clientes en grupos heterogéneos?
- ¿Existirán modelos de regresión logística robustos que proporcionen un puntaje para la clasificación efectiva de clientes como buenos o malos pagadores?
- ¿Es factible identificar grupos claramente diferenciados basándose en los puntajes proporcionados por las regresiones logísticas, permitiendo así una agrupación efectiva de los clientes en categorías de buenos y malos pagadores, desde una perspectiva estadística rigurosa y considerando medidas de validez y confiabilidad en el análisis de los resultados?
- ¿La implementación de los modelos de regresión logística y árboles de inferencia condicional, será capaz de reducir el valor de la prima de riesgo, considerando un objetivo de disminución de al menos un 20%, dado que experimentó un aumento del 50% en el último semestre del 2020?

¹ Es un indicador financiero utilizado por instituciones bancarias para evaluar el nivel de riesgo asociado a su cartera de clientes. Se calcula mediante la relación entre la suma de los saldos de los últimos 12 meses y el promedio de las provisiones correspondientes al mismo periodo

1.5. Objetivos

1.5.1. *Objetivo General*

Evaluar el valor de la prima de riesgo mediante la implementación de modelos de clasificación robustos, como regresión logística y árboles de inferencia condicional. Estos modelos permitirán la identificación de clientes como buenos o malos pagadores, basándose en grupos de riesgo definidos.

1.5.2. *Objetivo Específico*

- Identificar árboles de inferencia condicional que segmenten a los clientes en grupos heterogéneos, identificando comportamientos homogéneos dentro de cada grupo
- Desarrollar modelos de regresión logística robustos para clasificar a los clientes de manera efectiva como buenos o malos pagadores.
- Establecer grupos de riesgo a través de los puntajes generados por las regresiones logísticas.
- Evaluar la evolución de la Prima de Riesgo posterior a la implementación de los modelos de regresión logística y árboles de inferencia condicional, destacando las mejoras en la gestión del riesgo crediticio.

CAPITULO II: DEFINICIÓN DE TÉCNICAS

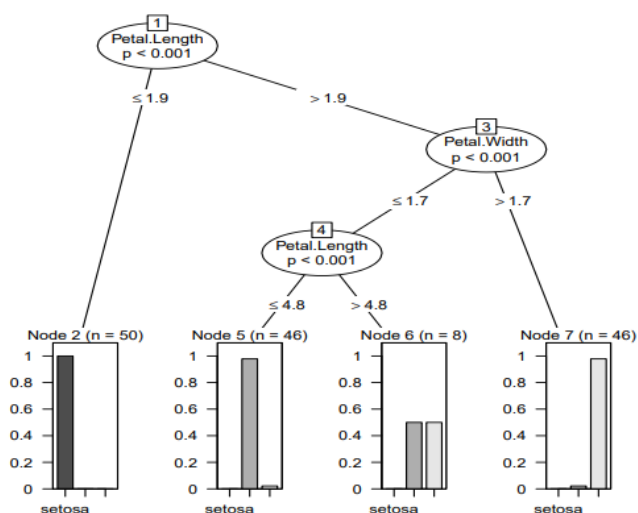
2.1 Árboles de Decisión

Es una técnica estadística que permite analizar decisiones secuenciales basada en el uso de resultados y probabilidades asociadas. El nombre de árbol de decisión proviene de la forma que adopta el modelo, semejante a un árbol. Está formado por múltiples nodos cuadrados, que representan los puntos de decisión, y de los cuales surgen ramas que representan las distintas alternativas.

Para el caso de este documento, la técnica utilizada fue “árboles de inferencia condicional”. De acuerdo con Strobl, Malley & Tutz (2009), esta técnica no paramétrica tiene como principal característica que el espacio ocupado por todas las variables predictivas es recursivamente particionado en un conjunto de áreas rectangulares. Este árbol se inclinará más en pruebas estadísticas robustas para una característica determinada (incumplimiento de pago - default) para determinar su significación estadística tal como se muestra en la Figura 1.

Figura 1

Ejemplo de Árbol de Decisión



En este análisis, se utiliza el software R, con la librería party documentada por Hothorn T. et al. (2012).

2.2 Clúster de Variables

El clúster de variables divide un conjunto de variables numéricas en clúster jerárquicos o también en disjuntos. Dentro cada clúster se encuentra una combinación lineal de las variables que pueden ser un componente principal o un componente centroide.

El clúster de variables puede ser usado como un método para reducir variables. Un conjunto grande de variables puede ser regularmente reemplazado por un conjunto de componentes clúster con una pequeña pérdida de información.

2.3 Análisis de Correlación

El análisis de correlación emplea métodos para medir el grado de asociación entre dos o más variables. El concepto de correlación está estrechamente vinculado al concepto de regresión, pues, para que una ecuación de regresión sea razonable, el ploteo de los puntos muestrales debe estar ceñido a la ecuación de regresión; además el coeficiente de correlación debe ser:

- Grande cuando el grado de asociación es alto, cerca de +1 o -1, y pequeño, cerca de cero, cuando es bajo.
- Independiente de las unidades en que se miden las variables.

2.4 Indicador Pesos de Evidencia o Weight of Evidence (WOE)

Este calcula el poder de predicción de cada atributo o grupo de atributos, diferenciando las proporciones de buenos y malos por atributo. Se define:

$$WOE_{ij} = 100 * \ln \left(\frac{\text{Distribución de buenos en el atributo } j \text{ de la característica } i}{\text{Distribución de malos en el atributo } j \text{ de la característica } i} \right)$$
$$WOE_{ij} = 100 * \ln \left(\frac{pb_{ij}}{pm_{ij}} \right) = 100 * \ln \left(\frac{b_{ij} * m_i}{m_{ij} * b_i} \right)$$

Cuando se obtienen valores negativos del WOE significa que se obtienen proporciones altas de malos sobre los buenos. De esta medida se toman en cuenta algunas consideraciones:

- El WOE_{ij} varía dependiendo de la forma en que se agrupan los atributos. Se acostumbra a ordenar de manera creciente al WOE.
- Para que el WOE_{ij} esté definido, ninguna de las clases debe estar formada únicamente por buenos o por malos.

2.5 Monotonidad

Este indicador refleja el caso en el que la relación entre dos factores (variable independiente PD y variable independiente X) siempre se mueve en la misma dirección, aunque a diferentes tasas o medidas y que, generalmente, se espera este comportamiento en las relaciones bivariadas entre la variable de respuesta y las variables independientes.

2.6 Information Value (IV)

Siddiqi (2006) indica que es una medida de entropía que aparece en la teoría de información y se define por:

$$IV = \sum_i \left(\frac{b_{ij}}{b_i} - \frac{m_{ij}}{m_i} \right) \ln \left(\frac{b_{ij} * m_i}{m_{ij} * b_i} \right)$$

Los valores que puede tomar el estadístico IV son no negativos, y es cero cuando $\frac{b_{ij}}{b_i} = \frac{m_{ij}}{m_i}$ lo que equivale a que $b_{ij} = b'_i m_{ij}$ o $m_{ij} = m'_i b_{ij}$, como se deduce directamente de la definición.

El IV es una medida del poder de predicción global de la característica. Para este trabajo se siguió lo definido por Siddiqi (2006) aplicándolo para variables con el siguiente criterio:

- Menor a 0.02 la variable no tiene poder predictivo y se descarta
- Entre 0.02 y 0.1 la variable posee un poder predictivo débil
- Entre 0.1 y 0.3 la variable posee un poder predictivo medio.
- Más de 0.3 la variable posee un poder predictivo alto.

Siddiqi (2006) indica que las características con IV por debajo de 0.02 deben ser excluidas del modelo y cuando se obtiene un IV mayor de 0.5 se dice que la característica está sobre prediciendo. En este trabajo se siguió el mismo criterio.

2.7 Transformación de Variables

Sirve para establecer una relación monotónica entre la variable regresora y la variable dependiente, ajustando la base o utilizando los coeficientes de partición para estandarizar para el tamaño, o un valor de referencia.

2.7.1 Truncamiento:

Es la eliminación de valores que se consideran anormales del conjunto de datos. En el caso de la puntuación de créditos puede ocurrir para casos que no son típicos de la actividad normal, sobre todo cuando se encuentran fuera del control de esa unidad de negocio. Para este trabajo no se ha usado esta técnica, ya que parafraseando Wicklin (2017) muchas estimaciones estadísticas se afectan por los valores extremos, pero estos valores extremos son tan importantes que eliminarlos o modificarlos invalida muchos de los análisis que se puedan hacer más adelante.

2.7.2 Reescalar por Rango:

Consiste en discretizar las variables, transformando variables continuas u ordinales en variables con categorías nominales que luego son incluidas en el modelo como dummies². Como dice Gupta (2019) este proceso puede ayudar a entender mejor el efecto de las variables en el modelo y reducir el impacto de las fluctuaciones de la información dentro de las variables.

Por otro lado, Tufféry (2011) ha demostrado que las categorías nominales tienen un mejor rendimiento en los modelos en comparación que las variables continuas en las regresiones logísticas.

2.8 Modelo Logit

La regresión logística es una técnica estadística ampliamente utilizada en los scores crediticios. En estos se considera un conjunto de variables independientes o

² Es una variable binaria que toma valores de 0 o 1, utilizada para representar categorías o grupos en análisis estadísticos y modelos, donde 0 y 1 representan la ausencia o presencia de una característica específica

variables predictoras y una variable dependiente con dos posibles valores, esta puede expresarse como:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k$$

Donde $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ son parámetros desconocidos, x_1, x_2, \dots, x_k son variables explicativas y p es la probabilidad del evento analizado que, en el ámbito de interés de este documento, es la probabilidad a priori de que un individuo incumpla su compromiso de pago. Es debido a esta definición que se dice que la variable p es una variable no observable que posee una función de distribución Bernoulli que toma los valores 0 y 1.

Es más, esta característica requiere que el proceso de estimación de los valores de los parámetros poblacionales sea el método de la máxima verosimilitud para evitar relaciones no lineales entre las variables explicativas y la explicada.

Así, en el caso de una variable dependiente dicotómica “Y”, se asume que los valores estimados por la función logística, o valores de los logits “L” de “Y”, se relacionan de forma lineal con las variables explicativas (x_i). En cambio, el ODDS y las probabilidades estimadas no son funciones lineales. Algunos supuestos para tener en cuenta en la construcción de modelos de regresión logística multivariable son: La independencia entre observaciones y la ausencia de error en la medida de las variables.

2.9 Multicolinealidad

La multicolinealidad se manifiesta cuando existe una fuerte relación lineal entre dos o más variables predictoras en un modelo de regresión. En otras palabras, una variable puede ser predicha a partir de las demás con una precisión considerable. Esta situación conlleva a problemas en la estimación de los coeficientes del modelo, ya que la variabilidad única de cada predictor se vuelve difícil de distinguir (Gujarati, 2003, p. 327-371).

Pero tal como dicen Das & Chatterjee (2011), este problema se puede solucionar de muchas maneras: aquí se está optando por retirar variables o por hacer transformaciones de estas.

2.10 Principio de Parsimonia

El principio de parsimonia (la navaja de Ockham) induce a optar por un modelo sencillo en vez de uno complicado. Dado un conjunto de posibles explicaciones igualmente buenas, la más sencilla se convierte en la mejor; cuantos menos parámetros intervengan en el modelo, relaciones lineales o con pocos factores señalan pistas que orientan nuestra búsqueda. Sin embargo, no exageremos en la sencillez del modelo.

2.11 Construcción del Scorecard (escalamiento)

El scorecard se construye con los estimadores de los parámetros de la regresión logística. Los puntajes del score son resultado de un reescalamiento y una traslación del modelo logístico, dado por la ecuación:

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

Donde offset es un término de traslación (o compensación) y Factor es un término de reescalamiento. Ellos deben satisfacer condiciones impuestas por el desarrollador del modelo. Este procedimiento permite la estandarización del score para que diferentes scorecard sean comparables. Los valores del scorecard son el resultado de una transformación de los coeficientes β_i del modelo de regresión logística. Se acostumbra a calibrar el scorecard de tal manera que cada cierto incremento en el puntaje (P_0), se obtenga el doble de la relación Bueno/Malo (Odds).

Para obtener los valores de offset y factor se resuelve el siguiente sistema de ecuaciones.

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

$$Score + P_0 = Offset + Factor * \ln(2 * Odds)$$

De aquí obtenemos:

$$Factor = \frac{P_0}{\ln(2)}$$

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

2.12 Indicadores de Discriminancia

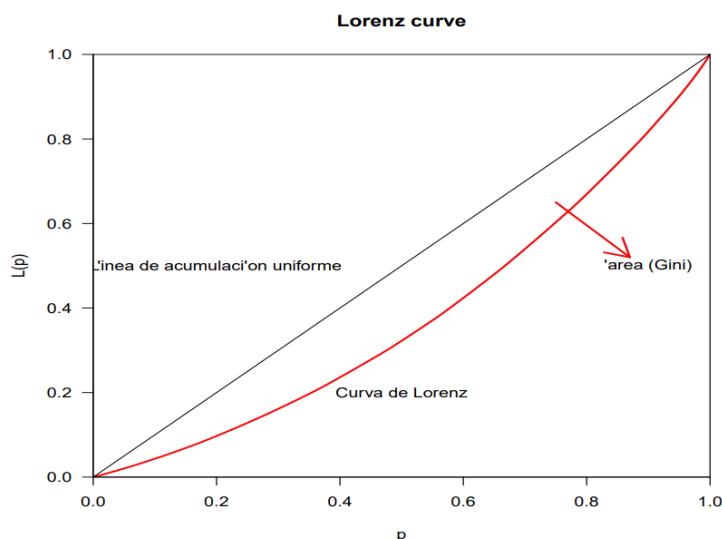
2.12.1 Coeficiente de Gini

El coeficiente de Gini es una medida de desigualdad de una variable respecto a otra, en un ordenamiento determinado. Para los fines de la evaluación de un modelo de score crediticio, este coeficiente busca medir qué tan desigual es la acumulación de los individuos aceptados y rechazados, ordenados por su probabilidad.

En la Figura 2 se muestra cómo se acumularían los individuos, en donde la línea recta indica una acumulación homogénea de ambos grupos y la curva de Lorenz, la manera como, efectivamente, se distribuyen.

Figura 2

Ejemplo de Acumulación de Grupos en Base a la Curva de Lorenz



Gráficamente, se puede deducir que la situación ideal se da cuando la curva de Lorenz toma la forma de una L invertida, es decir dos rectas: una pegada al eje de la acumulación de solicitantes buenos hasta tomar el valor de uno y la otra vertical en el eje de acumulación de solicitantes malos. Esto se debe a que, hasta cierto valor de probabilidad o nota score, se acumularán todos los solicitantes buenos y luego se empezarán a acumular los solicitantes malos. En este caso ideal, el área bajo la curva es igual a 0.

Es así como se espera que el coeficiente de Gini se acerque a 1, reflejando una buena discriminancia de los individuos. En la literatura, dicho coeficiente se calcula como:

$$Gini = \frac{0.5 - (Area_{Gini})}{0.5}$$

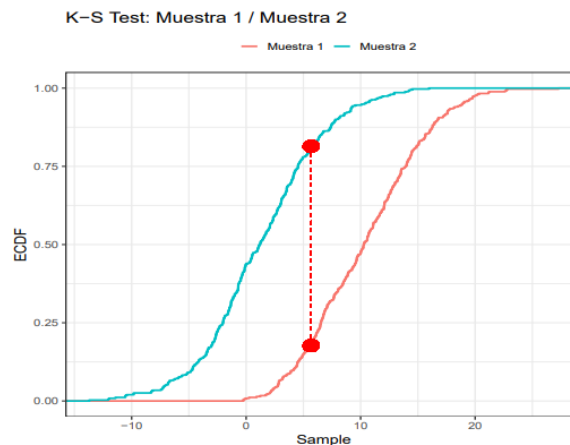
De acuerdo con los umbrales definidos en la política de gestión de riesgos de modelos de la entidad financiera, la capacidad de discriminación medida sobre la muestra test de los modelos según el tipo de modelo y el escenario es como sigue: Teniendo en cuenta que los resultados cumplan con a los umbrales definidos en la política de gestión de riesgos de modelos de la entidad.

2.12.2 Estadístico Kolmogorov-Smirnov (KS)

El estadístico Kolmogorov-Smirnov o de separación es una prueba no paramétrica empleada para la medición de la bondad de ajuste de dos distribuciones entre sí. Es así como, el KS busca verificar que la distribución de los solicitantes buenos o aceptados sea lo más distinta a la distribución de los solicitantes malos o rechazados. Bajo una situación de discriminancia perfecta, se espera que hasta cierta probabilidad se hayan distribuido todos los individuos buenos de la muestra y, a partir de ese punto, se acumulen los individuos malos; por lo que, la diferencia de las acumulaciones de deudores en esa probabilidad está en el rango de valores que van desde 0 hasta 1. Esta separación se muestra en la Figura 3 usando dos muestras que grafican la diferencia (Tineo Nieves, 2019, p. 23).

Figura 3

Ejemplo de Diferencia de Muestras como Indicador de Kolmogorov-Smirnov.



De lo anterior se entiende que cuanto más cercano a 1 sea el estadístico KS, mejor será la performance del modelo. Para este trabajo nos guiaremos de lo establecido por Halim & Humira (2014) que se muestra en el Tabla 2, considerando estos valores sólo para el modelo final.

Tabla 2

Criterios de Indicador K-S

Valor de KS	Credit Scoring
K-S < 15%	El modelo no es útil
15% <= K-S < 20%	Discriminancia pobre pero potencialmente útil, el impacto debe ser evaluado con cuidado
20% <= K-S < 28%	Discriminancia pobre pero útil
28% <= K-S < 35%	Discriminancia estándar, definitivamente útil
35% <= K-S < 45%	Alta discriminancia del modelo
K-S >= 45%	Muy alta calidad del modelo

Nota. Tomado de *Credit scoring modeling* (p.20), Halim, S., & Humira, Y. (2014).

Jurnal Teknik Industri

2.12.3 Curva ROC

La curva ROC (acrónimo de Receiver Operating Characteristic, o Característica Operativa del Receptor) es una representación gráfica de la sensibilidad frente a (1 – especificidad) para un sistema clasificador binario según se varía el umbral de discriminación. Otra interpretación de este gráfico es la representación de la razón o ratio de verdaderos positivos (VPR = Razón de Verdaderos Positivos) frente a la razón o ratio de falsos positivos (FPR = Razón de Falsos Positivos) también según se varía el umbral de discriminación (valor a partir del cual decidimos que un caso es un positivo). ROC también puede significar Relative Operating Characteristic (Característica Operativa Relativa) porque es una comparación de dos características operativas (VPR y FPR) según cambiamos el umbral para la decisión (Tineo Nieves, 2019, p. 24).

Para este trabajo, utilizamos el indicador que es el área bajo la curva ROC.

2.12.4 Divergencia

La divergencia mide la diferencia entre las medias de dos distribuciones estandarizadas usando las varianzas y tiene la siguiente expresión:

$$Divergencia = \frac{2(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$$

Cuando construimos un modelo logístico que clasifica dos poblaciones, se espera que los dos grupos estén estadísticamente bien separados; esto es, la diferencia entre sus medias sea importante. Entre más pequeña la divergencia nos estará diciendo que la distribución de cada población es parecida y no sabremos diferenciar un grupo del otro, es decir para un mismo puntaje de score tendremos cantidades similares de buenos y malos.

2.12.5 Estabilidad Poblacional

Mide la diferencia entre la muestra de análisis y otra muestra. La ecuación es la siguiente:

$$F = \sum_1^n \left[\left(\left(\frac{O_i}{\sum O} \right) - \left(\frac{E_i}{\sum E} \right) \right) * \left(\ln \left(\frac{\frac{O_i}{\sum O}}{\frac{E_i}{\sum E}} \right) \right) \right]$$

Donde O y E son observaciones (población reciente) y frecuencias esperadas (muestra de análisis) respectivamente. Para este trabajo se considerarán los criterios según la Tabla 3.

Tabla 3

Criterios de Indicador de Estabilidad Poblacional.

Indicador	Interpretación
Hasta 0.1	Población es similar, no requiere investigación
A partir de 0.1 hasta 0.25	Hay algunas diferencias en la población, requieren ser analizadas
A partir de 0.25	Cambio significativo en el portafolio, se requiere investigar

2.12.6 Bootstrap

El Bootstrap o Bootstrapping es una técnica de remuestreo no paramétrico. Se usa frecuentemente para aproximar el sesgo o la varianza de un análisis estadístico, así como para construir intervalos de confianza o realizar contrastes de hipótesis sobre parámetros de interés.

El Bootstrap se hace extrayendo un gran número de remuestras de tamaño “n” de la muestra original aleatoriamente y sin reemplazo. Así, aunque cada remuestra tendrá el mismo número de elementos que la muestra original, mediante el remuestreo con reposición cada remuestra podría tener algunos de los datos originales representados en ella más de una vez, y algunos datos podrían no aparecer. Por lo tanto, cada una de estas remuestras probablemente será leve y aleatoriamente diferente de la muestra original. Y como los elementos en estas remuestras varían levemente un estadístico calculado a partir de una de esas remuestras probablemente tomará un valor ligeramente diferente de los otros y del original.

En este trabajo se usará el Bootstrap de manera parecida a la que muestran en Adjei and Karim (2016) y Dilsha M. and Kiruthika (2014), pero con la diferencia que se generará la metodología en la muestra de validación con la intención de ver la distribución de cada coeficiente, luego en los casos donde se encuentre que más del 5% de la distribución presenta un signo cambiado o una falta de monotonidad en los valores de los coeficientes se procederá a unir las categorías de la variable hasta que la variable se muestre de manera adecuada o en su defecto se retirará la variable.

CAPITULO III: CONSTRUCCIÓN DEL MODELO

Para la construcción del modelo primero se realizó el tratamiento de información, luego se siguió un proceso iterativo que inicia con la selección de las variables iniciales y luego en cada modelo (incluyendo la creación de la segmentación).

3.1 Tratamiento de Información

El inicio de las actividades de desarrollo de modelos comprende una etapa de obtención y preparación de la información a fin de cumplir con los estándares de precisión, integridad, completitud y oportunidad de los datos empleados en el modelo.

3.1.1 Fuentes de Información

Con el objetivo de garantizar la calidad de los datos, se seleccionaron las siguientes fuentes de información crediticia.

A. Bases Internas: Base de datos de uso interno que contienen información sobre nombre, pagos, atrasos, fecha de constitución, accionistas, entre otros datos del universo de clientes del Banco Interamericano de Finanzas.

Tanto la información externa, como la interna se encuentra detallada en la sección 4.1.

B. Reporte Crediticio Consolidado (RCC): Es un reporte mensual que la Superintendencia de Banca, Seguros y AFPs (SBS) emite a todas las entidades financieras bancarias. El RCC contiene a todas las personas (jurídicas y naturales) que participan en el mercado crediticio, las cuales se identifican con un código único, y muestra la información sobre sus activos financieros para el cálculo del gasto de provisión que los bancos deben efectuar por mandato de la SBS.

La información es enviada mensualmente, con un desfase de hasta dos meses

3.1.2 Descripción del Universo de Observaciones

El universo consiste en todas las empresas que posean deudas crediticias directas o indirectas en la entidad financiera que no se encuentren bajo un proceso de cobranza, refinanciamiento o castigo. Se excluyeron aquellas observaciones que únicamente poseen deuda en sobregiros.

La información disponible corresponde a los cierres que van desde julio 2017 hasta junio 2019; esto da como resultado una base de registros con clientes que se repiten en cada cierre.

3.1.3 Ventanas de Tiempo

Para realizar el desarrollo de los modelos, se establecieron dos ventanas de tiempo:

a) La Ventana de Observación; en la cual se observa el comportamiento del individuo a través del análisis de un conjunto de variables. Aquí se consideran los clientes observados en los cierres de mes que van desde julio 2017 hasta diciembre 2018, haciendo un total de 18 meses, para definir la muestra de desarrollo. Por otro lado, las variables tienen diferentes ventanas de tiempo, considerando como tiempo límite de 48 meses para todos los productos en RCC y de 6 meses para información interna. Estos límites se establecen debido a la disponibilidad de datos. ´

b) La Ventana de Definición; en la cual se observa la performance del individuo en el pago de sus deudas. El número de meses que se consideran en esta ventana es de 12 meses, partiendo desde el siguiente mes posterior al cierre observado.

3.1.4 Definición de Incumplimiento de Pago de Crédito (Default)

Es en la “ventana de definición” en el cual se determina el incumplimiento de pago de crédito (default). El incumplimiento de pago de crédito (default) se define como la probabilidad que un crédito registre una de las siguientes condiciones:

- Clasificación como “deficiente”, “dudoso” o “pérdida”.

- Saldo en situación contable de judicial, refinanciado o castigo; observado en la entidad financiera o en el sistema financiero.
- Situación de cobranza (Banca de recuperaciones³).
- Deuda mayor a 60 días por un saldo mayor a S/ 5 mil.
- Proporción de la exposición directa mayor a 60 días en el sistema financiero, represente al menos un 5% de esta misma, sin incluir aquellos clientes que solo registren el producto sobregiro.

Es claro que la mayoría de estos comportamientos revelan un real incumplimiento de crédito para una empresa del sistema financiero y mucho más para un cliente de esta entidad financiera. Como sustento adicional se considera que, de acuerdo con el reglamento regulatorio peruano, las tasas de provisiones aplicadas a los créditos “deficiente”, “dudoso” y “pérdida” para el cálculo de provisiones por riesgo crediticio son de 25%, 60% y 100%, respectivamente.

En la Tabla 4 se puede observar la transición que sucede en los clientes del banco (persona jurídica) en el lapso de un año (desde diciembre 2018 hasta diciembre 2019), con el objetivo de tener una marca que indique cuando un cliente ya no puede recuperarse de un calificativo asignado.

Según el análisis realizado a los deudores en RCC en diciembre 2018, un total de 79% de deudores pasaron de un calificativo de “normal” o “CPP” a “deficiente”, “dudoso” o “pérdida” en el periodo evaluado después de doce meses. Más aún, un 96.8% de los que estaban en situación “pérdida” en el periodo evaluado mantuvieron esa clasificación luego de un año. El análisis usa un lapso de 12 meses debido a que la tasa de variación del porcentaje de deudores que caen en incumplimiento de pago de su crédito (default) ya no evidencia un crecimiento que justifique ampliar el horizonte.

³ Clientes con calificación deteriorada. Esta Banca tiene como objetivo definir la estrategia más adecuada para recuperar los créditos atrasados.

Tabla 4

Matriz de Transición de Clasificación SBS de Deudores (Diciembre 2018 vs Diciembre 2019)

Clasificación SBS en diciembre 2018	Clasificación SBS en diciembre 2019 (%)				
	Normal	CPP	Deficiente	Dudoso	Pérdida
Normal	86.3	4.8	2.3	3.1	3.4
CPP	32.0	27.8	13.8	12.4	14.0
Deficiente	14.9	6.4	18.1	18.7	41.8
Dudoso	9.8	0.7	5.2	13.9	70.4
Pérdida	1.2	0.3	0.5	1.2	96.8

3.1.5 Exclusiones del Modelo

Del total de observaciones identificadas (base tomada entre julio 2017 – diciembre 2018), se excluyen los clientes que cumplen con alguna de estas características (observadas en el mes de análisis):

- A. Calificativo deficiente, dudoso o pérdida en el último mes:** Estos clientes ya se encuentran en incumplimiento de pago de su crédito (default) en el Reporte Crediticio Consolidado (RCC), según la definición descrita para el modelo.
- B. No reportado en los últimos 24 meses:** Estos clientes no tienen historia crediticia que permita calcular una probabilidad de incumplimiento de pago de su crédito (default).
- C. Reportado sólo en el último mes:** Un mes no brinda suficiente información para poder estimar una probabilidad de incumplimiento de pago de su crédito (default) ni extraer conclusiones acerca del comportamiento de pago del individuo.
- D. Sin deuda en el periodo de desempeño:** Dificultad para determinar el comportamiento del cliente para cada periodo evaluado, por ende, no se puede definir una probabilidad de incumplimiento de pago de su crédito (default).

E. Pocos meses con deuda en el periodo de desempeño: Se establece que el cliente debe tener por lo menos 6 meses de comportamiento con deuda para cada periodo evaluado o caer en incumplimiento de pago de su crédito (default). Lo que suceda primero, debido a la dificultad en determinar el comportamiento en un corto periodo.

F. Default mes base: Se considera al cliente que registra algunas de las consideraciones descritas en la sección 3.1.4.

G. Solo Sobregiro: Esto es debido a que sobrevaloran el incumplimiento de pago de su crédito (default), al no poder hallar una responsabilidad de crédito específica en estas deudas.

A continuación, se presenta un cuadro resumen de las exclusiones del modelo para cada uno de los segmentos:

Tabla 5

Resumen de Caídas por Exclusiones - Segmento 1

Filtros	Clientes	Default	Caídas	Default%
Base Inicial (persona jurídica del periodo de evaluación)	24,095	1,415	0	5.87
Deficiente/Dudoso/Pérdida en el último mes	24,059	1,381	36	5.74
No reportado en los últimos 24 meses	24,056	1,380	3	5.74
Reportado solo en el último mes	24,056	1,380	0	5.74
Sin deuda en el periodo de desempeño	24,026	1,380	30	5.74
Pocos meses con deuda en el periodo de desempeño	23,782	1,374	244	5.78
Default mes base	53,564	1,167	218	4.85
Solo sobregiro	23,163	1,134	401	4.88

Tabla 6

Resumen de Caídas por Exclusiones - Segmento 2

Filtros	Clientes	Default	Caídas	Default%
Base Inicial (persona jurídica del periodo de evaluación)	18,916	2,816	0	14.89
Deficiente/Dudoso/Pérdida en el último mes	18,350	2,266	566	12.35
No reportado en los últimos 24 meses	18,310	2,251	40	12.29
Reportado solo en el último mes	18,302	2,250	8	12.29
Sin deuda en el periodo de desempeño	18,206	2,250	96	12.36
Pocos meses con deuda en el periodo de desempeño	17,507	2,236	699	12.77
Default mes base	17,191	1,935	316	11.26
Solo sobregiro	16,751	1,835	440	10.95

Tabla 7*Resumen de Caídas por Exclusiones - Segmento 3*

Filtros	Clientes	Default	Caídas	Default%
Base Inicial (persona jurídica del periodo de evaluación)	79,644	45,777	0	57.48
Deficiente/Dudoso/Pérdida en el último mes	54,036	21,592	25,608	39.96
No reportado en los últimos 24 meses	51,412	19,899	2,624	38.70
Reportado solo en el último mes	50,463	19,277	949	38.20
Sin deuda en el periodo de desempeño	49,760	19,274	703	38.73
Pocos meses con deuda en el periodo de desempeño	47,877	19,171	1,883	40.04
Default mes base	41,111	12,650	6,766	30.77
Solo sobregiro	33,366	8,975	7,745	26.90

En general las exclusiones se pueden resumir de la siguiente manera:

Tabla 8*Resumen de Caídas por Exclusiones – Base Total*

Filtros	Clientes	Default	Caídas	Default%
Base Inicial (persona jurídica del periodo de evaluación)	122,655	50,008	0	40.77
Deficiente/Dudoso/Pérdida en el último mes	96,445	25,239	26,210	26.17
No reportado en los últimos 24 meses	93,778	23,530	2,667	25.09
Reportado solo en el último mes	92,821	22,907	957	24.68
Sin deuda en el periodo de desempeño	91,992	22,904	829	24.90
Pocos meses con deuda en el periodo de desempeño	89,166	22,781	2,826	25.55
Default mes base	81,866	15,752	7,300	19.24
Solo sobregiro	73,280	11,941	8,586	16.30

3.1.6 Definición de Muestras de Trabajo

Para la elaboración de los modelos, se establecieron tres tipos de muestras de trabajo, según el universo en la sección 3.1.2:

- A. Muestra de desarrollo;** representa el 70% de la base tomada entre julio 2017 y diciembre 2018.
- B. Muestra de validación;** representa el 30% de la base tomada entre julio 2017 y diciembre 2018.
- C. Muestra de control (fuera de la muestra);** representa el 100% de la base tomada entre enero 2019 y junio 2019.

3.1.7 Definiciones

En esta sección se listan algunas definiciones relevantes para comprender las variables a partir de las cuales se realizaron transformaciones, truncamientos, agrupamientos o reagrupamientos para obtener las variables finalmente incluidas en los modelos.

- A. Entidades reportantes:** “Entidades del sistema financiero que reportan a un cliente con algún crédito activo o con alguna línea crediticia sin uso en el RCC” (Tineo Nieves, 2019).
- B. Saldo deudor:** “Es la deuda activa que tiene un cliente con alguna entidad del sistema financiero” (Tineo Nieves, 2019, p. 20).
- C. Calificativos de riesgo:** “Es la clasificación que determina la SBS de acuerdo con los días de atraso registrados por el cliente en cada cuenta de las entidades donde se mantenga saldo deudor” (Tineo Nieves, 2019, p. 20).
- D. Calificativo missing:** “El cliente tendrá un calificativo missing cuando no registre saldo deudor en el sistema financiero” (Tineo Nieves, 2019, p. 20).
- E. Calificativo normal:** “Registra el valor 1 si el individuo no presenta ninguna clasificación peor a normal en más del 5% de su saldo total en el sistema financiero” (Tineo Nieves, 2019, p. 20).
- F. Calificativo CPP:** “Registra el valor 2 si el individuo presenta clasificación CPP en más del 5% de su saldo total del sistema financiero y no tiene ninguna otra clasificación peor que represente más del 5% del total de su saldo deudor” (Tineo Nieves, 2019, p. 20).
- G. Calificativo deficiente:** “Registra el valor 3 si el individuo presenta clasificación deficiente en más del 5% de su saldo total del sistema financiero y no tiene ninguna otra clasificación peor que represente más del 5% del total de su saldo deudor” (Tineo Nieves, 2019, p. 20).
- H. Calificativo dudoso:** “Registra el valor 4, si el individuo presenta clasificación dudosa en más del 5% de su saldo total del sistema financiero y

no tiene ninguna otra clasificación peor que represente más del 5% del total de su saldo deudor” (Tineo Nieves, 2019, p. 20).

I. Calificativo pérdida: “Registra el valor 4, si el individuo presenta clasificación pérdida en más del 5% de su saldo total del sistema financiero” (Tineo Nieves, 2019, p. 20).

J. Periodos estables: Son periodos de dos meses consecutivos con una misma característica especial. Esta puede ser que el cliente sea reportado por la misma cantidad de entidades, de saldos, clasificación, o mantiene una misma ratio, etc. No se considera un periodo estable a dos meses consecutivos con la ausencia de la característica indicada por definición de la variable.

K. Préstamos a negocios: Se consideran a todos los clientes con deuda de microempresa y pequeña empresa.

L. Estados financieros: También denominados estados contables, son informes que emiten las empresas para dar a conocer su situación económica y financiera y los cambios que experimenta la misma a una fecha de corte determinado. Estos datos son enviados al banco al momento de iniciar operaciones con la empresa y en algunos casos se actualiza anualmente. Para el análisis realizado, se consideraron los cortes de diciembre más próximo.

M. Banca: Clasificación de empresas según el manual de Riesgos de la entidad financiera.

N. Banca corporativa: Clientes con ventas anuales mayores a US\$ 20 MM.

O. Banca empresa: Clientes con ventas anuales desde US\$ 4 MM a US\$ 20 MM.

P. Banca negocios: Clientes con ventas anuales desde US\$ 1 MM a US\$ 4 MM.

Q. Producto sobregiro: El sobregiro en cuenta corriente es un crédito directo que significa que el balance de la cuenta es negativo. El origen de los sobregiros es diverso:

- Montos no significativos por cargo de comisiones de mantenimiento, Bifnet, etc.

- Línea de crédito de avance en cuenta aprobada por riesgos.
- Límite de crédito ingresado por autonomía comercial.
- Ejecución de Cartas Fianzas, etc.

3.1.8 Construcción de la Base de Desarrollo

Para la base desarrollo del presente modelo, se determinó como insumo principal a los clientes identificados como persona jurídica con ruc que inicie en “20” y registren una deuda directa o indirecta mayor a cero, según el orden descrito en la sección 0.

3.1.9 Construcción de Variables

Para el desarrollo de las variables, en general, se usaron las fuentes internas, externas y según la definición especificada en la sección 0. Como se puede observar, la mayoría se extraen de la información de RCC y la información interna, sin algún cruce entre tablas, pero se tomaron especial consideración en las siguientes variables:

3.1.10 Ventas_Soles

Se identificaron las ventas utilizadas correspondientes al cierre de cada ejercicio, así mismo, dado que no todos los clientes registran ventas para el último periodo, se completó la información de ventas con cierres previos, teniendo una antigüedad máxima de cinco años. La fecha de corte de la información disponible al momento del desarrollo fue agosto 2020.

Como parte de la revisión en cuanto a la consistencia de la información, se revisó el cumplimiento del principio contable, dado por la siguiente formula:

$$Total Activo = Total Pasivo + Total Patrimonio$$

Se observó la existencia de clientes que no cumplen el principio indicado, para lo cual se revisó con mayor detalle y se identificó que la diferencia era debido a que se está desagregando un monto de pasivos (corriente y no corriente) en la cuenta “Ingresos Diferidos”, estos dos puntos totalizados en la cuenta “GANANCIAS DIFERIDAS”.

Así mismo con la finalidad de tener una consistencia, se pasa a explicar el caso del cliente “TELEFONICA DEL PERU SAA”. Se procedió a descargar la información

contable auditada de la Superintendencia de Mercado de Valores (SMV)⁴, tal como se muestra en la Figura 4.

Figura 4

Estado Financiero Auditados de la Superintendencia de Mercado de Valores (SMV)

ESTADO SEPARADO DE SITUACION FINANCIERA

ACTIVO	Nota	Al 31 de diciembre de		Al 01 de enero	PASIVO Y PATRIMONIO	Nota	Al 31 de diciembre de		Al 01 de enero
		2018	2017	de			2018	2017	de
		S/000	S/000	2017			S/000	S/000	2017
ACTIVO CORRIENTE					PASIVO CORRIENTE				
Efectivo y equivalente de efectivo	6	88,038	407,614	566,856	Otros pasivos financieros	16	459,730	396,637	857,640
Cuentas por cobrar comerciales, neto	7	1,525,749	1,266,883	1,409,966	Cuentas por pagar comerciales	17	2,365,433	2,177,215	2,335,683
Activos contractuales	9	267,299	275,446	370,070	Pasivos contractuales	9	162,252	190,706	191,577
Otras cuentas por cobrar a entidades relacionadas	21(e)	372	48,388	3,065	Otras cuentas por pagar	18	319,991	157,006	406,533
Otras cuentas por cobrar, neto	8	127,978	106,399	66,982	Otras cuentas por pagar a entidades relacionadas	21(e)	384	-	158
Inventarios	10	312,132	354,523	347,230	Otras provisiones	19	2,133,551	2,116,671	1,811,323
Otros activos no financieros	11	302,531	147,496	159,809	Ingresos diferidos	20	52,649	48,108	47,051
Activos por impuestos a las ganancias	22(f)	303,060	240,093	236,961	Total pasivos corrientes		5,483,990	5,086,343	5,649,965
Total activos corrientes		2,927,159	2,846,842	3,160,739					
ACTIVO NO CORRIENTE					PASIVO NO CORRIENTE				
Cuentas por cobrar comerciales	7	38,084	46,497	34,182	Otros pasivos financieros	16	1,689,598	1,508,799	982,086
Activos contractuales	9	17,867	27,274	41,638	Pasivos contractuales	9	61,205	30,421	34,362
Otras cuentas por cobrar a entidades relacionadas	21(e)	587	587	-	Otras provisiones	19	459,846	464,996	604,612
Otras cuentas por cobrar, neto	8	287,383	314,436	317,511	Otras cuentas por pagar	18	73,355	30,158	50,710
Inversiones en subsidiaria y asociada	12	12,200	34,554	36,946	Otras cuentas por pagar a entidades relacionadas	21(e)	3,390	-	-
Propiedades, planta y equipo, neto	13	6,729,637	6,651,712	6,744,795	Pasivos por impuestos diferidos, neto	22	-	110,688	77,626
Activos intangibles distintos de la plusvalía, neto	14	2,276,029	2,452,716	2,585,961	Ingresos diferidos	20	125,321	151,175	172,099
Activos por impuestos diferidos, neto	22	535	-	-	Total pasivos no corrientes		2,412,515	2,296,237	1,921,495
Plusvalía	15	1,224,379	1,224,379	1,224,379	Total pasivos		7,896,505	7,382,580	7,571,460
Otros activos no financieros	11	279,273	93,033	97,698					
Total activos no corrientes		10,865,974	10,845,188	11,083,110	PATRIMONIO	23			
					Capital emitido		2,876,152	2,876,152	2,876,152
					Primas de emisión		77,899	77,899	77,899
					Otras reservas de capital		576,453	584,857	545,435
					Resultados acumulados		2,368,477	2,778,645	3,179,697
					Otras reservas de patrimonio		(2,353)	(8,103)	(6,794)
					Total patrimonio		5,896,628	6,309,450	6,672,389
TOTAL DE ACTIVOS		13,793,133	13,692,030	14,243,849	TOTAL PATRIMONIO Y PASIVO		13,793,133	13,692,030	14,243,849

En relación con la información contable de la Superintendencia de Mercado de Valores (SMV), se observa la desagregación de la cuenta de Ganancias Diferidas de la siguiente manera:

Ganancias Diferidas

$$= \text{Ingresos Diferidos(Corriente)} + \text{Ingresos Diferidos(No Corriente)}$$

$$\text{Ganancias Diferidas} = S/52,649 + S/125,321 = S/177,970 \text{ miles de soles}$$

Con respecto a la información contable en la Entidad Financiera, se observa que la desagregación de las cuentas mencionadas en encuentran totalizadas en la cuenta “Ganancias Diferidas”. Ver Figura 5.

⁴ Página Web: <http://www.smv.gob.pe/>

Figura 5

Estado Financiero en el Sistema de la Entidad Financiera.

**** PROPUESTAS DE CREDITOS ****							16:12:45
ESTADOS FINANCIEROS							
TELEFONICA DEL PERU SAA			Formato: E EMPRESAS				
Auditores PWC			2018		SOL Miles		
	AUDITADO		AUDITADO		AUDITADO		
	DIC-16	%	DIC-17	%	DIC-18	%	
TOTAL PASIVO CORRIENTE	3681,403	26	2921,564	21	3297,790	24	
DEUDA BANCARIA A LARGO PLAZO	184,488	1	1508,799	11	1689,598	12	
OTRAS DEUDAS A LP	797,598	6	525,575	4	137,950	1	
PROVISION COMP. TIEMPO SERVS	2405,199	17	2116,671	15	2593,197	19	
CTAS POR PAGAR ACC/EMP.RELACIONADA							
IMPUESTO A LA RENTA DIFERIDO			110,688	1			
TOTAL PASIVO LARGO PLAZO	3387,285	24	4261,733	31	4420,745	32	
TOTAL PASIVOS	7068,688		7183,297	1	7718,535	5	
GANANCIAS DIFERIDAS	437,722	3	199,283	1	177,970	1	
PATRIMONIO							
CAPITAL SOCIAL	2876,152	21	2945,948	22	2954,051	21	
EXCEDENTE DE REVALUACION							
RESERVAS	623,492	4	584,857	4	574,100	+	

Con lo cual se observa que el principio contable cumple para el cliente analizado.

Por otro lado, se procedió a revisar la base definida de acuerdo al siguiente principio contable:

$$\text{Total Activo} = \text{Total Pasivo} + \text{Total Patrimonio} + \text{Ganancia Diferida}$$

Se analizó el periodo de agosto 2020, para aquellos deudores que registran un Balance de Comprobación al cierre de cada ejercicio a partir del 2011, registrándose un total de 57,017 registros, así mismo aquellos que registran una diferencia fueron 35 registros bajo el principio contable.

Tabla 9

Cantidad de Registros que no Cumplen el Principio Contable por año

Año de Estado Financiero	Registros
11	14
12	5
13	4
14	4
15	6
16	2

Con lo cual se concluye que la información de los Estados Financieros (EEFF) bajo el principio contable cumple las condiciones suficientes para ser utilizadas.

3.1.11 Grupo_Score_Empresa

Se tomó la información disponible en el momento del desarrollo del modelo (del periodo 201812 en adelante) y dado que hasta ese momento no se registraba la información histórica con cortes mensuales, para la muestra de desarrollo se utilizó la información del periodo 201812 y para control se utilizó la información mensual disponible.

Para definir el valor de esta variable, se realizaron los siguientes pasos:

- H.** Obtener los accionistas de cada empresa (cliente) que registren una participación de más del 5% (participación accionistas).
- I.** Considerar los cinco mayores accionistas.
- J.** Obtener la probabilidad de incumplimiento de pago teórica (PD), según el Score de Buro de Personas Naturales, de cada uno de los accionistas de cada empresa (PD Accionista).

Para los accionistas que no contaban con un valor de probabilidad de incumplimiento de pago teórica (PD) debido a la falta de información disponible en los meses de desarrollo, se aplicó un enfoque alternativo. Se calculó el promedio de la probabilidad de incumplimiento de pago (default) de los deudores que tenían un puntaje (score) entre 1 y 10 en junio de 2019, ya que esta fue la información disponible del Score de Buró de personas naturales en ese momento, para calcular la probabilidad de incumplimiento se necesita 12 meses de historial de desempeño. Los resultados de este cálculo se presentan en la Tabla 10.

Tabla 10*Distribución del Grupo de Riesgo - Junio 2019*

Grupo Score	Clientes	Default	Default%
1	662'824	8'490	1.28
2	658'546	14'031	2.13
3	694'743	24'592	3.54
4	630'697	31'980	5.07
5	751'233	48'964	6.52
6	673'266	56'168	8.34
7	665'181	77'909	11.71
8	741'397	115'256	15.55
9	681'830	156'640	22.97
10	671'375	342'318	50.99
93	98'365	18'699	19.01
Promedio grupo 1 al 10			12.80

También, se asignan los siguientes valores:

- Para los accionistas con un solo reporte (Score 93 - un solo reporte) se les asigna el valor observado 19.01%, ver Tabla 10.
- Aquellos accionistas que se encuentren en situación de dudoso o pérdida en el último mes, se les asigna una probabilidad de incumplimiento de pago teórica (PD) de 100%.

K. Realizar el cálculo de la probabilidad de incumplimiento de pago de la empresa (PD de la empresa), como un promedio ponderado, mediante la siguiente formula:

$$PD \text{ de la empresa} = \frac{\sum_{i=1}^{n=5} (\text{Participación accionista}_i * PD \text{ Accionistas}_i)}{\sum_{i=1}^{n=5} \text{Participación accionista}_i}$$

Donde: PD Accionista es la probabilidad de incumplimiento de pago teórica del accionista, definido con el Score de Buro de persona natural.

L. Una vez tenido el promedio ponderado de la PD de cada empresa, se procedió a aplicar los cortes de PD utilizados en el modelo de Score Buro para persona natural descritos la **Tabla 11**.

Tabla 11

Rangos de PD por Grupo de Riesgo - Base Desarrollo Score Buro para Persona

Natural - Abril 2018

Grupo Score	Rangos de PD
1	<0.0000000 - 0.0211968]
2	<0.0211976 - 0.0323708]
3	<0.0323708 - 0.0472088]
4	<0.0472093 - 0.0646692]
5	<0.0646693 - 0.0862198]
6	<0.0862202 - 0.1133800]
7	<0.1133810 - 0.1513860]
8	<0.1513870 - 0.2088480]
9	<0.2088480 - 0.3407520]
10	<0.3407530 - 1.0000000]

Cabe aclarar que, debido a la asignación de PD de cada accionista, se tienen los siguientes escenarios:

- Para las empresas sin accionistas registradas, se considera como si la empresa tuviera un solo accionista sin información, asignándosele el valor de 12.8% en su PD y por ende un grupo de riesgo final de 7.
- Para las empresas con un único accionista sin información se le asigna el valor de 12.8% en su PD y por ende un grupo de riesgo final de 7.
- Para las empresas con un único accionista con un solo reporte se le asigna el valor de 19% en su PD y por ende un grupo de riesgo final de 8.
- Para las empresas con un único accionista con el último mes en dudoso o pérdida se le asigna el valor de 100% en su PD y por ende un grupo de riesgo final de 10.

Finalmente, para asegurar la continuidad de los datos a nivel mensual, se realizó la solicitud al área de Información sobre Gestión de Riesgos para realizar una copia de la fuente al cierre de cada mes.

Para realizar otras consistencias de datos a las demás variables se procedió a realizar un análisis bivariado para cada una, descrito en esta sección.

3.2 Segmentación

Antes de desarrollar los modelos, se analizó la muestra para establecer grupos heterogéneos con comportamientos homogéneos dentro de cada grupo. Para establecer estas segmentaciones, se tomaron en cuenta tres frentes:

✓ Comportamiento en el sistema financiero, estableciendo 10 variables (descritas en la sección 0) que analizan los calificativos del cliente en los últimos 48 meses. A estas variables se les resumió en una sola, llamada Patrón Externo definida en la sección 4.1.1. Para esta segmentación se buscó establecer la mayor discriminancia posible, con la finalidad de tener grupos bien diferenciados por el riesgo.

✓ Comportamiento interno, estableciendo 10 variables (descritas en la sección 0) que analizan los pagos y atrasos del cliente en los últimos 12 meses. A estas variables se les resumió en una sola, llamada Patrón Interno definida en la sección 4.1.2. Para esta segmentación se buscó establecer la mayor discriminancia posible, con la finalidad de tener grupos bien diferenciados por el riesgo.

✓ Ventas, considerando que los clientes mantienen una diferenciación a nivel del volumen de ventas, se estableció una variable de segmentación llamada Ventas_Soles descrita en la sección 0. Esta variable forma parte de la información que proviene de los estados financieros disponibles. Para esta segmentación se buscó establecer la mayor discriminancia posible, buscando que cada grupo pueda reforzarse por variables específicas para su comportamiento.

3.3 Concentración y Penetración:

Se realizó un proceso iterativo donde se realizó al mismo tiempo un análisis de frecuencias (calculando percentiles de 6%) y un análisis de árboles de inferencia condicional. Los pasos de la iteración fueron los siguientes:

- a) Definición inicial de clases (sea por frecuencias o por árboles).
- b) Si alguna de las clases no cumple con el criterio de monotonicidad, se juntan las clases que tienen el riesgo más parecido. En este punto también intervienen los valores del WOE.
- c) Si alguna de las clases tiene una participación menor al 5%, se junta con otra clase con el riesgo más parecido.
- d) En este paso se puede descartar a una variable que tenga una clase con una participación mayor al 90%.
- e) Definición inicial de clases (sea por frecuencias o por árboles).

Una vez terminado este proceso, se siguió el criterio de Information Value (IV) definido en la sección 3.4, descartando las variables con poco poder predictivo.

3.4 Criterio Information Value (IV) y Pesos de Evidencia (WOE)

A las variables que cumplieron con la condición de concentración, se les calculó el índice Information Value (IV) teniendo en cuenta los criterios descritos en la siguiente tabla, según Siddiqi (2006):

Tabla 12

Rango del índice de Information Value (IV)

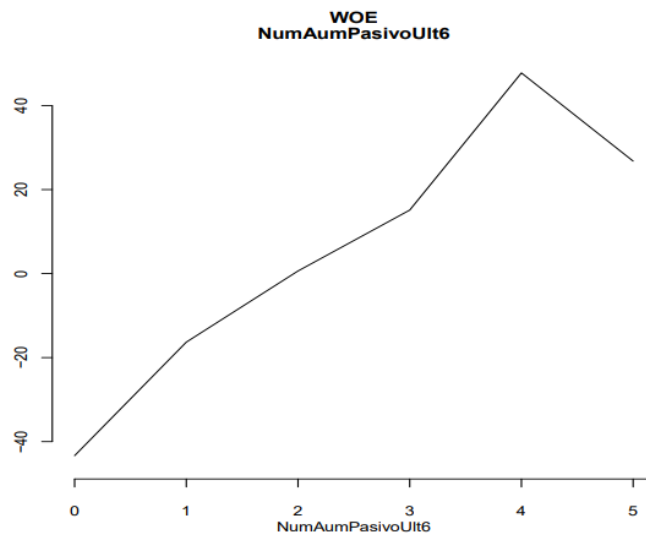
Indicador	Interpretación
$IV < 0.02$	La variable no tiene poder predictivo y se descarta.
$0.02 \leq IV < 0.1$	La variable posee un poder predictivo débil.
$0.02 \leq IV < 0.3$	La variable posee un poder predictivo medio.
$IV \geq 0.3$	La variable posee un poder predictivo alto.

A continuación, se pasó a graficar el WOE de cada variable a fin de verificar que estas cumplieran con el principio de monotonicidad. A las variables que no cumplieron con dicho principio se les aplicó el principio de transformación de variables mediante el acotamiento o el reescalamiento por rangos.

Por ejemplo, en el caso de la variable “NumAumPasivoUlt6”, se observa que la gráfica de la Figura 6 del WOE no cumple con el principio, ya que en cada punto sus pendientes varían en su signo.

Figura 6

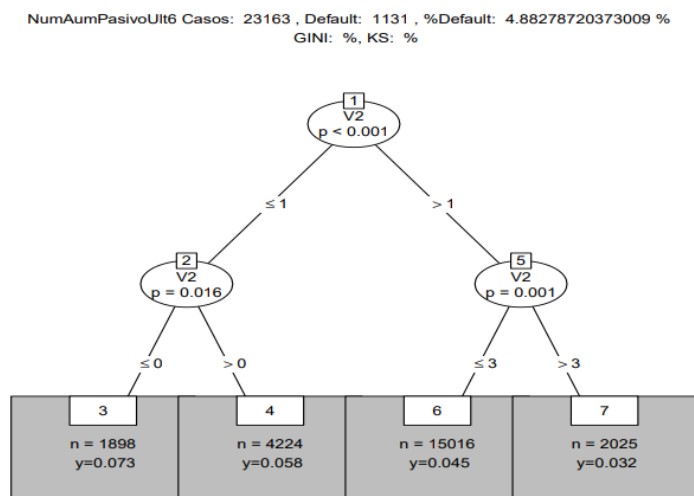
Ejemplo de Gráfica WOE para Variable no Monotónica.



Las variables similares fueron acotadas o reescaladas con la utilización de árboles de decisión, tal como se observa en la Figura 7.

Figura 7

Ejemplo Árbol de Decisión para Reescalar la Variable

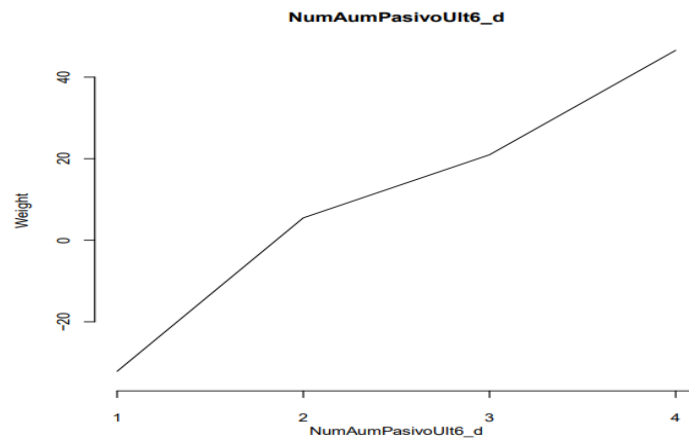


Estas transformaciones pasaron nuevamente por un estudio de concentración y de IV. “Este proceso fue iterativo, es decir, se continuaba con las particiones, hasta que la gráfica del WOE muestre el cumplimiento del principio de monotonidad” (Tineo Nieves, 2019, p. 32), o se descarta la variable por no cumplir con ambos criterios.

Volviendo al caso de la variable “NumAumPasivoUlt6”, luego de juntar los nodos que rompen el principio de monotonidad, se cumple con las condiciones de IV y de concentración, como se aprecia en la Figura 8.

Figura 8

Ejemplo de Gráfica WOE para Variable Transformada



3.5 Criterio de Correlación

El siguiente paso fue medir la correlación entre las variables por cada grupo. Para ello se tuvo en cuenta los rangos de correlación según Anderson (2007), ver Tabla 13.

Tabla 13

Rango de Correlaciones (r)

Indicador	Interpretación	Dirección
$r = 1$	Perfecto	Positivo
$0.9 < r < 1$	Fuerte	Positivo
$0.5 < r \leq 0.9$	Moderado	Positivo
$0.0 < r \leq 0.5$	Débil	Positivo
$r = 0$	No correlacionado	No correlacionado
$-0.5 \leq r \leq 0$	Débil	Negativo
$-0.9 \leq r \leq -0.5$	Moderado	Negativo
$-1.0 < r < -0.9$	Fuerte	Negativo
$r = -1$	Fuerte	Negativo

Se consideró que una o más variables están correlacionadas si su correlación es mayor a 70%. La elección de las variables que superaron los límites de correlación fue a través del mayor IV.

3.6 Significancia

A los tres grupos de cada segmento se les procedió a hacer las regresiones logísticas, aplicándoseles la técnica Stepwise⁵, arrojando cada grupo un modelo. Una vez obtenido el modelo en caso hubieran variables nominales tratadas como varias variables dummies que tengan niveles con coeficientes con signo cambiado (esto es, verificar que el beta de la variable arrojado por el modelo sea consistente con la pendiente del incumplimiento de pago de crédito) o que no cumplan con el criterio de monotonidad o que uno de los rangos de la variable figure como no significativo mientras que los demás niveles si pasaran el test de significancia; lo que se hace es juntar las clases según sea más conveniente. A estas variables se les llama variable reajustada.

3.7 Bootstrap

Aquí se procede con el modelo resultante y las variables que quedan a realizar el análisis Bootstrap definido en la sección 2.12.6. Aquellas variables que no pasan el criterio se descartan.

3.8 Aporte a Discriminancia

Por último, se revisa el aporte que puede efectuar cada variable a la discriminancia del modelo. Se observa el indicador de KS que resulta cuando se extrae una variable, luego se compara con el indicador inicial y la diferencia en porcentaje se le considera su aporte. Se retira a la variable que tenga un aporte de 0.5% o menos en todos los indicadores KS de la muestra de desarrollo.

3.9 Elección del Modelo

Para la elección del modelo se realizaron tres pasos:

- a) Se comparó los modelos definitivos de acuerdo a los siguientes indicadores:

⁵ Es un método de selección de variables en análisis estadísticos. Se realiza de manera iterativa, agregando o eliminando variables en cada paso basándose en criterios estadísticos, con el objetivo de encontrar el modelo más relevante o ajustado.

- GINI
 - Divergencia
 - AIC
 - ROC
 - Criterio de parsimonia, referido al número de variables involucradas en el modelo.
 - KS, bajo los criterios mencionados en la sección 2.12.2.
 - Estabilidad poblacional, bajo los criterios mencionados en la sección 2.12.5.
- b) Una vez seleccionado el mejor modelo, se aplicó la técnica del Bootstrap, para ver la estabilidad de las betas.
- c) Se tomó 1000 muestras sin reemplazo de la muestra de validación con la intención de ver la distribución de cada coeficiente, en los casos donde se encontró que más del 5% de la distribución presentaba un signo cambiado se procedió a unir las categorías de la variable, en el caso que todas las categorías de la variable presentan más del 5% de su distribución con un signo contrario se decidió retirar a la variable.
- d) Luego se procedió a realizar el escalamiento de los modelos, para lo que se consideró una relación de 15:1 a 660 doblando el odds cada 15 puntos. La ecuación de transformación fue la siguiente:

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

$$Factor = \frac{15}{\ln(2)} = 21.6404256$$

$$Offset = 660 - (21.6404256 * \ln(15)) = 601.396641$$

$$Score = 601.396641 - 21.6404256 * \ln(Odds)$$

CAPITULO IV: RESULTADOS

4.1 Segmentación

4.1.1 Patrón Externo

Se diseñó un conjunto de variables que reflejan el historial financiero del cliente durante los últimos 48 meses. Estas variables se dividen en dos grupos: El primer grupo consta de 7 variables que identifican a los clientes que han mantenido una clasificación "Normal" o que no han tenido una clasificación (missing) durante todo el período de evaluación. El segundo grupo incluye 3 variables que capturan el comportamiento de aquellos clientes que, durante el período de evaluación, han tenido al menos una clasificación distinta a "Normal" o han tenido algún estado en las categorías de CPP, Deficiente, Dudoso o Pérdida.

Figura 9

Esquema de Variables que Forman el Patrón Externo.

Periodo	Nombre variable	P1	P2	P3	P4	P5	P6	P7	P8		
		48 Norm	Hace 48m = 0	Hace 48m = 1	Hace 48m = 1	Hace 48m = 0	Hace 48m = 1	Hace 48m = 0	0 - 5		
			Último Mes = 1	Último Mes = 0	Último Mes = 1	Último Mes = 0	Último Mes = 0	Último Mes = 1	P8_1	P8_2	P8_3
t-1	clas_01	1	1	0	1	0	0	1	0-5	0-5	0-5
t-2	clas_02	1	1	0	0/1	0/1	0/1	0/1	0-5	0-5	0-5
t-3	clas_03	1	1	0	0/1	0/1	0/1	0/1	0-5	0-5	0-5
t-4	clas_04	1	1	0	0/1	0/1	0/1	0/1	0-5	0-5	0-5

t-45	clas_45	1	0	1	0/1	0/1	0/1	0/1	0-5	0-5	0-5
t-46	clas_46	1	0	1	0/1	0/1	0/1	0/1	0-5	0-5	0-5
t-47	clas_47	1	0	1	0/1	0/1	0/1	0/1	0-5	0-5	0-5
t-48	clas_48	1	0	1	1	0	1	0	0-5	0-5	0-5
Valores	flag	Num Norm	Num Norm	Num Norm	Num Norm	Num Norm	Num Norm	Num Norm	Num Norm	Max Clasif 48m	Hace cuanto <> Norm
	0-1	1-47	1-47	2-47	1-46	2-46	2-46	1-47	0-5	1-48	

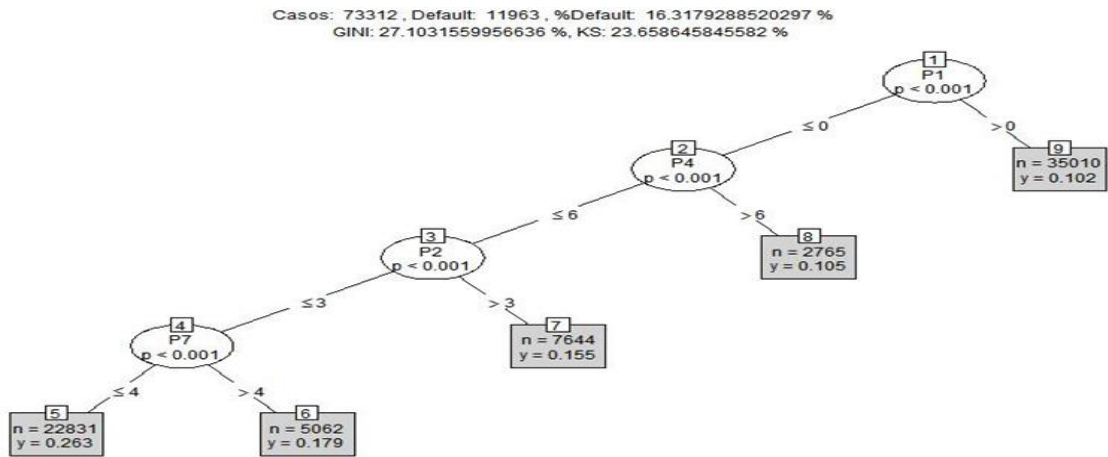
Nota. La descripción de la clasificación es: 0 es sin clasificación, 1 es "Normal", 2 es "Con Problemas Potenciales" (CPP), 3 es "Deficiente", 4 es "Dudoso" y 5 es "Pérdida".

Los patrones de comportamiento externo están descritos en la sección 0, pero en la Figura 9 se puede apreciar las combinaciones posibles entre todas las variables. Estas variables son: P1, P2, P3, P4, P5, P6, P7, P8_1, P8_2 y P8_3.

Con el fin de resumir los patrones de comportamiento, se realizó un análisis de árboles de inferencia condicional, según el mismo algoritmo, para observar cómo interactúan entre sí las variables de comportamiento externo. En la Figura 10 se puede observar el primer árbol generado.

Figura 10

Árbol Inicial con las Variables que Forman el Patrón Externo.



Debido a que se requería que el efecto de todas las variables se refleje en esta segmentación, se hicieron más árboles, esta vez aislando cada variable contra la variable objetivo (default).

Figura 11

Árbol de Inferencia Condicional de la Variable P1

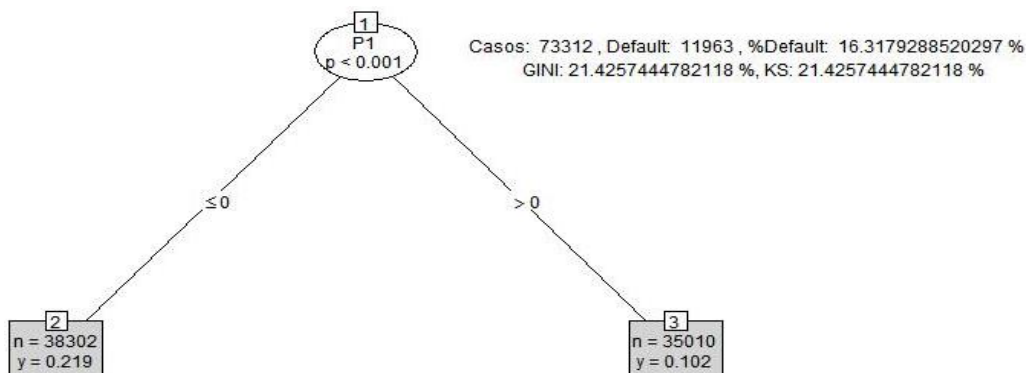


Figura 12

Árbol de Inferencia Condicional de la Variable P8_1

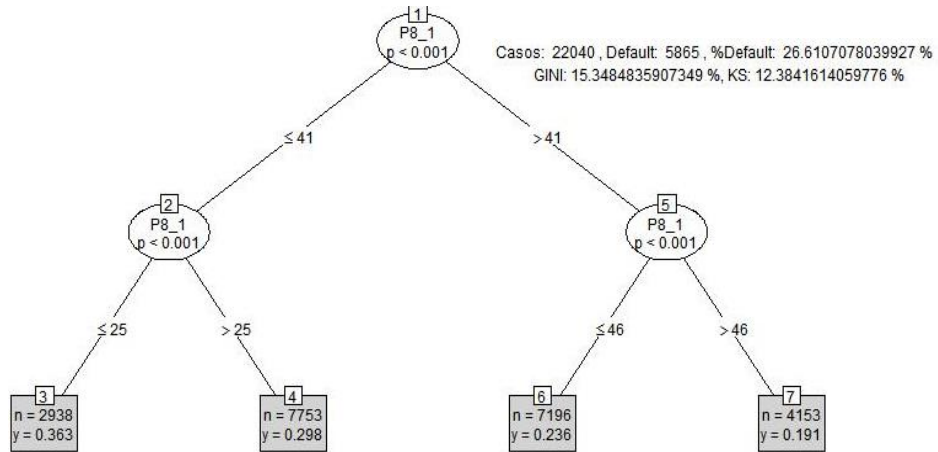


Figura 13

Árbol de Inferencia Condicional de la Variable P8_2

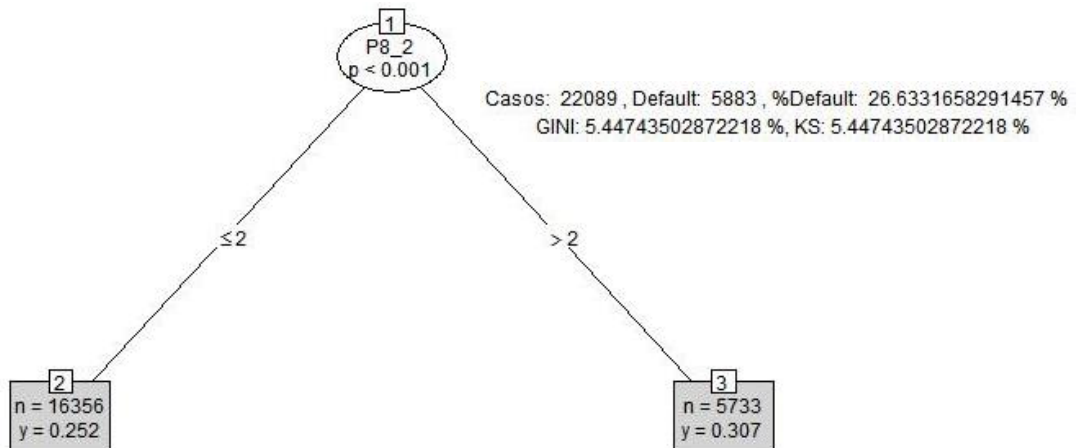
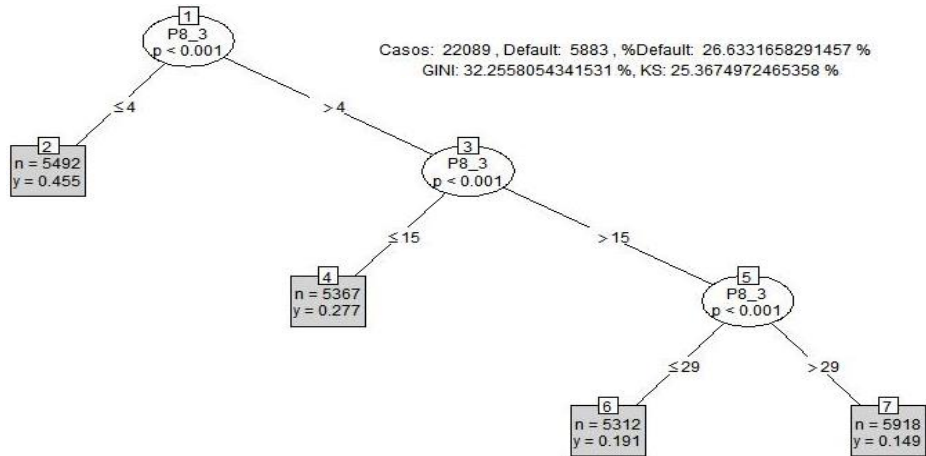


Figura 14

Árbol de Inferencia Condicional de la Variable P8_3



Estas variables por separado se ingresaron en un árbol de decisión y se obtuvo como resultado 3 segmentos (Bueno, Regular y Malo) según el nivel de default en cada grupo, tal como se observa en la Figura 15.

Figura 15

Distribución de Variables en el Patrón Externo Según Nivel de Riesgo y Participación.

Patrones	Casos	Default	Default %	Participación %
P1 > (0)	35,010	3,568	10.2%	47.8%
Patrones	Casos	Default	Default %	Participación %
P2 > (3)	7,644	1,184	15.5%	10.4%
P3 > (0)	38	6	15.8%	0.1%
P4 > (0)	2,774	295	10.6%	3.8%
P5 > (0)	232	42	18.1%	0.3%
P6 > (15)	43	2	4.7%	0.1%
P7 > (0)	5,144	926	18.0%	7.0%
P8_1 > (46) AND P8_2 <= (2) AND P8_3 > (15) and P8_3<=(29)	799	154	19.3%	1.1%
P8_1 > (46) AND P8_2 <= (2) AND P8_3 > (29)	1,965	286	14.6%	2.7%
P8_1 > (46) AND P8_2 > (2) AND P8_3 > (15) and P8_3<=(29)	110	1	0.9%	0.2%
P8_1 > (46) AND P8_2 > (2) AND P8_3 > (29)	160	13	8.1%	0.2%
Patrones	Casos	Default	Default %	Participación %
P2 <= (3) AND P2>0	195	54	27.7%	0.3%
P6 <= (15) AND P6>0	6	3	50.0%	0.0%
P8_1 <= (25) AND P8_2 <= (2) AND P8_3 > (15) and P8_3<=(29)	351	61	17.4%	0.5%
P8_1 <= (25) AND P8_2 <= (2) AND P8_3 > (29)	95	26	27.4%	0.1%
P8_1 <= (25) AND P8_2 > (2) AND P8_3 > (15) and P8_3<=(29)	378	70	18.5%	0.5%
P8_1 <= (25) AND P8_2 > (2) AND P8_3 > (29)	169	45	26.6%	0.2%
P8_1 > (25) and P8_1 <= (41) AND P8_2 <= (2) AND P8_3 > (4) and P8_3<=(15)	1,334	342	25.6%	1.8%
P8_1 > (25) and P8_1 <= (41) AND P8_2 <= (2) AND P8_3 > (15) and P8_3<=(29)	1,213	254	20.9%	1.7%
P8_1 > (25) and P8_1 <= (41) AND P8_2 <= (2) AND P8_3 > (29)	897	96	10.7%	1.2%
P8_1 > (25) and P8_1 <= (41) AND P8_2 > (2) AND P8_3 > (15) and P8_3<=(29)	599	153	25.5%	0.8%
P8_1 > (25) and P8_1 <= (41) AND P8_2 > (2) AND P8_3 > (29)	529	79	14.9%	0.7%
P8_1 > (41) and P8_1 <= (46) AND P8_2 <= (2) AND P8_3 > (4) and P8_3<=(15)	1,373	309	22.5%	1.9%
P8_1 > (41) and P8_1 <= (46) AND P8_2 <= (2) AND P8_3 > (15) and P8_3<=(29)	1,469	242	16.5%	2.0%
P8_1 > (41) and P8_1 <= (46) AND P8_2 <= (2) AND P8_3 > (29)	1,593	239	15.0%	2.2%
P8_1 > (41) and P8_1 <= (46) AND P8_2 > (2) AND P8_3 > (15) and P8_3<=(29)	393	81	20.6%	0.5%
P8_1 > (41) and P8_1 <= (46) AND P8_2 > (2) AND P8_3 > (29)	510	96	18.8%	0.7%
P8_1 > (46) AND P8_2 <= (2) AND P8_3 > (4) and P8_3<=(15)	640	161	25.2%	0.9%
P8_1 > (41) and P8_1 <= (46) AND P8_2 > (2) AND P8_3 > (4) and P8_3<=(15)	302	72	23.8%	0.4%
P8_1 <= (25) AND P8_2 <= (2) AND P8_3 <= (4)	734	291	39.6%	1.0%
P8_1 <= (25) AND P8_2 <= (2) AND P8_3 > (4) and P8_3<=(15)	488	150	30.7%	0.7%
P8_1 <= (25) AND P8_2 > (2) AND P8_3 <= (4)	412	238	57.8%	0.6%
P8_1 <= (25) AND P8_2 > (2) AND P8_3 > (4) and P8_3<=(15)	497	203	40.8%	0.7%
P8_1 > (25) and P8_1 <= (41) AND P8_2 <= (2) AND P8_3 <= (4)	1,788	764	42.7%	2.4%
P8_1 > (25) and P8_1 <= (41) AND P8_2 > (2) AND P8_3 <= (4)	709	372	52.5%	1.0%
P8_1 > (25) and P8_1 <= (41) AND P8_2 > (2) AND P8_3 > (4) and P8_3<=(15)	684	252	36.8%	0.9%
P8_1 > (41) and P8_1 <= (46) AND P8_2 <= (2) AND P8_3 <= (4)	1,332	569	42.7%	1.8%
P8_1 > (41) and P8_1 <= (46) AND P8_2 > (2) AND P8_3 <= (4)	224	87	38.8%	0.3%
P8_1 > (46) AND P8_2 > (2) AND P8_3 > (4) and P8_3<=(15)	49	-	0.0%	0.1%
P8_1 > (46) AND P8_2 > (2) AND P8_3 <= (4)	8	-	0.0%	0.0%
P8_1 > (46) AND P8_2 <= (2) AND P8_3 <= (4)	422	177	41.9%	0.6%

Tabla 14

Resumen de la Distribución de la Segmentación del Patrón Externo con GINI y KS.

Patrón	Casos	Default	Default%	Participación%
Bueno	34,994	3,557	10.2%	47.8
Regular	18,905	2,908	15.4%	25.8
Malo	19,381	5,476	28.3%	26.4%
Total	73,280	11,941	16.3%	100.0%

Gini	27.9
KS	23.2

4.1.2 Patrón Interno

Se desarrolló un conjunto de variables que representan el comportamiento del cliente en el Banco, en función a su comportamiento de pago de los últimos 12 meses. Este patrón está compuesto por 7 variables que identifican a clientes que no han presentado atraso en sus pagos en el Banco o no han presentado deuda en el Banco (missing) en el periodo de evaluación, y por 3 variables que recogen el comportamiento de clientes que en el periodo de evaluación han presentado al menos un atraso.

Los patrones de comportamiento interno están descritos en la sección 0, pero en la Figura 16 se puede apreciar las combinaciones posibles entre todas las variables. Estas variables son: INT_P1, INT_P2, INT_P3, INT_P4, INT_P5, INT_P6, INT_P7, INT_P8_1, INT_P8_2 y INT_P8_3.

Figura 16

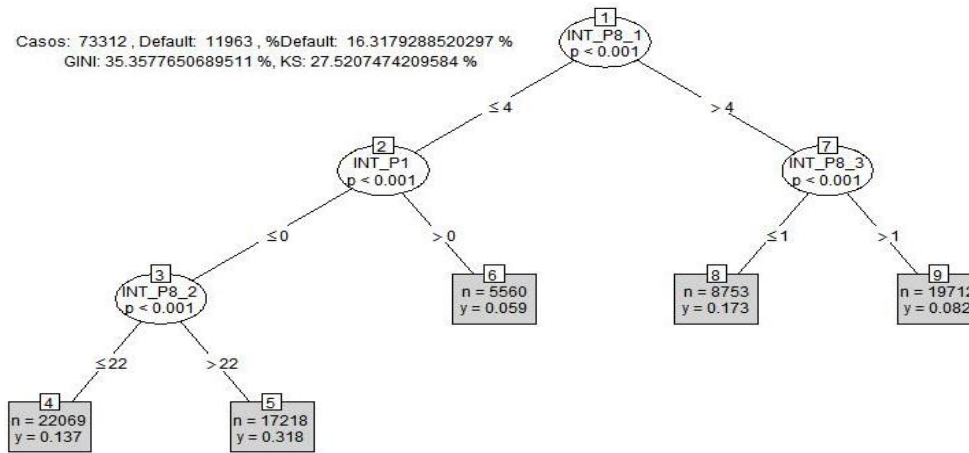
Esquema de Variables que Forman el Patrón Interno

Periodo	Nombre variable	P1	P2	P3	P4	P5	P6	P7	P8		
		12 meses al día	Hace 12m = null Ultimo Mes = 0	Hace 12m = 0 Ultimo Mes = null	Hace 12m = 0 Ultimo Mes = 0	Hace 12m = null Ultimo Mes = null	Hace 12m = 0 Ultimo Mes = null	Hace 12m = null Ultimo Mes = 0	null o > 0		
									P8_1	P8_2	P8_3
t-1	atraso_01	0	0	null	0	null	null	0	null o > 0	null o > 0	null o > 0
t-2	atraso_02	0	0	null	null/0	null/0	null/0	null/0	null o > 0	null o > 0	null o > 0
t-3	atraso_03	0	0	null	null/0	null/0	null/0	null/0	null o > 0	null o > 0	null o > 0
t-4	atraso_04	0	0	null	null/0	null/0	null/0	null/0	null o > 0	null o > 0	null o > 0
.
.
.
t-9	atraso_09	0	null	0	null/0	null/0	null/0	null/0	null o > 0	null o > 0	null o > 0
t-10	atraso_10	0	null	0	null/0	null/0	null/0	null/0	null o > 0	null o > 0	null o > 0
t-11	atraso_11	0	null	0	null/0	null/0	null/0	null/0	null o > 0	null o > 0	null o > 0
t-12	atraso_12	0	null	0	0	null	0	null	null o > 0	null o > 0	null o > 0
Valores	flag	Num meses al día	Num meses al día	Num meses al día	Num meses al día	Num meses al día	Num meses al día	Num meses al día	Num meses al día	Max Atraso 12m	Hace cuanto < 0
		0 - 1	1 - 11	1 - 11	2 - 11	1 - 10	2 - 10	2 - 10	1 - 11	> 0	1 - 12

Con el fin de resumir los patrones de comportamiento, se realizó un análisis de árboles de inferencia condicional, según el mismo algoritmo, para observar cómo interactúan entre sí las variables de comportamiento interno. En la Figura 17 se puede observar el primer árbol generado.

Figura 17

Árbol Inicial con las Variables que Forman el Patrón Interno



Debido a que se requería que el efecto de todas las variables se refleje en esta segmentación, se hicieron más árboles, esta vez aislando cada variable contra la variable objetivo (default).

Figura 18

Árbol de Inferencia Condicional de la Variable INT_P1

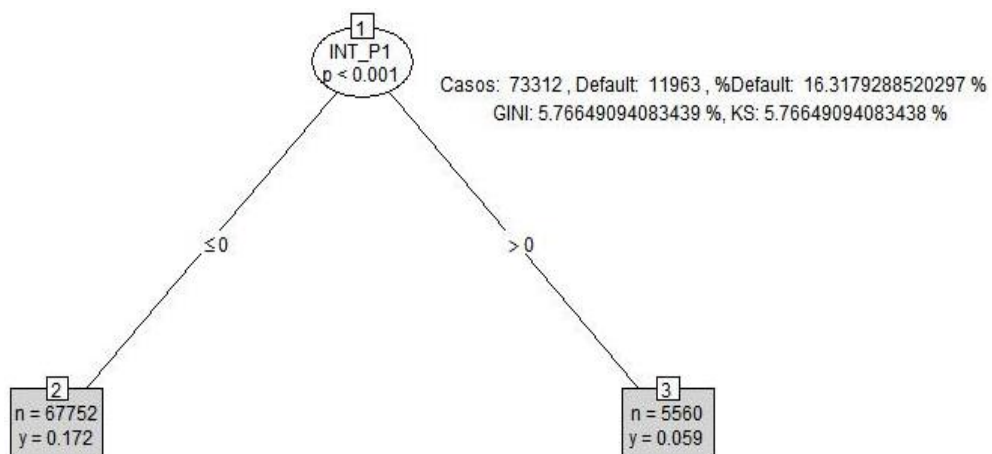


Figura 19

Árbol de Inferencia Condicional de la Variable INT_P8_1

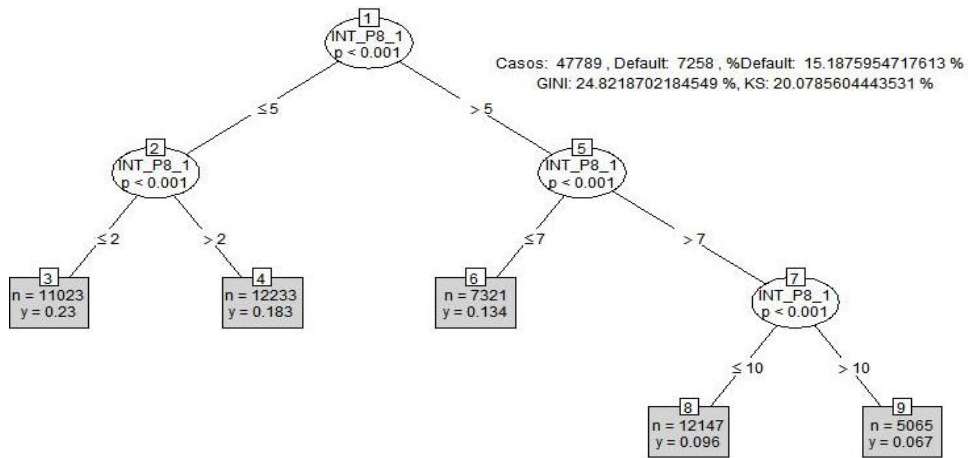


Figura 20

Árbol de Inferencia Condicional de la Variable INT_P8_2

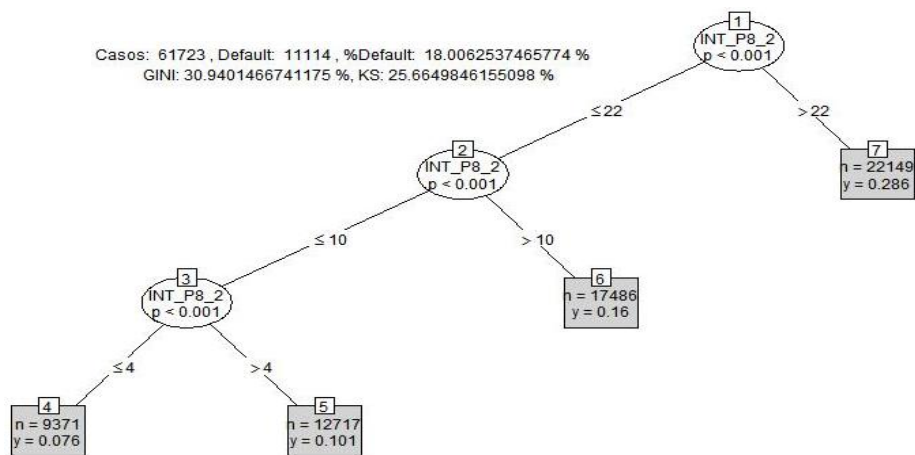
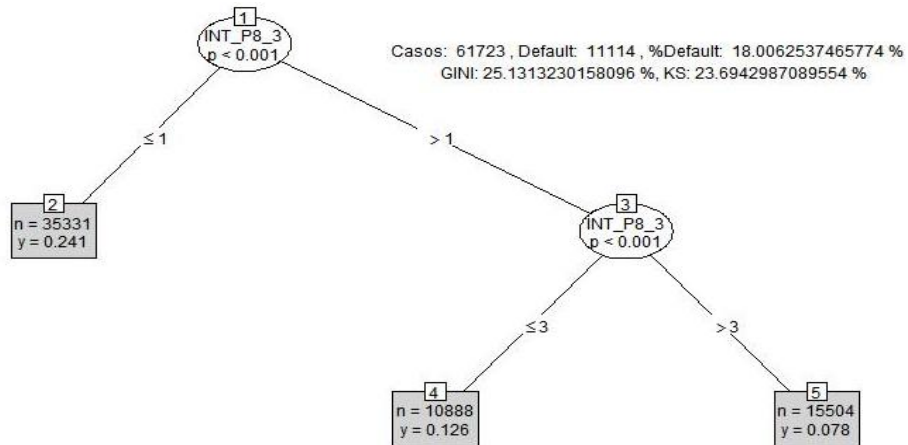


Figura 21

Árbol de Inferencia Condicional de la Variable INT_P8_3



Estas variables por separado se ingresaron en un árbol de decisión y se obtuvo como resultado 3 segmentos (Bueno, Regular y Malo) según el nivel de default en cada grupo, tal como se observa en Figura 22.

Figura 22

Distribución de variables en el patrón interno según nivel de riesgo y participación

Patrones	Casos	Default	Default %	Participación %
INT_P1 > (0)	5,560	330	5.9%	3.8%
INT_P2 > (3)	1,610	84	5.2%	1.1%
INT_P3 > (0)	141	7	5.0%	0.1%
INT_P4 > (0)	704	26	3.7%	0.5%
INT_P5 > (0)	237	3	1.3%	0.2%
INT_P6 > (0)	26	2	7.7%	0.0%
INT_P7 > (0)	149	4	2.7%	0.1%
INT_P8_1 > (7) AND INT_P8_1 <= (10) AND INT_P8_2 <= (4) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	879	74	8.4%	0.6%
INT_P8_1 > (7) AND INT_P8_1 <= (10) AND INT_P8_2 <= (4) AND INT_P8_3 > (3)	2,067	126	6.1%	1.4%
INT_P8_1 > (7) AND INT_P8_1 <= (10) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	1,072	92	8.6%	0.7%
INT_P8_1 > (7) AND INT_P8_1 <= (10) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 > (3)	2,030	152	7.5%	1.4%
INT_P8_1 > (7) AND INT_P8_1 <= (10) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 > (3)	1,284	107	8.3%	0.9%
INT_P8_1 > (10) AND INT_P8_2 <= (4) AND INT_P8_3 > (3)	3,288	180	5.5%	2.2%
INT_P8_1 <= (2) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	134	11	8.2%	0.1%
INT_P8_1 > (10) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 > (3)	543	32	5.9%	0.4%
INT_P8_1 > (10) AND INT_P8_2 <= (4) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	349	31	8.9%	0.2%
INT_P8_1 > (10) AND INT_P8_2 <= (4) AND INT_P8_3 <= (1)	249	27	10.8%	0.2%
INT_P2 <= 3 and INT_P2 > 0	963	111	11.5%	0.7%
INT_P8_1 <= (2) AND INT_P8_2 <= (4) AND INT_P8_3 <= (1)	2,450	341	13.9%	1.7%
INT_P8_1 <= (2) AND INT_P8_2 <= (4) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	151	13	8.6%	0.1%
INT_P8_1 <= (2) AND INT_P8_2 <= (4) AND INT_P8_3 > (3)	189	13	6.9%	0.1%
INT_P8_1 <= (2) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	704	93	13.2%	0.5%
INT_P8_1 <= (2) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 > (3)	264	37	14.0%	0.2%
INT_P8_1 <= (2) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 <= (1)	1,980	226	11.4%	1.4%
INT_P8_1 <= (2) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	397	31	7.8%	0.3%
INT_P8_1 <= (2) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 > (3)	308	30	9.7%	0.2%
INT_P8_1 > (10) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 <= (1)	32	9	28.1%	0.0%
INT_P8_1 > (10) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	36	6	16.7%	0.0%
INT_P8_1 > (10) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 > (3)	236	24	10.2%	0.2%
INT_P8_1 > (10) AND INT_P8_2 > (22) AND INT_P8_3 <= (1)	11	5	45.5%	0.0%
INT_P8_1 > (10) AND INT_P8_2 > (22) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	3	-	0.0%	0.0%
INT_P8_1 > (10) AND INT_P8_2 > (22) AND INT_P8_3 > (3)	66	5	7.6%	0.0%
INT_P8_1 > (10) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 <= (1)	118	11	9.3%	0.1%
INT_P8_1 > (2) AND INT_P8_1 <= (5) AND INT_P8_2 <= (4) AND INT_P8_3 <= (1)	199	33	16.6%	0.1%
INT_P8_1 > (2) AND INT_P8_1 <= (5) AND INT_P8_2 <= (4) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	153	20	13.1%	0.1%
INT_P8_1 > (2) AND INT_P8_1 <= (5) AND INT_P8_2 <= (4) AND INT_P8_3 > (3)	201	3	1.5%	0.1%
INT_P8_1 > (2) AND INT_P8_1 <= (5) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	1,034	158	15.3%	0.7%
INT_P8_1 > (2) AND INT_P8_1 <= (5) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 > (3)	546	45	8.2%	0.4%
INT_P8_1 > (2) AND INT_P8_1 <= (5) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	1,002	147	14.7%	0.7%
INT_P8_1 > (2) AND INT_P8_1 <= (5) AND INT_P8_2 > (22) AND INT_P8_3 > (3)	662	85	12.8%	0.5%
INT_P8_1 > (2) AND INT_P8_1 <= (5) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 <= (1)	1,651	224	13.6%	1.1%
INT_P8_1 > (2) AND INT_P8_1 <= (5) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	722	79	10.9%	0.5%
INT_P8_1 > (2) AND INT_P8_1 <= (5) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 > (3)	431	21	4.9%	0.3%
INT_P8_1 > (5) AND INT_P8_1 <= (7) AND INT_P8_2 <= (4) AND INT_P8_3 <= (1)	222	26	11.7%	0.2%
INT_P8_1 > (5) AND INT_P8_1 <= (7) AND INT_P8_2 <= (4) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	161	21	13.0%	0.1%
INT_P8_1 > (5) AND INT_P8_1 <= (7) AND INT_P8_2 <= (4) AND INT_P8_3 > (3)	275	8	2.9%	0.2%
INT_P8_1 > (5) AND INT_P8_1 <= (7) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 <= (1)	1,053	182	17.3%	0.7%
INT_P8_1 > (5) AND INT_P8_1 <= (7) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	686	98	14.3%	0.5%
INT_P8_1 > (5) AND INT_P8_1 <= (7) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 > (3)	663	63	9.5%	0.5%
INT_P8_1 > (5) AND INT_P8_1 <= (7) AND INT_P8_2 > (22) AND INT_P8_3 > (3)	597	59	9.9%	0.4%
INT_P8_1 > (5) AND INT_P8_1 <= (7) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 <= (1)	990	132	13.3%	0.7%
INT_P8_1 > (5) AND INT_P8_1 <= (7) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	668	61	9.1%	0.5%
INT_P8_1 > (5) AND INT_P8_1 <= (7) AND INT_P8_2 > (4) AND INT_P8_2 <= (10) AND INT_P8_3 > (3)	607	42	6.9%	0.4%
INT_P8_1 > (7) AND INT_P8_1 <= (10) AND INT_P8_2 <= (4) AND INT_P8_3 <= (1)	737	79	10.7%	0.5%
INT_P8_1 > (7) AND INT_P8_1 <= (10) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 <= (1)	776	120	15.5%	0.5%
INT_P8_1 > (7) AND INT_P8_1 <= (10) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	696	68	9.8%	0.5%
INT_P8_1 > (7) AND INT_P8_1 <= (10) AND INT_P8_2 > (22) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	300	32	10.7%	0.2%
INT_P8_1 > (7) AND INT_P8_1 <= (10) AND INT_P8_2 > (22) AND INT_P8_3 > (3)	760	62	8.2%	0.5%
INT_P8_1 > (7) AND INT_P8_1 <= (10) AND INT_P8_2 <= (4) AND INT_P8_2 <= (10) AND INT_P8_3 <= (1)	1,066	140	13.1%	0.7%
INT_P8_1 <= (2) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 <= (1)	6,765	1,180	17.4%	4.6%
INT_P8_1 <= (2) AND INT_P8_2 > (22) AND INT_P8_3 <= (1)	12,213	4,316	35.3%	8.3%
INT_P8_1 <= (2) AND INT_P8_2 > (22) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	1,248	266	21.3%	0.9%
INT_P8_1 <= (2) AND INT_P8_2 > (22) AND INT_P8_3 > (3)	487	123	25.3%	0.3%
INT_P8_1 > (2) AND INT_P8_1 <= (5) AND INT_P8_2 > (10) AND INT_P8_2 <= (22) AND INT_P8_3 <= (1)	2,711	600	22.1%	1.8%
INT_P8_1 > (2) AND INT_P8_1 <= (5) AND INT_P8_2 > (22) AND INT_P8_3 <= (1)	2,921	822	28.1%	2.0%
INT_P8_1 > (5) AND INT_P8_1 <= (7) AND INT_P8_2 > (22) AND INT_P8_3 <= (1)	906	216	23.8%	0.6%
INT_P8_1 > (5) AND INT_P8_1 <= (7) AND INT_P8_2 > (22) AND INT_P8_3 > (1) AND INT_P8_3 <= (3)	493	75	15.2%	0.3%
INT_P8_1 > (7) AND INT_P8_1 <= (10) AND INT_P8_2 > (22) AND INT_P8_3 <= (1)	480	114	23.8%	0.3%

Tabla 15

Resumen de la Distribución de Variables en el Patrón Interno Según Nivel de Riesgo y Participación con GINI y KS de la Segmentación Patrón Interno.

Patrón	Casos	Default	Default%	Participación%
Bueno	20,322	1,288	6.3%	27.7
Regular	24,766	2,963	12.0%	33.8
Malo	28,192	7,690	27.3%	38.5%
Total	73,280	11,941	16.3%	100.0%

Gini	34.8
KS	31

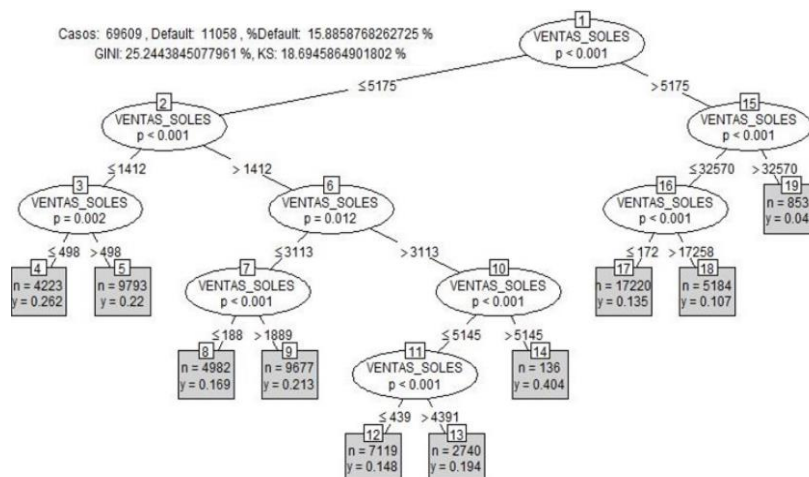
4.1.3 Nivel de Ventas (Ventas soles)

Para que el tratamiento de los registros sea diferenciado por el tipo de empresa, se consideró el nivel de ventas como una variable adicional de segmentación, usándola como aproximación para separar a las empresas grandes de las pequeñas.

Esta variable registra al cliente en cada cierre de ejercicio con una antigüedad máxima de cinco años. Se realizó una agrupación mediante un árbol de inferencia condicional, considerando a todos los clientes que en el mes de análisis tuvieran información de Nivel de Ventas. Tomando como variable objetivo el incumplimiento de pago de crédito (default) como se puede apreciar en la Figura 23 como se distribuyen las ventas diferenciadas por riesgo.

Figura 23

Árbol de Inferencia Condicional de la Variable Nivel de Ventas (Ventas_soles)



Los cortes que se observan en la Figura 23 se representan en la Tabla 16.

Tabla 16

Cortes Según el Árbol para la Variable Ventas Soles

Nivel ventas (S/.)	Clientes	Participación %	Default	Default%
<= 0.498 MM	4,223	6.10	1,106	26.20
< 0.498 – 1.412] MM	9,793	14.10	2,152	22.00
< 1.412 – 1.889] MM	4,982	7.2	842	16.90
< 1.889 – 3.113] MM	9,677	13.90	2,065	21.30
< 3.113 – 4.391] MM	7,119	10.20	1,053	14.80
< 4.394 – 5.145] MM	2,740	3.90	531	19.40
< 5.145 – 5.175] MM	136	0.20	55	40.40
< 5.175 – 17.258] MM	17,220	24.70	2,332	13.50
< 17.258 – 32.570] MM	5,184	7.40	554	10.70
> 32.570 MM	8,535	12.30	368	4.30

Luego se agruparon clases con niveles parecidos de defaults con la finalidad de reducir las que tenían menor participación y también buscando la monotonicidad de los rangos de la variable, tal como se aprecia en la Tabla 17.

Tabla 17

Cortes Agrupados para la Variable Nivel Ventas (Ventas_Soles)

Nivel ventas (S/.)	Clientes	Participación %	Default	Default%
<= 1.412 MM	14,016	20.10	3,258	22.00
< 1.412 – 3.113] MM	14,459	21.10	2,907	14.80
< 3.113 – 5.175] MM	9,995	14.40	1,639	40.40
< 5.175 – 32.570] MM	22,404	32.20	2,886	10.70
> 32.570 MM	8,535	12.30	368	4.30

Se redondearon los valores de los límites de los intervalos. Así se llegaron a los siguientes cortes:

- Hasta S/ 1.5 millones
- Más de S/ 1.5 millones hasta S/ 3 millones.
- Más de S/ 3 millones hasta S/ 5 millones.
- Más de S/ 5 millones hasta S/ 30 millones.
- Más de S/ 30 millones.

Estos cortes se acercan a la clasificación que tiene el banco para sus clientes empresariales. En la Tabla 18 se puede observar cómo los créditos con ventas más altas están más representados por las bancas Corporativa y Banca Empresa.

Tabla 18

Porcentaje de Distribución de los Niveles de Venta por Banca.

Banca	Nivel de Ventas (S/.)				
	> 30MM	<5-30]MM	<3-5]MM	<1.5-3]MM	<=1.5MM
Corporativa	72.02	16.16	1.19	1.67	8.96
Empresa	28.98	48.08	6.32	6.36	10.27
Negocios	3.10	28.91	16.48	21.68	29.83
Inmobiliaria	8.97	40.22	11.96	11.96	26.90
Otros	0.00	0.00	0.00	0.00	100

La distribución de las clases resultantes se observa en la Tabla 19.

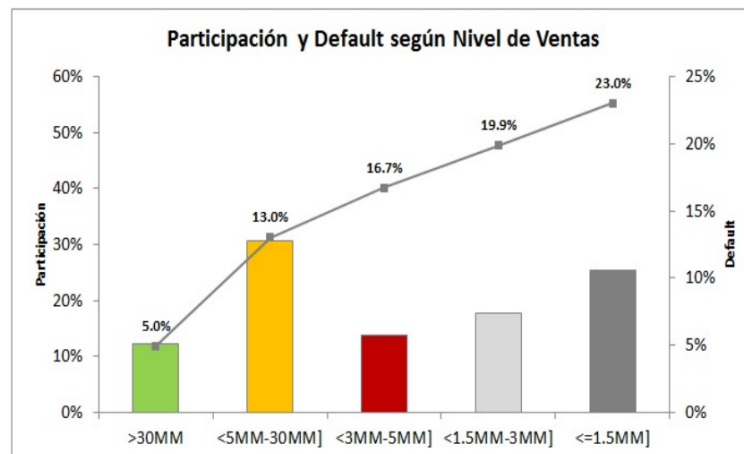
Tabla 19

Cortes Finales para la Variable Ventas Soles

Nivel ventas (S/)	Clientes	Participación %	Default	Default%
> 30 MM	9,055	13.0	449	5
< 5 – 30] MM	22,509	32.3	2,936	13
< 3 – 5] MM	10,073	14.5	1,685	16.7
< 1.5 – 3] MM	13,038	18.7	2,592	19.9
<= 1.5 MM	14,934	21.5	3,396	22.7

Figura 24

Participación de las Clases de la Variable Nivel de Ventas por Default.



4.1.4 Segmentos Finales

Dado que las variables Patrón Interno, Patrón Externo y Nivel de Ventas han sido categorizadas previamente, se procedió a realizar una segmentación utilizando los valores de la probabilidad de incumplimiento (PD) estimados mediante el modelo presentado. El objetivo es identificar segmentos de riesgo, clasificándolos en las categorías de alto, medio y bajo riesgo.

Tabla 20

Porcentaje de Incumplimiento de Pago de Crédito (Default) de la Combinación de Variables que Forman los Segmentos Finales.

Nivel Ventas (S/. MM)	Patrón Externo								
	1			2			3		
	Patrón Interno			Patrón Interno			Patrón Interno		
	1	2	3	1	2	3	1	2	3
>30	2.4%	4.1%	7.8%	2.0%	1.9%	8.1%	3.7%	2.5%	19.8%
<5 – 30]	3.2%	6.4%	16.5%	5.1%	10.3%	23.0%	11.3%	13.0%	33.1%
<3 – 5]	5.3%	9.4%	19.5%	5.7%	11.6%	21.7%	10.5%	17.4%	36.6%
<1.5 – 3]	8.3%	13.6%	24.8%	4.9%	11.4%	22.9%	8.2%	19.6%	40.8%
<=1.5	11.1%	9.5%	22.9%	9.7%	20.2%	32.9%	13.0%	22.4%	43.4%

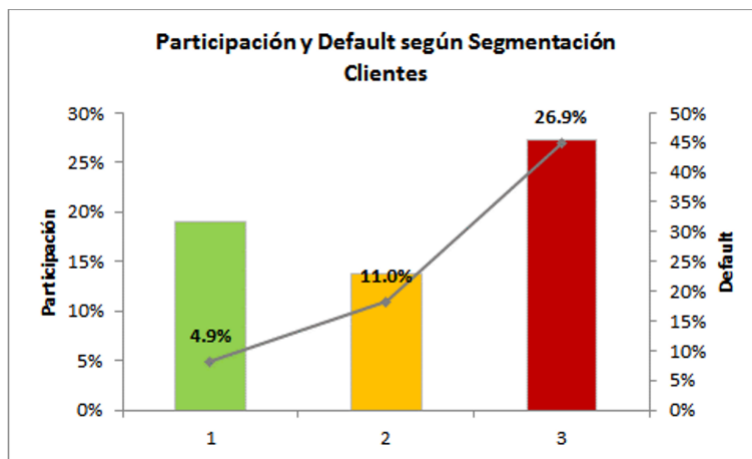
En la Tabla 20 se observan las variables Patrón Externo, Patrón Interno y Nivel de ventas categorizadas de la siguiente manera:

- Riesgo bajo: Segmento 1 (color verde)
- Riesgo medio: Segmento 2 (color ámbar)
- Riesgo alto: Segmento 3 (color rojo)

A continuación, se muestran los resultados de la segmentación tanto en participación de la población

Figura 25

Participación e Incumplimiento de Pago de Crédito (Default) de los Segmentos Finales



4.2 Modelo para el Segmento 1 (Riesgo bajo)

La ecuación resultante con las variables que intervienen en el modelo final se resume en la Tabla 21. Puede apreciarse que para este segmento hay 3 variables que influyen positivamente y 3 que influyen negativamente en la probabilidad de default (Pd). Las variables con mayor influencia positiva en la Pd son: EntRepPromUlt24m_d⁶ y AtrasoPromPondUlt24m_d2⁷, mientras que las variables que mayor impacto negativo sobre la Pd son: Antigüedad_Empresa_d2⁸ y Ventas_Soles_d2⁹.

En cuanto al impacto estadístico y el signo de los coeficientes (betas), todas las variables presentan signo coherente de influencia en la variable explicativa (default) y p-value inferiores al nivel crítico fijado en 0.05. Por tanto, todas las variables independientes del modelo influyen en el comportamiento de pago del deudor.

⁶ Número de entidades reportantes en promedio en los últimos 24 meses.

⁷ Promedio ponderado - ponderando según la antigüedad - de los días de mora, en créditos directos, en los últimos 24 meses.

⁸ Antigüedad de la empresa, tomando en consideración la fecha de constitución.

⁹ Información que proviene de los estados financieros disponibles, tomando en consideración el monto al cierre de cada ejercicio.

Indicadores adicionales del modelo como su nivel de significancia, los ratios de ODDS de cada variable y la matriz de correlaciones se pueden revisar en la sección 0 del presente documento.

Tabla 21

Análisis de Estimaciones de Máxima Verosimilitud.

Parameter	Levels	DF	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq
(Intercept)		1	-3.6153	0.1512	571.98	0.0000
Grupo_Score_Empresa_New_d2	4	1	0.6879	0.0957	51.63	0.0000
Grupo_Score_Empresa_New_d2	6	1	1.1169	0.1044	114.48	0.0000
AtrasoPromPondUlt24m_d2	2	1	0.6964	0.1516	21.11	0.0000
AtrasoPromPondUlt24m_d2	4	1	1.0404	0.1500	48.13	0.0000
AtrasoPromPondUlt24m_d2	8	1	1.2170	0.1738	49.06	0.0000
EntRepPromUlt24m_d	2	1	0.2466	0.0826	8.90	0.0028
EntRepPromUlt24m_d	3	1	1.8465	0.1577	137.16	0.0000
Antigüedad_Empresa_d2	5	1	-0.4461	0.0903	24.43	0.0000
Antigüedad_Empresa_d2	6	1	-0.8583	0.1828	22.05	0.0000
RatioUlt1MaxUlt6Pasivo_d2	4	1	-0.4390	0.0767	32.75	0.0000
Ventas_Soles_d2	3	1	-0.8496	0.1161	53.58	0.0000

Para la validación del modelo se muestran los valores de Gini, KS y ROC del modelo en la Tabla 22, estos valores están dentro del rango aceptable de modelos considerados como Muy Buenos, tal como se describió en la metodología de construcción.

Tabla 22

Indicadores de Discriminancia

Medidas de separación	Muestra de Análisis	Muestra de Validación	Total
Gini	44.21	44.36	44.24281
KS	32.92	32.05	32.43752
ROC	72.11	72.18	72.12141
Divergencia	0.68	0.70	0.68911
Estabilidad Poblacional			0.01677

También se analizó el aporte a la discriminancia que tenía cada una de las variables involucradas en el modelo. El aporte se calculó, extrayendo una variable, midiendo el modelo sin ella y comparando en cuanto decrecía el indicador KS del modelo en la muestra de análisis. Esto se aprecia en la Tabla 23.

Tabla 23

Aporte de las Variables al KS del Modelo en la Muestra de Análisis.

Variable	Aporte de la variable %
AtrasoPromPondUlt24m_d2	13.18
Grupo_Score_Empresa_New_d2	13.10
EntRepPromUlt24m_d	11.10
Antiguedad_Empresa_d2	4.36
Ventas_Soles_d2	4.25
RatioUlt1MaxUlt6Pasivo_d2	4.17

Para hacer la validación del modelo de clasificación de buenos y malos se calcula la distribución de los clientes de acuerdo con el score asignado, tal como se aprecia en la Figura 26, este modelo presenta un nivel de divergencia de 0.69 que es bueno para un modelo de clasificación.

Figura 26

Distribución de Clientes que no han Incumplido el Pago de su Crédito (Buenos) y

Clientes que han Incumplido el Pago de su Crédito (Malos) en la Muestra de Análisis

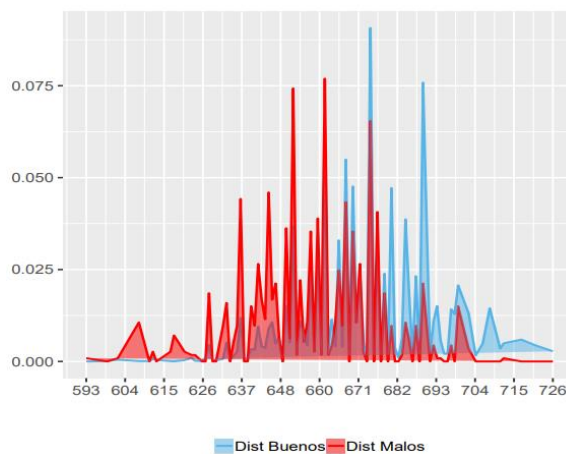


Figura 27

Distribución Acumulada de Buenos y Malos

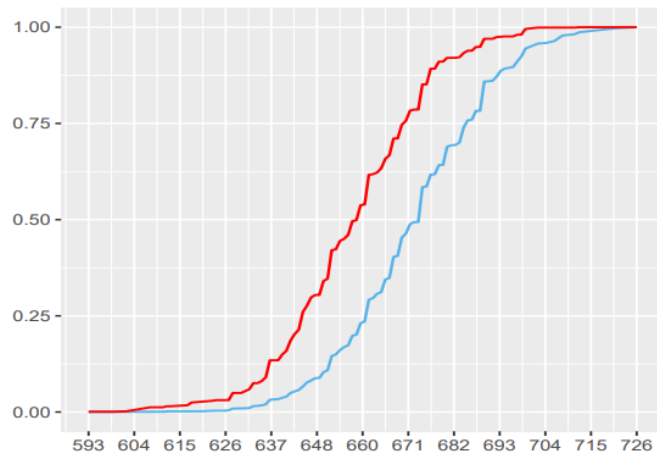
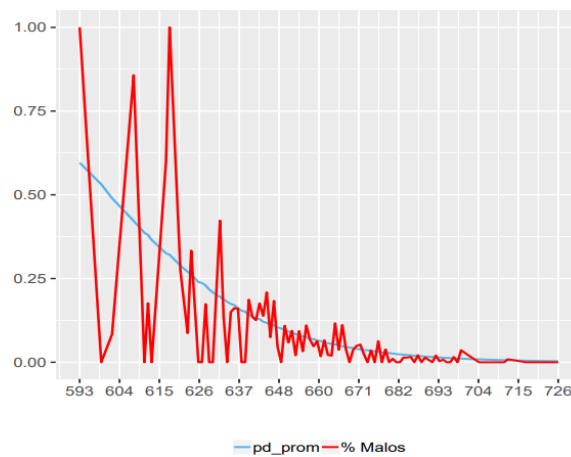


Figura 28

Promedio Teórico del Incumplimiento de Pago de Crédito del Modelo (pd_prom) y

Porcentaje de Clientes en Incumplimiento de Pago de Crédito Real (% malos)



En cuanto a la validación muestra de control, obtenemos los siguientes resultados:

Tabla 24

Indicadores de discriminancia - Muestra de control (Fuera de la muestra)

KS	GINI	ROC	Divergencia	Estabilidad Poblacional
22.48	30.94	65.47	0.34	0.06180

Figura 29

Distribución de Clientes que no han Incumplido el Pago de su Crédito (Buenos) y Clientes que han Incumplido el Pago de su Crédito (Malos) en la Muestra de Control.

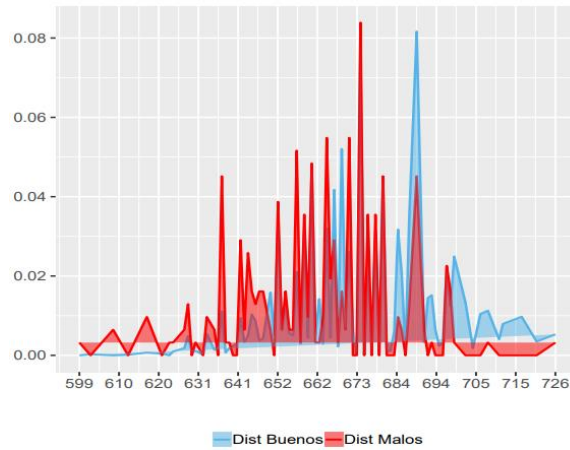


Figura 30

Distribución acumulada de Buenos y Malos - Desarrollo vs Fuera de Muestra

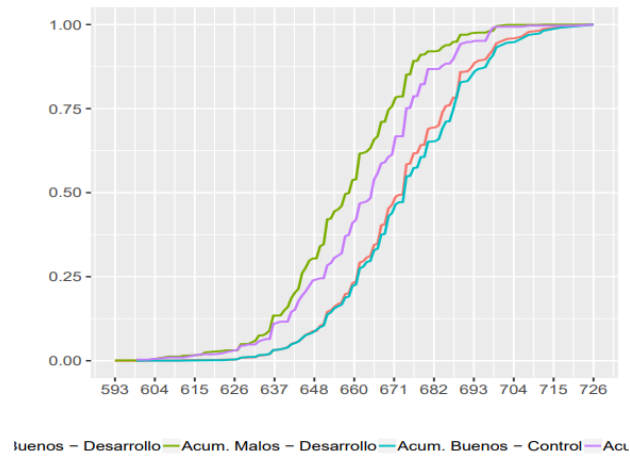
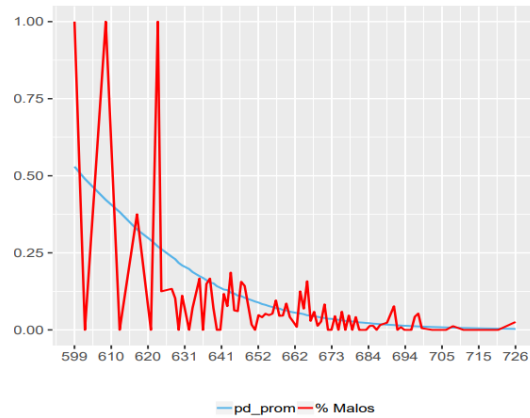


Figura 31

Promedio Teórico del Incumplimiento de Pago de Crédito del Modelo (pd_prom) y Porcentaje de Clientes en Incumplimiento de Pago de Crédito Real (% malos) en la



4.2.1 Grupos de Riesgo

Finalmente, el modelo puede presentarse estableciendo grupos de riesgo divididos tal como se observa en la Tabla 25. Esta partición fue aplicada a la muestra de análisis.

Tabla 25

Grupos de Riesgo por Deciles en la Muestra de Análisis.

	Grupos de Score	Clientes	Default	Default %	PD Prom %
1	(697 - 999]	1,414	19	1.34	0.86
2	(689 - 697]	828	10	1.21	1.36
3	(681 - 689]	2,594	43	1.66	1.90
4	(676 - 681]	1,173	18	1.53	2.67
5	(671 - 676]	2,069	90	4.35	3.30
6	(667 - 671]	1,356	50	3.69	4.11
7	(661 - 667]	1,782	75	4.21	4.95
8	(657 - 661]	1,568	97	6.19	6.14
9	(650 - 657]	1,604	121	7.54	8.14
10	[0 - 650]	1,827	282	15.44	14.34

Aplicando los cortes determinados a la muestra de control, podemos observar cómo se comportan los diferentes grupos de riesgo en el tiempo. Ver Tabla 26.

Tabla 26*Grupos de Riesgo por Deciles en la Muestra de Control*

	Grupos de Score	Clientes	Default	Default %	PD Prom %
1	(697 - 999]	760	8	1.05	0.83
2	(689 - 697]	491	10	2.04	1.35
3	(681 - 689]	1,291	23	1.78	1.88
4	(676 - 681]	590	25	4.24	2.69
5	(671 - 676]	814	37	4.55	3.31
6	(667 - 671]	675	25	3.70	4.11
7	(661 - 667]	759	37	4.87	5.00
8	(657 - 661]	653	30	4.59	6.17
9	(650 - 657]	673	39	5.79	8.14
10	[0 - 650]	793	76	9.58	14.43

4.3 Modelo para el Segmento 2 (Riesgo medio)

La ecuación resultante con las variables que intervienen en el modelo final se resume en la Tabla 27. Puede apreciarse que para este segmento hay 2 variables que influyen positivamente y 3 que influyen negativamente en la probabilidad de default (Pd). Las variables con mayor influencia positiva en la Pd son: *AtrasoPromPondUlt6m_d2*¹⁰ y *Grupo_Score_Empresa_New*¹¹, mientras que las variables que mayor impacto negativo sobre la Pd son: *Antigüedad_Empresa_d2*¹² y *PromPasivosUlt3_d2*¹³.

En cuanto al impacto estadístico y el signo de los coeficientes (betas), todas las variables presentan signo coherente de influencia en la variable explicativa (default) y p-value inferiores al nivel crítico fijado en 0.05. Por tanto, todas las variables independientes del modelo influyen en el comportamiento de pago del deudor.

Indicadores adicionales del modelo como su nivel de significancia, los ratios de ODDS de cada variable y la matriz de correlaciones se pueden revisar en la sección 0 del presente documento.

¹⁰ Promedio ponderado ponderando según la antigüedad de los días de mora, en créditos directos, en los últimos 6 meses.

¹¹ Promedio ponderado entre la participación de acciones de los accionistas de la empresa por el score de buro de persona natural vigente.

¹² Antigüedad de la empresa, tomando en consideración la fecha de constitución.

¹³ Promedio del saldo pasivo de los últimos 3 meses.

Tabla 27*Análisis de Estimaciones de Máxima Verosimilitud*

Parameter	Levels	DF	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq
(Intercept)		1	-2.0480	0.0733	780.81	0.0000
AtrasoPromPondUlt6m_d2	3	1	0.4595	0.0889	26.74	0.0000
AtrasoPromPondUlt6m_d2	4	1	0.5723	0.0700	66.91	0.0000
Antiguedad_Empresa_d2	4	1	-0.2523	0.0762	10.95	0.0000
Antiguedad_Empresa_d2	6	1	-0.6901	0.0813	71.97	0.0000
PromPasivosUlt3_d2	7	1	-0.4233	0.0789	28.80	0.0000
PromPasivosUlt3_d2	8	1	-0.4863	0.0864	31.67	0.0000
RatioUlt1MaxUlt6Pasivo_d	2	1	-0.1808	0.0785	5.30	0.0214
RatioUlt1MaxUlt6Pasivo_d	3	1	-0.3440	0.0774	19.77	0.0000
Grupo_Score_Empresa_New_d	2	1	0.3318	0.0740	20.11	0.0000
Grupo_Score_Empresa_New_d	3	1	0.7349	0.1019	52.01	0.0000
Grupo_Score_Empresa_New_d	4	1	1.0272	0.0829	53.70	0.0000

Para la validación del modelo se muestran los valores de Gini, KS y ROC del modelo en la Tabla 28, estos valores están dentro del rango aceptable de modelos considerados como Muy Buenos, tal como se describió en la metodología de construcción.

Tabla 28*Indicadores de Discriminancia*

Medidas de separación	Muestra de Análisis	Muestra de Validación	Total
Gini	35.38	36.99	35.85303
KS	25.18	26.76	25.52766
ROC	67.69	68.49	67.92652
Divergencia	0.42	0.46	0.43208
Estabilidad Poblacional			0.01666

También se analizó el aporte a la discriminancia que tenía cada una de las variables involucradas en el modelo. El aporte se calculó, extrayendo una variable, midiendo el modelo sin ella y comparando en cuanto decrecía el indicador KS del modelo en la muestra de análisis. Esto se aprecia en la Tabla 29.

Tabla 29

Aporte de las Variables al KS del Modelo

Variable	Aporte de la variable %
PromPasivosUlt3_d2	5.73
Antigüedad_Empresa_d2	5.35
Grupo_Score_Empresa_New_d	5.26
RatioUlt1MaxUlt6Pasivo_d	3.51
AtrasoPromPondUlt6m_d2	0.66

Para hacer la validación del modelo de clasificación de buenos y malos se calcula la distribución de los clientes de acuerdo con el score asignado, tal como se aprecia en el Figura 32, este modelo presenta un nivel de divergencia de 0.43 que es bueno para un modelo de clasificación.

Figura 32

Distribución de Clientes que no han Incumplido el Pago de su Crédito (Buenos) y Clientes que han Incumplido el Pago de su Crédito (Malos) en la Muestra de

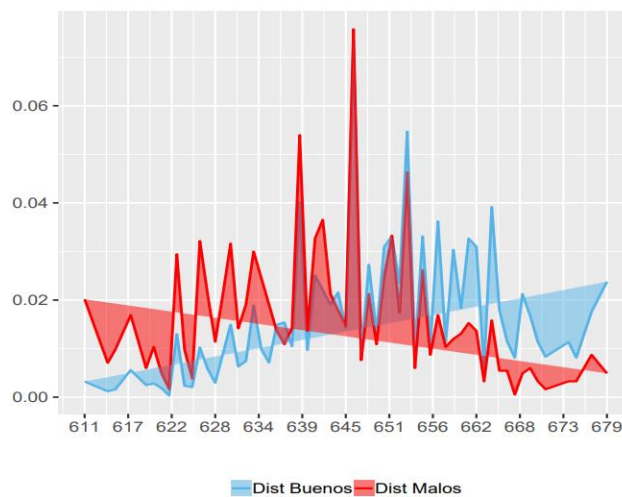


Figura 33

Distribución acumulada de Buenos y Malos

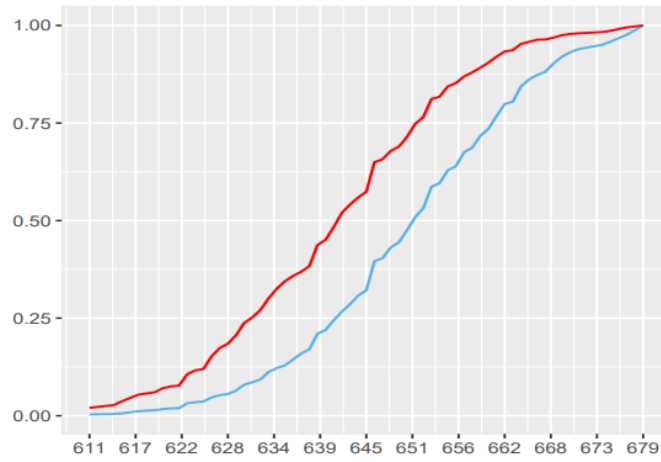


Figura 34

Promedio Teórico del Incumplimiento de Pago de Crédito del Modelo (pd_prom) y Porcentaje de Clientes en Incumplimiento de Pago de Crédito Real (% malos) en la

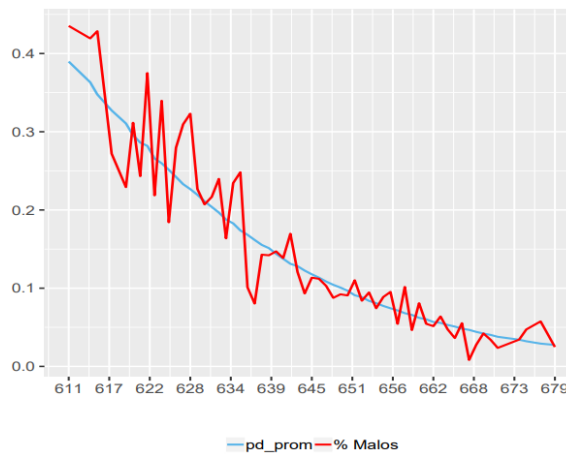


Tabla 30

Indicadores de Discriminancia - Muestra de Control (Fuera de la Muestra)

KS	GINI	ROC	Divergencia	Estabilidad Poblacional
26.91	32.08	66.04	0.31	0.04319

Figura 35

Distribución de Clientes que no han Incumplido el Pago de su Crédito (Buenos) y Clientes que han Incumplido el Pago de su Crédito (Malos) en la Muestra de Control.

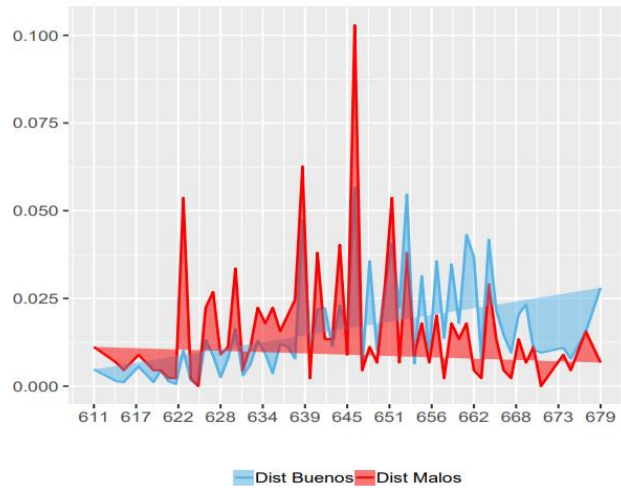


Figura 36

Distribución acumulada de Buenos y Malos - Desarrollo vs Fuera de Muestra

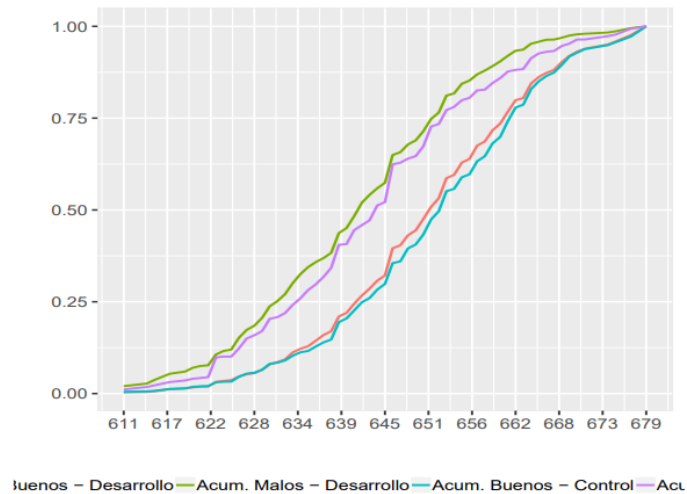
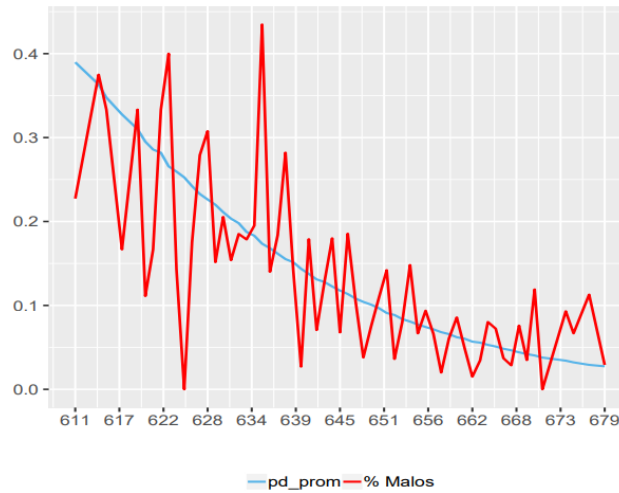


Figura 37

Promedio Teórico del Incumplimiento de Pago de Crédito del Modelo (pd_prom) y

Porcentaje de Clientes en Incumplimiento de Pago de Crédito Real (% Malos) en la



4.3.1 Grupos de Riesgo

Finalmente, el modelo puede presentarse estableciendo grupos de riesgo divididos tal como se observa en la Tabla 31. Esta partición fue aplicada a la muestra de análisis.

Tabla 31

Grupos de Riesgo por Deciles en la Muestra de Análisis.

	Grupos de Score	Clientes	Default	Default %	PD Prom %
1	(668 - 999]	1,068	37	3.46	3.38
2	(662 - 668]	1,115	46	4.13	5.00
3	(658 - 662]	1,227	72	5.87	6.11
4	(653 - 658]	1,129	88	7.79	7.42
5	(650 - 653]	1,293	131	10.13	8.69
6	(646 - 650]	885	84	9.49	10.12
7	(643 - 646]	1,289	139	10.78	11.58
8	(639 - 643]	938	141	15.03	13.41
9	(631 - 639]	1,548	244	15.76	16.73
10	[0 - 631]	1,215	316	26.01	25.64

Aplicando los cortes determinados a la muestra de control, podemos observar cómo se comportan los diferentes grupos de riesgo en el tiempo. Ver Tabla 32.

Tabla 32

Grupos de Riesgo por Deciles en la Muestra de Control

	Grupos de Score	Clientes	Default	Default %	PD Prom %
1	(668 - 999]	400	24	6.00	3.41
2	(662 - 668]	442	29	6.56	5.01
3	(658 - 662]	497	24	4.83	6.11
4	(653 - 658]	365	25	6.85	7.39
5	(650 - 653]	464	44	9.48	8.72
6	(646 - 650]	301	22	7.31	10.13
7	(643 - 646]	407	68	16.71	11.64
8	(639 - 643]	264	30	11.36	13.46
9	(631 - 639]	482	88	18.26	16.51
10	[0 - 631]	393	93	23.66	25.76

4.4 Modelo para el Segmento 3 (Riesgo alto)

La ecuación resultante con las variables que intervienen en el modelo final se resume en la Tabla 33. Puede apreciarse que para este segmento hay 2 variables que influyen positivamente y 5 que influyen negativamente en la probabilidad de default (Pd). Las variables con mayor influencia positiva en la Pd son: *AtrasoPromPondIntUlt6m_d2*¹⁴ y *Grupo_Score_Empresa_New_d2*¹⁵, mientras que las variables que mayor impacto negativo sobre la Pd son: *NumNorUlt6m_d2*¹⁶ e *IncPromPasivoUlt6_d2*¹⁷.

En cuanto al impacto estadístico y el signo de los coeficientes (betas), todas las variables presentan signo coherente de influencia en la variable explicativa (default) y p-value inferiores al nivel crítico fijado en 0.05. Por tanto, todas las variables independientes del modelo influyen en el comportamiento de pago del deudor.

¹⁴ Promedio ponderado - ponderando según la antigüedad - de los días de atraso interno, en los últimos 6 meses.

¹⁵ Promedio ponderado entre la participación de acciones de los accionistas de la empresa por el score de buro de persona natural vigente.

¹⁶ Número de calificaciones normales en los últimos 6 meses.

¹⁷ Incremento promedio de pasivos en los últimos 6 meses.

Indicadores adicionales del modelo como su nivel de significancia, los ratios de ODDS de cada variable y la matriz de correlaciones se pueden revisar en la sección 0 del presente documento.

Tabla 33

Análisis de Estimaciones de Máxima Verosimilitud

Parameter	Levels	D F	Estimate	Standar d Error	Wald Chi- Square	Pr>Chi Sq
(Intercept)		1	-0.4556	0.0514	78.59	0.0000
IncPromPasivoUlt6_d2	4	1	-0.6371	0.0672	89.81	0.0000
AtrasoPromPondIntUlt6m_d2	6	1	0.3799	0.0449	71.68	0.0000
AtrasoPromPondIntUlt6m_d2	9	1	0.4527	0.0700	41.86	0.0000
AtrasoPromPondIntUlt6m_d2	10	1	0.6924	0.0649	113.98	0.0000
AtrasoPromPondIntUlt6m_d2	11	1	0.7679	0.0654	138.06	0.0000
AtrasoPromPondIntUlt6m_d2	12	1	0.8062	0.0634	161.68	0.0000
AtrasoPromPondIntUlt6m_d2	13	1	1.0431	0.0517	406.48	0.0000
NumNorUlt6m_d2	4	1	-0.6183	0.0366	285.81	0.0000
Antiguedad_Empresa_d2	6	1	-0.2166	0.0348	38.74	0.0000
Antiguedad_Empresa_d2	9	1	-0.4071	0.0597	46.52	0.0000
Pasivos_1_d2	2	1	-0.3887	0.0339	131.12	0.0000
Pasivos_1_d2	4	1	-0.6818	0.0876	60.61	0.0000
Ventas_Soles_d2	2	1	-0.2778	0.0424	42.84	0.0000
Ventas_Soles_d2	6	1	-0.3697	0.0729	25.69	0.0000
Grupo_Score_Empresa_New_d2	3	1	0.3273	0.0519	39.72	0.0000
Grupo_Score_Empresa_New_d2	4	1	0.5868	0.0470	155.81	0.0000
Grupo_Score_Empresa_New_d2	5	1	0.9637	0.0443	473.47	0.0000

Para la validación del modelo se muestran los valores de Gini, KS y ROC del modelo en la Tabla 34, estos valores están dentro del rango aceptable de modelos considerados como Muy Buenos, tal como se describió en la metodología de construcción.

Tabla 34

Indicadores de Discriminancia

Medidas de separación	Muestra de Análisis	Muestra de Validación	Total
Gini	41.91	42.96	42.23027
KS	30.91	32.03	31.06622
ROC	70.95	71.48	71.11513
Divergencia	0.61	0.64	0.61888
Estabilidad Poblacional			0.01445

También se analizó el aporte a la discriminancia que tenía cada una de las variables involucradas en el modelo. El aporte se calculó, extrayendo una variable, midiendo el modelo sin ella y comparando en cuanto decrecía el indicador KS del modelo en la muestra de análisis. Esto se aprecia en la Tabla 35.

Tabla 35

Aporte de las Variables al KS del Modelo

Variable	Aporte de la variable %
Grupo_Score_Empresa_New_d2	15.45
AtrasoPromPondIntUlt6m_d2	10.42
NumNorUlt6m_d2	3.95
IncPromPasivoUlt6_d2	2.69
pasivos_1_d2	2.20
Antiguedad_Empresa_d2	1.99
Ventas_Soles_d2	0.55

Para hacer la validación del modelo de clasificación de buenos y malos se calcula la distribución de los clientes de acuerdo con el score asignado, tal como se aprecia en el Figura 38, este modelo presenta un nivel de divergencia de 0.62 que es bueno para un modelo de clasificación.

Figura 38

Distribución de Clientes que no han Incumplido el Pago de su Crédito (Buenos) y Clientes que han Incumplido el Pago de su Crédito (Malos) en la Muestra de Análisis

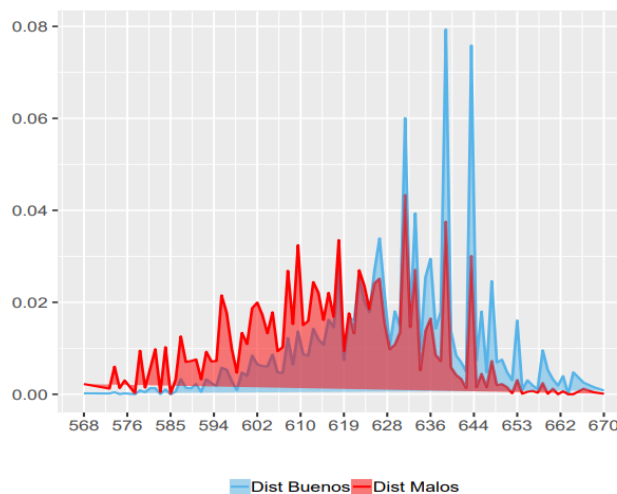


Figura 39

Distribución acumulada de Buenos y Malos

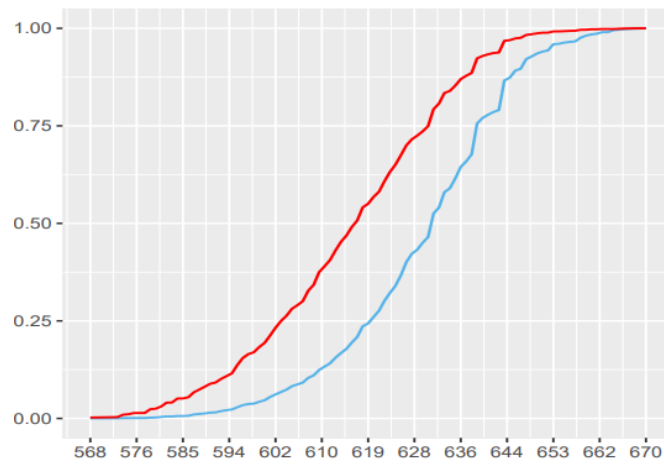


Figura 40

Promedio Teórico del Incumplimiento de Pago de Crédito del Modelo (pd_prom) y Porcentaje de Clientes en Incumplimiento de Pago de Crédito Real (% Malos) en la

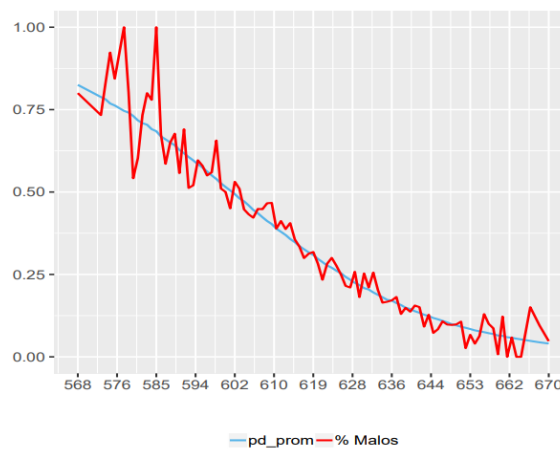


Tabla 36

Indicadores de Discriminancia - Muestra de Control (Fuera de la muestra)

KS	GINI	ROC	Divergencia	Estabilidad Poblacional
26.98	37.51	68.75	0.49	0.04524

Figura 41

Distribución de clientes que no han incumplido el pago de su crédito (Buenos) y clientes que han incumplido el pago de su crédito (Malos) en la muestra de control.

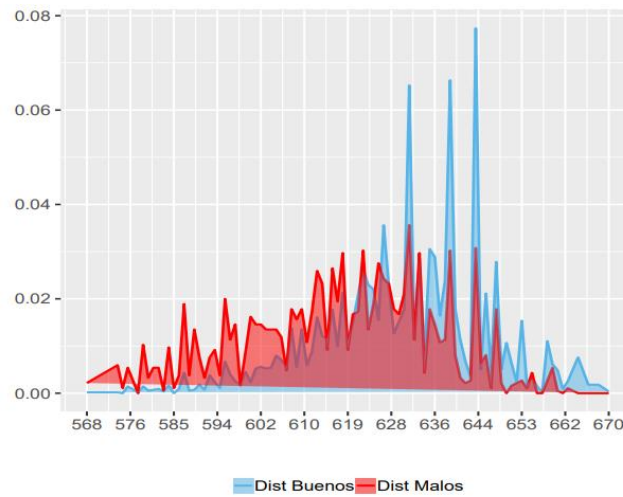


Figura 42

Distribución Acumulada de Buenos y Malos - Desarrollo vs Fuera de Muestra

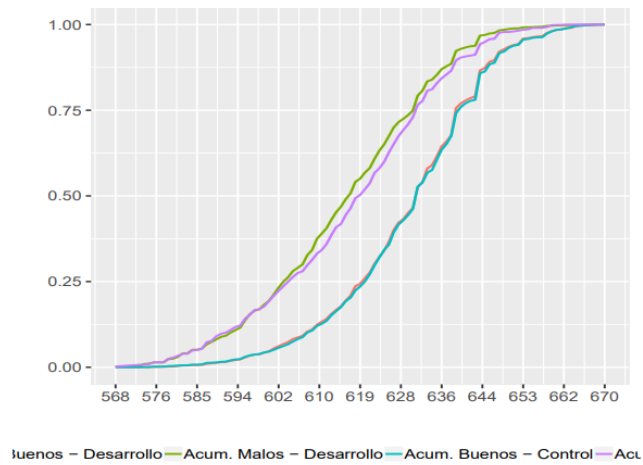


Figura 43

Promedio Teórico del Incumplimiento de Pago de Crédito del Modelo (pd_prom) y Porcentaje de Clientes en Incumplimiento de Pago de Crédito Real (% Malos) en la Muestra de Control



4.4.1 Grupos de Riesgo

Finalmente, el modelo puede presentarse estableciendo grupos de riesgo divididos tal como se observa en la Tabla 37. Esta partición fue aplicada a la muestra de análisis.

Tabla 37

Grupos de Riesgo por Deciles en la Muestra de Análisis.

	Grupos de Score	Clientes	Default	Default %	PD Prom %
1	(646 - 999]	2,022	167	8.26	8.50
2	(639 - 646]	2,633	326	12.43	12.59
3	(636 - 639]	2,247	337	15.00	15.23
4	(632 - 636]	2,175	392	18.02	17.88
5	(628 - 632]	2,371	512	21.59	20.59
6	(623 - 628]	2,445	575	23.52	24.52
7	(618 - 623]	2,007	565	28.15	28.40
8	(611 - 618]	2,691	924	34.34	34.24
9	(602 - 611]	2,215	1,005	45.37	43.03
10	[0 - 602]	2,496	1,438	57.61	59.12

Aplicando los cortes determinados a la muestra de control, podemos observar cómo se comportan los diferentes grupos de riesgo en el tiempo. Ver Tabla 38.

Tabla 38

Grupos de Riesgo por Deciles en la Muestra de Control.

	Grupos de Score	Clientes	Default	Default %	PD Prom %
1	(646 - 999]	720	79	10.97	8.49
2	(639 - 646]	904	114	12.61	12.63
3	(636 - 639]	687	97	14.12	15.32
4	(632 - 636]	653	123	18.84	17.79
5	(628 - 632]	767	157	20.47	20.60
6	(623 - 628]	809	207	25.59	24.38
7	(618 - 623]	692	161	23.27	28.44
8	(611 - 618]	820	280	34.15	34.51
9	(602 - 611]	609	221	36.29	43.22
10	[0 - 602]	726	413	56.89	60.02

4.5 Modelo Total

4.5.1 Resumen de Variables

Los tres modelos y sus variables se pueden resumir en la Tabla 39.

Tabla 39

Resumen de uso de Variables por Modelo

Variables	Descripción de variables	Segmento	Segmento	Segmento
		1	2	3
Antigüedad_Empresa	Antigüedad de la Empresa	✓	✓	✓
AtrasoPromPondIntUlt6m	Atraso Promedio Ponderado Interno - 6M			✓
AtrasoPromPondUlt24m	Atraso Promedio Ponderado SF últimos 24 meses	✓		
AtrasoPromPondUlt6m	Atraso Promedio Ponderado SF últimos 6 meses		✓	
EntRepPromUlt24m	Promedio de Entidades Reportadas últimos 24 meses	✓		
Grupo_Score_Empresa_New	Grupo Score Empresa -Accionista	✓	✓	✓
IncPromPasivoUlt6	Incremento Promedio de Pasivos últimos 6 meses			✓
NumNorUlt6m	Número de Normales últimos 6 meses			✓
Pasivos_1	Saldo Pasivos último mes			✓
PromPasivosUlt3	Promedio de Pasivos -3M		✓	
RatioUlt1MaxUlt6Pasivo	División saldo pasivo sobre máximo saldo pasivo de los últimos 6 meses	✓	✓	
Ventas_Soles	Ventas al cierre del ejercicio	✓		✓

4.5.2 Grupos de Riesgo

Finalmente, el modelo puede presentarse estableciendo grupos de riesgo divididos tal como se observa en la Tabla 40. Esta partición fue aplicada a la muestra de análisis.

Tabla 40

Grupos de Riesgo por Deciles en la Muestra de Análisis.

	Grupos de Score	Clientes	Default	Default %	PD Prom %
1	(680 - 999]	4,891	72	1.47	1.51
2	(669 - 680]	4,661	165	3.54	3.21
3	(660 - 669]	5,743	269	4.68	5.13
4	(652 - 660]	4,582	319	6.96	7.35
5	(644 - 652]	5,533	581	10.50	10.26
6	(639 - 644]	3,782	521	13.78	12.97
7	(632 - 639]	6,294	1,020	16.21	16.56
8	(624 - 632]	5,326	1,185	22.25	22.20
9	(613 - 624]	4,988	1,500	30.07	30.33
10	[0 - 613]	5,414	2,712	50.09	49.75

Aplicando los cortes determinados a la muestra de control, podemos observar cómo se comportan los diferentes grupos de riesgo en el tiempo. Ver Tabla 41.

Tabla 41

Grupos de Riesgo por Deciles en la Muestra de Control.

	Grupos de Score	Clientes	Default	Default %	PD Prom %
1	(680 - 999]	2,549	41	1.61	1.47
2	(669 - 680]	1,992	102	5.12	3.20
3	(660 - 669]	2,419	102	4.22	5.10
4	(652 - 660]	1,741	129	7.41	7.30
5	(644 - 652]	2,001	205	10.24	10.21
6	(639 - 644]	1,280	154	12.03	12.98
7	(632 - 639]	1,980	324	16.36	16.50
8	(624 - 632]	1,749	389	22.24	22.18
9	(613 - 624]	1,609	441	27.41	30.17
10	[0 - 613]	1,581	722	45.67	50.08

CONCLUSIONES

- Basándonos en el análisis realizado utilizando la probabilidad de incumplimiento de pago (PD) junto con las variables patrón interno, patrón externo y nivel de ventas, se han identificado tres categorías de riesgo para los clientes: Estos grupos, que se han segmentado utilizando árboles de inferencia condicional, demuestran una clara distinción entre clientes buenos y malos pagadores.
 - Riesgo bajo: 1
 - Riesgo medio: 2
 - Riesgo alto: 3

Estos grupos, que se han segmentado utilizando árboles de inferencia condicional, demuestran una clara distinción (discriminación) entre clientes buenos y malos pagadores

- Se desarrolló modelos de regresión logística dentro de cada uno de los tres segmentos definidos por los árboles de inferencia condicional. Estos modelos fueron seleccionados en función de sus indicadores de rendimiento, específicamente el índice de Gini y el valor de KS. Los modelos finales seleccionados en cada segmento exhibieron indicadores de performance destacados.

Este análisis respalda la existencia de modelos de regresión logística robustos que proporcionan puntajes diferenciados para la clasificación de clientes como buenos o malos pagadores.

- Se ha demostrado la factibilidad de identificar grupos claramente diferenciados basándose en los puntajes proporcionados por las regresiones logísticas. Mediante el uso de árboles de inferencia condicional y regresión logística, se construyeron tres submodelos, cada uno capaz de asignar un puntaje (score) a cada cliente, clasificándolos según su probabilidad de incumplimiento de pago.

Como resultado de este análisis, se identificaron diez grupos de riesgo distintos.

- Se ha evidenciado una notable disminución del 30% en la prima de riesgo en los primeros meses del 2021, gracias a la implementación exitosa de los modelos de regresión logística y árboles de inferencia condicional. Este resultado es particularmente significativo, ya que supera el objetivo de reducción del 20% establecido anteriormente, y se produce en un contexto en el que la prima de riesgo había experimentado un aumento del 50% en el último semestre del 2020.

La implementación de estos modelos estadísticos ha demostrado ser una herramienta efectiva que mejora significativamente la gestión del riesgo de los clientes, lo que se refleja en la reducción de la prima de riesgo. Este hallazgo subraya el impacto positivo de los modelos en la gestión del riesgo crediticio.

- La implementación de modelos de clasificación robustos como la regresión logística y los árboles de inferencia condicional ha sido fundamental en la mejora de la gestión del riesgo crediticio en el contexto financiero. Estos modelos han permitido identificar grupos de clientes con diferentes niveles de riesgo, clasificándolos de manera efectiva como buenos o malos pagadores. Además, han contribuido significativamente a la reducción del valor de la prima de riesgo, mejorando así la gestión del riesgo crediticio en su conjunto.

En conjunto, los hallazgos mencionados respaldan la eficacia de los modelos de clasificación utilizados para mejorar la gestión del riesgo crediticio y reducir el valor de la prima de riesgo en el contexto financiero. Este estudio subraya la importancia de utilizar modelos estadísticos avanzados para abordar los desafíos en la gestión del riesgo crediticio, lo que puede tener un impacto significativo en la estabilidad financiera y el éxito de las instituciones financieras.

RECOMENDACIONES

- Implementar un sistema de monitoreo trimestral para verificar que los indicadores de discriminancia de los modelos de la entidad financiera se mantengan dentro de los límites establecidos por la política de la entidad financiera.
- Se sugiere incluir variables macroeconómicas como el Producto Bruto Interno (PBI), la Inflación, Tipo de Cambio, Empleo Formal, entre otras, durante el proceso de calibración del modelo.
- Desarrollar una plantilla en Latex¹⁸ con R¹⁹ que nos permita documentar los procedimientos y metodologías utilizadas en el desarrollo de modelos predictivos. Esto facilitaría las futuras revisiones y auditorías.
- Explorar nuevas técnicas estadísticas como el XGBoost (refuerzo de gradientes extremo) que es una implementación eficiente y escalable de algoritmos de aumento de gradiente, utilizado para problemas de aprendizaje automático y análisis predictivo. Destaca por su rendimiento y capacidad para manejar conjuntos de datos grandes y complejos (Chen & Guestrin, 2016).
- En las variables de atraso realizar un análisis de materialidad para no castigar a empresas por deudas pequeñas en atraso que pueden ser causadas por un sobregiro producto de una mala gestión.

¹⁸ Es un sistema de preparación de documentos que facilita la creación de documentos de alta calidad, especialmente útil para la redacción de artículos científicos, tesis y documentos técnicos en informática y otras disciplinas.

¹⁹ Es un lenguaje de programación y entorno de software utilizado para estadísticas y análisis de datos.

REFERENCIAS BIBLIOGRÁFICAS

- Adjei, I., & Karim, R. (2016). An Application of Bootstrapping in Logistic Regression Model. *Scientific Research*, 3, 1-9. doi:10.4236/oalib.1103049
- Anderson, R. (2007). *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. Oxford University Press.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (págs. 785-794). Obtenido de <https://doi.org/10.1145/2939672.2939785>
- Das, S., & Chatterjee, S. (2011). Multicollinearity Problem – Root Cause, Diagnostics and Way Outs. *SSRN*, 17. Obtenido de https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1830043
- Gujarati, D. (2003). *Econometría*. McGraw-Hill Interamericana.
- Gupta, R. (6 de 12 de 2019). *An Introduction to Discretization An Introduction to Discretization*. Obtenido de Towards data science: <https://towardsdatascience.com/an-introduction-to-discretization-in-data-science-55ef8c9775a2>
- Halim, S., & Humira, Y. (2014). Credit scoring modeling. *Jurnal Teknik Industri*, 16, 17-24. doi:10.9744/jti.16.1.17-24
- M., D., & Kiruthika. (2014). Model Validation of a Credit Scorecard Using Bootstrap Method. *IOSR Journal of Economics and Finance*, 3, 64-68. doi:10.9790/5933-0336468
- Siddiqi, N. (2006). *Credit risk scorecards: developing and implementing intelligent credit*

scoring. John Wiley & Sons, Inc.

Strobl, C., Malley, J., & Tutz, G. (2009). An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. *Psychological Methods*, 4, 323-348.

doi:10.1037/a0016973

Tineo Nieves, A. K. (2019). *Modelos de clasificación de clientes bancarizados de los niveles socioeconómicos "C" y "D"*. [Tesis de maestría. Graduate School of Business - ESAN].

Tufféry, S. (2011). *Data Mining and Statistics for Decision Making*. John Wiley & Sons, Ltd.

Wicklin, R. (08 de 02 de 2017). *Winsorization: The good, the bad, and the ugly*. Obtenido de [Entrada de blog]: <https://blogs.sas.com/content/iml/2017/02/08/winsorization-good-bad-and-ugly.html>

ANEXOS

Anexo 1:	Segmento 1, según Patrón Interno / Patrón Externo / Nivel de Ventas	1
Anexo 2:	Segmento 2, según Patrón Interno / Patrón Externo / Nivel de Ventas	10
Anexo 3:	Segmento 3, según Patrón Interno / Patrón Externo / Nivel de Ventas	18

Anexo 1:

Segmento 1, según Patrón Interno / Patrón Externo / Nivel de Ventas

Análisis de Estabilidad y Correlación de Variables - Segmento 1

Luego para analizar la consistencia de los coeficientes se procedió a aplicar la técnica del Bootstrap, se tomó 1000 muestras sin reemplazo de la muestra de validación con la intención de ver la distribución de cada coeficiente, en los casos donde se encontró que más del 5% de la distribución presentaba un signo cambiado se procedió a unir las categorías de la variable, en el caso que todas las categorías de la variable presentan más del 5% de su distribución con un signo contrario se decidió retirar a la variable.

A continuación, se muestra la distribución de cada coeficiente del modelo final:

Cuantiles de los Estimadores Generados por el Método Bootstrap

Variabes	1%	5%	10%	25%	50%	75%	90%	95%	99%
(Intercept)	-5.05167	-4.7895	-4.63118	-4.47375	-4.30818	-4.13546	-4.02741	-3.96388	-3.85075
Grupo_Score_Empresa_New_d24	0.33841	0.4288	0.48175	0.57368	0.66752	0.763	0.8404	0.88571	0.97056
Grupo_Score_Empresa_New_d26	0.68052	0.78317	0.84824	0.93743	1.031	1.12121	1.20906	1.26757	1.35313
AtrasoPromPondUlt24m_d22	0.55102	0.68878	0.75326	0.87561	1.03119	1.19517	1.38379	1.50545	1.73044
AtrasoPromPondUlt24m_d24	0.84129	0.95333	1.01103	1.13988	1.29472	1.4614	1.63381	1.73077	1.99356
AtrasoPromPondUlt24m_d28	1.08524	1.21098	1.2885	1.42047	1.59549	1.77502	1.95598	2.0664	2.31972
EntRepPromUlt24m_d2	0.35682	0.42284	0.45562	0.50896	0.58136	0.6581	0.73301	0.78023	0.85362
EntRepPromUlt24m_d3	1.41131	1.56876	1.63789	1.75965	1.90292	2.04544	2.16365	2.23281	2.35645
Antigüedad_Empresa_d25	-0.52366	-0.44535	-0.4025	-0.33042	-0.24674	-0.16188	-0.08693	-0.04586	0.0253
Antigüedad_Empresa_d26	-1.10912	-0.83909	-0.75515	-0.57156	-0.40661	-0.27125	-0.14445	-0.07857	0.03494
RatioUlt1MaxUlt6Pasivo_d24	-0.50704	-0.44912	-0.40706	-0.34781	-0.27974	-0.20656	-0.15011	-0.10651	-0.01986
VENTAS_SOLES_d23	-1.42038	-1.27555	-1.21977	-1.11608	-1.01518	-0.9187	-0.82968	-0.7933	-0.70732

Para las variables finales dentro del modelo se presenta a continuación la matriz de correlación, en donde, se han resaltado de negrita los valores que exceden el límite máximo establecido (70% de correlación).

Matriz de Correlación de Variables Finales del Segmento 1

Variables	Grupo_Score_Empresa_New_d2	AtrasoPromPondUlt24m_d2	EntRepPromUlt24m_d	Antiguedad_Empresa_d2	RatioUlt1MaxUlt6Pasivo_d2	VENTAS_SOLES_d2
Grupo_Score_Empresa_New_d2	1.00	0.04	-0.04	-0.07	-0.06	-0.17
AtrasoPromPondUlt24m_d2	0.04	1.00	0.35	0.04	-0.07	0.14
EntRepPromUlt24m_d	-0.04	0.35	1.00	0.18	-0.04	0.34
Antiguedad_Empresa_d2	-0.07	0.04	0.18	1.00	0.06	0.22
RatioUlt1MaxUlt6Pasivo_d2	-0.06	-0.07	-0.04	0.06	1.00	0.06
VENTAS_SOLES_d2	-0.17	0.14	0.34	0.22	0.06	1.00

Análisis de Variables por Incumplimiento de Pago(Default) - Segmento 1

A cada una de las variables se le realizó un análisis contra la variable objetivo (variable default), para verificar su poder de discriminancia individual. Las variables con un sufijo “_d” al final indican que han sido transformadas, mientras que las variables con el sufijo “_d2” indican que han sido reajustadas.

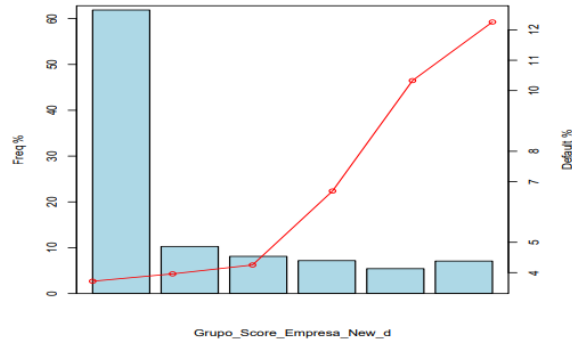
A continuación, se muestra el análisis bivariado por cada variable final del modelo.

a) Grupo_Score_Empresa_New

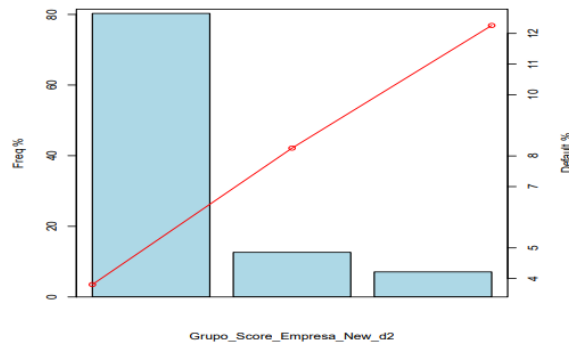
Análisis de la Variable Grupo_Score_Empresa_New_d vs Default - Segmento 1

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
4>= Grupo_Score_Empresa_New or Grupo_Score_Empresa_New is null	1	10'034	61.88	373	3.72	9'661	373	0.63	0.46	30.23	0.05
Grupo_Score_Empresa_New >4 and 5>= Grupo_Score_Empresa_New	2	1'667	10.28	66	3.96	1'601	66	0.10	0.08	23.68	0.01
Grupo_Score_Empresa_New >5 and 6>= Grupo_Score_Empresa_New	3	1'317	8.12	56	4.25	1'261	56	0.08	0.07	16.24	0.00
Grupo_Score_Empresa_New >6 and 7>= Grupo_Score_Empresa_New	4	1'166	7.19	78	6.69	1'088	78	0.07	0.10	-31.65	0.01
Grupo_Score_Empresa_New >7 and 8>= Grupo_Score_Empresa_New	5	881	5.43	91	10.33	790	91	0.05	0.11	-79.08	0.05
Grupo_Score_Empresa_New >8	6	1'150	7.09	141	12.26	1'009	141	0.07	0.18	-98.40	0.11
Grupo_Score_Empresa_New_d		16'215		805		15'410	805				0.22

Participación de Grupo_Score_Empresa_New_d por Default - Segmento 1



Participación de Grupo_Score_Empresa_New_d2 por Default - Segmento 1



Análisis de la Variable Grupo_Score_Empresa_New_d2 vs Default - Segmento 1

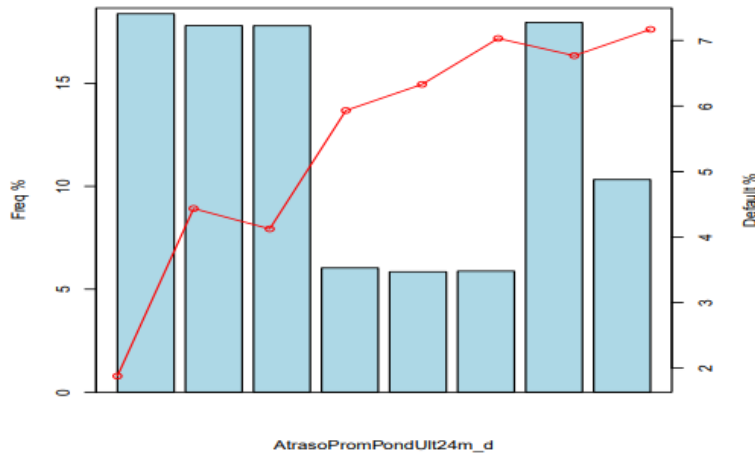
	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= Grupo_Score_Empresa_New_d <=3	1	13'018	80.28	495	3.80	12'523	495	0.81	0.61	27.88	0.06
4<= Grupo_Score_Empresa_New_d <=5	4	2'047	12.62	169	8.26	1'878	169	0.12	0.21	-54.39	0.05
6<= Grupo_Score_Empresa_New_d <=6	6	1'150	7.09	141	12.26	1'009	141	0.07	0.18	-98.40	0.11
Grupo_Score_Empresa_New_d2		16'215		805		15'410	805				0.21

b) AtrasoPromPondUlt24m

Análisis de la Variable AtrasoPromPondUlt24m_d vs Default - Segmento 1

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
0.04>= AtrasoPromPondUlt24m	1	2'980	18.38	56	1.88	2'924	56	0.19	0.07	100.34	0.12
AtrasoPromPondUlt24m >0.04 and 0.51>= AtrasoPromPondUlt24m	2	2'886	17.80	128	4.44	2'758	128	0.18	0.16	11.83	0.00
AtrasoPromPondUlt24m >0.51 and 1.57>= AtrasoPromPondUlt24m	3	2'885	17.79	119	4.12	2'766	119	0.18	0.15	19.41	0.01
AtrasoPromPondUlt24m >1.57 and 2.08>= AtrasoPromPondUlt24m	4	978	6.03	58	5.93	920	58	0.06	0.07	-18.80	0.00
AtrasoPromPondUlt24m >2.08 and 2.76>= AtrasoPromPondUlt24m	5	948	5.85	60	6.33	888	60	0.06	0.07	-25.73	0.00
AtrasoPromPondUlt24m >2.76 and 3.65>= AtrasoPromPondUlt24m	6	953	5.88	67	7.03	886	67	0.06	0.08	-36.99	0.01
AtrasoPromPondUlt24m >3.65 and 8.86>= AtrasoPromPondUlt24m	7	2'911	17.95	197	6.77	2'714	197	0.18	0.24	-32.90	0.02
AtrasoPromPondUlt24m >8.86 or AtrasoPromPondUlt24m is null	8	1'674	10.32	120	7.17	1'554	120	0.10	0.15	-39.08	0.02
AtrasoPromPondUlt24m_d		16'215		805		15'410	805				0.19

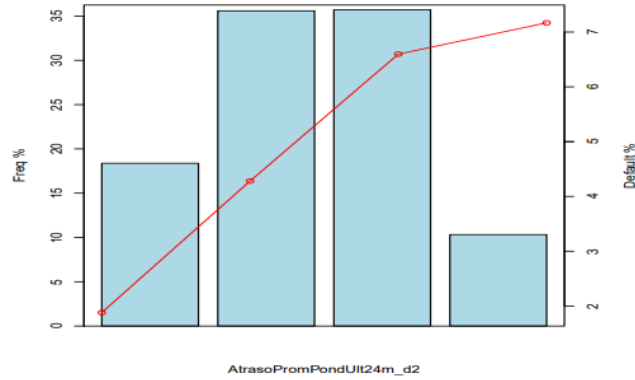
Participación de AtrasoPromPondUlt24m_d por Default - Segmento 1



Análisis de la Variable AtrasoPromPondUlt24m_d2 vs Default - Segmento 1

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= AtrasoPromPondUlt24m_d <=1	1	2'980	18.38	56	1.88	2'924	56	0.19	0.07	100.34	0.12
2<= AtrasoPromPondUlt24m_d <=3	2	5'771	35.59	247	4.28	5'524	247	0.36	0.31	15.55	0.01
4<= AtrasoPromPondUlt24m_d <=7	4	5'790	35.71	382	6.60	5'408	382	0.35	0.47	-30.17	0.04
8<= AtrasoPromPondUlt24m_d <=8	8	1'674	10.32	120	7.17	1'554	120	0.10	0.15	-39.08	0.02
AtrasoPromPondUlt24m_d2		16'215		805		15'410	805				0.18

Participación de AtrasoPromPondUlt24m_d2 por Default - Segmento 1

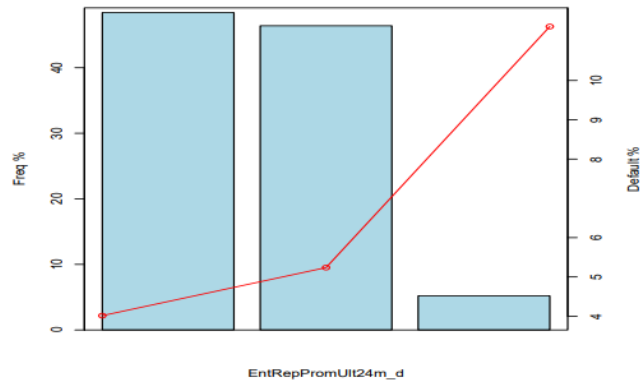


c) EntRepPromUlt24m

Análisis de la Variable EntRepPromUlt24m_d vs Default - Segmento 1

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
3.54 >= EntRepPromUlt24m or EntRepPromUlt24m is null	1	7'848	48.40	315	4.01	7'533	315	0.49	0.39	22.25	0.02
EntRepPromUlt24m >3.54 and 7.12 >= EntRepPromUlt24m	2	7'523	46.40	394	5.24	7'129	394	0.46	0.49	-5.64	0.00
EntRepPromUlt24m >7.12	3	844	5.21	96	11.37	748	96	0.05	0.12	-89.89	0.06
EntRepPromUlt24m_d		16'215		805		15'410	805				0.09

Participación de EntRepPromUlt24m_d por Default - Segmento 1

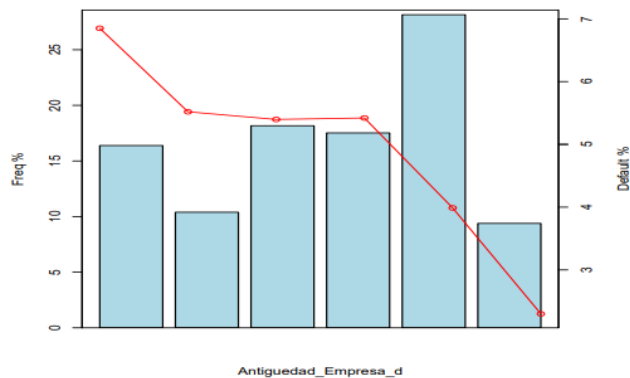


d) Antigüedad_Empresa

Análisis de la Variable Antigüedad_Empresa_d vs Default - Segmento 1

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
6<= Antigüedad_Empresa or Antigüedad_Empresa is null	1	2'657	16.39	182	6.85	2'475	182	0.16	0.23	-34.19	0.02
Antigüedad_Empresa >6 and 8<= Antigüedad_Empresa	2	1'685	10.39	93	5.52	1'592	93	0.10	0.12	-11.18	0.00
Antigüedad_Empresa >8 and 12<= Antigüedad_Empresa	3	2'946	18.17	159	5.40	2'787	159	0.18	0.20	-8.81	0.00
Antigüedad_Empresa >12 and 16<= Antigüedad_Empresa	4	2'841	17.52	154	5.42	2'687	154	0.17	0.19	-9.27	0.00
Antigüedad_Empresa >16 and 34<= Antigüedad_Empresa	5	4'564	28.15	182	3.99	4'382	182	0.28	0.23	22.93	0.01
Antigüedad_Empresa >34	6	1'522	9.39	35	2.30	1'487	35	0.10	0.04	79.72	0.04
Antigüedad_Empresa_d		16'215		805		15'410	805				0.08

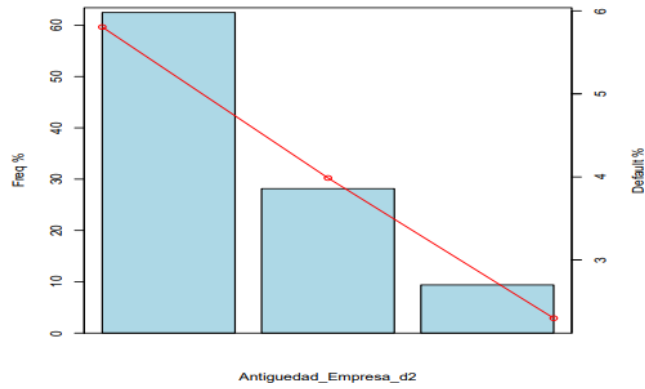
Participación de Antigüedad_Empresa_d por Default - Segmento 1



Análisis de la Variable Antigüedad_Empresa_d2 vs Default - Segmento 1

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= Antigüedad_Empresa_d <=4	1	10'129	62.47	588	5.81	9'541	588	0.62	0.73	-16.53	0.02
5<= Antigüedad_Empresa_d <=5	5	4'564	28.15	182	3.99	4'382	182	0.28	0.23	22.93	0.01
6<= Antigüedad_Empresa_d <=6	6	1'522	9.39	35	2.30	1'487	35	0.10	0.04	79.72	0.04
Antigüedad_Empresa_d2		16'215		805		15'410	805				0.07

Participación de Antigüedad_Empresa_d2 por Default - Segmento 1

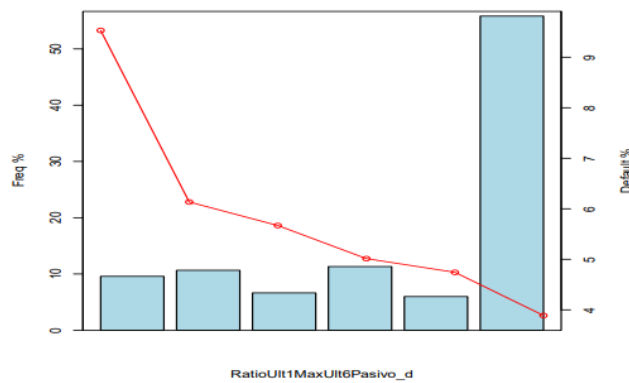


e) RatioUlt1MaxUlt6Pasivo

Análisis de la Variable RatioUlt1MaxUlt6Pasivo_d vs Default - Segmento 1

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
0.01>= RatioUlt1MaxUlt6Pasivo or RatioUlt1MaxUlt6Pasivo is null	1	1'553	9.58	148	9.53	1'405	148	0.09	0.18	-70.13	0.06
RatioUlt1MaxUlt6Pasivo >0.01 and 0.06>= RatioUlt1MaxUlt6Pasivo	2	1'728	10.66	106	6.13	1'622	106	0.11	0.13	-22.40	0.01
RatioUlt1MaxUlt6Pasivo >0.06 and 0.11>= RatioUlt1MaxUlt6Pasivo	3	1'076	6.64	61	5.67	1'015	61	0.07	0.08	-14.02	0.00
RatioUlt1MaxUlt6Pasivo >0.11 and 0.24>= RatioUlt1MaxUlt6Pasivo	4	1'835	11.32	92	5.01	1'743	92	0.11	0.11	-1.04	0.00
RatioUlt1MaxUlt6Pasivo >0.24 and 0.33>= RatioUlt1MaxUlt6Pasivo	5	970	5.98	46	4.74	924	46	0.06	0.06	4.81	0.00
RatioUlt1MaxUlt6Pasivo >0.33	6	9'053	55.83	352	3.89	8'701	352	0.56	0.44	25.56	0.03
RatioUlt1MaxUlt6Pasivo_d		16'215		805		15'410	805				0.11

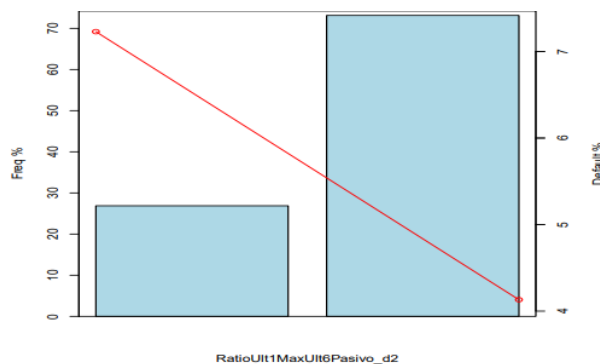
Participación de RatioUlt1MaxUlt6Pasivo_d por Default - Segmento 1



Análisis de la Variable RatioUlt1MaxUlt6Pasivo_d2 vs Default - Segmento 1

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= RatioUlt1MaxUlt6Pasivo_d <=3	1	4'357	26.87	315	7.23	4'042	315	0.26	0.39	-40.00	0.05
4<= RatioUlt1MaxUlt6Pasivo_d <=6	4	11'858	73.13	490	4.13	11'368	490	0.74	0.61	19.22	0.02
RatioUlt1MaxUlt6Pasivo_d2		16'215		805		15'410	805				0.08

Participación de RatioUlt1MaxUlt6Pasivo_d2 por Default - Segmento 1

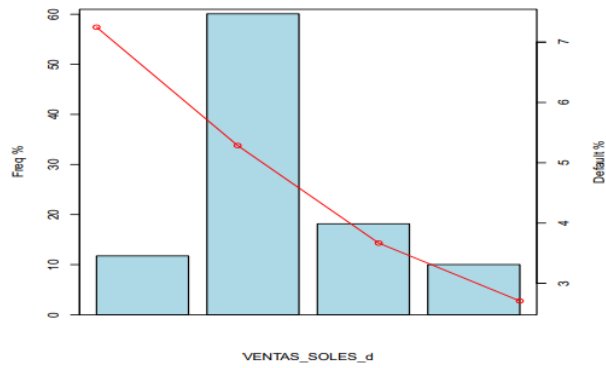


f) Ventas_Soles

Análisis de la Variable Ventas_Soles_d vs Default - Segmento 1

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
3.14e+03>= VENTAS_SOLES	1	1'904	11.74	138	7.25	1'766	138	0.11	0.17	-40.27	0.02
VENTAS_SOLES >3.14e+03 and 4.4e+04>= VENTAS_SOLES	2	9'743	60.09	515	5.29	9'228	515	0.60	0.64	-6.61	0.00
VENTAS_SOLES >4.4e+04 and 1.66e+05>= VENTAS_SOLES	3	2'944	18.16	108	3.67	2'836	108	0.18	0.13	31.61	0.02
VENTAS_SOLES >1.66e+05	4	1'624	10.02	44	2.71	1'580	44	0.10	0.05	62.91	0.03
VENTAS_SOLES_d		16'215		805		15'410	805				0.07

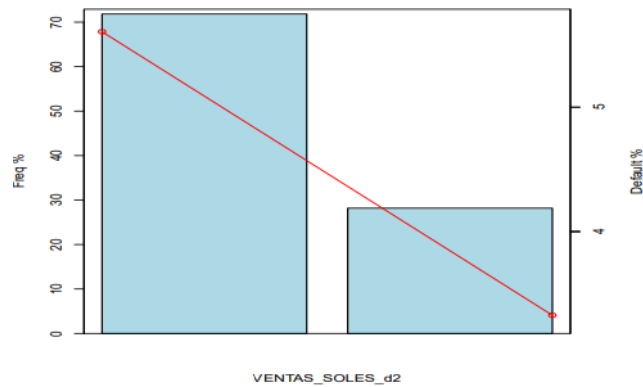
Participación de Ventas_Soles_d por Default - Segmento 1



Análisis de la Variable Ventas_Soles_d2 vs Default - Segmento 1

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= VENTAS_SOLES_d <=2	1	11'647	71.83	653	5.61	10'994	653	0.71	0.81	-12.84	0.01
3<= VENTAS_SOLES_d <=4	3	4'568	28.17	152	3.33	4'416	152	0.29	0.19	41.72	0.04
VENTAS_SOLES_d2		16'215		805		15'410	805				0.05

Participación de Ventas_Soles_d2 por Default - Segmento 1



Anexo 2:

Segmento 2, según Patrón Interno / Patrón Externo / Nivel de Ventas

Análisis de Estabilidad y Correlación de Variables - Segmento 2

Luego para analizar la consistencia de los coeficientes se procedió a aplicar la técnica del Bootstrap, se tomó 1000 muestras sin reemplazo de la muestra de validación con la intención de ver la distribución de cada coeficiente, en los casos donde se encontró que más del 5% de la distribución presentaba un signo cambiado se procedió a unir las categorías de la variable, en el caso que todas las categorías de la variable presentan más del 5% de su distribución con un signo contrario se decidió retirar a la variable.

Cuantiles de los Estimadores Generados por el Método Bootstrap

VARIABLES	1%	5%	10%	25%	50%	75%	90%	95%	99%
(Intercept)	-2.27602	-2.24013	-2.21101	-2.17065	-2.12587	-2.08506	-2.04875	-2.02724	-1.99187
AtrasoPromPondUlt6m_d23	0.01829	0.08355	0.1198	0.16231	0.21677	0.26957	0.31275	0.33194	0.38181
AtrasoPromPondUlt6m_d24	0.54306	0.58947	0.61287	0.64332	0.67693	0.71618	0.74439	0.76682	0.79999
Antigüedad_Empresa_d24	-0.43701	-0.38798	-0.36704	-0.33072	-0.28771	-0.24564	-0.21134	-0.18928	-0.1455
Antigüedad_Empresa_d26	-0.56087	-0.5124	-0.48563	-0.45047	-0.41096	-0.37108	-0.33599	-0.31109	-0.27708
PromPasivosUlt3_d27	-0.32575	-0.27307	-0.25057	-0.21252	-0.17586	-0.13515	-0.10048	-0.07894	-0.04476
PromPasivosUlt3_d28	-0.59913	-0.54356	-0.5149	-0.47298	-0.42157	-0.37769	-0.33639	-0.3098	-0.26368
RatioUlt1MaxUlt6Pasivo_d2	-0.66168	-0.61783	-0.60018	-0.56212	-0.51778	-0.47376	-0.43549	-0.41066	-0.37315
RatioUlt1MaxUlt6Pasivo_d3	-0.79531	-0.75165	-0.72503	-0.68539	-0.6427	-0.60334	-0.56429	-0.54004	-0.50795
Grupo_Score_Empresa_New_d2	0.38948	0.4223	0.43898	0.47305	0.51351	0.55133	0.58468	0.60681	0.648
Grupo_Score_Empresa_New_d3	0.39251	0.45336	0.48778	0.54207	0.60231	0.65377	0.70201	0.74153	0.78546
Grupo_Score_Empresa_New_d4	1.2018	1.2364	1.25661	1.29513	1.3397	1.38251	1.42248	1.44458	1.48848

Para las variables finales dentro del modelo se presenta a continuación la matriz de correlación, en donde, se han marcado en negrita los valores que exceden el límite máximo establecido (70% de correlación).

Matriz de Correlación de Variables - Segmento 2

Variables	AtrasoPromPondUlt6m_d2	Antiguedad_Empresa_d2	PromPasivosUlt3_d2	RatioUlt1MaxUlt6Pasivo_d	Grupo_Score_Empresa_New_d
AtrasoPromPondUlt6m_d2	1.00	0.07	-0.02	-0.04	0.10
Antiguedad_Empresa_d2	0.07	1.00	0.06	0.05	-0.06
PromPasivosUlt3_d2	-0.02	0.06	1.00	0.38	-0.03
RatioUlt1MaxUlt6Pasivo_d	-0.04	0.05	0.38	1.00	-0.05
Grupo_Score_Empresa_New_d	0.10	-0.06	-0.03	-0.05	1.00

Análisis de Variables por Default - Segmento 2

A cada una de las variables se le realizó un análisis contra la variable objetivo (variable default), para verificar su poder de discriminancia individual. Las variables con un sufijo “_d” al final indican que han sido transformadas, mientras que las variables con el sufijo “_d2” indican que han sido reajustadas.

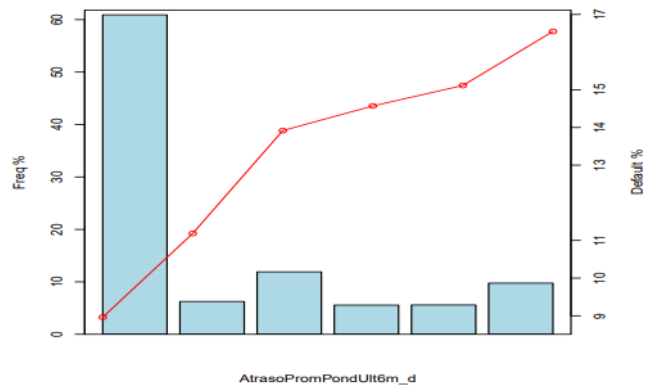
A continuación, se muestra el análisis bivariado por cada variable final del modelo.

a) AtrasoPromPondUlt6m

Análisis de la variable AtrasoPromPondUlt6m_d vs Default - Segmento 2

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
0.57>= AtrasoPromPondUlt6m or AtrasoPromPondUlt6m is null	1	7'131	60.91	639	8.96	6'492	639	0.62	0.49	23.66	0.03
AtrasoPromPondUlt6m >0.57 and 1.05>= AtrasoPromPondUlt6m	2	733	6.26	82	11.19	651	82	0.06	0.06	-1.01	0.00
AtrasoPromPondUlt6m >1.05 and 2.86>= AtrasoPromPondUlt6m	3	1'394	11.91	194	13.92	1'200	194	0.12	0.15	-25.96	0.01
AtrasoPromPondUlt6m >2.86 and 4.52>= AtrasoPromPondUlt6m	4	652	5.57	95	14.57	557	95	0.05	0.07	-31.32	0.01
AtrasoPromPondUlt6m >4.52 and 7.14>= AtrasoPromPondUlt6m	5	655	5.59	99	15.11	556	99	0.05	0.08	-35.62	0.01
AtrasoPromPondUlt6m >7.14	6	1'142	9.75	189	16.55	953	189	0.09	0.15	-46.40	0.03
AtrasoPromPondUlt6m_d		11'707		1'298		10'409	1'298				0.08

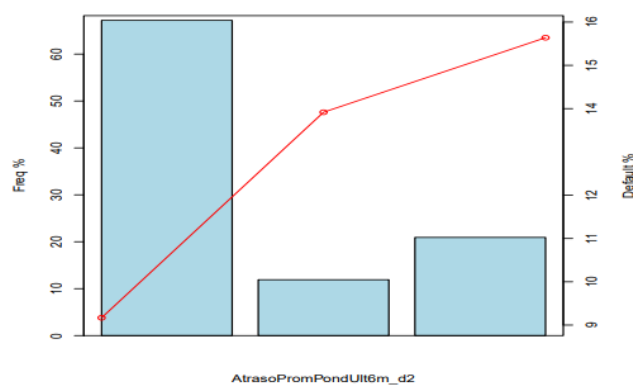
Participación de *AtrasoPromPondUlt6m_d* por Default - Segmento 2



Análisis de la Variable *AtrasoPromPondUlt6m_d2* vs Default - Segmento 2

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= <i>AtrasoPromPondUlt6m_d</i> <=2	1	7'864	67.17	721	9.17	7'143	721	0.69	0.56	21.14	0.03
3<= <i>AtrasoPromPondUlt6m_d</i> <=3	3	1'394	11.91	194	13.92	1'200	194	0.12	0.15	-25.96	0.01
4<= <i>AtrasoPromPondUlt6m_d</i> <=6	4	2'449	20.92	383	15.64	2'066	383	0.20	0.30	-39.65	0.04
<i>AtrasoPromPondUlt6m_d2</i>		11'707		1'298		10'409	1'298				0.07

Participación de *AtrasoPromPondUlt6m_d2* por Default - Segmento 2

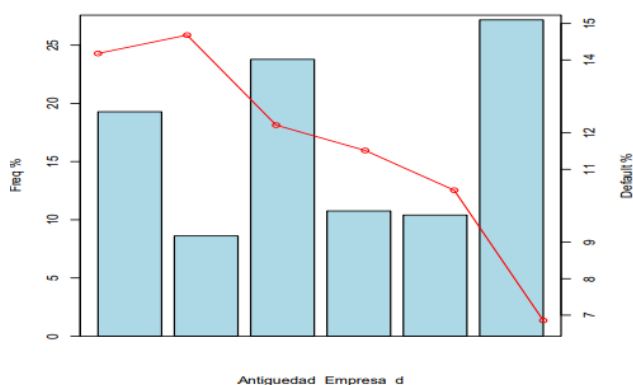


b) Antigüedad_Empresa

Análisis de la Variable Antigüedad_Empresa_d vs Default - Segmento 2

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
4>= Antigüedad_Empresa or Antigüedad_Empresa is null	1	2'257	19.28	320	14.18	1'937	320	0.19	0.25	-28.13	0.02
Antigüedad_Empresa >4 and 5>= Antigüedad_Empresa	2	1'008	8.61	148	14.68	860	148	0.08	0.11	-32.21	0.01
Antigüedad_Empresa >5 and 8>= Antigüedad_Empresa	3	2'784	23.78	340	12.21	2'444	340	0.23	0.26	-10.94	0.00
Antigüedad_Empresa >8 and 10>= Antigüedad_Empresa	4	1'259	10.75	145	11.52	1'114	145	0.11	0.11	-4.29	0.00
Antigüedad_Empresa >10 and 13>= Antigüedad_Empresa	5	1'218	10.40	127	10.43	1'091	127	0.10	0.10	6.88	0.00
Antigüedad_Empresa >13	6	3'181	27.17	218	6.85	2'963	218	0.28	0.17	52.76	0.06
Antigüedad_Empresa_d		11'707		1'298		10'409	1'298				0.09

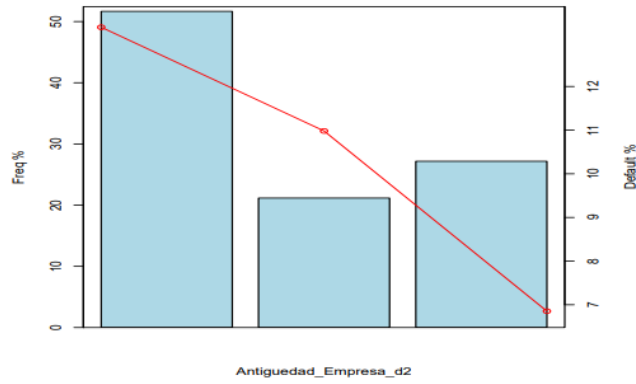
Participación de Antigüedad_Empresa_d por Default - Segmento 2



Análisis de la Variable Antigüedad_Empresa_d2 vs Default - Segmento 2

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= Antigüedad_Empresa_d <=3	1	6'049	51.67	808	13.36	5'241	808	0.50	0.62	-21.21	0.03
4<= Antigüedad_Empresa_d <=5	4	2'477	21.16	272	10.98	2'205	272	0.21	0.21	1.08	0.00
6<= Antigüedad_Empresa_d <=6	6	3'181	27.17	218	6.85	2'963	218	0.28	0.17	52.76	0.06
Antigüedad_Empresa_d2		11'707		1'298		10'409	1'298				0.09

Participación de Antigüedad_Empresa_d2 por Default - Segmento 2

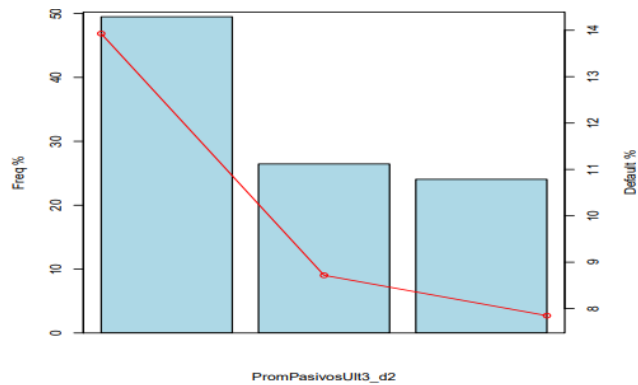


c) PromPasivosUlt3

Análisis de la Variable PromPasivosUlt3_d vs Default - Segmento 2

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
100>= PromPasivosUlt3 or PromPasivosUlt3 is null	1	2'139	18.27	347	16.22	1'792	347	0.17	0.27	-44.01	0.04
PromPasivosUlt3 >100 and 368>= PromPasivosUlt3	2	597	5.10	86	14.41	511	86	0.05	0.07	-29.98	0.01
PromPasivosUlt3 >368 and 870>= PromPasivosUlt3	3	607	5.18	89	14.66	518	89	0.05	0.07	-32.05	0.01
PromPasivosUlt3 >870 and 1.61e+03>= PromPasivosUlt3	4	594	5.07	83	13.97	511	83	0.05	0.06	-26.43	0.00
PromPasivosUlt3 >1.61e+03 and 4.47e+03>= PromPasivosUlt3	5	1'245	10.63	145	11.65	1'100	145	0.11	0.11	-5.55	0.00
PromPasivosUlt3 >4.47e+03 and 6.97e+03>= PromPasivosUlt3	6	612	5.23	57	9.31	555	57	0.05	0.04	19.41	0.00
PromPasivosUlt3 >6.97e+03 and 4.91e+04>= PromPasivosUlt3	7	3'097	26.45	270	8.72	2'827	270	0.27	0.21	26.67	0.02
PromPasivosUlt3 >4.91e+04 and 2.61e+05>= PromPasivosUlt3	8	1'820	15.55	154	8.46	1'666	154	0.16	0.12	29.94	0.01
PromPasivosUlt3 >2.61e+05	9	996	8.51	67	6.73	929	67	0.09	0.05	54.76	0.02
PromPasivosUlt3_d		11'707		1'298		10'409	1'298				0.11

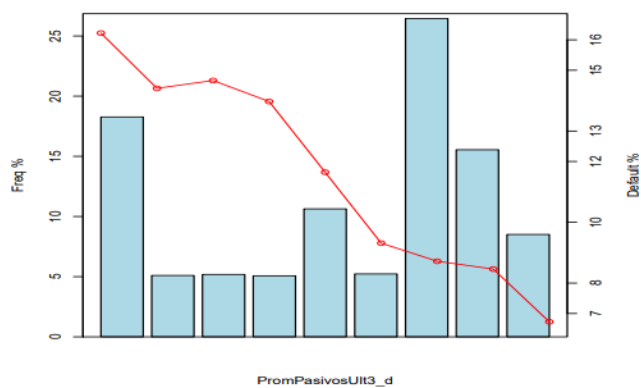
Participación de PromPasivosUlt3_d por Default - Segmento 2



Análisis de la Variable PromPasivosUlt3_d2 vs Default - Segmento 2

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1 <= PromPasivosUlt3_d <= 6	1	5'794	49.49	807	13.93	4'987	807	0.48	0.62	-26.06	0.04
7 <= PromPasivosUlt3_d <= 7	7	3'097	26.45	270	8.72	2'827	270	0.27	0.21	26.67	0.02
8 <= PromPasivosUlt3_d <= 9	8	2'816	24.05	221	7.85	2'595	221	0.25	0.17	38.13	0.03
PromPasivosUlt3_d2		11'707		1'298		10'409	1'298				0.08

Participación de PromPasivosUlt3_d2 por Default - Segmento 2

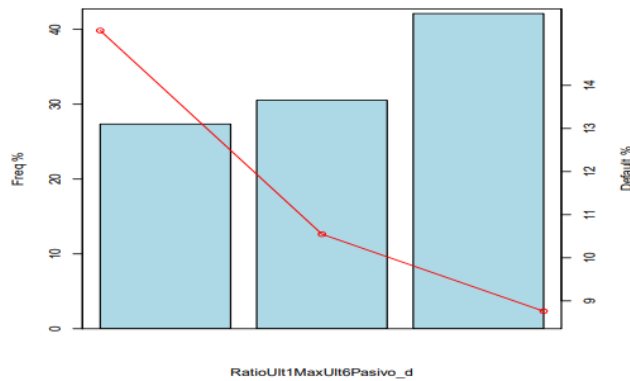


d) RatioUlt1MaxUlt6Pasivo

Análisis de la Variable RatioUlt1MaxUlt6Pasivo_d vs Default - Segmento 2

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
0.03>= RatioUlt1MaxUlt6Pasivo or RatioUlt1MaxUlt6Pasivo is null	1	3'202	27.35	489	15.27	2'713	489	0.26	0.38	-36.84	0.04
RatioUlt1MaxUlt6Pasivo >0.03 and 0.5>= RatioUlt1MaxUlt6Pasivo	2	3'574	30.53	377	10.55	3'197	377	0.31	0.29	5.59	0.00
RatioUlt1MaxUlt6Pasivo >0.5	3	4'931	42.12	432	8.76	4'499	432	0.43	0.33	26.13	0.03
RatioUlt1MaxUlt6Pasivo_d		11'707		1'298		10'409	1'298				0.07

Participación de RatioUlt1MaxUlt6Pasivo_d por Default - Segmento 2

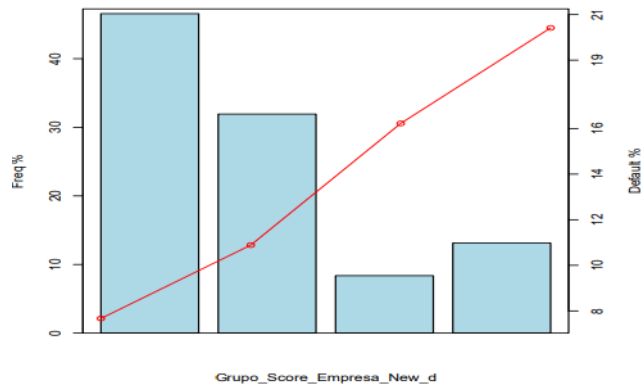


e) Grupo_Score_Empresa_New

Análisis de la Variable Grupo_Score_Empresa_New_d vs Default - Segmento 2

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
4>= Grupo_Score_Empresa_New or Grupo_Score_Empresa_New is null	1	5'451	46.56	418	7.67	5'033	418	0.48	0.32	40.64	0.07
Grupo_Score_Empresa_New >4 and 7>= Grupo_Score_Empresa_New	2	3'738	31.93	407	10.89	3'331	407	0.32	0.31	2.04	0.00
Grupo_Score_Empresa_New >7 and 8>= Grupo_Score_Empresa_New	3	980	8.37	159	16.22	821	159	0.08	0.12	-44.02	0.02
Grupo_Score_Empresa_New >8	4	1'538	13.14	314	20.42	1'224	314	0.12	0.24	-72.14	0.09
Grupo_Score_Empresa_New_d		11'707		1'298		10'409	1'298				0.17

Participación de Grupo_Score_Empresa_New_d por Default - Segmento 2



Anexo 3:

Segmento 3, según Patrón Interno / Patrón Externo / Nivel de Ventas

Análisis de Estabilidad y Correlación de Variables - Segmento 3

Luego para analizar la consistencia de los coeficientes se procedió a aplicar la técnica del Bootstrap, se tomó 1000 muestras sin reemplazo de la muestra de validación con la intención de ver la distribución de cada coeficiente, en los casos donde se encontró que más del 5% de la distribución presentaba un signo cambiado se procedió a unir las categorías de la variable, en el caso que todas las categorías de la variable presentan más del 5% de su distribución con un signo contrario se decidió retirar a la variable.

A continuación, se muestra la distribución de cada coeficiente del modelo final:

Cuantiles de los Estimadores Generados por el Método Bootstrap

Variables	1%	5%	10%	25%	50%	75%	90%	95%	99%
(Intercept)	-0.62388	-0.54058	-0.51392	-0.45763	-0.38862	-0.31508	-0.26559	-0.23275	-0.1671
IncPromPasivoUlt6_d24	-0.81248	-0.7403	-0.70255	-0.62696	-0.54616	-0.46235	-0.39682	-0.35356	-0.28517
AtrasoPromPondIntUlt6m_d26	0.21023	0.27641	0.30443	0.35748	0.41934	0.47206	0.52325	0.55127	0.60963
AtrasoPromPondIntUlt6m_d29	0.23106	0.30362	0.35389	0.43356	0.51331	0.59395	0.67588	0.71546	0.77407
AtrasoPromPondIntUlt6m_d210	0.3969	0.46507	0.51034	0.58302	0.66249	0.74362	0.8116	0.84998	0.91634
AtrasoPromPondIntUlt6m_d211	0.61027	0.68181	0.71782	0.78307	0.86311	0.93856	1.00747	1.05033	1.11519
AtrasoPromPondIntUlt6m_d212	0.59075	0.70135	0.746	0.81788	0.89975	0.98922	1.05769	1.11208	1.18472
AtrasoPromPondIntUlt6m_d213	0.96345	1.03054	1.06226	1.124	1.19502	1.25007	1.31576	1.35223	1.41678
NumNorUlt6m_d24	-0.85789	-0.81939	-0.79713	-0.75884	-0.71115	-0.66184	-0.62255	-0.59733	-0.56142
Antigüedad_Empresa_d26	-0.43177	-0.39706	-0.37344	-0.33481	-0.29013	-0.24626	-0.20752	-0.18149	-0.12711
Antigüedad_Empresa_d29	-0.70337	-0.63225	-0.58362	-0.51698	-0.44511	-0.3712	-0.31083	-0.26524	-0.18104
pasivos_1_d22	-0.49202	-0.45865	-0.43827	-0.39769	-0.35727	-0.31319	-0.27431	-0.2515	-0.20517
pasivos_1_d24	-1.14904	-1.01384	-0.94771	-0.85088	-0.73617	-0.62899	-0.53479	-0.47938	-0.39654
VENTAS_SOLES_d22	-0.43084	-0.38466	-0.35383	-0.30385	-0.2534	-0.19698	-0.15046	-0.11459	-0.08066
VENTAS_SOLES_d26	-0.66614	-0.59319	-0.54313	-0.45489	-0.36324	-0.27662	-0.20556	-0.16232	-0.07159
Grupo_Score_Empresa_New_d23	0.00121	0.06764	0.1033	0.16993	0.24323	0.30849	0.36659	0.40915	0.46646
Grupo_Score_Empresa_New_d24	0.40193	0.46963	0.49598	0.55267	0.61137	0.66811	0.71719	0.74926	0.79936
Grupo_Score_Empresa_New_d25	0.62823	0.68749	0.71606	0.76819	0.82058	0.87938	0.9273	0.95601	1.0035

Para las variables finales dentro del modelo se presenta a continuación la matriz de correlación, en donde, se han marcado los valores que exceden el límite máximo establecido (70% de correlación).

Matriz de correlación de variables - Segmento 3

Variables	IncPromPasivoUlt6_d2	AtrasoPromPondIntUlt6m_d2	NumNorUlt6m_d2	Antiguedad_Empresa_d2	pasivos_1_d2	VENTAS_SOLES_d2	Grupo_Score_Empresa_New_d2
IncPromPasivoUlt6_d2	1.00	-0.05	0.03	0.01	0.32	0.14	-0.08
AtrasoPromPondIntUlt6m_d2	-0.05	1.00	-0.15	0.05	-0.09	0.10	0.12
NumNorUlt6m_d2	0.03	-0.15	1.00	0.06	0.08	0.11	-0.11
Antiguedad_Empresa_d2	0.01	0.05	0.06	1.00	0.09	0.29	-0.07
pasivos_1_d2	0.32	-0.09	0.08	0.09	1.00	0.16	-0.11
VENTAS_SOLES_d2	0.14	0.10	0.11	0.29	0.16	1.00	-0.03
Grupo_Score_Empresa_New_d2	-0.08	0.12	-0.11	-0.07	-0.11	-0.03	1.00

Análisis de Variables por Default - Segmento 3

A cada una de las variables se le realizó un análisis contra la variable objetivo (variable default), para verificar su poder de discriminancia individual. Las variables con un sufijo “_d” al final indican que han sido transformadas, mientras que las variables con el sufijo “_d2” indican que han sido reajustadas.

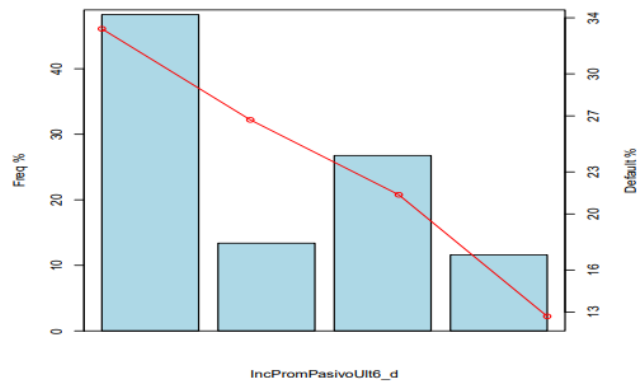
A continuación, se muestra el análisis bivariado por cada variable final del modelo.

a) IncPromPasivoUlt6

Análisis de la Variable IncPromPasivoUlt6_d vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
3.82e+03>= IncPromPasivoUlt6 or IncPromPasivoUlt6 is null	1	11'239	48.25	3'733	33.21	7'506	3'733	0.44	0.60	-30.66	0.05
IncPromPasivoUlt6 >3.82e+03 and 1.08e+04>= IncPromPasivoUlt6	2	3'118	13.39	833	26.72	2'285	833	0.13	0.13	0.40	0.00
IncPromPasivoUlt6 >1.08e+04 and 7.88e+04>= IncPromPasivoUlt6	3	6'233	26.76	1'332	21.37	4'901	1'332	0.29	0.21	29.77	0.02
IncPromPasivoUlt6 >7.88e+04	4	2'702	11.60	343	12.69	2'359	343	0.14	0.05	92.32	0.08
IncPromPasivoUlt6_d		23'292		6'241		17'051	6'241				0.15

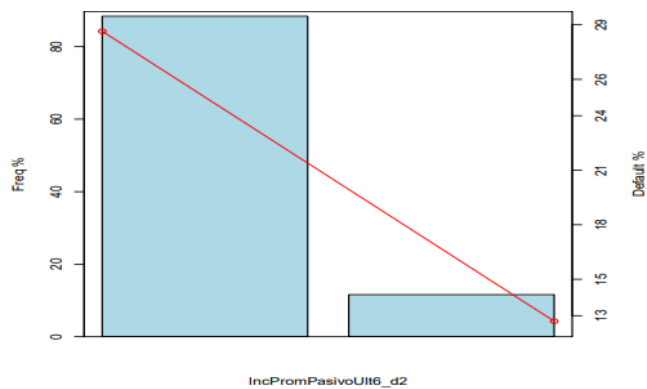
Participación de IncPromPasivoUlt6_d por Default - Segmento 3



Análisis de la Variable IncPromPasivoUlt6_d2 vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= IncPromPasivoUlt6_d <=3	1	20'590	88.40	5'898	28.64	14'692	5'898	0.86	0.95	-9.24	0.01
4<= IncPromPasivoUlt6_d <=4	4	2'702	11.60	343	12.69	2'359	343	0.14	0.05	92.32	0.08
IncPromPasivoUlt6_d2		23'292		6'241		17'051	6'241				0.08

Participación de IncPromPasivoUlt6_d2 por Default - Segmento 3

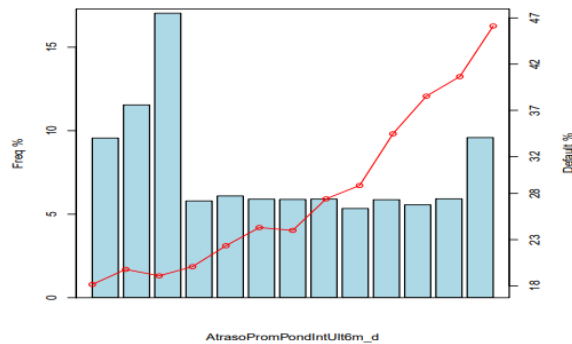


b) AtrasoPromPondIntUlt6m

Análisis de la Variable AtrasoPromPondIntUlt6m_d vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1.1>= AtrasoPromPondIntUlt6m or AtrasoPromPondIntUlt6m is null	1	2'227	9.56	404	18.14	1'823	404	0.11	0.06	50.18	0.02
AtrasoPromPondIntUlt6m >1.1 and 3.84>= AtrasoPromPondIntUlt6m	2	2'689	11.54	532	19.78	2'157	532	0.13	0.09	39.48	0.02
AtrasoPromPondIntUlt6m >3.84 and 7.19>= AtrasoPromPondIntUlt6m	3	3'968	17.04	757	19.08	3'211	757	0.19	0.12	43.99	0.03
AtrasoPromPondIntUlt6m >7.19 and 8.19>= AtrasoPromPondIntUlt6m	4	1'350	5.80	271	20.07	1'079	271	0.06	0.04	37.66	0.01
AtrasoPromPondIntUlt6m >8.19 and 9.33>= AtrasoPromPondIntUlt6m	5	1'419	6.09	317	22.34	1'102	317	0.06	0.05	24.09	0.00
AtrasoPromPondIntUlt6m >9.33 and 10.6>= AtrasoPromPondIntUlt6m	6	1'374	5.90	334	24.31	1'040	334	0.06	0.05	13.08	0.00
AtrasoPromPondIntUlt6m >10.6 and 12.1>= AtrasoPromPondIntUlt6m	7	1'371	5.89	329	24.00	1'042	329	0.06	0.05	14.78	0.00
AtrasoPromPondIntUlt6m >12.1 and 14.1>= AtrasoPromPondIntUlt6m	8	1'375	5.90	377	27.42	998	377	0.06	0.06	-3.16	0.00
AtrasoPromPondIntUlt6m >14.1 and 16.6>= AtrasoPromPondIntUlt6m	9	1'244	5.34	359	28.86	885	359	0.05	0.06	-10.28	0.00
AtrasoPromPondIntUlt6m >16.6 and 20>= AtrasoPromPondIntUlt6m	10	1'367	5.87	471	34.46	896	471	0.05	0.08	-36.20	0.01
AtrasoPromPondIntUlt6m >20 and 24.6>= AtrasoPromPondIntUlt6m	11	1'295	5.56	499	38.53	796	499	0.05	0.08	-53.81	0.02
AtrasoPromPondIntUlt6m >24.6 and 31.6>= AtrasoPromPondIntUlt6m	12	1'380	5.92	561	40.65	819	561	0.05	0.09	-62.67	0.03
AtrasoPromPondIntUlt6m >31.6	13	2'233	9.59	1'030	46.13	1'203	1'030	0.07	0.17	-84.98	0.08
AtrasoPromPondIntUlt6m_d		23'292		6'241		17'051	6'241				0.21

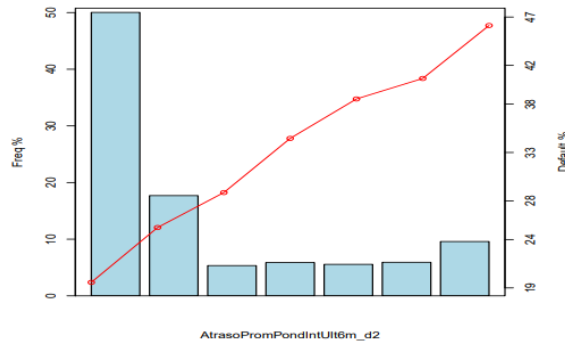
Participación de AtrasoPromPondIntUlt6m_d por Default - Segmento 3



Análisis de la Variable AtrasoPromPondIntUlt6m_d2 vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= AtrasoPromPondIntUlt6m_d <=5	1	11'653	50.03	2'281	19.57	9'372	2'281	0.55	0.37	40.80	0.08
6<= AtrasoPromPondIntUlt6m_d <=8	6	4'120	17.69	1'040	25.24	3'080	1'040	0.18	0.17	8.06	0.00
9<= AtrasoPromPondIntUlt6m_d <=9	9	1'244	5.34	359	28.86	885	359	0.05	0.06	-10.28	0.00
10<= AtrasoPromPondIntUlt6m_d <=10	10	1'367	5.87	471	34.46	896	471	0.05	0.08	-36.20	0.01
11<= AtrasoPromPondIntUlt6m_d <=11	11	1'295	5.56	499	38.53	796	499	0.05	0.08	-53.81	0.02
12<= AtrasoPromPondIntUlt6m_d <=12	12	1'380	5.92	561	40.65	819	561	0.05	0.09	-62.67	0.03
13<= AtrasoPromPondIntUlt6m_d <=13	13	2'233	9.59	1'030	46.13	1'203	1'030	0.07	0.17	-84.98	0.08
AtrasoPromPondIntUlt6m_d2		23'292		6'241		17'051	6'241				0.21

Participación de AtrasoPromPondIntUlt6m_d2 por Default - Segmento 3

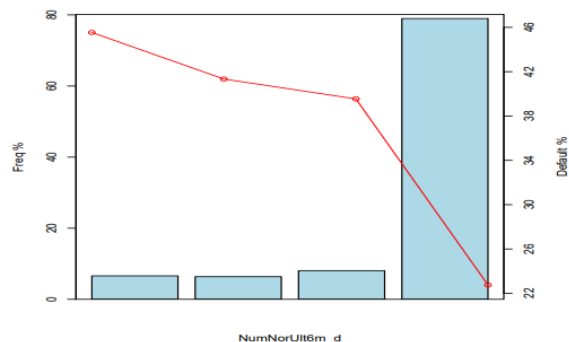


c) NumNorUlt6m

Análisis de la Variable NumNorUlt6m_d vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
2>= NumNorUlt6m	1	1'535	6.59	699	45.54	836	699	0.05	0.11	-82.61	0.05
NumNorUlt6m >2 and 4>= NumNorUlt6m	2	1'483	6.37	613	41.34	870	613	0.05	0.10	-65.49	0.03
NumNorUlt6m >4 and 5>= NumNorUlt6m	3	1'879	8.07	743	39.54	1'136	743	0.07	0.12	-58.05	0.03
NumNorUlt6m >5	4	18'395	78.98	4'186	22.76	14'209	4'186	0.83	0.67	21.71	0.04
NumNorUlt6m_d		23'292		6'241		17'051	6'241				0.15

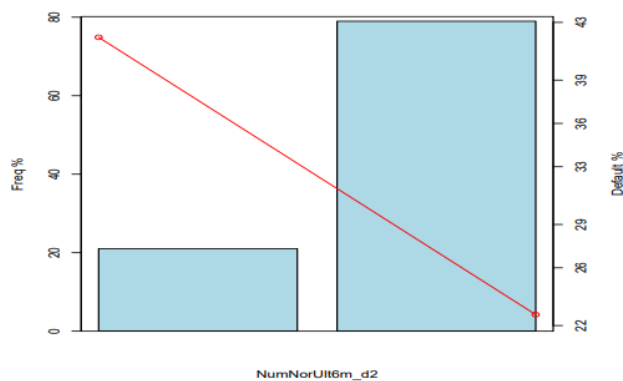
Participación de NumNorUlt6m_d por Default - Segmento 3



Análisis de la Variable NumNorUlt6m_d2 vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= NumNorUlt6m_d <=3	1	4'897	21.02	2'055	41.96	2'842	2'055	0.17	0.33	-68.08	0.11
4<= NumNorUlt6m_d <=4	4	18'395	78.98	4'186	22.76	14'209	4'186	0.83	0.67	21.71	0.04
NumNorUlt6m_d2		23'292		6'241		17'051	6'241				0.15

Participación de NumNorUlt6m_d2 por Default - Segmento 3

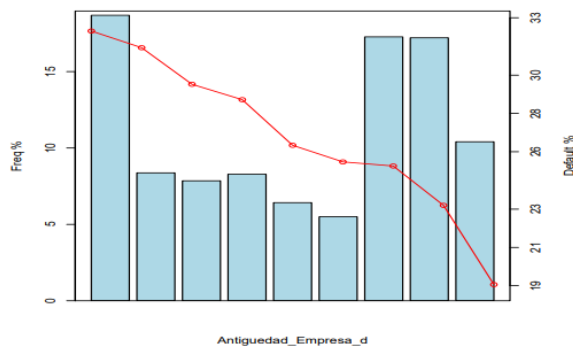


d) Antigüedad_Empresa

Análisis de la Variable Antigüedad_Empresa_d vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
4>= Antigüedad_Empresa	1	4'352	18.68	1'406	32.31	2'946	1'406	0.17	0.23	-26.54	0.01
Antigüedad_Empresa >4 and 5>= Antigüedad_Empresa	2	1'950	8.37	613	31.44	1'337	613	0.08	0.10	-22.52	0.00
Antigüedad_Empresa >5 and 6>= Antigüedad_Empresa	3	1'829	7.85	540	29.52	1'289	540	0.08	0.09	-13.50	0.00
Antigüedad_Empresa >6 and 7>= Antigüedad_Empresa	4	1'929	8.28	554	28.72	1'375	554	0.08	0.09	-9.60	0.00
Antigüedad_Empresa >7 and 8>= Antigüedad_Empresa	5	1'496	6.42	394	26.34	1'102	394	0.06	0.06	2.35	0.00
Antigüedad_Empresa >8 and 9>= Antigüedad_Empresa	6	1'280	5.50	326	25.47	954	326	0.06	0.05	6.87	0.00
Antigüedad_Empresa >9 and 13>= Antigüedad_Empresa	7	4'023	17.27	1'016	25.25	3'007	1'016	0.18	0.16	8.00	0.00
Antigüedad_Empresa >13 and 21>= Antigüedad_Empresa	8	4'009	17.21	930	23.20	3'079	930	0.18	0.15	19.21	0.01
Antigüedad_Empresa >21 or Antigüedad_Empresa is null	9	2'424	10.41	462	19.06	1'962	462	0.12	0.07	44.11	0.02
Antigüedad_Empresa_d		23'292		6'241		17'051	6'241				0.05

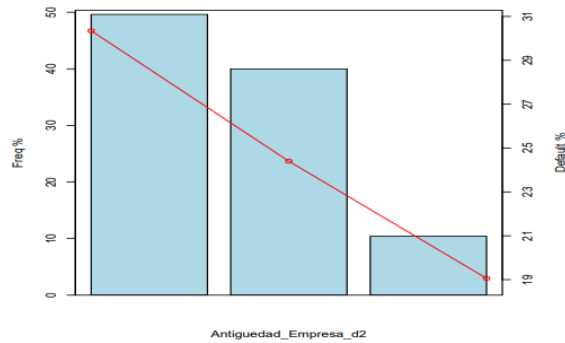
Participación de Antigüedad_Empresa_d por Default - Segmento 3



Análisis de la Variable Antigüedad_Empresa_d2 vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= Antigüedad_Empresa_d <=5	1	11'556	49.61	3'507	30.35	8'049	3'507	0.47	0.56	-17.43	0.02
6<= Antigüedad_Empresa_d <=8	6	9'312	39.98	2'272	24.40	7'040	2'272	0.41	0.36	12.59	0.01
9<= Antigüedad_Empresa_d <=9	9	2'424	10.41	462	19.06	1'962	462	0.12	0.07	44.11	0.02
Antigüedad_Empresa_d2		23'292		6'241		17'051	6'241				0.04

Participación de Antigüedad_Empresa_d2 por Default - Segmento 3

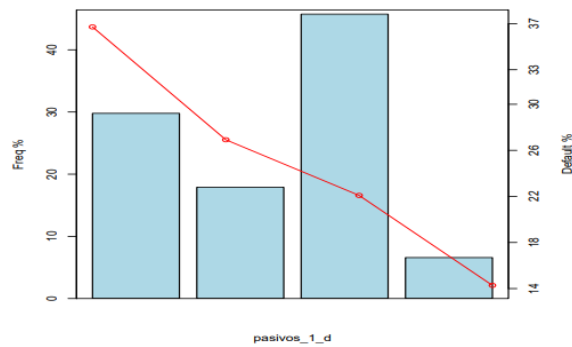


e) Pasivos_1

Análisis de la Variable Pasivos_1_d vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
92.1>= pasivos_1 or pasivos_1 is null	1	6'937	29.78	2'547	36.72	4'390	2'547	0.26	0.41	-46.07	0.07
pasivos_1 >92.1 and 1.31e+03>= pasivos_1	2	4'168	17.89	1'122	26.92	3'046	1'122	0.18	0.18	-0.64	0.00
pasivos_1 >1.31e+03 and 2.64e+05>= pasivos_1	3	10'653	45.74	2'353	22.09	8'300	2'353	0.49	0.38	25.55	0.03
pasivos_1 >2.64e+05	4	1'534	6.59	219	14.28	1'315	219	0.08	0.04	78.75	0.03
pasivos_1_d		23'292		6'241		17'051	6'241				0.13

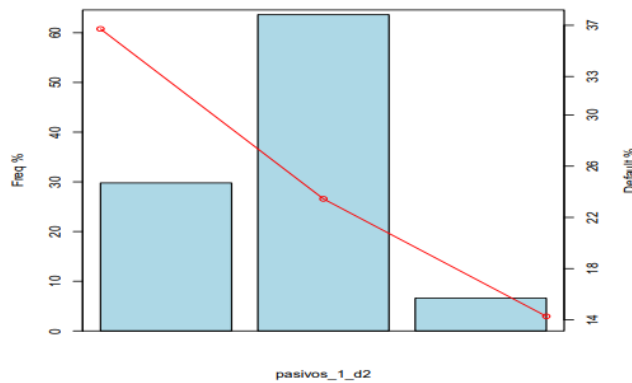
Participación de Pasivos_1_d por Default - Segmento 3



Análisis de la Variable Pasivos_1_d2 vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= pasivos_1_d <=1	1	6'937	29.78	2'547	36.72	4'390	2'547	0.26	0.41	-46.07	0.07
2<= pasivos_1_d <=3	2	14'821	63.63	3'475	23.45	11'346	3'475	0.67	0.56	17.82	0.02
4<= pasivos_1_d <=4	4	1'534	6.59	219	14.28	1'315	219	0.08	0.04	78.75	0.03
pasivos_1_d2		23'292		6'241		17'051	6'241				0.12

Participación de Pasivos_1_d2 por Default - Segmento 3

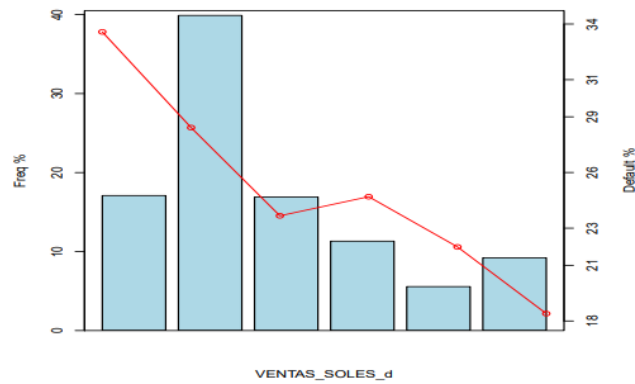


f) Ventas_Soles

Análisis de la Variable Ventas_Soles_d vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
652>= VENTAS_SOLES or VENTAS_SOLES is null	1	3'982	17.10	1'337	33.58	2'645	1'337	0.16	0.21	-32.28	0.02
VENTAS_SOLES >652 and 3.15e+03>= VENTAS_SOLES	2	9'293	39.90	2'641	28.42	6'652	2'641	0.39	0.42	-8.13	0.00
VENTAS_SOLES >3.15e+03 and 5.69e+03>= VENTAS_SOLES	3	3'938	16.91	932	23.67	3'006	932	0.18	0.15	16.60	0.00
VENTAS_SOLES >5.69e+03 and 1.03e+04>= VENTAS_SOLES	4	2'636	11.32	651	24.70	1'985	651	0.12	0.10	10.98	0.00
VENTAS_SOLES >1.03e+04 and 1.46e+04>= VENTAS_SOLES	5	1'296	5.56	285	21.99	1'011	285	0.06	0.05	26.11	0.00
VENTAS_SOLES >1.46e+04	6	2'147	9.22	395	18.40	1'752	395	0.10	0.06	48.46	0.02
VENTAS_SOLES_d		23'292		6'241		17'051	6'241				0.05

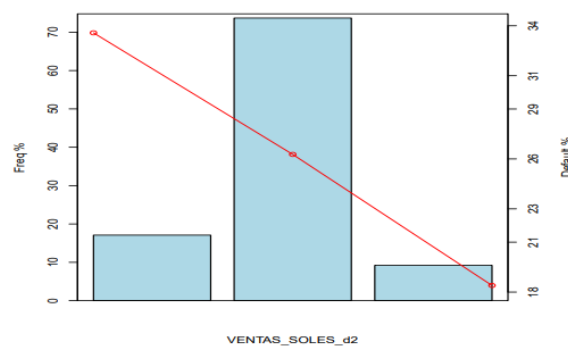
Participación de Ventas_Soles_d por Default - Segmento 3



Análisis de la Variable Ventas_Soles_d2 vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= VENTAS_SOLES_d <=1	1	3'982	17.10	1'337	33.58	2'645	1'337	0.16	0.21	-32.28	0.02
2<= VENTAS_SOLES_d <=5	2	17'163	73.69	4'509	26.27	12'654	4'509	0.74	0.72	2.68	0.00
6<= VENTAS_SOLES_d <=6	6	2'147	9.22	395	18.40	1'752	395	0.10	0.06	48.46	0.02
VENTAS_SOLES_d2		23'292		6'241		17'051	6'241				0.04

Participación de Ventas_Soles_d2 por Default - Segmento 3

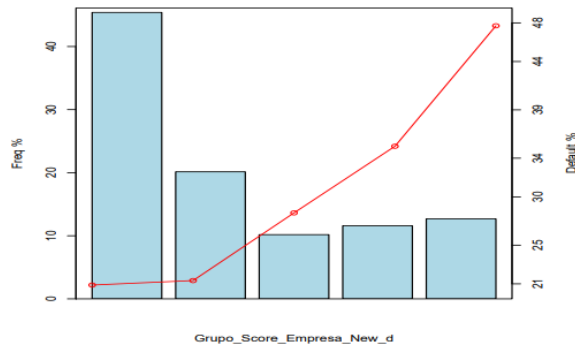


g) Grupo_Score_Empresa_New

Análisis de la Variable Grupo_Score_Empresa_New_d vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
5>= Grupo_Score_Empresa_New or Grupo_Score_Empresa_New is null	1	10'577	45.41	2'208	20.88	8'369	2'208	0.49	0.35	32.74	0.04
Grupo_Score_Empresa_New >5 and 7>= Grupo_Score_Empresa_New	2	4'691	20.14	1'001	21.34	3'690	1'001	0.22	0.16	29.96	0.02
Grupo_Score_Empresa_New >7 and 8>= Grupo_Score_Empresa_New	3	2'371	10.18	672	28.34	1'699	672	0.10	0.11	-7.75	0.00
Grupo_Score_Empresa_New >8 and 9>= Grupo_Score_Empresa_New	4	2'699	11.59	951	35.24	1'748	951	0.10	0.15	-39.64	0.02
Grupo_Score_Empresa_New >9	5	2'954	12.68	1'409	47.70	1'545	1'409	0.09	0.23	-91.29	0.12
Grupo_Score_Empresa_New_d		23'292		6'241		17'051	6'241				0.21

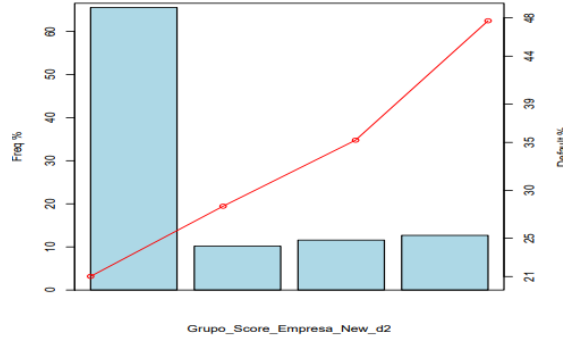
Participación de Grupo_Score_Empresa_New_d por Default - Segmento 3



Análisis de la Variable Grupo_Score_Empresa_New_d2 vs Default - Segmento 3

	Nivel	Casos	Part. %	Default	Default %	Buenos	Malos	Dist. Buenos	Dist. Malos	WOE	IV
1<= Grupo_Score_Empresa_New_d <=2	1	15'268	65.55	3'209	21.02	12'059	3'209	0.71	0.51	31.88	0.06
3<= Grupo_Score_Empresa_New_d <=3	3	2'371	10.18	672	28.34	1'699	672	0.10	0.11	-7.75	0.00
4<= Grupo_Score_Empresa_New_d <=4	4	2'699	11.59	951	35.24	1'748	951	0.10	0.15	-39.64	0.02
5<= Grupo_Score_Empresa_New_d <=5	5	2'954	12.68	1'409	47.70	1'545	1'409	0.09	0.23	-91.29	0.12
Grupo_Score_Empresa_New_d2		23'292		6'241		17'051	6'241				0.21

Participación de Grupo_Score_Empresa_New_d2 por Default - Segmento 3



Lista de Variables

Definición de las Variables Finales

Variable	Descripción de variables
Ventas_Soles	Información que proviene de los estados financieros disponibles, tomando en consideración el monto al cierre de cada ejercicio.
Patron_Interno	Patrón que mezcla los patrones INT_1 hasta INT_P8_3. Representan el comportamiento del cliente en el Banco, en función a su comportamiento de pago de los últimos 12 meses.
INT_P1	Marca que indica que el cliente no ha tenido atraso en su pago en los últimos 12 meses
INT_P2	Número de meses sin atraso del patrón que cumple: Hace 12 meses sin deuda y último mes sin atraso
INT_P3	Número de meses sin atraso del patrón que cumple: Hace 12 meses sin atraso y último mes sin deuda.
INT_P4	Número de meses sin atraso del patrón que cumple: Hace 12 meses sin atraso, último mes sin atraso y valores intermitentes de sin atraso y sin deuda.
INT_P5	Número de meses sin atraso del patrón que cumple: Hace 12 meses sin deuda, último mes sin deuda y valores intermitentes de sin atraso y sin deuda.
INT_P6	Número de meses sin atraso del patrón que cumple: Hace 12 meses sin atraso, último mes sin deuda y valores intermitentes de sin atraso y sin deuda
INT_P7	Número de meses sin atraso del patrón que cumple: Hace 12m al día, último mes sin atraso y valores intermitentes de sin atraso y sin deuda.
INT_P8_1	Número de meses sin atraso del patrón que cumple: el cliente tiene al menos un mes con atraso en el pago en los últimos 12 meses
INT_P8_2	El atraso máximo en los últimos 12 meses
INT_P8_3	Cuantos meses han pasado desde su último mes con

Variable	Descripción de variables
	atraso en el pago
Patron_Externo	Patrón que mezcla los patrones P1 hasta P8_3
P1	Patrón full normal en los últimos 48 meses
P2	Cuenta el número de normales del patrón que cumple: Hace 48 meses no tiene clasificación (0) y último mes tiene clasificación Normal (1).
P3	Cuenta el número de normales del patrón que cumple: Hace 48m tiene clasificación Normal (1) y en el último mes no tiene clasificación.
P4	Cuenta el número de normales del patrón que cumple: Hace 48 meses tiene clasificación Normal (1), el último mes tiene clasificación Normal (1) y tiene valores intermitentes de sin clasificación (0) y Normal (1)
P5	Cuenta el número de normales del patrón que cumple: Hace 48 meses no tiene clasificación (0), el último mes sin clasificación (0) y valores intermitentes de sin clasificación (0) y Normal (1).
P6	Cuenta el número de normales del patrón que cumple: Hace 48 meses tiene clasificación Normal, el último mes no tiene clasificación (0) y valores intermitentes de sin clasificación (0) y Normal (1)
P7	Cuenta el número de normales del patrón que cumple: Hace 48 meses no tiene clasificación, el último mes tiene clasificación Normal y valores intermitentes de sin clasificación (0) y Normal (1)
P8_1	Cuenta el número de normales del patrón que cumple: El cliente tiene al menos un calificativo peor que normal en los últimos 48 meses
P8_2	La máxima clasificación en los últimos 48 meses
P8_3	Cuantos meses han pasado desde su última clasificación peor que Normal
Antigüedad_Empresa	Antigüedad de la empresa, tomando en consideración la fecha de constitución.
AtrasoPromPondIntUlt6m	Promedio ponderado ponderando según la antigüedad de los días de atraso interno, en los últimos 6 meses.
AtrasoPromPondUlt24m	Promedio ponderado ponderando según la antigüedad de los días de mora, en créditos directos, en los últimos 24 meses.
AtrasoPromPondUlt6m	Promedio ponderado ponderando según la antigüedad de los días de mora, en créditos directos, en los últimos 6 meses.
EntRepPromUlt24m	Número de entidades reportantes en promedio en los últimos 24 meses
IncPromPasivoUlt6	Incremento Promedio de Pasivos en los últimos 6 meses
Grupo_Score_Empresa_New	Promedio ponderado entre la participación de acciones de los accionistas de la empresa por el

Variable	Descripción de variables
	score de buro PN vigente.
NumNorUlt6m	Número de normales en los últimos 6 meses.
Pasivos_1	Saldo pasivo en el último mes.
PromPasivosUlt3	Promedio del saldo pasivo de los últimos 3 meses.
RatioUlt1MaxUlt6Pasivo	División entre el saldo pasivo en el mes actual y el máximo saldo pasivo en los últimos 6 meses.
IncPromPasivoUlt6	Incremento Promedio de Pasivos en los últimos 6 meses.