

UNIVERSIDAD ESAN



**Modelos de clasificación de clientes bancarizados de los niveles
socioeconómicos “C” y “D”**

**Tesis presentada para cumplir con uno de los requisitos para la obtención
del grado académico de Maestro en Finanzas por:**

Anthony Kevin Tineo Nieves _____

Programa de la Maestría en Finanzas

Lima, 12 de marzo de 2019

Esta tesis

**Modelos de clasificación de clientes bancarizados de los niveles
socioeconómicos “C” y “D”**

ha sido aprobada.

.....
Alfredo Mendiola Cabrera (Jurado)

.....
Javier Rojas Chang (Jurado)

.....
Luis Chavez-Bedoya Mercado (Asesor)

Universidad Esan

2019

A Dios por ser mi guía espiritual y a mi familia, por su infinito apoyo.

Anthony Kevin Tineo Nieves

ÍNDICE GENERAL

CAPITULO I. PLANTEAMIENTO DEL PROBLEMA	1
1.1. Análisis del Sector Microfinanciero	3
1.1.1. Análisis de los Créditos de Consumo y Microempresa en el Perú	4
1.1.2. Nivel de morosidad de los Créditos Consumo y Microempresa	6
1.2. Problemática	8
1.3. Objetivos	10
1.3.1. Objetivo principal	10
1.3.2. Objetivos específicos	10
1.4. Alcance y Limitaciones	10
1.4.1. Alcance	10
1.4.2. Limitaciones	11
CAPITULO II. TRATAMIENTO DE LA INFORMACIÓN	12
2.1. Fuentes de información	12
2.1.1. Reporte Crediticio Consolidado (RCC)	12
2.1.2. Información socio-demográfica	13
2.2. Descripción del universo de observaciones	13
2.3. Ventanas de tiempo	14
2.4. Exclusiones en la Base de Datos	15
2.5. Definición del default	16
2.6. Variables independientes	17
2.7. Definición de muestras de trabajo	20
CAPITULO III. MARCO TEÓRICO	21
3.1. Riesgo de crédito	21
3.2. Indicadores de discriminancia y ajuste	22
3.2.1. Coeficiente de Gini	22
3.2.2. Estadístico Kolmogorov-Smirnov	23
3.2.3. Curva ROC	24
3.3. Modelos de clasificación	25
3.3.1. Modelo logit	25
3.3.2. Árbol de decisión	27
3.3.3. Redes Neuronales	28
CAPITULO IV. METODOLOGÍA DE SELECCIÓN DE VARIABLES	30
4.1. Análisis de concentración (Frecuencia)	31
4.2. Indicadores <i>Information value (IV)</i> y <i>Weight of evidence (WOE)</i>	31
4.3. Análisis de Correlación Bivariada	33
CAPITULO V. DESARROLLO DE MODELOS DE CLASIFICACIÓN	35
5.1. Modelo de Regresión logística	35
5.1.1. Interpretación de las variables independientes	35
5.1.2. Análisis de Significancia Individual	40
5.1.3. Análisis de Ajuste y Discriminancia	40
5.2. Modelo de Árbol de Decisión	42
5.2.1. Interpretación de las variables independientes	43
5.2.2. Análisis de Significancia Individual	46
5.2.3. Análisis de Ajuste y Discriminancia	47
5.3. Modelo de Redes Neuronales	48
5.3.1. Parámetros e interpretación del modelo	48
5.3.2. Análisis de importancia de las variables independientes	49

5.3.3. <i>Análisis de Ajuste y Discriminancia</i>	50
5.4. Comparación de los modelos desarrollados	51
CAPITULO VI.CONCLUSIONES	53
BIBLIOGRAFÍA	57

LISTA DE GRÁFICOS

Gráfico I.1: Colocaciones al 31 de diciembre 2017.....	5
Gráfico I.2: Evolución de Créditos Microempresa y Consumo.....	6
Gráfico II.1: Ventanas para la creación de variables.....	15
Gráfico III.1: Acumulación de grupos.....	22
Gráfico III.2: Acumulación de distribución de solicitantes buenos y malos.....	24
Gráfico III.3: Curva ROC.....	25
Gráfico III.4: Árbol de decisión.....	27
Gráfico III.5: Redes Neuronales Artificiales.....	28
Gráfico III.6: Arquitectura de una RNA.....	29
Gráfico V.1: Análisis del Árbol de Decisión.....	44

LISTA DE TABLAS

Tabla I.1: Principales diferencias entre Financieras, CMACs, CRACs y EDPYMES ..	4
Tabla I.2: Nivel de morosidad (%) según tipo crédito e institución financiera	7
Tabla II.1: Exclusiones en la Base de Datos	15
Tabla II.2: Variables independientes	17
Tabla IV.1: Criterios del índice IV	32
Tabla IV.2: Rango de correlaciones.....	33
Tabla IV.3: Variables correlacionadas.....	33
Tabla IV.4: Resultado de la Metodología de la Selección de Variables.....	34
Tabla V.1 - Resultados de la Regresión Logística	35
Tabla V.2: Criterios KS	41
Tabla V.3: Resumen Estadístico del Modelo Logit	41
Tabla V.4: Resumen del Modelo de Árbol de Decisión	43
Tabla V.5: Resumen Estadístico del Modelo de Árbol de Decisión.....	47
Tabla V.6: Análisis de sensibilidad de las variables dependientes	50
Tabla V.7: Resumen Estadístico del Modelo de Redes Neuronales	50
Tabla V.8: Variables independientes de cada modelo	52
Tabla V.9: Indicadores de nivel de ajuste y discriminancia de cada modelo	52

ANEXOS

Anexo I: Clasificación regulatoria de créditos Resolución SBS N° 11356 - 2010.....	59
Anexo II: Definiciones relevantes.....	60
Anexo III: Variables con una concentración mayor al 90% en valores perdidos	61
Anexo IV: Variables con una concentración mayor al 90% en algún valor o clase	62
Anexo V: Índice <i>Information Value</i>	64
Anexo VI: <i>Weight of Evidence</i> (WOE).....	65
Anexo VII: Principio de Monotonicidad	67
Anexo VIII: Transformación de variables	68
Anexo IX: Lista de variables después del filtro IV y WOE.....	72
Anexo X: Correlación de variables	73
Anexo XI: Resultados de la correlación de variables	75
Anexo XII: Modelo de Árbol de Decisión.....	77
Anexo XIII: Modelo de Redes Neuronales.....	85
Anexo XIV: Pesos Sinápticos del Modelo de Redes Neuronales.....	86

GLOSARIO DE TÉRMINOS

AIC	Criterio de Información de Akaike
APEIM	Asociación Peruana de Empresas de Investigación de Mercados
AUCROC	Área bajo la curva de ROC
CHAID	Detector de interacción automático Chi-cuadrado
CMAC	Caja municipal de ahorro y crédito
CPP	Clientes con problemas potenciales
CRAC	Caja rural de ahorro y crédito
EDPYME	Entidad de desarrollo de la micro y pequeña empresa
ENAHO	Encuesta nacional de hogares
IMF	Institución microfinanciera
IV	Indicador information value (valor de información)
KS	Estadístico Kolmogorov-Smirnov
MES	Créditos microempresa
MLP	Perceptrón multicapa
NSE	Nivel socioeconómico
PP	Préstamos personales
RCC	Reporte crediticio consolidado
RNA	Redes neuronales artificiales
SBS	Superintendencia de Banca, Seguros y AFP
TC	Préstamos de tarjetas de crédito
WOE	Indicador weight of evidence (peso de la evidencia)

Expresar mi más sincero agradecimiento a mi asesor Luis Chávez-Bedoya (PhD.) por su gran apoyo y dedicación en el desarrollo de la presente tesis.

Asimismo, expresar mi agradecimiento a todos los docentes del Programa de la Maestría en Finanzas por sus enseñanzas.

Finalmente mi agradecimiento a la UNIVERSIDAD ESAN, por ser mi casa de estudios.

ANTHONY KEVIN TINEO NIEVES

Magíster de Finanzas de ESAN, Economista de la Universidad de Piura. Experiencia en análisis de información, construcción de indicadores de desempeño, finanzas corporativas, proyección de estados financieros, control y gestión. Adicionalmente, experiencia en análisis macroeconómico y del sector real. Nivel de inglés avanzado.

FORMACIÓN

2016-2019 **Universidad ESAN**

Magíster de Finanzas.

2010-2015 **Universidad de Piura**

Licenciado en Economía.

EXPERIENCIA

Mar 2019 **Protecta Security**

Actualmente Analista de Inversiones Inmobiliarias

- Evaluación de proyectos inmobiliarios.
- Estructuración y negociación de operaciones inmobiliarias con *brokers* e inversionistas.
- Análisis de riesgo y del mercado inmobiliario.

Oct 2017 **Ministerio de la Producción**

Jul 2018 Consultor Económico Financiero

- Evaluación de proyectos de inversión privada.
- Elaboración de proyecciones de estados financieros.
- Análisis de sensibilidad y riesgos.
- Evaluación de instrumentos para aumentar la formalización y productividad empresarial.
- Desarrollo de modelos econométricos.
-

Oct 2015	Ministerio de Economía y Finanzas
Oct 2017	Analista Económico
	<ul style="list-style-type: none"> - Análisis económico de propuestas y reformas sectoriales. - Elaboración del presupuesto de la Estrategia de Formalización. - Construcción de indicadores de desempeño (KPIs). - Implementación de <i>dashboards</i> para la balanza comercial. - Control y gestión de las reuniones del Grupo de Finanzas del APEC 2015.

CURSOS Y PROGRAMAS DE ESPECIALIZACIÓN

- Programa de Alta Especialización en Inteligencia de Negocios (2016 – Universidad ESAN).
- Curso de Técnicas de Minería de Datos (2016 - Universidad de los Andes, Bogotá).
- Curso de Uso y Explotación de Bases de Datos (2016 - Instituto Nacional de Estadística e Informático del Perú).
- Curso de Actualización en Economía (2015 - Banco Central de Reserva del Perú).

Resumen ejecutivo

El objetivo de la presente investigación es desarrollar tres modelos de clasificación de deudores a fin de determinar la probabilidad de incumplimiento asociada a los créditos “retail” solicitados a instituciones microfinancieras. Los modelos utilizados son: una función logística binaria, árbol de decisión y redes neuronales artificiales.

Asimismo, esta investigación centra sus esfuerzos en: determinar las variables relevantes asociadas a la probabilidad de incumplimiento, identificar el nivel de ajuste y discriminancia en cada modelo desarrollado, y conocer las principales fortalezas y debilidades de cada modelo.

Los resultados muestran que el modelo de redes neuronales artificiales presenta mejor nivel de ajuste y discriminancia en relación a los modelos de función logística binaria y árbol de decisión. Adicionalmente, se ha encontrado que las variables con mayor significancia se encuentran asociadas al historial crediticio del cliente, tales como: el calificaciones normales obtenidas durante el periodo de análisis y en los últimos 6 meses, los días de atraso en los últimos 6 y 18 meses, créditos microempresa y disposiciones de efectivo en los últimos 6 meses, entidades financieras que reportan saldo deudor en el último mes y la antigüedad del cliente en el sistema financiero.

Finalmente, los modelos de función logística binaria y árbol de decisión también muestran un buen nivel de ajuste y discriminancia, motivo por lo cual soportar una decisión de clasificación de clientes considerando la complementariedad de dichos modelos resulta optima, ya que dichos modelos consideran adicionalmente otras variables de historial crediticio relevantes y de tipo demográficas relevantes para la toma de decisiones.

CAPITULO I. PLANTEAMIENTO DEL PROBLEMA

La producción de mayores bienes y servicios por parte de las personas residentes de un país permite a los mismos incrementar principalmente su poder adquisitivo y de esa manera el bienestar social, no obstante, con frecuencia las personas enfrentan restricciones monetarias que les impiden la mayor producción de bienes y servicios.

En atención a lo expuesto, es relevante mencionar el rol de las instituciones financieras, puesto que son estas quienes se encargan de captar los excedentes de dinero de los ahorristas y colocar dicho excedente en personas con necesidades de financiamiento, lo cual permite efectivizar la ejecución de proyectos de inversión por parte de los inversionistas o en otros casos, permite dinamizar el consumo de productos. De manera particular, resulta preciso señalar el rol de las instituciones microfinancieras (en adelante, IMFs) en la economía peruana, quienes tienen como foco de atención los niveles socioeconómicos¹ (en adelante, NSE) “C” y “D” para la colocación de créditos retail².

Sobre la demanda de créditos de los NSE “C” y “D” en el Perú, esta se encuentra compuesta por emprendedores o trabajadores por cuenta propia, empleadores microempresarios y asalariados (OCDE/CEPAL/CAF, 2016). De otro lado, la Corporación Andina de Fomento (2013) señala que los emprendedores, los cuales en su mayoría son personas de bajos recursos, identifican oportunidades de negocios y se ocupan de obtener los recursos necesarios a través de créditos microempresa³ (entre otras opciones de financiamiento, tales como: prestamistas, familiares, entre otros) para llevar a cabo emprendimientos. Asimismo, Agapito (2010) afirma que también se encuentran las personas con necesidades de consumo de bienes (productos y/o servicios) cuyos ingresos se sustentan en actividades dependientes de un empleador.

¹ Según la Asociación Peruana de Empresas de Investigación de Mercados (APEIM), en el año 2017, el NSE “C” y “D” representan el 25.7% y 24.1% respectivamente del total de la población peruana, la cual asciende 31.8 millones de personas para el año en mención (data de la Encuesta Nacional de Hogares - ENAHO 2017).

² Hace referencia los créditos de tarjeta de crédito, personales y microempresas.

³ Son aquellos créditos destinados al financiamiento actividades de producción, comercialización o prestación de servicios de pequeña escala.

Adicionalmente, resulta relevante tener en consideración los altos niveles de morosidad en los créditos consumo y microempresa otorgados por las IMFs respecto de la banca comercial (ver Tabla I.2), motivo por el cual se sustenta el interés por parte de las IMFs en conocer la probabilidad de impago, es decir, aquella probabilidad de incumplimiento de las personas de los NSE “C” y “D” respecto de los diversos productos que ofrecidos, para lo cual hacen uso de información histórica sobre el comportamiento crediticio y de diferentes técnicas estadísticas y econométricas para determinar dicha probabilidad.

De acuerdo con el párrafo previo, se destaca a Coussement, Van den Bossche y De Bock (2014), quienes nos respaldan el interés por parte de las IMFs en desarrollar diversas técnicas para determinar la probabilidad de impago (pág. 2753), ya que de esa manera se podría realizar una mejor gestión sobre la reducción de pérdidas asociadas al castigo de las carteras excesivamente morosas, mejorar la colocación a clientes de bajo riesgo, adecuada fijación de precios de los productos ofrecidos, entre otras.

Por otro lado, es preciso señalar que el acuerdo internacional sobre regulación y supervisión bancaria denominado “Nuevo Acuerdo de Capital”⁴, aprobado en el año 2004, exhorta a las IMFs a la revisión de sus dotaciones de capital asociada a cubrir riesgos. En ese sentido, las IMFs están condicionadas a adoptar procesos internos para la medición del riesgo crediticio (Trivelli et al., 2004), tal como la implementación de modelos de clasificación crediticia, con la finalidad de discriminar los clientes según su probabilidad de riesgo. Asimismo, es relevante mencionar que la normativa peruana vigente⁵, reglamentada a través de la Resolución SBS N° 14354-2009 permite a las IMFs la implementación de modelos internos de manera adicional al modelo estándar para la gestión de crédito.

De acuerdo con lo expuesto, el presente documento de investigación se encuentra compuesto por cinco capítulos. El primer capítulo, aborda el planteamiento del

⁴ Convergencia Internacional de medidas y normas de capital, Comité de Supervisión Bancaria de Basilea (2004). También conocido con Basilea II.

⁵ Artículo 188 del Decreto Legislativo 1028, Decreto Legislativo que modifica la Ley General del Sistema Financiero y del Sistema de Seguros y Orgánica de la Superintendencia de la Banca y Seguros.

problema a través del análisis del nivel de colocaciones y morosidad del sector microfinanciero en el Perú. De la misma manera, el capítulo expone las consecuencias de los altos niveles de morosidad, objetivos y alcance de esta investigación.

El segundo capítulo contiene el tratamiento de la información, detalle de las fuentes utilizadas, descripción del universo de observaciones y las exclusiones consideradas para la selección de la base de datos. Asimismo, el capítulo define las ventanas de tiempo y la probabilidad de incumplimiento. Finalmente, el capítulo contiene una sección sobre la construcción de variables y otra sobre la creación de la muestra de entrenamiento y validación.

El tercer capítulo sobre marco teórico incluye la definición de riesgo de crédito, descripción de los indicadores de discriminancia y nivel ajuste (coeficiente de Gini, el estadístico de Kolmogorov-Smirnov y la Curva de ROC). Seguidamente, el capítulo detalla los modelos a desarrollar en la presente investigación (regresión logística, árbol de decisión y redes neuronales artificiales).

El siguiente capítulo aborda la elección de las variables en función a tres filtros: (1) análisis de concentración, (2) medición del nivel predictivo y verificación del principio de monotonicidad, y (3) análisis de correlación bivariada. Finalmente, en el último capítulo se detallan las conclusiones obtenidas luego de desarrollar los modelos de regresión logística, árbol de decisión y redes neuronales artificiales.

1.1. Análisis del Sector Microfinanciero

El Sector Microfinanciero en Perú se encuentra compuesto principalmente por (i) Empresas financieras, (ii) Cajas Municipales de Ahorro y Crédito (CMACs), (iii) Cajas Rurales de Ahorro y Crédito (CRACs) y (iv) Entidades de Desarrollo de la Micro y Pequeña Empresa (EDPYMES). Al respecto, en la Tabla I.1, se presentan las principales características establecidas por Arana , Núñez , Ore y Velazco (2017) entre CMACs, CRACs y EDPYMES. Adicionalmente, se agrega a la Tabla I.1, las principales características de las Empresas financieras con el objeto de poder contar con un alcance global para cada tipo de institución microfinanciera.

Tabla I.1: Principales diferencias entre Financieras, CMACs, CRACs y EDPYMES

Característica	Financieras	CMACs	CRACs	EDPYMES
Capital	Aportación de los socios en dinero, bienes o servicios.	Aportación de una Municipalidad Provincial en dinero y bienes.	Aportación de los socios en dinero, bienes o servicios.	Aportación de los socios en dinero, bienes o servicios.
Poder de decisión	La Junta General de Accionistas, el Directorio y la Gerencia General con poder.	El único accionista es una Municipalidad Provincial, el Directorio y la Gerencia Mancomunada con poder.	La Junta General de Accionistas, el Directorio y la Gerencia General con poder.	La Junta General de Accionistas, el Directorio y la Gerencia General con poder.
Clientes	Micro y pequeñas empresas, independientes y asalariados.	Micro y pequeños empresarios, en zonas urbanas.	Agricultores, ganaderos, micro y pequeños empresarios, en zonas rurales.	Micro y pequeños empresarios en zonas urbanas
Productos en ahorros	Autorizados a captar depósitos de ahorro y a plazo fijo.	Autorizados a captar depósitos de ahorro y a plazo fijo.	Autorizados a captar depósitos de ahorro y a plazo fijo.	Imposibilitados de captar depósitos del público.
Tamaño del crédito	En promedio es superior a los reportados por las CMACs.	Se mantiene por debajo del promedio en el sistema financiero	El crédito promedio es inferior a los reportados por las CMACs.	Se mantiene por debajo del promedio en el sistema financiero.

Fuente: Arana et al., 2017, pág. 5

De otro lado, es relevante tener en consideración que a marzo 2018, según la Superintendencia de Banca, Seguros y AFP (SBS), existen 39 instituciones microfinancieras: 11 Empresas Financieras, 12 CMACs, 6 CRACs, 9 EDPYMES y un banco. No obstante, es preciso señalar que la Banca múltiple en Perú, posee áreas especializadas para la atención de créditos para pequeñas y micro empresas.

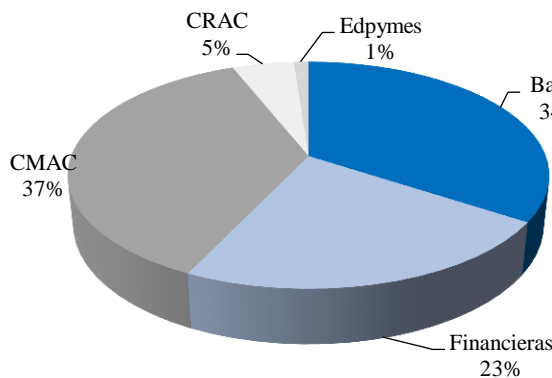
1.1.1. Análisis de los Créditos de Consumo y Microempresa en el Perú

Como se puede apreciar en el Gráfico I.1.A, a diciembre de 2017, las CMACs representan el 37% del mercado de créditos microempresa⁶, seguido por la Banca Múltiple y las Empresas Financieras que concentran el 34% y 23% respectivamente.

⁶ Tarjetas de crédito, descuentos, préstamos, factoring, comercio exterior, arrendamiento financiero y lease-back.

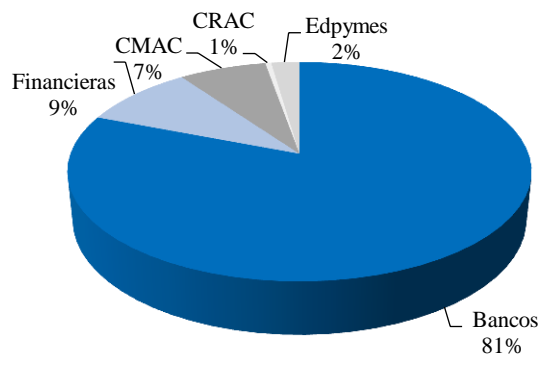
Gráfico I.1: Colocaciones al 31 de diciembre 2017

A. Créditos Microempresa



Elaboración Propia
Fuente: SBS (diciembre 2017)

B. Créditos Consumo



Elaboración Propia
Fuente: SBS (diciembre 2017)

Además, se puede observar que las CRACs y EDPYMES mantienen una participación conjunta del 6%. Por otro lado, el Gráfico I.1.B muestra la cuota de mercado de las colocaciones de créditos para consumo⁷, del cual se puede inferir que son los Bancos quienes lideran el segmento con una participación del 81% y que las IMFs representan el 19% restante.

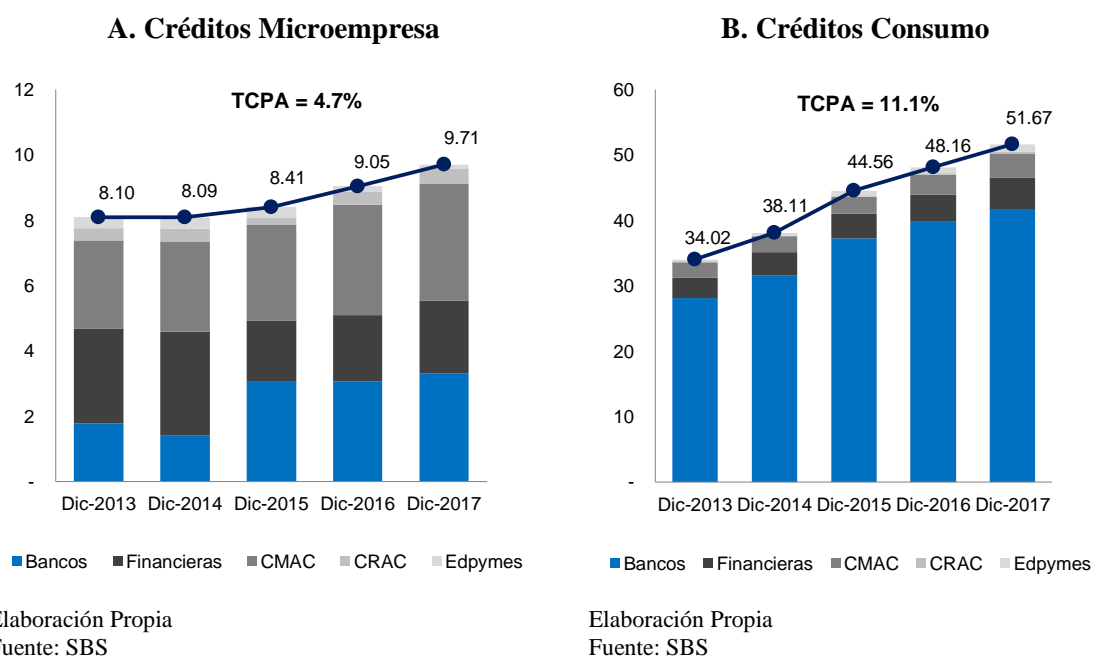
En efecto, del Gráfico I.1.A se puede concluir que las colocaciones de créditos microempresa tiene como posición predominante a las IMFs (Financieras, CMACs, CRACs y EDPYMES) con el 66% y por el contrario, la colocaciones de créditos consumo se encuentra como actor predominante a la Banca Múltiple, quien mantiene una cuota de mercado de 81%. No obstante, resulta preciso señalar que la Banca Múltiple también atiende a los NSE “A” y “B”, y que son las IMFs, las que concentran mayor esfuerzo en atender a los NSE “C” y “D” (Agapito, 2010).

Con respecto a la evolución de ambos segmentos, se puede apreciar que los créditos microempresa vienen creciendo a una tasa promedio anual de 4.7% durante el periodo 2013 – 2017 y actualmente se condice con un volumen que asciende a 9.71 miles de millones de soles. Por otro lado, los créditos consumo muestran una tasa de

⁷ Tarjetas de crédito, préstamos revolventes y no revolventes, préstamos para autos, arrendamiento financiero y lease-back.

crecimiento anual promedio (TCPA) de 11.1% para el mismo periodo de análisis y un total de 51 miles de millones de soles. (Ver Gráfico I.2.B).

Gráfico I.2: Evolución de Créditos Microempresa y Consumo



Sobre el párrafo previo, se puede concluir que el segmento de créditos consumo muestran mayor crecimiento respecto de los créditos microempresa y que el total de colocaciones de créditos consumo es 5.3 veces mayor al de créditos microempresarios. Sin embargo, es preciso señalar que los Bancos no tienen como foco de atención los NSE “C” y “D” y que el monto de créditos consumo colocado por las IMFs asciende a 9.89 miles de millones de soles.

1.1.2. Nivel de morosidad de los Créditos Consumo y Microempresa

Al respecto, se consideran dos definiciones al momento de analizar los niveles de morosidad, ambas en función de los días de incumplimiento. En ese sentido, la primera se encuentra acorde a los estándares internacionales establecidos en Basilea II mediante el cual se considera un crédito como moroso (vencido) cuando supera los 90 días de incumplimiento. De otro lado, según el criterio contable de la SBS, la morosidad se encuentra en función al tipo de crédito, por lo cual el plazo para

considerar la totalidad del crédito como vencido es después de transcurridos 15 días calendario para créditos de deudores no minoristas; 30 días calendario para créditos a pequeñas empresas y a microempresas; y para el caso de créditos de consumo, se existe un tratamiento especial, el cual considera vencida sólo la porción no pagada en los 30 días de la fecha pactada, mientras que después de los 90 días del primer incumplimiento, se considerará la totalidad de la deuda.

En el presente trabajo, se tomará el criterio contable utilizado por la SBS para comparar los distintos niveles de morosidad según el tipo de crédito (consumo y microempresa) y actores del sistema financiero. En ese sentido, en el Tabla I.2.A se puede apreciar la evolución del nivel de morosidad de los créditos microempresa en las IMFs, así, se puede evidenciar que las EDPYMES presentan el mayor porcentaje de morosidad en el año 2017, las CMACs muestran un nivel de morosidad constante en el tiempo y cercano al 6%. Adicionalmente, se puede observar que las CRACs han reducido notablemente la morosidad de sus créditos microempresa (8.69% en el 2013 vs 5.98 en el 2017) y que son los Bancos quienes poseen menor mora en el segmento de análisis.

Tabla I.2: Nivel de morosidad (%) según tipo crédito e institución financiera

A. Créditos Microempresa

	2013	2014	2015	2016	2017	
Bancos	3.64	3.93	2.94	3.08	3.22	
Financieras	4.45	4.68	4.79	4.47	4.19	
CMAC	6.15	5.98	6.00	5.86	6.09	
CRAC	8.69	11.69	9.10	5.92	5.98	
Edpymes	5.20	4.83	4.64	5.78	7.09	

Elaboración Propia
Fuente: SBS

B. Créditos Consumo

	2013	2014	2015	2016	2017	
Bancos	3.37	3.27	3.28	3.58	3.64	
Financieras	4.20	4.43	4.04	4.92	4.91	
CMAC	2.38	2.38	2.89	2.96	2.83	
CRAC	4.54	6.53	4.78	3.19	3.07	
Edpymes	3.84	3.04	3.25	3.12	3.61	

Elaboración Propia
Fuente: SBS

Con respecto a la Tabla I.2.B, este muestra el nivel de morosidad de los créditos de consumo, en el cual se aprecian niveles inferiores respecto de los créditos microempresa, lo cual se podría explicar al analizar la demanda dichos productos, ya que los créditos de consumo (préstamos personales, tarjetas de crédito, entre otros)

son productos demandados independientemente del NSE. Otro factor relevante para tener en cuenta, es el elevado riesgo que conlleva la ejecución de proyectos por parte de microempresarios mientras que la obtención de un crédito de consumo podría ser para cubrir una eventual necesidad con montos por lo general menores a los créditos microempresa. Finalmente, para realizar un mejor análisis de los préstamos de consumo se debería solo considerar aquellos productos demandados por los NSE “C” y “D”, ya que la gran mayoría de personas que demandan créditos microempresa se encuentran dentro de los estratos señalados.

1.2. Problemática

Los altos niveles de morosidad se encuentran asociados a mayores provisiones por parte de las entidades financieras, es decir, conforme aumenta el número de días de atraso, la institución se encuentra en la obligación de aumentar el porcentaje de provisión conforme lo establecido por la SBS⁸.

Lo expuesto, socava la viabilidad financiera de las instituciones microfinancieras, toda vez que deteriora la calidad de los activos dentro del balance de cada institución, es decir, aumenta la proporción de cartera atrasada⁹ y alto riesgo¹⁰. Asimismo, los altos niveles de provisiones dificultan la generación de ingresos (colocación de créditos) y aumenta los gastos operativos asociados a monitorear y hacer seguimiento a los créditos atrasados, lo cual incide en menores ratios de rentabilidad para los accionistas (Trivelli et al., 2004).

De otro lado, Addae-Korankye (2014) señala que la causa de los altos niveles de morosidad podría ser atribuible a la mala gestión de riesgo crediticio¹¹ en las IMFs, lo cual se pueda deber a un deficiente conocimiento sobre la probabilidad de incumplimiento y clasificación de los demandantes.

⁸ Ver apartado 2.Tasas de Provisiones del Capítulo III sobre Exigencia de Provisiones del Reglamento para la Evaluación y Clasificación del Deudor y la Exigencia de Provisiones aprobado por Resolución SBS N° 11356.

⁹ La cartera atrasada se obtiene del ratio de colocaciones vencidas más aquellas en cobranza judicial sobre el total de colocaciones.

¹⁰ La cartera de alto riesgo se obtiene del ratio de colocaciones vencidas más cobranza judicial, refinanciadas y reestructuradas sobre el total de colocaciones.

¹¹ La definición de riesgo de crédito se encuentra dentro de la sección 3.1 del Capítulo III. Marco Teórico.

Finalmente, es preciso traer a colación lo señalado por Ledgerwood (1998), quien sostiene que los altos niveles de morosidad traen consigo problemas para el adecuado funcionamiento de las instituciones microfinancieras, como por ejemplo: elevados gastos operativos para hacer seguimiento a los créditos reportados con atraso, lo cual podría afectar la liquidez de la empresa. De otro lado, los altos niveles de morosidad generan un efecto adverso sobre las ganancias de capital y rentabilidad (pág. 244).

Adicionalmente, Westley y Shaffer (1997) exponen que los altos niveles de morosidad generan afectación a la relación de largo plazo de las IMFs y sus clientes, puesto que menoscaba la lealtad de los últimos y además, conlleva a un efecto contagio en cuanto a la adopción de una posición de no pago (pág. 5).

En conclusión, los altos niveles de morosidad son un problema que afecta tanto a la oferta como a la demanda de créditos. En ese sentido, resulta necesario y contar con herramientas que permitan a las IMFs conocer la probabilidad de incumplimiento de los clientes, con el objeto de poder clasificarlos y en efecto, mitigar los niveles de morosidad en las carteras.

Al respecto, el presente trabajo plantea tres métodos de clasificación de clientes bancarizados de los NSE “C” y “D” atendidos por IMFs del Perú como contribución a los esfuerzos de la gestión de riesgo crediticio de contar con una adecuada clasificación de clientes buenos y malos. En ese sentido, se plantea un modelo *credit scoring* correspondiente con una regresión logística que permita determinar las variables relevantes y estimar la probabilidad de incumplimiento (default) de créditos microempresa y consumo en conjunto. Asimismo, se plantea un modelo de árbol de decisión que ayude a conocer cuáles son las variables determinantes y reglas sucesivas para la clasificación crediticia y finalmente se desarrolla un modelo de redes neuronales artificiales (en adelante, RNA) como herramienta complementaria y alternativa a las dos anteriores, la cual forma parte del set de herramientas de inteligencia artificial asociadas al aprendizaje automático y retroalimentación, permitiendo comparar el nivel predictivo de cada modelo propuesto.

1.3. Objetivos

1.3.1. Objetivo principal

Evaluar la gestión de riesgo crediticio de las IMFs a través del desarrollo de modelos de clasificación para clientes bancarizados de los niveles socioeconómicos “C” y “D” de la economía peruana.

1.3.2. Objetivos específicos

- Determinación de variables explicativas relevantes acorde con las mejores prácticas estadísticas. Contar con indicadores de discriminancia y ajuste aceptables para los modelos.
- Construir un modelo de regresión logística, un modelo de árbol de decisión y un modelo de redes neuronales para conocer la probabilidad de incumplimiento y clasificar a los clientes como buenos y malos.
- Comparar los modelos propuestos para identificar cual es el mejor predictor de riesgo de crédito, así como estudiar sus fortalezas y debilidades.

1.4. Alcance y Limitaciones

1.4.1. Alcance

El presente documento desarrolla modelos de clasificación crediticia para clientes de NSE “C” y “D” que se encuentran bancarizados y han obtenido un crédito “*retail*” (crédito consumo o microempresa) en alguna IMF.

La data de la presente investigación contiene solo una muestra representativa de las colocaciones otorgadas Empresas Financieras, CMACs, CRACs y EDPYMES.

1.4.2. Limitaciones

Las limitaciones y dificultades en la elaboración de modelos de calificación crediticia de clientes se originan principalmente por la obtención de las bases de datos, la construcción de las variables, utilización de programas estadísticos acorde al volumen de datos, selección adecuada de las potenciales variables relevantes para cada modelo, entre otras. En ese sentido, la presente investigación supera las limitaciones anteriormente comentadas y genera un aporte al desarrollo de modelos de calificación.

CAPITULO II. TRATAMIENTO DE LA INFORMACIÓN

En el presente capítulo contiene las fuentes de información utilizadas para la creación de variables, la descripción del universo de datos para el desarrollo de modelos de clasificación. Asimismo, se define las ventanas de tiempo y se detalla aquellas exclusiones relevantes que se realizaron sobre el universo de datos. Posteriormente, se aborda la definición de probabilidad de incumplimiento y se construyen las variables independientes que podrían determinar el comportamiento de pago de los clientes de los NSE “C” y “D”.

Finalmente, el capítulo detalla los porcentajes referentes a las muestras de entrenamiento y validación, las cuales serán utilizadas para desarrollar y medir el desempeño de los modelos de clasificación.

2.1. Fuentes de información

Para asegurar la calidad de la información, se utilizaron las siguientes fuentes de información:

2.1.1. Reporte Crediticio Consolidado (RCC)

El RCC es un reporte con periodicidad mensual que la SBS remite a todas las entidades del sector financiero. Dentro del RCC se encuentra información correspondiente a las personas que participan en el mercado de créditos. De manera más detallada, el RCC está compuesto por dos tablas, en la primera se puede encontrar información sobre: el tipo de persona (natural o jurídica), el número de empresas que ha reportado cada cliente y saldo deudor según el tipo de calificación (normal, cliente con problemas potenciales, deficiente, dudoso o perdida). De otro lado, en la segunda tabla, se encuentra información sobre: la institución financiera donde se registra la deuda, el tipo de crédito, la condición del crédito en cuanto a días de atraso, saldo deudor y la clasificación del cliente por parte de la institución financiera ((normal, cliente con problemas potenciales, deficiente, dudoso o perdida).

Finalmente, resulta relevante señalar que para la presente investigación se han empleado 42 RRCs, asociados al periodo entre julio 2014 y diciembre 2016.

2.1.2. Información socio-demográfica

Adicionalmente, se construyeron variables de naturaleza cualitativa y cuantitativa a partir de información socio-demográfica de los participantes del mercado de créditos que conforman universo de observaciones (707 mil registros aproximadamente). Al respecto, se crearon variables, tales como: edad, género, estado civil, número de hijos, entre otras.

2.2. Descripción del universo de observaciones

El sistema financiero en el Perú atiende a individuos con requerimientos de deuda crediticia directa o indirecta, la cual se podría segmentar según su destinatario, siendo empresas o personas. En efecto, según el Reglamento para la Evaluación y Clasificación del Deudor y la Exigencia de Provisiones aprobado por Resolución SBS N° 11356-2010, las deudas destinadas a empresas se subdividen en cinco categorías para permitir una mayor diferenciación en el nivel de endeudamiento; mientras que, para los créditos a personas se separan en créditos de consumo revolvente (tarjetas de créditos) y no revolventes (Ver Anexo I).

En relación a lo expuesto, la base de datos contiene observaciones de créditos “*retail*”, es decir, conformada por los siguientes tipos de créditos:

- Préstamos personales (PP): considera a todos los clientes con deuda de consumo no-revolvente sin considerar los créditos pignoraticios¹² en los últimos 30 meses, es decir desde julio 2014 hasta diciembre 2016. Esto no implica que el cliente haya tenido deuda en los 30 meses.

¹²

Son préstamos prendarios en donde se deja como garantía un objeto de valor.

- Para préstamos a negocios (MES): Se consideran a todos los clientes con préstamos a microempresa en los últimos 30 meses, es decir desde julio 2014 hasta diciembre 2016. Esto no implica que el cliente haya tenido deuda en los 24 meses.
- Para tarjetas de crédito (TC): Se consideran a todos los clientes con por lo menos una línea de consumo revolvente reportada en el último mes (diciembre 2016) con o sin deuda.

Con estas definiciones, queda claro que no necesariamente todos los clientes definidos en el universo tienen saldo deudor al cierre de diciembre 2016.

Con respecto a la cantidad de observaciones, se cuenta con un universo de aproximadamente 707 mil registros, del cual solo se utilizará una muestra representativa de 319 mil observaciones, la cual es resultado de la aplicación de dos filtros relevantes para el desarrollo de la presente investigación. El primer filtro se encuentra relacionado con el NSE “C” y “D”, el selecciona aquellos créditos otorgados por Empresas financieras, CMACs, CRACs y EDPYMES. Por otro lado, el segundo filtro se basa realiza según el tipo de crédito, seleccionado aquellos créditos consumo y microempresa (Ver Anexo I sobre el tipo de créditos).

2.3. Ventanas de tiempo

Para definir la probabilidad de *default*¹³, se establecieron dos ventanas de tiempo:

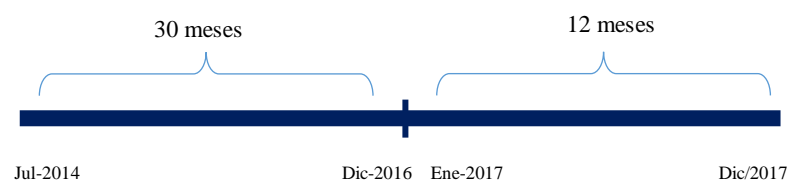
- La ventana de observación, mediante la cual se observa el comportamiento del individuo durante 30 meses correspondiente con el análisis del conjunto de variables explicativas construidas a partir de los RCCs.
- La ventana de definición, a través de la cual se observa la desempeño del individuo en el pago de sus deudas. Esta ventana considera un horizonte de 12 meses, con el objeto de contar con un amplio panorama del desempeño.

¹³

En el siguiente acápite del presente Capítulo se define el *default* (incumplimiento).

Es relevante señalar que la construcción de la variable *default* (variable dependiente) considera 12 meses, toda vez que se considera poco probable que la tasa de deudores que caen en *default* tenga un crecimiento que justifique ampliar el horizonte (Gráfico II.1).

Gráfico II.1: Ventanas para la creación de variables



Elaboración Propia

2.4. Exclusiones en la Base de Datos¹⁴

Del total de observaciones, se excluyen los clientes que cumplen con alguna de las siguientes características durante el periodo de observación de 30 meses:

Tabla II.1: Exclusiones en la Base de Datos

Restricción	Descripción	Motivo	Aplica
R1	Si registra calificativo “dudoso” o “pérdida” en el último reporte.	Este tipo de clientes ya se encuentran en default, según la definición en el acápite anterior.	PP, MES y TC
R2	No registrar saldo deudor en los últimos 24 meses.	No existe historial crediticio para calcular una probabilidad de default.	TC
R3	Si registra solo un mes con saldo en 24 meses.	Insuficiente información para extraer conclusiones.	PP, MES y TC
R4	DNI ¹⁵ asociado a más de 1 código SBS.	Dificultad en la consolidación de la información.	PP, MES y TC
R5	RUC ¹⁶ asociado a más de 1 código SBS.	Motivo: Dificultad en la consolidación	PP, MES y TC
R6	Tener un préstamo con convenio	No presenta un comportamiento real de la voluntad de pago del crédito.	PP

¹⁴ Anexo 2 sobre definiciones relevantes para un mejor entendimiento de las variables.

¹⁵ Documento de Identidad.

¹⁶ El Registro Único de Contribuyente es un código identificador de los contribuyentes en el sistema tributario peruano administrado por la Superintendencia Nacional de Aduanas y Administración Tributaria (SUNAT).

2.5. Definición del default

Con respecto a la definición de incumplimiento o impago, es relevante traer a colación lo expuesto por Rayo, Lara y Camino (2010), ya que establecen criterios que se deben tener en cuenta al momento definir “incumplimiento”, los cuales se encuentran asociados a los días atraso de un crédito, toda vez que los mismo suponen un coste para la organización.

En ese sentido, Rayo et al. (2010) afirman lo siguiente:

- “a. El atraso percibido ha de ser real y no estimado, según fechas concretas marcadas en la contratación del crédito, en función del método estipulado para su amortización por las partes contratantes.
- b. El atraso ha de producirse en, al menos, una cuota de amortización del microcrédito.
- c. El atraso considerado ha de suponer un incremento en el coste para la entidad más que proporcional al habitual en caso de no sucederse esta contingencia. Generalmente, estos incrementos suelen darse en términos de costes administrativos debido al incremento monetario que supone realizar un seguimiento y gestionar el pago de un crédito cuyo reembolso mantiene un retraso considerable (pág. 97).”

De otro lado, se podría definir impago como un evento adverso asociado una probabilidad de ocurrencia. De lo expuesto y para efectos de la presente investigación, se define el impago como la probabilidad de que un crédito registre un calificativo “dudoso” o “pérdida” (al 5%) en alguno de los 12 meses del periodo de definición.

La justificación del párrafo anterior, versa en tanto los créditos con calificaciones de “dudoso” y “pérdida” están asociados con un nivel de exigencia de provisiones regulatorias por encima del 60% del saldo deudor. Según el marco normativo peruano, las tasas de provisiones¹⁷ para de calificativo “dudoso” y “pérdida” son de 60% y 100%, respectivamente.

¹⁷ Las tasas de provisiones indicadas son para créditos sin ningún tipo de garantía vinculada.

2.6. Variables independientes

En este apartado, se construyeron las variables independientes conforme con las fuentes de información de apartado 2.1 de presente capítulo (RCCs y base socio-demográfica). Dichas variables se construyeron con el objeto de poder explicar el comportamiento de pago de los clientes. En ese sentido, las variables en cuestión son potencialmente relevantes para determinar la probabilidad de incumplimiento y son de naturaleza cuantitativa y cualitativa.

El número total de variables construidas fue de 113, siendo las diez primeras variables de la Tabla II.2 construidas sobre la base socio-demográfica y el resto sobre la base de los RCCs.

Tabla II.2: Variables independientes

Nº	Variables	Descripción
1	Edad	Edad de la persona
2	Sexo	Sexo de la persona (Femenino o Masculino)
3	Estado Civil	Estado Civil de la Persona (Soltero, Conviviente, Casado, etc)
4	Flg_hijos	Dummy si tiene o no hijos
5	Num_hijos	Número de hijos
6	Flg_auto	Dummy si tiene o no auto
7	Num_autos	Número de autos
8	Lima	Si el deudor pertenece a Lima o a provincia
9	Departamento	Departamento al que pertenece el deudor
10	Provincia	Provincia al que pertenece el deudor
11	Nro_ENT_1m	Número de entidades en el último mes
12	Nro_ENT_6m	Número de entidades en los últimos 6 meses
13	Nro_ENT_12m	Número de entidades en los últimos 12 meses
14	Nro_ENT_18m	Número de entidades en los últimos 18 meses
15	Nro_ENT_24m	Número de entidades en los últimos 24 meses
16	Clasificacion_6m	Clasificación por la SBS en los últimos 6 meses
17	Clasificacion_12m	Clasificación por la SBS en los últimos 12 meses
18	Clasificacion_18m	Clasificación por la SBS en los últimos 18 meses
19	Clasificacion_24m	Clasificación por la SBS en los últimos 24 meses
20	Clasificacion_30m	Clasificación por la SBS en los últimos 30 meses
21	Clasificacion_6m_m	Clasificación por la SBS en los últimos 6 meses de microempresa
22	Clasificacion_12m_m	Clasificación por la SBS en los últimos 12 meses de microempresa

23	Clasificacion_18m_m	Clasificación por la SBS en los últimos 18 meses de microempresa
24	Clasificacion_24m_m	Clasificación por la SBS en los últimos 24 meses de microempresa
25	Clasificacion_30m_m	Clasificación por la SBS en los últimos 30 meses de microempresa
26	condicion_6m	Días de atraso en los últimos 6 meses
27	condicion_12m	Días de atraso en los últimos 12 meses
28	condicion_18m	Días de atraso en los últimos 18 meses
29	condicion_24m	Días de atraso en los últimos 24 meses
30	condicion_30m	Días de atraso en los últimos 30 meses
31	condicion_6m_m	Días de atraso en los últimos 6 meses de microempresa
32	condicion_12m_m	Días de atraso en los últimos 12 meses de microempresa
33	condicion_18m_m	Días de atraso en los últimos 18 meses de microempresa
34	condicion_24m_m	Días de atraso en los últimos 24 meses de microempresa
35	condicion_30m_m	Días de atraso en los últimos 30 meses de microempresa
36	Nro_Normal_6m	Clasificación normal en los últimos 6 meses
37	Nro_Normal_12m	Clasificación normal en los últimos 12 meses
38	Nro_Normal_18m	Clasificación normal en los últimos 18 meses
39	Nro_Normal_24m	Clasificación normal en los últimos 24 meses
40	Nro_Normal_30m	Clasificación normal en los últimos 30 meses
41	Peor_6m	Peor calificación en los últimos 6 meses
42	Peor_12m	Peor calificación en los últimos 12 meses
43	Peor_18m	Peor calificación en los últimos 18 meses
44	Peor_24m	Peor calificación en los últimos 24 meses
45	Peor_30m	Peor calificación en los últimos 30 meses
46	Nro_cast_6m	Total de castigos en los últimos 6 meses
47	Nro_cast_12m	Total de castigos en los últimos 12 meses
48	Nro_cast_18m	Total de castigos en los últimos 18 meses
49	Nro_cast_24m	Total de castigos en los últimos 24 meses
50	Nro_cast_30m	Total de castigos en los últimos 30 meses
51	Nro_ref_6m	Total de contratos refinanciados en los últimos 6 meses
52	Nro_ref_12m	Total de contratos refinanciados en los últimos 12 meses
53	Nro_ref_18m	Total de contratos refinanciados en los últimos 18 meses
54	Nro_ref_24m	Total de contratos refinanciados en los últimos 24 meses
55	Nro_ref_30m	Total de contratos refinanciados en los últimos 30 meses
56	Nro_venc_6m	Total de contratos vencidos en los últimos 6 meses
57	Nro_venc_12m	Total de contratos vencidos en los últimos 12 meses
58	Nro_venc_18m	Total de contratos vencidos en los últimos 18 meses
59	Nro_venc_24m	Total de contratos vencidos en los últimos 24 meses
60	Nro_venc_30m	Total de contratos vencidos en los últimos 30 meses

61	Nro_jud_6m	Total de contratos judiciales en los últimos 6 meses
62	Nro_jud_12m	Total de contratos judiciales en los últimos 12 meses
63	Nro_jud_18m	Total de contratos judiciales en los últimos 18 meses
64	Nro_jud_24m	Total de contratos judiciales en los últimos 24 meses
65	Nro_jud_30m	Total de contratos judiciales en los últimos 30 meses
66	Disp_6m	Número de disposición de efectivo durante los últimos 6 meses
67	Disp_12m	Número de disposición de efectivo durante los últimos 12 meses
68	Disp_18m	Número de disposición de efectivo durante los últimos 18 meses
69	Disp_24m	Número de disposición de efectivo durante los últimos 24 meses
70	Disp_30m	Número de disposición de efectivo durante los últimos 30 meses
71	Monto_max_Dip_6m	Monto máximo disponible en efectivo en los últimos 6 meses
72	Monto_max_Dip_12m	Monto máximo disponible en efectivo en los últimos 12 meses
73	Monto_max_Dip_18m	Monto máximo disponible en efectivo en los últimos 18 meses
74	Monto_max_Dip_24m	Monto máximo disponible en efectivo en los últimos 24 meses
75	Monto_max_Dip_30m	Monto máximo disponible en efectivo en los últimos 30 meses
76	Tiene_prestamo_1m	Clasificación de préstamos en el último mes
77	Tiene_prestamo_6m	Clasificación de préstamos en los últimos 6 mes
78	Tiene_prestamo_12m	Clasificación de préstamos en los últimos 12 meses
79	Tiene_prestamo_18m	Clasificación de préstamos en los últimos 18 meses
80	Tiene_prestamo_24m	Clasificación de préstamos en los últimos 24 meses
81	Tiene_pignoraticio_1m	Clasificación de producto joyas en el último mes
82	Tiene_pignoraticio_6m	Clasificación de producto joyas en los últimos 6 meses
83	Tiene_pignoraticio_12m	Clasificación de producto joyas en los últimos 12 meses
84	Tiene_pignoraticio_18m	Clasificación de producto joyas en los últimos 18 meses
85	Tiene_pignoraticio_24m	Clasificación de producto joyas en los últimos 24 meses
86	Tiene_DE_1m	Clasificación de disponibilidad en efectivo en el último mes
87	Tiene_DE_6m	Clasificación de disponibilidad en efectivo en los últimos 6 meses
88	Tiene_DE_12m	Clasificación de disponibilidad en efectivo en los últimos 12 meses
89	Tiene_DE_18m	Clasificación de disponibilidad en efectivo en los últimos 18 meses
90	Tiene_DE_24m	Clasificación de disponibilidad en efectivo en los últimos 24 meses
91	DE_Porc_1m	Porcentaje de disposición en efectivo en el último mes
92	DE_Porc_6m	Porcentaje de disposición en efectivo en los últimos 6 meses
93	DE_Porc_12m	Porcentaje de disposición en efectivo en los últimos 12 meses
94	DE_Porc_18m	Porcentaje de disposición en efectivo en los últimos 18 meses
95	DE_Porc_24m	Porcentaje de disposición en efectivo en los últimos 24 meses
96	LU_porc_1m	Porcentaje de línea utilizada en el último mes
97	LU_porc_6m	Porcentaje de línea utilizada en los últimos 6 meses
98	LU_porc_12m	Porcentaje de línea utilizada en los últimos 12 meses

99	LU_porc_18m	Porcentaje de línea utilizada en los últimos 18 meses
100	LU_porc_24m	Porcentaje de línea utilizada en los últimos 24 meses
101	Nro_emp_6m_m	Número de empresas en los últimos 6 meses de microempresa
102	Nro_emp_12m_m	Número de empresas en los últimos 12 meses de microempresa
103	Nro_emp_18m_m	Número de empresas en los últimos 18 meses de microempresa
104	Nro_emp_24m_m	Número de empresas en los últimos 24 meses de microempresa
105	Nro_emp_30m_m	Número de empresas en los últimos 30 meses de microempresa
106	Tiene_hip_1m	Clasificación de hipotecario en el último mes
107	Tiene_hip_6m	Clasificación de hipotecario en los últimos 6 meses
108	Tiene_hip_12m	Clasificación de hipotecario en los últimos 12 meses
109	Tiene_hip_18m	Clasificación de hipotecario en los últimos 18 meses
110	Tiene_hip_24m	Clasificación de hipotecario en los últimos 24 meses
111	NroNromales	Total de calificación Normal
112	Antigüedad	Tiempo en meses del cliente
113	D_Dispon_6m	Dummy indica si tiene alguna disposición de efectivo en los últimos 6 meses o presenta valor nulo

2.7. Definición de muestras de trabajo

Para el desarrollo de los modelos se definieron dos tipos de muestras de trabajo:

- Muestra de entrenamiento: representa el 80% de la base de datos definido desde julio 2014 hasta diciembre 2016.
- Muestra de validación: representa el 20% de la base de datos definido desde julio 2014 hasta diciembre 2016.

CAPITULO III. MARCO TEÓRICO

En la primera sección del presente capítulo se define riesgo de crédito en función al riesgo de incumplimiento y de mercado. Posteriormente, se describen los indicadores de discriminancia y nivel ajuste, los cuales permiten evaluar el desempeño estadístico de los modelos de clasificación de deudores. Así, se especifica el coeficiente de Gini, el estadístico de Kolmogorov-Smirnov y la Curva de ROC.

Finalmente, este capítulo aborda el marco teórico de los tres modelos de clasificación que son objeto de estudio en la presente investigación. De esta manera, en la última sección se detallan los modelos de regresión logística, árbol de decisión y redes neuronales artificiales.

3.1. Riesgo de crédito

El riesgo de crédito se define como el potencial de que un prestatario bancario o contraparte incumpla el pago de una obligación parcial o total según los términos acordados, el cual puede darse por factores intrínsecos al cliente (riesgo de incumplimiento) o algún efecto adverso en el ámbito económico (riesgo de mercado). Si el riesgo de crédito obedece a factores propios del cliente, se debe contar con una valoración objetiva de la probabilidad de no pago, la cual puede emanar del hecho de que la contraparte no esté dispuesta a pagar o no pueda cumplir con sus obligaciones financieras. Por otro lado, si el riesgo de crédito deriva de un efecto adverso en la economía, se debe evaluar la pérdida potencial de la cartera de créditos de una institución financiera.

La presente investigación tendrá en consideración el riesgo de crédito asociado al riesgo de incumplimiento por factores intrínsecos a los clientes, falta de disposición a pagar o incapacidad para cumplir con las obligaciones financieras acordadas.

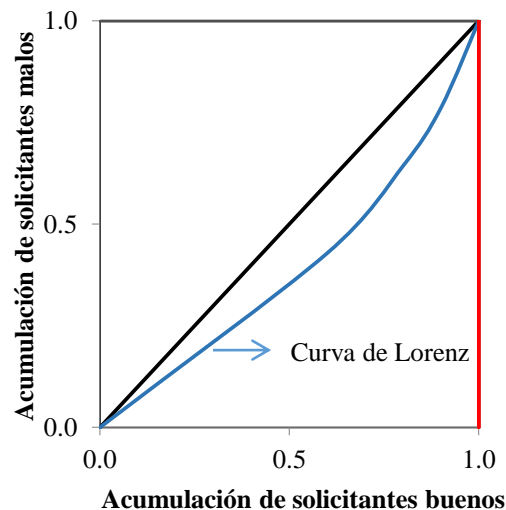
3.2. Indicadores de discriminancia y ajuste

3.2.1. Coeficiente de Gini

El coeficiente de Gini es un método que permite la evaluación de un modelo de score crediticio, para lo cual mide la desigualdad entre dos variables en un determinado ordenamiento. De manera particular, el coeficiente en cuestión mide qué tan desigual es la acumulación de los individuos aceptados y rechazados, ordenados por su probabilidad.

Como se puede apreciar en la Gráfico III.1, la línea recta indica una acumulación homogénea de ambos grupos y la curva de Lorenz¹⁸ muestra como realmente se distribuyen.

Gráfico III.1: Acumulación de grupos



Elaboración Propia

De manera más detallada, se puede evidenciar en el Gráfico III.1 que la mejor situación se materializa cuando la curva de Lorenz toma la forma de una L invertida, constituida por el eje de la acumulación de solicitantes buenos hasta tomar el valor de uno y luego el eje de acumulación de solicitantes malos, siendo el área debajo de la

¹⁸ Hace referencia a la representación gráfica utilizada para analizar la distribución relativa de una variable.

curva de Lorenz “B” igual a 0 y el coeficiente de Gini igual a 1. En efecto, el coeficiente de análisis se calcula de la siguiente manera:

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

Al respecto, Peussa (2016) al medir el desempeño de un modelo estadístico de calificación crediticia a través del coeficiente de Gini concluyó que contar con un valor de 44.8% en la muestra de entrenamiento representa un valor alto para el indicador en cuestión. Asimismo, señaló que un valor de 37.2% para la muestra de validación aún continua siendo bueno al ser comparado con el valor observado del modelo no estadístico basado en expectativas racionales utilizado por la institución bancaria que analizó.

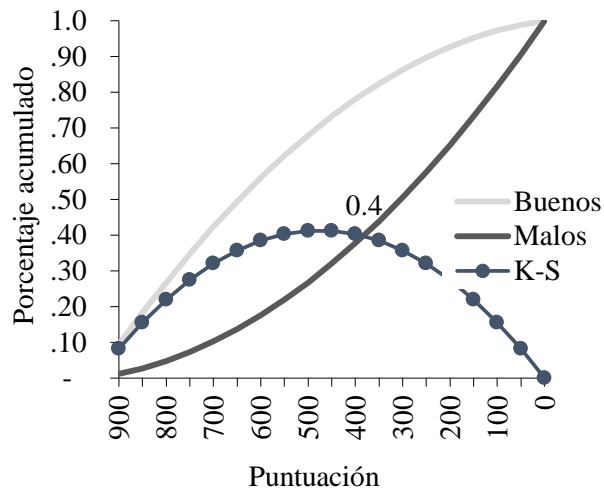
3.2.2. Estadístico Kolmogorov-Smirnov

El estadístico Kolmogorov-Smirnov (en adelante, KS) o también conocido como “de separación” es una prueba no paramétrica utilizada principalmente para medir la bondad de ajuste de dos distribuciones entre sí. En efecto, el KS tiene como objetivo verificar que la distribución de los solicitantes buenos (aceptados) es lo más distinta posible a la distribución de los solicitantes malos (rechazados).

En virtud de lo expuesto, en un escenario de discriminancia perfecta, se esperaría que hasta determinada probabilidad se hayan distribuido todos los solicitantes buenos de la muestra y, a partir de esa probabilidad, se acumulen los individuos malos; por lo que, el diferencial de las acumulaciones de clientes buenos y malos en esa probabilidad es equivalente a 1. De otro lado, un escenario realista sería el planteado en el Gráfico III.2, asociado a discriminancia imperfecta, en el cual se puede apreciar que la curva KS se obtiene del diferencial de la curva de buenos y malos.

De manera adicional, se puede agregar que cuanto el KS es más cercano a 1 mejor será el desempeño del modelo. Así, Halim y Vina Humira (2014), señalaron que un modelo de puntuación crediticia tiene un buen poder de discriminación cuando el valor del KS se encuentra entre 28% y 35%.

Gráfico III.2: Acumulación de distribución de solicitantes buenos y malos



Elaboración Propia

3.2.3. Curva ROC

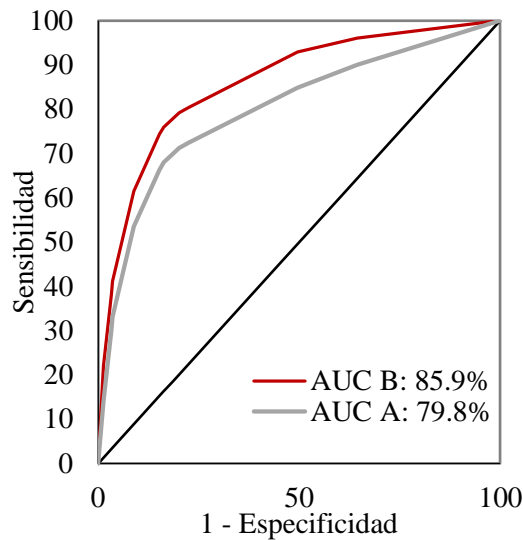
La curva ROC (Característica Operativa del Receptor¹⁹) es un método estadístico para determinar la sensibilidad ($1 - \text{especificidad}$) en un sistema clasificador binario según la variación del umbral de discriminación. La curva en mención representa el ratio de verdaderos positivos (VPR = Razón de Verdaderos Positivos) frente al ratio de falsos positivos (FPR = Razón de Falsos Positivos), considerando la variación del umbral de discriminación, valor a partir del cual decidimos que un caso es un positivo (Tseng-Chung y Li-Chiu, 2005).

Como se puede apreciar en el Gráfico III.3, el área bajo la Curva de ROC (AUCROC) “A” tiene un área mayor respecto de la Curva “B”, lo cual permite concluir que la primera curva es más efectiva al momento de clasificar, puesto que existe una probabilidad de 85.9% de que se clasificador diagnostique de manera adecuada.

¹⁹

Toma dicho nombre puesto que es una comparación de dos características operativas (VPR y FPR).

Gráfico III.3: Curva ROC



Elaboración Propia

3.3. Modelos de clasificación

3.3.1. Modelo Logit

Probablemente, la técnica más común para la predicción es la regresión logística. Se emplea en la solución de problemas de asignación de probabilidad a un evento, es este caso, probabilidad de incumplimiento, siendo la variable dependiente de naturaleza binaria (Nikolic, Zarkic-Joksimovic, Stojanovski y Joksimovic, 2013).

Para esta investigación en particular la regresión logística será de gran utilidad, puesto que se busca obtener una probabilidad que indique la probabilidad de impago de un individuo. Adicionalmente, se puede señalar que la regresión logística es una técnica de segmentación popular (Siddiqi, 2006) y se produce con frecuencia principalmente por tres razones: es una técnica conceptualmente simple, considera variables independientes de distinta naturaleza (cualitativa y cuantitativa) y proporciona resultados rápidos y robustos en comparación con las otras técnicas²⁰.

²⁰ K. Coussement, F. Van den Bossche and K. De Bock. (2014).

Este tipo de regresión toma en cuenta un conjunto de variables independientes (predictoras) y una variable dependiente con dos posibles valores, esta puede expresarse de la siguiente manera:

$$\log\left(\frac{p}{1-p}\right) = \alpha + \omega_1 x_1 + \omega_2 x_2 + \dots + \omega_n x_n$$

En donde “p” es la probabilidad a priori de que un individuo incumpla su compromiso de pago. En virtud de ello, la probabilidad “p” resulta no observable y por ende, posee una función de distribución Bernoulli que toma los valores 0 y 1, cuando paga o no paga respectivamente. Con respecto a la estimación de los parámetros, esta se rige por el método de la máxima verosimilitud para evitar relaciones no lineales entre las variables explicativas y la dependiente.

Para seleccionar el conjunto de variables explicativas que mejor se ajustan a la variable dependiente en el modelo de regresión logística se utilizó la metodología *stepwise*, la cual consiste en la combinación de los métodos *forward* y *backward*.

Al respecto, el método *forward* empieza sin ninguna variable independiente, después se añade la primera variable con mayor coeficiente de correlación respecto de la variable dependiente. De esta manera, se incorporan al modelo las variables que presenten un mayor coeficiente de correlación con la variable dependiente. Finalmente, la incorporación de variables explicativas se detiene cuando el incremento en el coeficiente de determinación es relativamente marginal.

Por otro lado, el método *backward* incluye al modelo todas las variables independientes disponibles con el objeto de ir descartando una a una de acuerdo a su coeficiente de correlación con la variable dependiente, o menor *z value*. Dicho método, finaliza con la eliminación de una variable independiente que represente un descenso demasiado pequeño en el coeficiente de determinación.

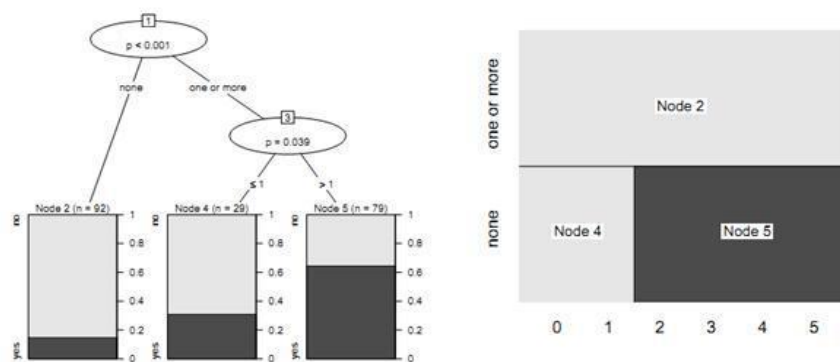
De acuerdo con lo expuesto de forma previa, la metodología *stepwise* versa en la combinación de los métodos *forward* y *backward*, es decir, se procede como el

método *forward*, sin embargo, tiene como característica relevante que es posible la expulsión de una variable explicativa siempre que la inclusión de una nueva variable refleje un comportamiento redundante de la variable expulsada

3.3.2. *Árbol de decisión*²¹

Es un modelo predictivo utilizado en diferentes ámbitos, el cual permite analizar datos de forma secuencial mediante reglas que representan y categorizan un conjunto de condiciones asociadas a la variable dependiente. Esta técnica estadística contiene múltiples nodos, los cuales representan los puntos de decisión entre las ramas, asociadas a distintas alternativas.

Gráfico III.4: Árbol de decisión



Fuente: Strobl, Malley y Tutz, 2009, pág. 326.

De acuerdo con Strobl, Malley y Tutz (2009), esta técnica no paramétrica tiene como principal característica que el espacio ocupado por todas las variables predictoras es recursivamente particionado en un conjunto de áreas rectangulares (pág. 325), tal como se observa en el Gráfico III.4, en el cual se aprecia al lado izquierdo, la representación del árbol de ejemplo, y esta se explica en la partición rectangular a la derecha.

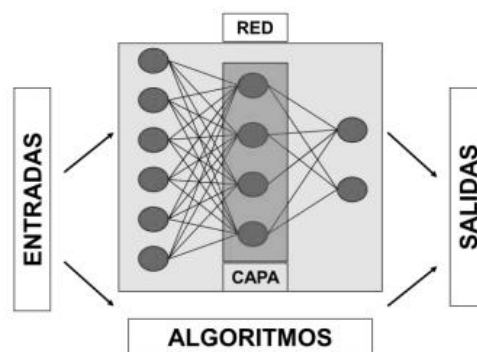
²¹ La técnica hace referencia a la forma del modelo equivalente a un árbol.

3.3.3. *Redes Neuronales*

Las redes neuronales artificiales (en adelante, RNA) es un técnica no paramétrica diseñada para simular el funcionamiento cerebral al momento de procesar información, construyendo un modelo basado en las relaciones entre los valores disponibles (Villamil Bahamón, 2013).

Las RNA pueden ser utilizadas para la predicción, clasificación y solución de problemas complejos en diversas áreas de conocimiento, en este caso para tomar conocimiento de la probabilidad de no pago de los clientes bancarizados, para lo cual simula el comportamiento de las neuronas biológicas mediante la creación de capas ocultas que se interconectan con las variables de entrada y la variable de salida (Ver Gráfico III.5).

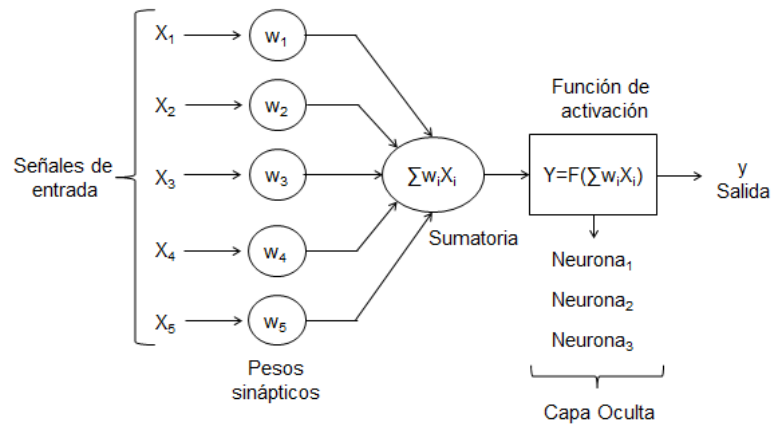
Gráfico III.5: Redes Neuronales Artificiales



Fuente: (Villamil Bahamón, 2013, pág. 64).

De manera más detallada, la RNA se encuentra compuesta principalmente por X_i (vector de variables de entrada) y w_i (vector de pesos sinápticos), los cuales se apoyan en una función $Y=F(\sum w_i X_i)$ también conocida como función de activación para la neuronas artificiales agrupadas a través de capas ocultas. Asimismo, las neuronas artificiales obtenidas en las capas ocultas junto con su vector de pesos y función de activación correspondientes derivan en la obtención del vector de resultados comparable con la variable de salida (ver Gráfico III.6).

Gráfico III.6: Arquitectura de una RNA



Elaboración Propia

Este modelo simula un proceso de aprendizaje humano mediante el reentrenamiento. No obstante, presenta desventajas asociadas al tiempo de aprendizaje, el elevado número de datos para el entrenamiento y flexibilidad para adaptarse al cambio de condiciones.

CAPITULO IV. METODOLOGÍA DE SELECCIÓN DE VARIABLES

El presente capítulo se aborda la elección de las variables con el objeto de seleccionar las variables adecuadas para el desarrollo de modelos. En ese sentido, se inició con un universo de 113 variables, de acuerdo con el apartado 2.6 del Capítulo II. Dichas variables fueron sometidas a tres filtros: (1) concentración (frecuencia), (2) indicadores *information value* (IV) e indicador *weight of evidence* (WOE) y finalmente (3) análisis de correlación bivariada.

El análisis de concentración se justifica conforme con lo expuesto por Pardo y Ruiz (2005), quienes sostienen que se debe realizar un exploración minuciosa de los datos de manera previa a la realización de cualquier inferencia estadística, toda vez que dicha exploración permite identificar problemas asociados pautas extrañas en los datos (abundante repetición de valores o ausencia de valores) o variabilidad no esperada (excesiva concentración en una determinada clase o valor).

Por otro lado, Zeng (2013) señala que el indicador *information value* (IV) es un medida ampliamente utilizada para la clasificación crediticia, cuyo aporte se fundamenta en otorgar un valor numérico que determina el poder predictivo de una variable independiente en relación a la variable dependiente binaria. Asimismo, afirma que el indicador *weight of evidence* (WOE) contribuye en la medición de la fuerza de cada atributo de la variable independiente, de acuerdo a la connotación de bueno y malo otorgada por la variable dependiente, permitiendo verificar la monotonicidad de la variables.

De otro lado, Yoo et al., (2014) indican la relevancia del análisis de correlación bivariada para detectar problemas de multicolinealidad y redundancia en las regresiones multivariadas, resaltando los efectos adversos que dicho problema acarrea en la determinación de coeficientes de las variables independientes.

En virtud de lo expuesto, la primera sección del presente capítulo aborda el análisis de concentración para verificar que las variables mantengan niveles de frecuencia apropiados en los valores, clases y valores perdidos. Posteriormente, la segunda

sección se enfoca en los indicadores de *information value* para medir el nivel predictivo de las variables y en el indicador *weight of evidence* para validar el principio de monotonicidad. Finalmente, la última sección de este capítulo, realiza un análisis de correlación bivariada para evitar posteriores problemas de multicolinealidad y redundancia.

4.1. Análisis de concentración (Frecuencia)

La primera prueba, y filtro, para las 113 variables propuestas en el apartado 2.6 del Capítulo II fue la prueba de frecuencia, tomándose como criterio de descarte, establecido por el autor del presente documento, no permitir que las variables alcancen una participación de mayor al 90% en algún determinado valor, clase o en valores perdidos. Lo expuesto, se justifica, toda vez que una variable con dichos niveles de participación está muy concentrada en ella, reduciendo su grado de variabilidad y por ende, su poder de predicción (Ver Anexo III y IV).

Como se puede apreciar en el Anexo III, 40 de las 113 variables concentran un porcentaje mayor al 90% en los valores nulos, motivo por el cual se descartan dichas variables. Asimismo, al observar el Anexo IV, se puede verificar que 25 variables muestran una frecuencia por encima del 90% en algún valor o clase.

Conforme con lo expuesto, se concluye que después del primer filtro quedan 48 variables, las cuales deberán pasar los demás filtros.

4.2. Indicadores *Information value (IV)* y *Weight of evidence (WOE)*

Como segundo filtro para las 48 variables que cumplieron con la condición de concentración (frecuencia), se calculó el índice *Information Value (IV)* para determinar el nivel predictivo de cada variable, conforme con los criterios establecidos por Siddiqi (2006), los cuales se resumen en la siguiente tabla:

Tabla IV.1: Criterios del índice IV

Intervalo	Criterio
I.V. < 0.02	La variable no tiene poder predictivo y se descarta.
0.02 < I.V < 0.1	La variable posee un poder predictivo débil.
0.1 < I.V < 0.3	La variable posee un poder predictivo medio.
I.V > 0.3	La variable posee un poder predictivo alto.

Conforme con la Tabla IV.1, se encontró que 17 variables no cuentan con poder predictivo en relación a la variable default, motivo por el cual se procedió a descartar dichas variables (Ver Anexo V).

A continuación, se pasó a graficar el WOE de las 31 variables restantes (Ver Anexo VI) a fin de verificar el cumplimiento del principio de monotonicidad. En el Anexo VII se puede apreciar las 22 variables que no cumplieron con dicho principio, a las cuales se les aplicó el principio de transformación de variables mediante el acotamiento o el reescalamiento por rangos con la utilización de la técnica árboles de decisión, toda vez que la aplicación de dicha técnica resulta óptima para el tratamiento del problema, conforme con Bijak y Thomas (2012).

Este proceso fue iterativo, es decir, se continuaba con las particiones, hasta que la gráfica del WOE muestre el cumplimiento del principio de monotonicidad. En ese sentido, es importante señalar que 3 variables²² no lograron superar el principio mencionado. En el Anexo VIII se puede visualizar las nuevas 19 variables, las cuales cumplen con el índice IV y el principio de Monotonicidad (WOE).

Por otro lado, es importante señalar que 2 de las 19 variables²³ que lograron cumplir ambos supuestos, mediante la transformación a variables categóricas, lo lograron a través de la inclusión de los valores perdidos como característica.

Finalmente, después análisis del indicador IV y WOE quedaron 28 variables, siendo 19 las variables transformadas (Ver Anexo IX).

²² Las variables eliminadas fueron: Disp_18m, Disp_24m y Disp_30m.

²³ Las variables con transformación a categóricas fueron: Disp_6m y Disp_12m.

4.3. Análisis de Correlación Bivariada

El siguiente y último filtro fue analizar la correlación bivariada de las 28 variables. Para ello se tuvo en cuenta los rangos expuestos por Anderson (2007) sobre correlación:

Tabla IV.2: Rango de correlaciones

Rango	Fuerza y dirección	
$r = +1.0$	Perfecto	Positivo
$+0.9 < r < +1.0$	Fuerte	
$+0.5 < r \leq +0.9$	Moderado	
$0.0 < r \leq 0.5$	Débil	
$r = 0$	No correlacionado	
$-0.5 \leq r < 0.0$	Débil	Negativo
$0.9 \leq r < -0.5$	Moderado	
$-1 < r < -0.9$	Fuerte	
$r = -1.0$	Perfecto	

No obstante, se consideró el criterio de Yoo et al., (2014), el cual versa en que una o más variables están correlacionadas si su correlación es mayor o igual al 70% (Ver Anexo X).

La elección de las variables que superaron los límites de correlación fue a través de los indicadores IV y Gini. Por ejemplo, las variables que observamos en la Tabla IV.3 son *condicion_6m* y *Clasificacion_6m*, las cuales poseen una correlación superior al 70%, y se optó por mantener la primera variable y descartar la otra, por tener un mayor Gini y IV.

Tabla IV.3: Variables correlacionadas

Variables	Correlación	IV	Gini	Resultado
<i>condicion_6m</i>	.835	0.152	0.182	Vence
<i>Clasificacion_6m</i>		0.112	0.106	Eliminar

Finalmente, luego de verificar el cumplimiento de los filtros de concentración, indicadores IV – WOE y la no superación de los límites de correlación quedaron 18 variables, las cuales se pueden observar en la Tabla IV.4.

Tabla IV.4: Resultado de la Metodología de la Selección de Variables

Nro.	Variable	IV	Gini	Chi-2
1	Antigüedad	0.072	0.168	2,117.64
2	Clasificacion_18m	0.056	0.139	1,015.71
3	condicion_6m	0.152	0.182	4,123.45
4	condicion_12m	0.089	0.190	1,850.87
5	condicion_18m	0.065	0.193	1,150.72
6	condicion_24m_m	0.027	0.081	185.06
7	condicion_30m_m	0.021	0.082	130.84
8	N_Displ_6m	0.038	0.221	1,131.40
9	Nro_emp_6m_m	0.023	0.177	251.94
10	Nro_ENT_1m	0.025	0.286	709.97
11	Nro_Normal_6m	0.127	0.102	4,005.85
12	Nro_Normal_12m	0.063	0.097	1,603.31
13	Nro_Normal_18m	0.051	0.095	1,111.32
14	Nro_Normal_24m	0.049	0.098	960.68
15	Nro_Normal_30m	0.056	0.101	998.25
16	NroNromales	0.155	0.251	3,928.44
17	Peor_24m	0.051	0.896	1,124.48
18	Peor_30m	0.057	0.892	1,146.76

CAPITULO V. DESARROLLO DE MODELOS DE CLASIFICACIÓN

5.1. Modelo de Regresión logística

Tomando en cuenta los resultados obtenidos en el Capítulo IV sobre la selección de variables, se procedió a realizar regresiones logísticas, aplicando la técnica *stepwise* y al mismo tiempo se procedió a eliminar las variables con signos contrarios a lo esperado, con lo cual se encontró un modelo con 9 variables explicativas significativas y con los siguientes coeficientes:

Tabla V.1 - Resultados de la Regresión Logística

Variable	Coefficient	Std. Error	z value	Pr(> z) 1/
C	-1.088	0.123	-8.868	0.000 ***
condicion_6m	0.318	0.020	15.762	0.000 ***
condicion_12m	0.193	0.016	12.284	0.000 ***
Clasificacion_18m (2)	0.233	0.032	7.386	0.000 ***
Clasificacion_18m (3)	0.335	0.051	6.516	0.000 ***
Peor_30m (1)	0.132	0.035	3.783	0.000 ***
Peor_30m (2)	0.308	0.058	5.338	0.000 ***
Peor_30m (3)	0.491	0.070	6.977	0.000 ***
Peor_30m (4)	0.620	0.098	6.345	0.000 ***
Nro_Normal_6m	-0.231	0.027	-8.451	0.000 ***
Nro_Normal_24m	-0.087	0.010	-8.363	0.000 ***
Nro_ENT_1m	0.170	0.011	15.757	0.000 ***
N_Displ_6m (2)	0.462	0.030	15.322	0.000 ***
N_Displ_6m (3)	0.758	0.026	29.502	0.000 ***
Antigüedad	-0.498	0.015	-32.257	0.000 ***

1/ Nivel de Significación: 0 (***); 0.001 (**); 0.01 (*); 0.05 (.)

Elaboración Propia

5.1.1. Interpretación de las variables independientes

- Días de atraso en los últimos 6 meses (Condicion_6m)

La variable de análisis brinda información sobre el número de días de atraso de los créditos *retail* (consumo y microempresa) otorgados por las entidades microfinancieras representativas en los últimos 6 meses. El coeficiente

Logit obtenido es de (+) 0.318, dicho signo sugiere que conforme el número de días de atraso aumente, la probabilidad de no pago.

- Días de atraso en los últimos 12 meses (Condicion_12m)

De forma similar a la variable anterior, al examinar la acumulación de días de atraso de los créditos *retail* en los últimos 12 meses se espera una relación positiva para la variable “Condicion_12m”, toda vez que el incremento de los días de atraso conlleva a un aumento de la probabilidad de no pago. En efecto, el coeficiente Logit obtenido es de (+) 0.193, lo cual corresponde a lo esperado.

De otro lado, al comparar los aportes de las variables “Condicion_6m” y “Condicion_12m” al modelo, se puede apreciar que la primera variable contiene un mayor peso respecto de la segunda, respectivamente. No obstante, ambas variables son significativas dentro del modelo.

- Clasificación por la SBS en los últimos 18 meses (Clasificacion_18m)

La variable “Clasificacion_18m” contiene la clasificación otorgada por la SBS en los últimos 18 meses. No obstante, resulta relevante señalar que la variable en cuestión es categórica y fue transformada en el “Capítulo IV” con el objeto de validar el cumplimiento del principio de monotonía.

En ese sentido, la variable “Clasificacion_18m” presenta un valor de 2 cuando las calificaciones otorgadas por la SBS son clientes con problemas potenciales (CPP) y dudoso. Asimismo, otorga un valor 3 cuando las calificaciones otorgadas por la SBS se condicionan con deficiente y pérdida.

Tomando en cuenta lo expuesto en el párrafo precedente, el coeficiente Logit estimado cuando la variable toma el valor de 2 es 0.233 y 0.335 para el valor de 3. Finalmente, resulta preciso señalar que el aporte de calificativo otorgado por la SBS como normal asociado con un valor de 0

(cumple puntualmente con el pago de sus obligaciones) es considerado dentro de la constante.

- Peor calificación en los últimos 30 meses (Peor_30m)

La variable “Peor_30m” es categórica y hace referencia a la peor calificación obtenida por un cliente en los últimos 30 meses en un crédito *retail*, tomando valores de 0 a 4, lo cuales se condicen con una calificación normal (0), clientes con problemas potenciales (1), deficiente (2), dudoso (3) y perdida (4).

De acuerdo con el párrafo previo, se espera un coeficiente positivo y creciente, toda vez que una mayor categoría es equivalente a una peor calificación en los últimos 30 meses y por ende una mayor probabilidad de no pago.

Al respecto, se puede observar en la Tabla V.1 que los coeficientes Logit son 0.132; 0.308; 0.491 y 0.620 cuando la variable asume las calificaciones de 1 a 4 respectivamente. De otro lado, resulta preciso señalar que el aporte del calificativo normal asociado con un valor de 0 (cumple puntualmente con el pago de sus obligaciones) es considerado dentro de la constante.

- Número de calificaciones normales en los últimos 6 meses (Nro_Normal_6m)

La variable “Nro_Normal_6m” mide el número de calificaciones normales alcanzadas en los últimos 6 meses. El coeficiente Logit obtenido es 0.231 y tiene signo negativo, el cual sugiere que a mayor acumulación de calificaciones normales se reduce la probabilidad de no pago, ya que la calificación normal se encuentra relacionada a un buen desempeño en el pago de las deudas, con lo cual contar con una alta frecuencia de calificaciones normales se asocia con una baja probabilidad de no pago.

- Número de calificaciones normales en los últimos 24 meses (Nro_Normal_24m)

La variable “Nro_Normal_24m” muestra el número de calificaciones normales en los últimos 24 meses. El coeficiente Logit obtenido es 0.087 y tiene signo negativo, el cual sugiere que a mayor acumulación de calificaciones normales se reduce la probabilidad de no pago, de forma similar a la variable “Nro_Normal_6m”.

- Número de entidades en el último mes (Nro_ENT_1m)

La variable “Nro_ENT_1m” proporciona información sobre el número de entidades que reportan al cliente con saldo deudor en el último mes. Dicha variable asume que a mayor número de entidades incrementa la probabilidad de no pago ya que la capacidad de pago del cliente podría verse limitada.

En efecto, el coeficiente Logit obtenido es 0.170 y posee signo positivo, lo cual valida el comportamiento esperado para la variable de análisis.

- Número de disposición de efectivo en los últimos 6 meses (N_Dis_6m)

La variable “N_Dis_6m” es categórica y fue creada a partir de la variable “Disp_6m”, la cual mide el número de disposición de efectivo en los últimos 6 meses. La transformación se realizó con el objeto de validar el principio de monotonía, de esta manera contiene en su primera categoría la agrupación de valores perdidos.

De otro lado, la segunda categoría toma en cuenta disposiciones de efectivo menor igual a 5 y la tercera categoría valores mayores a 5. En ese sentido, los coeficientes Logit estimados para la segunda y tercera categoría son de 0.462 y 0.758 respectivamente, ambas con signo positivo, lo cual verifica lo el comportamiento esperado, toda vez que el incremento del número de

disposiciones de efectivo conlleva a un aumento de la probabilidad de no pago.

- Tiempo de antigüedad en meses del cliente (Antigüedad)

La variable “Antigüedad” es una variable numérica mensual y hace referencia a la experiencia crediticia del cliente en el sistema financiero, lo cual se verifica a través del tiempo transcurrido desde la fecha en que accedió al primer crédito.

El coeficiente Logit estimado es 0.498 con signo negativo, lo cual verifica que a mayor antigüedad del cliente en el sistema financiero, menor es la probabilidad de no pago.

- Constante

Se espera un signo positivo para la constante del modelo ante la ausencia de variables categóricas, toda vez que si las entidades microfinancieras no realizaran ningún esfuerzo, gestión o análisis del comportamiento del cliente se esperaría que este no cumpliera con sus obligaciones de pago, ya no existen incentivos para un comportamiento diferente.

No obstante, la constante del modelo “C” tiene un coeficiente Logit 1.088 con signo negativo, ya que recoge la información de las siguientes variables categóricas:

- Clasificacion_18m (1), la cual contiene las calificaciones normales otorgadas por la SBS en los últimos 18 meses, la cual sugiere a reducción de la probabilidad de no pago, ya que la calificación normal es otorgada aquellos clientes que cumplen de manera puntual el pago de sus deudas.

- Peor_30m (0), contiene como peor calificación en los últimos 30 meses los calificativos normales, lo cual supone un buen comportamiento de pago por parte de cliente, por ende reduce la probabilidad de no pago.
- N_Dispos_6m (1), contiene aquellos valores perdidos del número de disposiciones de efectivo en los últimos 6 meses, lo cual se reduce la probabilidad de no pago en relación a las otras categorías de la variable que verifican un valor específico para las disposiciones de efectivo.

5.1.2. Análisis de Significancia Individual

Para determinar la significancia individual de cada una de las variables independientes del modelo de regresión logística se emplea el test Wald (test chi-cuadrado), el cual calcula los p-values observados en la última columna de la Tabla V.1 y realiza un contraste entre dos hipótesis (H_0 nula y H_1 alternativa), con la finalidad de verificar el aporte individual de cada una de las variables independientes.

$$H_0: \beta_k = 0$$

$$H_1: \beta_k \neq 0$$

La regla de decisión para rechazar la H_0 se da cuando el p-valor es menor a 0.05, lo cual se entiende como tener un coeficiente distinto de cero y por ende un variable con influencia en la determinación de la variable dependiente.

Como se puede apreciar en la Tabla V.1, el modelo logit propuesto se encuentra compuesto por variables independientes significativas con respecto de la variable default.

5.1.3. Análisis de Ajuste y Discriminancia

Al respecto, se consideraron los indicadores de Gini, Kolmogorov-Smirnov (KS), Curva de ROC para los modelos de regresión logística, árbol de decisión y redes neuronales artificiales, conforme con lo expuesto por Kočenda y Vojtek (2009) y

adicionalmente, el Criterio de Información de Akaike (AIC) para la regresión logística. Así, se eligió el modelo con mayor Gini y ROC. Asimismo, se optó por un modelo con un menor valor de AIC.

Con respecto, al estadístico KS se consideraron las reglas de evaluación para modelos de puntuación de créditos expuestas por Halim Vina Humira (2014), conforme la tabla siguiente:

Tabla V.2: Criterios KS

Valor KS	Nivel de Predicción
KS < 15%	No aceptable
15% <= KS < 20%	Deficiente
20% <= KS < 28%	Bajo
28% <= KS < 35%	Bueno
35% <= KS < 45%	Muy Bueno
KS => 45%	Excelente

Fuente: Halim y Vina Humira (2014, pàg. 20).

En la Tabla V.3, se puede observar un resumen estadístico del modelo seleccionado y desempeño en la muestra de validación:

Tabla V.3: Resumen Estadístico del Modelo Logit

Muestra	Nro Variables	AIC	AUCROC	KS	Gini	Nro Obs
Modelo	9	63984	71.04%	31.29%	42.08%	255,011
Validación	-	-	71.13%	31.81%	42.26%	64,053

De acuerdo con lo expuesto en la Tabla V.2, se puede verificar que el modelo propuesto en la Tabla V.1 presenta un buen nivel de predicción, toda vez que el área debajo de la Curva de ROC (AUCROC) es de 71% tanto en la muestra de entrenamiento como validación, lo cual indica que la regresión logística clasifica de manera adecuada el 71% de las veces a los clientes en ambas muestras. De similar forma, al observar el valor obtenido para el indicador de Gini²⁴ de 42% para ambas muestras, se puede deducir que el modelo propuesto tiene un buen nivel de predicción,

²⁴ Se obtiene de la siguiente manera: $2 \times \text{AUCROC} - 1$.

ya que el indicador en cuestión mide la eficiencia del *scorecard* al contrastar el porcentaje de clientes buenos y malos.

Con respecto al nivel de discriminancia, se debe analizar el estadístico KS, ya que este indicador determina cuan diferentes son las funciones de distribución de clientes buenos y malos. El indicador en análisis muestra un valor de 31.3% para la muestra de entrenamiento y 31.8% para la muestra de validación, con lo cual se verifica una ligera mejora al momento de validar el poder de discriminancia del modelo. Asimismo, al considerar los criterios de la Tabla V.2, se puede verificar que el modelo presenta un buen nivel de predicción, ya que supera el umbral del 30%.

Por otro lado, resulta relevante señalar que el Criterio de Información de Akaike (AIC) del modelo es el valor más bajo obtenido al momento de la elección del modelo, teniendo en cuenta los signos esperados para cada variable, la significancia individual y global del modelo.

De acuerdo con lo expuesto en los párrafos previos, se puede verificar a través de los estadísticos Curva de ROC (AUCROC), KS y Gini que el modelo de regresión logística propuesto presenta buenos niveles de ajuste y discriminancia.

5.2. Modelo de Árbol de Decisión

Para el desarrollo de la técnica de partición recursiva, se utilizaron las 48 variables que superaron el primer filtro (Concentración) del Capítulo VI, el cual descarta 65 de 113 variables, específicamente 40 por concentrar un porcentaje mayor al 90% en los valores nulos y 25 por mostrar una frecuencia por encima del 90% en algún valor o clase.

Sobre la validación del modelo, se generó una muestra de entrenamiento (80% de los datos) y se reservó una muestra de validación (20% restante). Por otro lado, se eligió una profundidad máxima del árbol de decisión de 7 por debajo del nodo raíz, toda vez que la técnica de árbol de decisión tiene como objetivo encontrar variables adicionales y relevantes para la clasificación de clientes. Adicionalmente, se consideró oportuno

contar con un número de casos mínimo de 2550 por nodo padre y 1275 para el nodo hijo, ya que dicho número para el nodo padre representa aproximadamente el 5% de la muestra de entrenamiento.

5.2.1. Interpretación de las variables independientes

El modelo de árbol de decisión resultante consta de 29 variables relevantes, 88 nodos terminales, 164 nodos totales y una profundidad de 7 (Ver Tabla V.4). Al respecto, resulta relevante señalar que dicha técnica toma en cuenta 8 variables consideradas en el modelo de regresión logística y adicionalmente 21 variables que resultan importantes para la clasificación de clientes, tales como: Edad, Sexo, Antigüedad, entre otras.

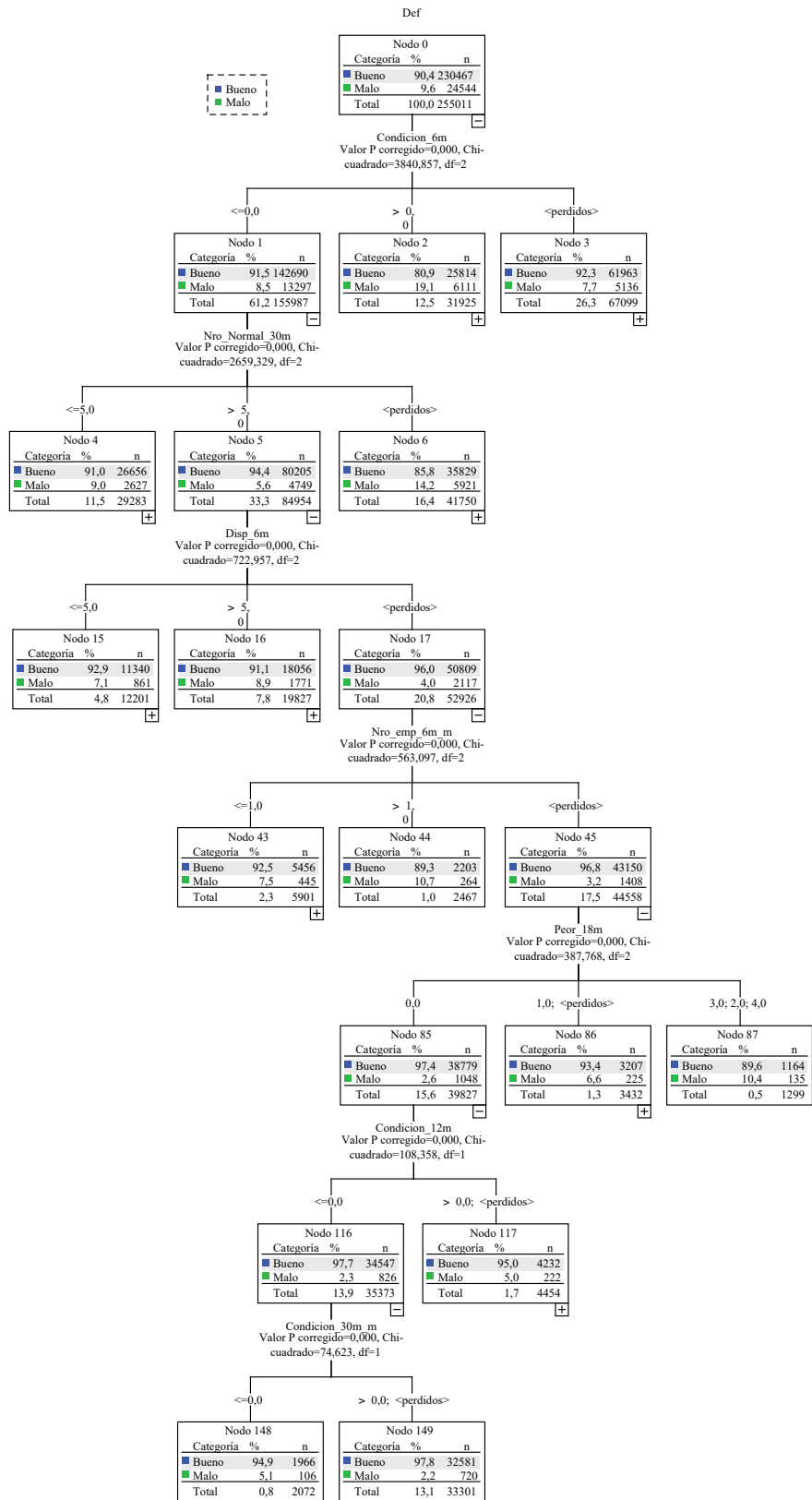
Tabla V.4: Resumen del Modelo de Árbol de Decisión

Variables independientes	1.- Condicion_6m*	11.- Nro_Normal_24m*	21.- Nro_ENT_6m
	2.- Nro_Normal_30m	12.- Peor_30m*	22.- Nro_Normal_6m*
	3.- Disp_6m	13.- N_Dis_6m*	23.- Clasificacion_6m
	4.- NroNromales	14.- Nro_emp_6m_m	24.- Peor_12m
	5.- Condicion_24m	15.- Condicion_18m*	25.- Nro_emp_24m_m
	6.- Nro_ENT_24m	16.- Peor_18m	26.- Nro_emp_12m_m
	7.- Edad	17.- Condicion_12m*	27.- Nro_Normal_12m
	8.- Clasificacion_12m	18.- Condicion_30m_m	28.- Nro_ENT_18m
	9.- Antigüedad*	19.- Sexo	29.- Nro_Normal_18m
	10.- Nro_ENT_1m*	20.- N_Dis_12m	
Número de nodos	164		
Nodos terminales	88		
Profundidad	7		

*/ Variables de la regresión logística

Sobre la interpretación, se procederá a analizar la rama principal del árbol de decisión (ver Gráfico V.1), la cual se puede replicar al resto de ramas (para mayor detalle ver Anexo XII).

Gráfico V.1: Análisis de la rama principal del árbol de decisión



En ese sentido, se puede apreciar que el nodo 0 describe la variable default, en la cual se puede observar que el 90% de los datos pertenecen a clientes clasificados como buenos, motivo por el cual se realizará el análisis sobre las variables relevantes para clasificar a un cliente como bueno. En segundo lugar, se puede observar que la variable default se ramifica en tres nodos pertenecientes a la variable Condicion_6m, la cual resulta ser la principal variable predictora.

Los nodos de la variable Condicion_6m son: nodo 1 cuando la variable toma valores de 0, nodo 2 cuando la variable toma valores mayor a 0 y nodo 3 que agrupa a los valores perdidos.

Como se puede notar, la mayor concentración de datos está en el nodo 1 con el 61.2% de los cuales el 91.5% son clasificados como buenos, motivo por el cual se prefiere el nodo 1 respecto del nodo 2 y 3.

Asimismo, se puede observar que el nodo 1 se vuelve a ramificar en tres nodos pertenecientes a la variable Nro_Normal_30m, los cuales son: nodo 4 cuando la variable toma valores menor igual a 5, nodo 5 cuando la variable toma valores mayor a 5 y nodo 6 que agrupa a los valores perdidos.

En este punto, se puede verificar que existe un 94.4% de clientes buenos en el nodo 5 frente al 91% y 85.8% de los nodos 4 y 6 respectivamente. Adicionalmente, la mayor concentración de valores se encuentra en el nodo 5 con 33.3%

En tercer lugar, se puede constatar que la tercera variable relevante dentro de la rama de análisis es Disp_6m, la cual se ramifica en el nodo 15 (valores menores iguales a 5), nodo 16 (valores mayores a 5) y nodo 17 (valores perdidos), siendo el ultimo nodo el que concentra (20.8%) la mayor cantidad valor y de clientes calificados como buenos (96%).

En cuarto lugar, se puede observar a la variable Nro_emp_6m_m, la cual de forma muy similar a la variable anterior se segmenta en tres nodos (nodo 43, 44 y 45), siendo

el nodo 45 el que concentra la mayor cantidad de valores (17.5%) y clientes clasificados como buenos (96.8%).

En quinto lugar, se puede apreciar a la variable Peor_18m, la cual se divide en tres nodos: nodo 85 cuando la variable toma valores igual a 0, nodo 86 cuando la variable toma valores de 1 o perdidos y nodo 87 cuando la variable toma valores de 2, 3 o 4.

Así, se puede percibir que cuando la variable toma valores de 0, la mayor proporción de clientes son buenos (97.4% vs 93.4% vs 89.6%). De igual manera, sucede cuando la variable Condición_12m toma valores de 0 (97.7% vs 95%). Finalmente, la última variable dentro de esta rama Condición_30m_m estable que la mayor proporción de clientes buenos se encuentran cuando estos toman valor mayores a 0 o perdidos.

A manera de resumen, del análisis de la rama principal del árbol, las variables que influyen en que un cliente sea calificado como bueno son: Condicion_6m igual a 0; Nro_Normal_30m mayor a 5; Disp_6m cuando toma valores perdidos; Nro_emp_6m_m cuando toma valores perdidos; Peor_18m igual a 0; Condicion_12m igual a 0 y Condicion_30m_m valores mayores a 0 o perdidos.

5.2.2. Análisis de Significancia Individual

Con respecto a la significancia individual, se utilizó el algoritmo de CHAID (Chi-square automatic interaction detector), el cual segmenta las variables independientes tomando como referencia interacciones mediante Chi-cuadrado, eligiendo como primera división la variable independiente de mayor interacción con la variable de respuesta (dependiente), seguidamente selecciona la segunda variable con mayor significancia, y así sucesivamente.

El proceso señalado en el párrafo precedente, se utiliza hasta obtener aquellas variables relevantes en cuanto a interacción con la variable dependiente. En ese sentido, se obtuvo un modelo con 29 variables predictoras significativas. (Ver Tabla V.4).

5.2.3. Análisis de Ajuste y Discriminancia

Al respecto, se consideraron los indicadores de Gini, Kolmogorov-Smirnov (KS) y Curva de ROC (AUCROC). Así, en la Tabla V.5, se puede observar un resumen estadístico del árbol de decisión y desempeño en la muestra de validación:

Tabla V.5: Resumen Estadístico del Modelo de Árbol de Decisión

Muestra	Nro Variables	AUCROC	KS	Gini	Nro Obs
Modelo	29	70.6%	30.0%	41.20%	255,011
Validación	-	70.2%	30.0%	40.40%	64,053

De acuerdo con lo expuesto en la Tabla V.5, se puede verificar que el modelo propuesto en la Anexo XII presenta un buen nivel de ajuste, ya que presenta una Curva de ROC superior al 70% tanto en la muestra de entrenamiento como validación, lo cual indica que el árbol de decisión clasifica de manera adecuada el 70.6% de las veces en la muestra de entrenamiento y 70.2% en la validación, por ende muestra buen nivel predictivo.

En adición al párrafo previo, al observar los valores obtenidos para los coeficientes de Gini²⁵ de 41% - 40% para las muestras de entrenamiento y validación respectivamente, se puede deducir que el modelo propuesto tiene un buen nivel discriminancia, ya que el indicador en cuestión mide el desempeño del *scorecard* al comparar el porcentaje acumulado de clientes buenos y malos.

Por otro lado, el estadístico KS -el cual mide cuan distintas son las funciones de distribución de buenos y malos como indicador de discriminancia- alcanza un valor del 30% para ambas muestras (entrenamiento y validación), motivo por el cual se considera que el modelo desarrollado refleja un buen nivel de predictivo, ya que dicho valor maximiza la diferencia entre ambas distribuciones.

²⁵ Se obtiene de la siguiente manera: $2 \times \text{AUCROC} - 1$.

En conclusión y de acuerdo con lo expuesto, se puede verificar a través de los estadísticos Curva de ROC (AUCROC), KS y Gini que el modelo de árbol de decisión propuesto presenta buenos niveles de ajuste y discriminancia.

5.3. Modelo de Redes Neuronales

5.3.1. *Parámetros e interpretación del modelo*

Para el desarrollo de las RNA, se tuvo en cuenta las 18²⁶ variables que superaron los tres filtros (Concentración, IV-WOE y Correlación) del Capítulo VI, descartando 65 variables en el filtro de concentración, 20 variables en el filtro de IV-WOE y 10 en el último filtro.

Con respecto al modelo de RNA, se utilizó una red neuronal artificial denominada Perceptrón Multicapa (MLP), la cual forma múltiples capas, dándole la capacidad al modelo de resolver problemas no lineales. Asimismo, el modelo utilizado en este documento tiene en consideración un proceso de aprendizaje de tipo supervisado, el cual hace referencia al tipo de entrenamiento mediante el cual el sistema neuronal cuenta con información en las entradas y la salida esperada.

Por otro lado, el algoritmo utilizado en el modelo propuesto es de corrección de error (propagación hacia atrás), también conocido como regla delta, el cual hace referencia a la optimización a través de la minimización de la función de error de los pesos sinápticos del sistema neuronal.

Adicionalmente, resulta relevante señalar que el modelo propuesto contiene dos capas ocultas, las cuales se rigen por una función de activación de tangente hiperbólica, la cual tiene la siguiente forma:

$$\gamma(c) = \tanh(c) = (e^c - e^{-c}) / (e^c + e^{-c})$$

²⁶ La variable Nro_emp_6m_m fue transformada en categórica con la finalidad de considerar los valores perdidos dentro del análisis.

Dicha función permite un aprendizaje más rápido en términos del número de iteraciones requerido²⁷. Asimismo, se considera una función de activación SoftMax para la capa de salida, de la siguiente forma:

$$\gamma(c_k) = \exp(c_k) / \sum \exp(c_j)$$

Dicha función se condice con la forma dicotómica de la variable dependiente default, ya que dicha función representa una generalización de la función logística.

De otro lado, se consideró una muestra de entrenamiento con el 80% de los datos y una muestra de validación con el 20% restante. Así, se obtiene como resultado el modelo del Anexo XIII, el cual se encuentra compuesto por dos variables cualitativas y seis cuantitativas de acuerdo con su aporte individual y global dentro del modelo. De esta manera, se puede apreciar las diversas interacciones entre las variables de entrada, derivando en la generación de ocho y seis neuronas para la primera y segunda capa respectivamente.

Así, el modelo de RNA Perceptrón Multicapa del Anexo XIII estima como pesos óptimos para la minimización de la función de error los valores expuestos en el Anexo XIV, los cuales se obtienen para las capas ocultas y la de salida.

5.3.2. Análisis de importancia de las variables independientes

Se realiza un análisis de sensibilidad para determinar la fuerza de cada variable independiente dentro de modelo de RNA, el cual se basa en las muestras de entrenamiento y validación combinadas (ver Tabla V.6).

En ese sentido, se puede apreciar que la variable más significativa es NroNromales con un aporte del 18.3% al modelo, seguida de condición_6m con 15.6%. Al respecto resulta relevante, señalar que la tercera variable de mayor contribución es

²⁷ Mejía Sánchez J. Arturo (2004). Sistema de detección de intrusos en redes de comunicaciones utilizando redes neuronales.

Nro_emp_6m_m_1, la cual fue transformada durante el proceso a categórica con la finalidad de considerar dentro del análisis los valores perdidos.

Por otro lado, resulta importante señalar que el modelo propuesto comparte cinco variables independientes con el modelo de regresión logística. No obstante, el modelo de RNA es el modelo con menor número de variables en relación a los dos modelos desarrollados de manera previa.

Tabla V.6: Análisis de sensibilidad de las variables dependientes

Variabes	Importancia
NroNromales	18.3%
condicion_6m*	15.6%
Nro_emp_6m_m_1	13.8%
N_Dis_6m*	13.8%
Antigüedad*	12.1%
Nro_ENT_1m*	10.3%
condicion_18m	9.3%
Nro_Normal_6m*	7.0%
Total	100%

*/ Variables de la regresión logística

5.3.3. Análisis de Ajuste y Discriminancia

Al respecto, se consideraron los indicadores de Gini, Kolmogorov-Smirnov (KS) y Curva de ROC (AUCROC) para verificar el ajuste del modelo RNA Perceptrón Multicapa. Así, en la Tabla V.7, se puede observar un resumen estadístico del modelo para la muestra de entrenamiento (80%) y validación (20%):

Tabla V.7: Resumen Estadístico del Modelo de Redes Neuronales

Muestra	Nro Variables	AUCROC	KS	Gini	Nro Obs
Modelo	8	71.85%	32.37%	43.70%	255,577
Validación	-	72.40%	34.24%	44.80%	63,487

Así, el modelo propuesto en la Anexo XIII, considerando los pesos sinápticos del Anexo XIV, presenta un nivel de predicción bueno, ya que presenta una Curva de ROC de 71.9% para la muestra de entrenamiento, lo cual indica que el modelo clasifica de

manera adecuada el 71.9% de las veces. Adicionalmente, el mismo estadístico para la muestra de validación es 72.4%, lo cual presenta una ligera mejora y valida el valor obtenido mediante la muestra de entrenamiento.

Adicionalmente, al observar los valores obtenidos para los coeficientes de Gini²⁸ de 43.7% - 44.8% para las muestras de entrenamiento y validación respectivamente, se puede deducir que el modelo RNA tiene un buen nivel de predicción, toda vez que el indicador en cuestión mide el desempeño al momento de comparar el porcentaje de clientes buenos y malos.

De otro lado, se puede observar que el estadístico KS, el cual determina si la muestra de buenos y malos difieren significativamente, es de 32.4% y 34.2% para la muestra de entrenamiento y validación respectivamente. Sobre los valores hallados para el estadístico KS, se puede concluir que un valor por encima de 30% muestra un buen poder de discriminación entre las distribuciones de probabilidad de clientes.

En conclusión, se puede verificar mediante los estadísticos Curva de ROC (AUCROC), KS y Gini que el modelo RNA Perceptrón Multicapa propuesto presenta buenos niveles de ajuste y discriminancia.

5.4. Comparación de los modelos desarrollados

Como se puede apreciar en la Tabla V.8, el modelo de RNA utiliza 8 variables, mientras que los modelos de regresión logística y el árbol de decisión consideran 9 y 29 respectivamente. También se puede verificar que hay variables que son tomadas en cuenta por los tres modelos, tales como: “Antigüedad”, “Condicion_6m”, “N_Displ_6m”, “Nro_Ent_1m” y “Nro_Normal_6m”. En ese sentido, se puede deducir la superior relevancia de dichas variables respecto del resto.

Adicionalmente, se puede ver que el modelo de RNA y árbol de decisión comparten las variables: “Condición_18m”, “Nro_emp_6m_m” y “NroNromales”; las cuales no considera la regresión logística. De igual manera, se puede observar que la regresión

²⁸ Se obtiene de la siguiente manera: $2 \times \text{AUCROC} - 1$.

logística y el árbol de decisión comparten tres variables adicionales: “Condicion_12m”, “Peor_30m” y “Nro_Normal_24m”.

Tabla V.8: Variables independientes de cada modelo

V. Independientes Redes Neuronales	V. Independientes Regresión Logit	V. Independientes Árbol de Decisión		
Antigüedad	Antigüedad	Antigüedad	Nro_Normal_24m	Nro_Normal_30m
Condicion_6m	Condicion_6m	Condicion_6m	Clasificacion_12m	Clasificacion_6m
N_Displ_6m	N_Displ_6m	N_Displ_6m	Sexo	Peor_12m
Nro_ENT_1m	Nro_ENT_1m	Nro_ENT_1m	Edad	Nro_emp_24m_m
Nro_Normal_6m	Nro_Normal_6m	Nro_Normal_6m	Nro_ENT_24m	Nro_emp_12m_m
Condicion_18m	Condicion_12m	Condicion_18m	Nro_ENT_6m	Nro_Normal_12m
Nro_emp_6m_m_1	Peor_30m	Nro_emp_6m_m	Peor_18m	Nro_ENT_18m
NroNromales	Nro_Normal_24m	NroNromales	Condicion_14m	Nro_Normal_18m
	Clasificacion_18m	Condicion_12m	Condicion_30m_m	Disp_6m
		Peor_30m	N_Displ_12m	

Por otro lado, en la Tabla V.9 se puede contrastar los niveles de ajuste y discriminancia de las muestras de entrenamiento y validación de cada modelo. Así, se puede establecer la superioridad del modelo RNA en los indicadores AUCROC, KS y Gini) y en los dos tipos de muestra. Además, resulta relevante tener en consideración que el modelo RNA se encuentra construido con menos variables independientes que los otros dos modelos, por lo cual se presenta como una mejor solución.

Tabla V.9: Indicadores de nivel de ajuste y discriminancia de cada modelo

	Muestra	Nro Variables	AUCROC	KS	Gini	Nro Obs
Regresión logística	Entrenamiento	9	71.04%	31.29%	42.08%	255,011
	Validación		71.13%	31.81%	42.26%	64,053
Árbol de Decisión	Entrenamiento	29	70.60%	30.00%	41.20%	255,011
	Validación		70.20%	30.00%	40.40%	64,053
RNA	Entrenamiento	8	71.85%	32.37%	43.70%	255,577
	Validación		72.40%	34.24%	44.80%	63,487

En conclusión, al observar los niveles de ajuste presentados y el número de variables de cada modelo se determina la superioridad de modelo RNA frente a los modelos de regresión logística y árbol de decisión.

CAPITULO VI. CONCLUSIONES

De la presente investigación enmarcada en el desarrollo de tres técnicas de clasificación de deudores de créditos “*retail*” se concluye que es el modelo de redes neuronales de perceptrón multicapa el que mejor discrimina y se ajusta a los datos. Asimismo, se puede inferir que las ocho variables relevantes de acuerdo con el modelo de redes neuronales para determinar la probabilidad de incumplimiento en el pago de un crédito “*retail*” son: NroNromales, condicion_6m, Nro_emp_6m_m, N_Dis_6m, Antigüedad, Nro_ENT_1m, condicion_18m y Nro_Normal_6m.

Sobre las variables relevantes para el modelo RNA, se puede agregar que ninguna de ellas pertenece al conjunto de variables demográficas. En efecto, las ocho variables encontradas son intrínsecas al historial crediticio de los clientes, las cuales se encuentran relacionadas al número total de calificaciones normales obtenidas durante el periodo de análisis (NroNromales), los días de atraso en los últimos seis meses (condicion_6m), número de créditos microempresa en los últimos seis meses (Nro_emp_6m_m), número de disposiciones de efectivo en los últimos seis meses (N_Dis_6m), número de meses dentro del sistema financiero (Antigüedad), número de entidades financieras que reportan saldo deudor en el último mes (Nro_ENT_1m), días de atraso en los últimos dieciocho meses (condicion_18m) y el número de calificaciones normales en los últimos seis meses (Nro_Normal_6m).

De acuerdo con el párrafo anterior, se puede concluir que las variables asociadas al historial crediticio de los clientes poseen poder predictivo relevante en relación a las variables demográficas al momento de determinar la probabilidad de pago de los clientes y clasificación de los mismos.

Adicionalmente, resulta fundamental señalar que las ocho variables utilizadas dentro del modelo RNA pasaron por un proceso de transformación con el objeto de mejorar su poder de predicción, siendo dos de ellas (N_Dis_6m y Nro_emp_6m_m) transformadas en variables categóricas, toda vez que a través de dicho proceso se incorporaron los valores perdidos dentro de una categoría. El resto de variables fueron

transformadas sobre el agrupamiento de los valores expuestos por cada variable original.

Con respecto al nivel de ajuste del modelo RNA, se puede verificar que dicho modelo presenta una Curva de ROC de 71.9% correspondiente con la muestra de entrenamiento (255 mil registros aproximadamente), siendo dicho indicador superior a los obtenidos en el modelo de regresión logística (71%) y en el árbol de decisión (70%). En ese sentido, se puede deducir que el modelo RNA clasifica a los clientes de manera adecuada el 71.9% de las veces, siendo el modelo con mejor nivel predictivo.

Los resultados obtenidos con la muestra de entrenamiento se pueden garantizar a través de la clasificación de la muestra de validación (64 mil registros), la cual tiene como objeto de verificar si el nivel de ajuste obtenido para cada modelo es consistente al momento de clasificar a clientes diferentes a los utilizados por la muestra de entrenamiento. En efecto, se ha constatado la superioridad del modelo RNA (72.4%) respecto del modelo de regresión logística (71.1%) y árbol de decisión (70.2%) para clasificar de forma correcta a los clientes.

De igual forma, el nivel de ajuste se puede corroborar a través del estadístico de Kolmogorov-Smirnov (KS), el cual muestra la máxima diferencia entre las distribuciones de clientes buenos y malos, siendo mejor el modelo que contenga un KS mayor y superior al 30%. En ese sentido, se pudo observar que los tres modelos desarrollados cuentan con un estadístico KS mayor al 30%, lo cual permite señalar el buen nivel predictivo de cada uno. No obstante, el modelo RNA muestra valores de 32.4% para la muestra de entrenamiento y 34.2% para la muestra de validación, mientras que los modelos de regresión logística y árbol de decisión muestran valores inferiores. En ese sentido, se puede concluir que el modelo RNA se ajusta mejor para la clasificación de clientes.

Por otro lado, teniendo en cuenta el principio parsimonia mediante el cual se establece que la solución más simple es mejor, resulta importante señalar que el modelo RNA considera relevante 8 variables, mientras que los modelos de regresión logística y árbol de decisión estiman necesarias 9 y 29 variables respectivamente, motivo por el

cual se deduce que el modelo RNA resulta mejor en relación a los otros dos modelos. No obstante, se debe tener en cuenta que los modelos de regresión logística y árbol de decisión cumplen con los criterios de nivel de ajuste y discriminancia, por lo cual soportar una decisión de clasificación de clientes considerando la complementariedad de dichos modelos resulta optima, ya que dichos modelos consideran otras variables de historial crediticio y demográficas que podrían resultar relevantes para mayor alcance, tales como: la peor calificación en los últimos 30 meses, el sexo y edad de cliente.

Sobre las ventajas y desventajas de los modelos desarrollados en el presente documento, se puede señalar que el modelo de regresión logit es una herramienta tradicional para la modelización no lineal y flexible en cuanto a la naturaleza de las variables explicativas, ya que estas pueden ser escalares o categóricas. Asimismo, la regresión logística no realiza supuestos sobre el tipo de distribución de las variables independientes. Por otro lado, la estimación de los parámetros resulta compleja en comparación a modelos lineales y en presencia de multicolinealidad en las variables independientes, se podrían obtener coeficientes de estimación sesgados.

Con respecto a la técnica de árbol de decisión, se puede señalar que es una técnica no paramétrica basada en la interacción de los datos, la cual resulta robusta ante la presencia de datos atípicos. Asimismo, dicha técnica brinda soluciones de fácil interpretación e independientemente de la naturaleza de la variable dependiente (escalar, ordinal o nominal). Sin embargo, la técnica en cuestión presenta desventajas, las cuales se encuentran asociadas a la elección del árbol optimo, ausencia de una función para las variables y la presencia de una gran cantidad de registros que garanticen la robustez de la técnica.

Finalmente, el modelo de RNA es una técnica de inteligencia artificial que presenta una capacidad de autoaprendizaje adaptativo a través del reentrenamiento y de auto-organización que permite responder apropiadamente a los datos. Asimismo, las RNA muestran como característica relevante la tolerancia de fallos, ya que aprende a reconocer los patrones con ruidos y distorsionados. Por otro lado, el modelo de RNA presente desventajas en cuanto a la determinación adecuada de neuronas ocultas, ya

que una incorrecta asignación de neuronas podría generar un mal desempeño de la red. Además, la inclusión de variables de entrada irrelevantes en el modelo conlleva a la reducción de la capacidad de generalización del modelo, lo cual deriva en la reducción del nivel predictivo del modelo.

BIBLIOGRAFÍA

- Addae-Korankye, A. (2014). Causes and Control of Loan Default/Delinquency in Microfinance Institutions in Ghana. *American International Journal of Contemporary Research* , 36-45.
- Agapito Mesta, G. (2010). Determinantes del crédito de consumo por tipo de institución financiera en el Perú. *Repositorio Institucional PIRHUA – Universidad de Piura*, 1-118.
- Anderson, R. (2007). *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. Oxford: Oxford University Press.
- Arana Pacheco , R., Núñez Porras , J., Ore Mego, J., & Velazco Infante, D. (2017). Planeamiento Estratégico del Sistema de Cajas Rurales de Ahorro y Crédito en el Perú. *Pontificia Universidad Católica del Perú (PUCP)*, 1-138.
- Bijak, K., & Thomas, L. C. (2012). *Does segmentation always improve model performance in credit scoring?* Southampton: School of Management, University of Southampton.
- Corporación Andina de Fomento. (2013). *Emprendimientos en América Latina: Desde la subsistencia hacia la transformación productiva*. Bogotá, Colombia.: CAF.
- Coussement, K., Van den Bossche, F., & De Bock, K. (2014). Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees. *Journal of Business Research*, 2751–2758.
- Halim, S., & Vina Humira, Y. (2014). Credit Scoring Modeling. *Jurnal Teknik Industri*, 17-24.
- Kočenda, E., & Vojtek , M. (2009). *Default Predictors and Credit Scoring Models for Retail Banking*. Munich: CESifo Working Paper No. 2862.
- Ledgerwood, J. (1998). *Microfinance handbook: an institutional and financial perspective*. Washington, DC: World Bank.
- Nikolic, N., Zarkic-Joksimovic, N., Stojanovski, D., & Joksimovic, I. (2013). The application of brute force logistic regression to corporate credit scoring models: Evidence from Serbian financial statements. *Expert Systems with Applications*, 5932–5944.
- OCDE/CEPAL/CAF. (2016). *Perspectivas económicas de América Latina 2017: Juventud, competencias y emprendimiento*. Paris: OECD Publishing.
- Pardo Merino, A., & Ruiz Díaz, M. (2005). *Análisis de datos con SPSS 13 Base*. Madrid: McGraw-Hil Interamericana de España S.L.

- Peussa, A. (2016). Credit risk scorecard estimation by logistic regression. *University of Helsinki, Faculty of Science, Department of Mathematics and Statistics*, 1-33.
- Rayo Cantón, S., Lara Rubio, J., & Camino Blasco, D. (2010). A Credit Scoring Model for Institutions of Microfinance under the Basel II Normative. *Journal of Economics, Finance and Administrative Science*, 89-124.
- Siddiqi, N. (2006). *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. JohnWiley & Sons.
- 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*, 323-348.
- Trivelli, C., Portocarrero, F., Byrne, G., Torero, M., Field, E., Aguilar, G., y otros. (2004). *Mercado y gestión del microcrédito en el Perú*. Lima: IEP. Consorcio de Investigación Económica y Social .
- Tseng-Chung, T., & Li-Chiu, C. (2005). Predicting multilateral trade credit risks: comparisons of Logit and Fuzzy Logic models using ROC curve analysis. *Expert Systems with Applications*, 547-556.
- Villamil Bahamón, R. (2013). *Modelo Predictivo Neuronal para la Evaluación del Riesgo Crediticio*. Bogotá: Universidad Nacional de Colombia.
- Westley, G., & Shaffer, S. (1997). *Credit Union Policies and Performance in Latin America*. Washington, DC: Working paper #355, Office of the Chief Economist, Inter-American Development Bank.
- Yoo, W., Mayberry, R., Bae, S., Singh, K., Qinghua, P., & Lillard Jr., J. (2014). A Study of Effects of MultiCollinearity in the Multivariable Analysis. *International Journal of Applied Science and Technology*, 9-19.
- Zeng, G. (2013). Metric Divergence Measures and Information Value in Credit Scoring. *Journal of Mathematics*, 1-10.

Anexo I: Clasificación regulatoria de créditos Resolución SBS N° 11356 - 2010

Tipo de crédito	Nivel de ventas	Endeudamiento en el SF 1/
Créditos Corporativos	Mayor a S/. 200 millones en los dos (2) últimos años	-
Créditos a Grandes Empresas	Mayor a S/. 20 millones en los dos (2) últimos años	-
Créditos a Medianas Empresas	Menor a S/. 20 millones en los dos (2) últimos años	Mayor a S/. 300 mil en los últimos seis (6) meses 2/
Créditos a Pequeñas Empresas	-	Mayor a S/. 20 mil en los últimos seis (6) meses
Créditos a Microempresas	-	Menor a S/. 20 mil en los últimos seis (6) meses
Créditos de Consumo Resolvente 3/	-	-
Créditos de Consumo No Resolvente 3/	-	-
Créditos de Hipotecarios para Vivienda	-	-

Nota: en el caso de los créditos de consumo (revolvente y no revolvente) e hipotecarios para vivienda no toman en cuenta parámetros sobre el nivel de ventas y endeudamiento (ver 3/), puesto que se encuentran asociados al objeto del crédito, como por ejemplo: la atención de pagos de bienes, servicios o gastos no relacionados con actividad empresarial, adquisición de viviendas, entre otros.

1/ Sistema Financiero

2/ También se considera el endeudamiento en créditos categorizados como pequeñas empresas y microempresas.

3/ Solo si el cliente tiene un crédito pequeña empresa o microempresa y su endeudamiento es mayor a S/. 300.000 por seis meses consecutivos, se reclasifica a mediana empresa.

Anexo II: Definiciones relevantes

Nombre	Definición
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.
Saldo deudor	Es la deuda activa que tiene un cliente con alguna entidad del sistema financiero.
Entidades con saldo	Refiere al número de entidades que reportan al cliente con saldo deudor.
Saldo total	Es la sumatoria del saldo deudor del individuo
Calificativos de riesgo	Es la clasificación que determina la SBS de acuerdo a los días de atraso registrados por el cliente en cada cuenta de las entidades donde se mantenga saldo deudor.
Calificativo missing	El cliente tendrá un calificativo missing cuando no registre saldo deudor en el sistema financiero.
Calificativo normal	Registra el valor 0 si el individuo no presenta ninguna clasificación peor a normal en más del 5% de su saldo total en el sistema financiero.
Calificativo CPP	Registra el valor 1 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.
Calificativo deficiente	Registra el valor 2 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.
Calificativo dudoso	Registra el valor 3, si el individuo presenta clasificación dudoso 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.
Calificativo pérdida	Registra el valor 4, si el individuo presenta clasificación pérdida en más del 5% de su saldo total del sistema financiero.

Anexo III: Variables con una concentración mayor al 90% en valores perdidos

Variable	Válido	Perdidos	% Perdidos
DE_Porc_1m	-	319,064	100%
DE_Porc_6m	257	318,807	100%
DE_Porc_12m	248	318,816	100%
DE_Porc_18m	326	318,738	100%
DE_Porc_24m	303	318,761	100%
LU_porc_1m	-	319,064	100%
LU_porc_6m	3,289	315,775	99%
LU_porc_12m	3,362	315,702	99%
LU_porc_18m	3,532	315,532	99%
LU_porc_24m	3,132	315,932	99%
Monto_max_Dip_6m	20,976	298,088	93%
Monto_max_Dip_12m	19,806	299,258	94%
Monto_max_Dip_18m	17,881	301,183	94%
Monto_max_Dip_24m	15,912	303,152	95%
Monto_max_Dip_30m	14,587	304,477	95%
Nro_cast_6m	2,629	316,435	99%
Nro_cast_12m	3,671	315,393	99%
Nro_cast_18m	5,119	313,945	98%
Nro_cast_24m	6,563	312,501	98%
Nro_cast_30m	10,892	308,172	97%
Nro_jud_6m	43	319,021	100%
Nro_jud_12m	95	318,969	100%
Nro_jud_18m	183	318,881	100%
Nro_jud_24m	207	318,857	100%
Nro_jud_30m	270	318,794	100%
Nro_ref_6m	1,029	318,035	100%
Nro_ref_12m	1,437	317,627	100%
Nro_ref_18m	1,762	317,302	99%
Nro_ref_24m	1,967	317,097	99%
Nro_ref_30m	2,243	316,821	99%
Nro_venc_6m	8,733	310,331	97%
Nro_venc_12m	10,532	308,532	97%
Nro_venc_18m	11,715	307,349	96%
Nro_venc_24m	12,350	306,714	96%
Nro_venc_30m	12,784	306,280	96%
Tiene_pignoraticio_6m	3,972	315,092	99%
Tiene_pignoraticio_1m	4,739	314,325	99%
Tiene_pignoraticio_12m	3,724	315,340	99%
Tiene_pignoraticio_18m	3,546	315,518	99%
Tiene_pignoraticio_24m	3,648	315,416	99%

Anexo IV: Variables con una concentración mayor al 90% en algún valor o clase

Clasificacion_6m_m	Frec	% válido
0	96,734	96.2
1	2,833	2.8
2	535	.5
3	349	.3
4	120	.1
Sub-Total	100,571	100.0
Perdidos	218,493	
Total	319,064	

Clasificacion_12m_m	Frec	% válido
0	80,556	95.4
1	2,615	3.1
2	548	.6
3	368	.4
4	310	.4
Sub-Total	84,397	100.0
Perdidos	234,667	
Total	319,064	

Clasificacion_18m_m	Frec	% válido
0	71,323	94.6
1	2,596	3.4
2	636	.8
3	387	.5
4	476	.6
Sub-Total	75,418	100.0
Perdidos	243,646	
Total	319,064	

Clasificacion_24m_m	Frec	% válido
0	64,645	93.5
1	2,710	3.9
2	740	1.1
3	498	.7
4	525	.8
Sub-Total	69,118	100.0
Perdidos	249,946	
Total	319,064	

Clasificacion_30m_m	Frec	% válido
0	59,733	92.7
1	2,766	4.3
2	773	1.2
3	559	.9
4	605	.9
Sub-Total	64,436	100.0
Perdidos	254,628	
Total	319,064	

condicion_6m_m	Frec	% válido
0	90,676	90.2
1-750	9,886	9.8
750-1500	7	.0
1500-3000	-	.0
3000-4600	8,898	.0
Sub-Total	100,571	100.0
Perdidos	218,493	
Total	319,064	

Peor_6m	Frec	% válido
0	286,147	89.8
1	22,888	7.2
2	5,082	1.6
3	3,015	.9
4	1,445	.5
Sub-Total	318,577	100.0
Perdidos	487	
Total	319,064	

Tiene_DE_1m	Frec	% válido
1	63,936	100.0
Perdidos	255,128	
Total	319,064	

Tiene_DE_6m	Frec	% válido
1	54,777	100.0
Perdidos	264,287	
Total	319,064	

condicion_12m_m	Frec	% válido
0	76,345	90
750-1500	8,030	10
1500-7080	22	0
Sub-Total	84,397	100
Perdidos	234,667	
Total	319,064	

condicion_18m_m	Frec	% válido
0	67,638	90
750-1500	7,741	10
1500-7000	39	0
Sub-Total	75,418	100
Perdidos	243,646	
Total	319,064	

num_hijos	Frec	% válido
0	287,405	90
1	17,806	6
2	10,443	3
3	2,760	1
4	650	0
Total	319,064	100

Tiene_prestamo_24m	Frec	% válido
1	97,096	100.0
Perdidos	221,968	
Total	319,064	

Tiene_DE_18m	Frec	% válido
1	42,136	100.0
Perdidos	276,928	
Total	319,064	

Tiene_DE_24m	Frec	% válido
1	37,898	100.0
Perdidos	281,166	
Total	319,064	

Tiene_hip_1m	Frec	% válido
0	306,181	96.0
1	12,883	4.04
Total	319,064	100

Tiene_DE_12m	Frec	% válido
1	50,033	100.0
Perdidos	269,031	
Total	319,064	

Tiene_hip_6m	Frec	% válido
0	306,784	96.2
1	12,280	3.85
Total	319,064	100

Tiene_hip_12m	Frec	% válido
0	306,932	96.2
1	12,132	3.80
Total	319,064	100

Tiene_hip_18m	Frec	% válido
0	307,398	96.3
1	11,666	3.66
Total	319,064	100

Tiene_hip_24m	Frec	% válido
0	307,710	96.4
1	11,354	3.56
Total	319,064	100

Tiene_prestamo_1m	Frec	% válido
1	148,721	100.0
Perdidos	170,343	
Total	319,064	

Tiene_prestamo_6m	Frec	% válido
1	122,224	100.0
Perdidos	196,840	
Total	319,064	

Tiene_prestamo_12m	Frec	% válido
1	112,035	100.0
Perdidos	207,029	
Total	319,064	

Tiene_prestamo_18m	Frec	% válido
1	102,985	100.0
Perdidos	216,079	
Total	319,064	

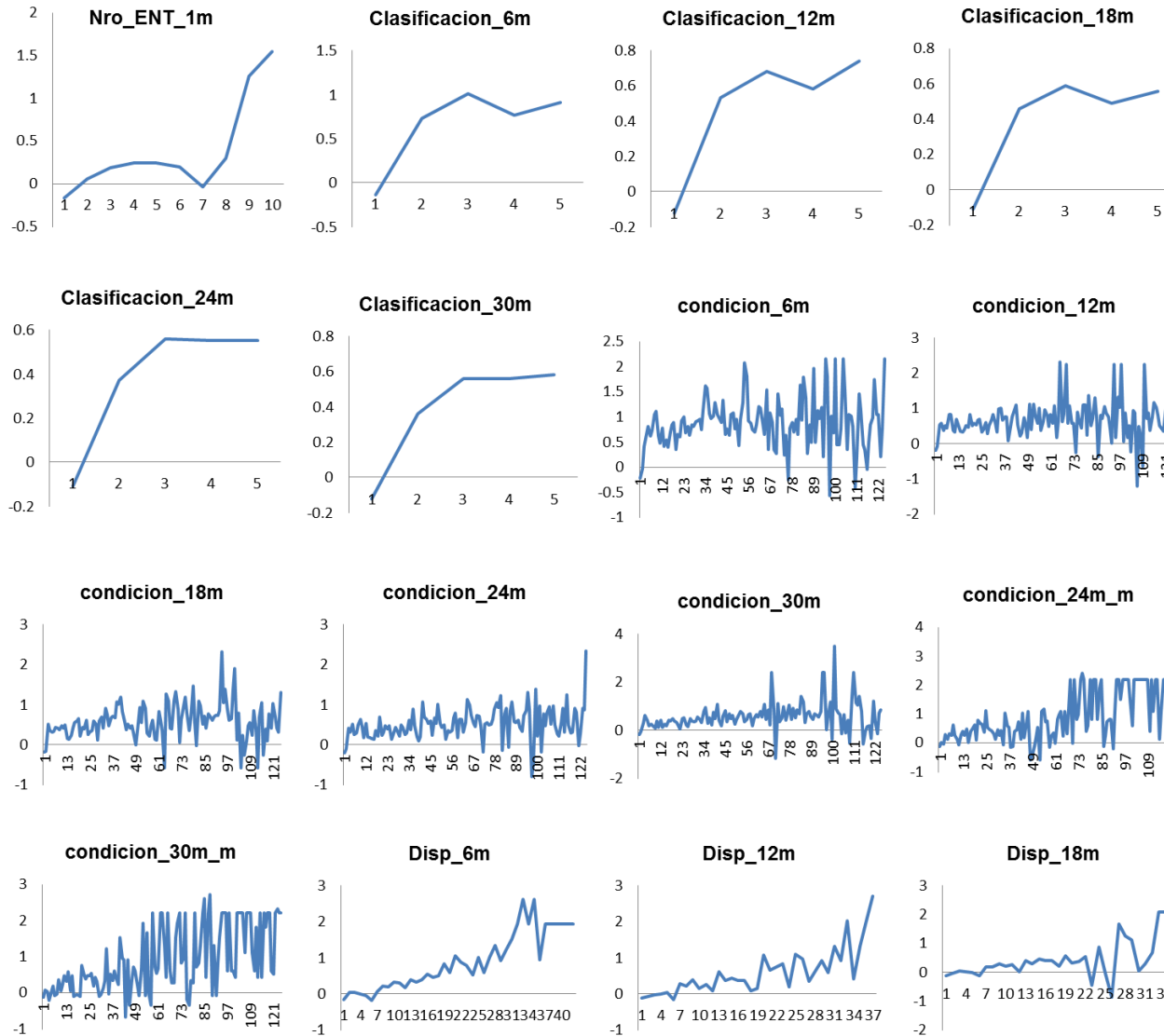
Anexo V: Índice Information Value

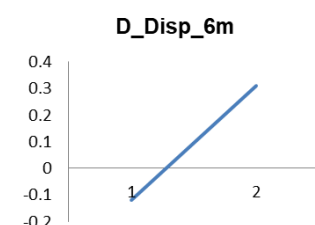
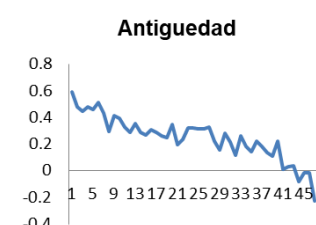
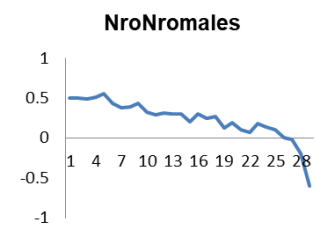
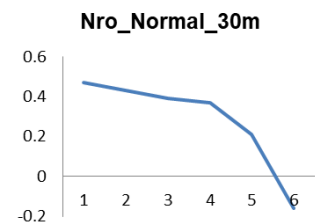
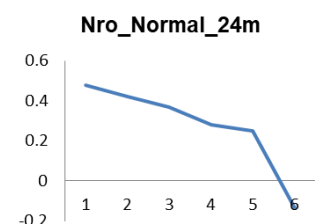
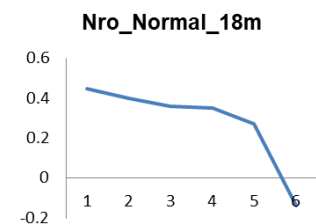
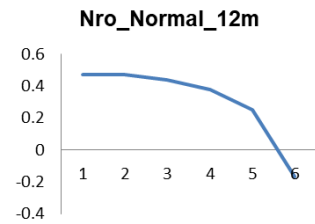
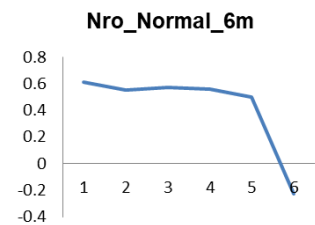
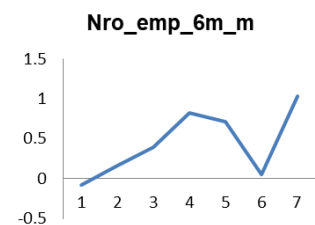
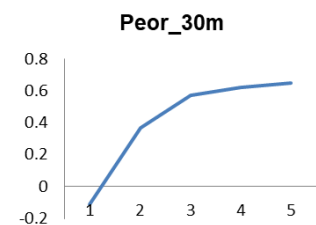
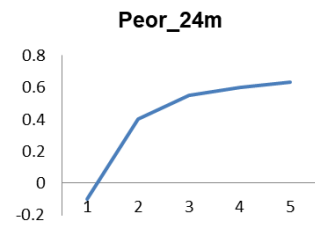
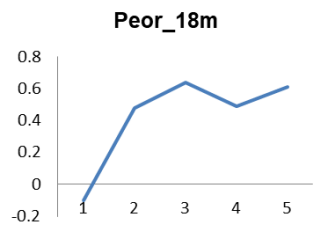
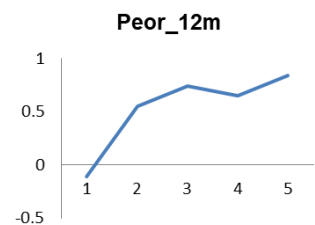
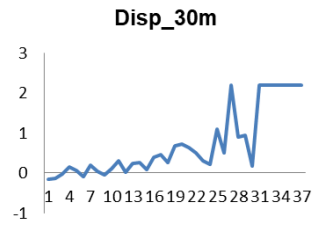
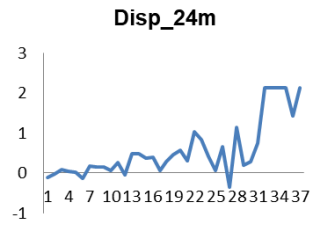
Nro	Variable	IV
1	Flg_hijos	0.000
2	Flg_autos	0.000
3	Num_autos	0.000
4	Sexo	0.000
5	Lima	0.000
6	Estado Civil	0.000
7	Nro_ENT_18m	0.001
8	Departamento	0.001
9	Edad	0.002
10	Nro_ENT_24m	0.002
11	Nro_ENT_12m	0.002
12	Nro_emp_24m_m	0.004
13	Nro_ENT_6m	0.005
14	Nro_emp_18m_m	0.007
15	Nro_emp_30m_m	0.009
16	Nro_emp_12m_m	0.011

Nro	Variable	IV
17	Provincia	0.013
18	Nro_emp_6m_m	0.025
19	Nro_ENT_1m	0.026
20	Disp_24m	0.026
21	Disp_18m	0.028
22	Disp_30m	0.030
23	D_Dis_6m	0.037
24	Disp_12m	0.039
25	Disp_6m	0.045
26	Nro_Normal_24m	0.049
27	Nro_Normal_18m	0.051
28	Peor_24m	0.051
29	Clasificacion_24m	0.052
30	Peor_18m	0.054
31	Clasificacion_18m	0.056
32	Nro_Normal_30m	0.056

Nro	Variable	IV
33	Clasificacion_30m	0.057
34	Peor_30m	0.057
35	Nro_Normal_12m	0.063
36	Clasificacion_12m	0.069
37	Peor_12m	0.070
38	Antiguedad	0.074
39	Clasificacion_6m	0.113
40	Nro_Normal_6m	0.127
41	condicion_12m	0.155
42	NroNromales	0.156
43	condicion_18m	0.158
44	condicion_24m_m	0.162
45	condicion_30m	0.183
46	condicion_24m	0.185
47	condicion_6m	0.196
48	condicion_30m_m	0.197

Anexo VI: Weight of Evidence (WOE)





Anexo VII: Principio de Monotonicidad

A. Variables que cumplen

Nro.	Variable
1	Nro_ENT_1m
2	Clasificacion_6m
3	Clasificacion_12m
4	Clasificacion_18m
5	condicion_6m
6	condicion_12m
7	condicion_18m
8	condicion_24m
9	condicion_30m
11	condicion_30m_m
12	Disp_6m
13	Disp_12m
14	Disp_18m
15	Disp_24m
16	Disp_30m
17	Peor_12m
18	Peor_18m
19	Nro_emp_6m_m
20	Nro_Normal_6m
21	NroNromales
22	Antiguedad

B. Variables que no cumplen

Nro.	Variable
1	Clasificacion_24m
2	Clasificacion_30m
3	Peor_24m
4	Peor_30m
5	Nro_Normal_12m
6	Nro_Normal_24m
7	Nro_Normal_18m
8	Nro_Normal_30m
9	D_Displ_6m

Anexo VIII: Transformación de variables

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
Nro_ENT_1m	1	<= 1,0	141326	12810	154136	0.083	0.417	0.491	0.164	0.012	
	2	(1,0, 2,0]	75149	8544	83693	0.102	0.278	0.261	- 0.063	0.001	
	3	(2,0, 3,0]	40304	5182	45486	0.114	0.169	0.140	- 0.186	0.005	
	4	> 3,0	31006	4188	35194	0.119	0.136	0.108	- 0.235	0.007	
	Total			287785	30724	318509					

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
Clasificacion_6m	1	,0	189983	19036	209019	0.091	0.782	0.901	0.142	0.017	
	2	1,0; 3,0	16881	4079	20960	0.195	0.168	0.080	- 0.739	0.065	
	3	2,0; 4,0	3930	1217	5147	0.236	0.050	0.019	- 0.987	0.031	
	Total			210794	24332	235126	0.103				

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
Clasificacion_12m	1	,0	165514	15050	180564	0.083	0.774	0.872	0.120	0.012	
	2	1,0; 3,0	18861	3301	22162	0.149	0.170	0.099	- 0.535	0.038	
	3	2,0; 4,0	5345	1098	6443	0.170	0.056	0.028	- 0.695	0.020	
	Total			189720	19449	209169	0.093				

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
Clasificacion_18m	1	,0	150304	12916	163220	0.079	0.767	0.858	0.112	0.010	
	2	1,0; 3,0	18481	2830	21311	0.133	0.168	0.106	- 0.465	0.029	
	3	2,0; 4,0	6356	1093	7449	0.147	0.065	0.036	- 0.581	0.017	
	Total			175141	16839	191980	0.088				

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
condicion_6m	1	≤ 0	178363	16690	195053	0.086	0.686	0.846	0.210	0.034	
	2	$(0, 10,0]$	13642	2742	16384	0.167	0.113	0.065	- 0.555	0.027	
	3	$> 10,0$	18789	4900	23689	0.207	0.201	0.089	- 0.815	0.091	
	Total		210794	24332	235126	0.103				0.152	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
condicion_12m	1	≤ 0	155512	13520	169032	0.080	0.695	0.820	0.165	0.021	
	2	$(0, 15,0]$	16703	2578	19281	0.134	0.133	0.088	- 0.409	0.018	
	3	$> 15,0$	17505	3351	20856	0.161	0.172	0.092	- 0.625	0.050	
	Total		189720	19449	209169	0.093				0.089	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
condicion_18m	1	≤ 0	140755	11729	152484	0.077	0.697	0.804	0.143	0.015	
	2	$(0, 16,0]$	18318	2449	20767	0.118	0.145	0.105	- 0.330	0.013	
	3	$> 16,0$	16068	2661	18729	0.142	0.158	0.092	- 0.544	0.036	
	Total		175141	16839	191980	0.088				0.065	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
condicion_24m	1	$\leq 1,0$	132618	10488	143106	0.073	0.699	0.808	0.144	0.016	
	2	$(1,0, 20,0]$	15839	2108	17947	0.117	0.141	0.097	- 0.376	0.017	
	3	$> 20,0$	15669	2399	18068	0.133	0.160	0.095	- 0.516	0.033	
	Total		164126	14995	179121	0.084				0.066	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
condicion_30m	1	≤ 0	120289	9108	129397	0.070	0.670	0.775	0.146	0.015	
	2	$(0, 1,0]$	4999	442	5441	0.081	0.033	0.032	- 0.009	0.000	
	3	$(1,0, 22,0]$	15234	1862	17096	0.109	0.137	0.098	- 0.333	0.013	
	4	$> 22,0$	14653	2180	16833	0.130	0.160	0.094	- 0.530	0.035	
	Total		155175	13592	168767	0.081				0.063	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
condicion_24m_m	1	$\leq 2,0$	56698	5626	62324	0.090	0.854	0.907	0.060	0.003	
	2	$> 2,0$	5834	960	6794	0.141	0.146	0.093	- 0.446	0.023	
	Total		62532	6586	69118	0.095				0.027	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
condicion_30m_m	1	$\leq 3,0$	52872	5071	57943	0.088	0.857	0.904	0.053	0.003	
	2	$> 3,0$	5644	849	6493	0.131	0.143	0.096	- 0.397	0.019	
	Total		58516	5920	64436	0.092				0.021	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
Peor_18m	1	$\leq ,0$	207680	17621	225301	0.078	0.789	0.875	0.103	0.009	
	2	$> ,0$	29708	4710	34418	0.137	0.211	0.125	- 0.522	0.045	
	Total		237388	22331	259719	0.086				0.054	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
NroNromales	1	$\leq 5,0$	24439	4281	28720	0.149	0.144	0.086	- 0.511	0.030	
	2	(5,0, 11,0]	28590	4374	32964	0.133	0.147	0.101	- 0.376	0.017	
	3	(11,0, 17,0]	30055	4166	34221	0.122	0.140	0.106	- 0.277	0.009	
	4	(17,0, 25,0]	54555	6621	61176	0.108	0.223	0.193	- 0.145	0.004	
	5	(25,0, 27,0]	25305	2641	27946	0.095	0.089	0.089	0.006	0.000	
	6	(27,0, 28,0]	25968	2243	28211	0.080	0.075	0.092	0.196	0.003	
	7	$> 28,0$	94049	5394	99443	0.054	0.181	0.332	0.605	0.091	
	Total		282961	29720	312681	0.095				0.155	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
Antigüedad	1	$\leq 9,0$	26691	4529	31220	0.145	0.147	0.093	- 0.465	0.025	
	2	(9,0, 34,0]	55604	7807	63411	0.123	0.254	0.193	- 0.275	0.017	
	3	(34,0, 46,0]	30882	3557	34439	0.103	0.116	0.107	- 0.077	0.001	
	4	$> 46,0$	175152	14842	189994	0.078	0.483	0.607	0.229	0.029	
	Total		288329	30735	319064	0.096				0.072	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
Nro_Normal_6m	1	$\leq 3,0$	26514	5015	31529	0.159	0.164	0.092	- 0.575	0.041	
	2	(3,0, 5,0]	37809	6799	44608	0.152	0.222	0.131	- 0.524	0.047	
	3	$> 5,0$	223222	18805	242027	0.078	0.614	0.776	0.234	0.038	
	Total		287545	30619	318164					0.127	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
Peor_12m	1	$\leq ,0$	228525	20222	248747	0.081	0.793	0.888	0.113	0.011	
	2	$> ,0$	28873	5274	34147	0.154	0.207	0.112	- 0.612	0.058	
	Total		257398	25496	282894					0.069	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
Nro_emp_6m_m	1	$\leq 1,0$	69220	7608	76828	0.099	0.710	0.770	0.081	0.005	
	2	(1,0, 2,0]	16440	2296	18736	0.123	0.214	0.183	- 0.159	0.005	
	3	$> 2,0$	4201	806	5007	0.161	0.075	0.047	- 0.476	0.014	
	Total		89861	10710	100571	0.106				0.023	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
N_Displ_6m	1	Null	219640	20748	240388	0.086	0.675	0.762	0.121	0.010	
	2	≤ 5	29304	4133	33437	0.124	0.134	0.102	- 0.280	0.009	
	3	> 5	39385	5854	45239	0.129	0.190	0.137	- 0.332	0.018	
	Total		288329	30735	319064	0.096				0.038	

Variable	Nodo	Regla	Bueno	Malo	Total	R. Malos	D. de malos	D. de buenos	WOE	IV	Graf. WOE
N_Displ_12m	1	Null	225969	22473	248442	0.090	0.731	0.784	0.069	0.004	
	2	≤ 6	48828	5917	54745	0.108	0.193	0.169	- 0.128	0.003	
	3	> 6	13532	2345	15877	0.148	0.076	0.047	- 0.486	0.014	
	Total		288329	30735	319064	0.096				0.021	

Anexo IX: Lista de variables después del filtro IV y WOE

Nro.	Variable	Transformada	IV
1	Antigüedad	Si	7.16%
2	Clasificacion_6m	Si	11.25%
3	Clasificacion_12m	Si	6.91%
4	Clasificacion_18m	Si	5.60%
5	Clasificacion_24m	No	5.23%
6	Clasificacion_30m	No	5.66%
7	condicion_6m	Si	15.17%
8	condicion_12m	Si	8.87%
9	condicion_18m	Si	6.48%
10	condicion_24m	Si	6.56%
11	condicion_30m	Si	6.32%
12	condicion_24m_m	Si	2.65%
13	condicion_30m_m	Si	2.11%
14	D_Displ_6m	No	3.74%
15	N_Displ_12m	Si	2.09%
16	N_Displ_6m	Si	3.76%
17	Nro_emp_6m_m	Si	2.34%
18	Nro_ENT_1m	Si	2.52%
19	Nro_Normal_6m	Si	12.66%
20	Nro_Normal_12m	No	6.34%
21	Nro_Normal_18m	No	5.10%
22	Nro_Normal_24m	No	4.89%
23	Nro_Normal_30m	No	5.62%
24	NroNromales	Si	15.51%
25	Peor_12m	Si	6.86%
26	Peor_18m	Si	5.36%
27	Peor_24m	No	5.13%
28	Peor_30m	No	5.66%

Anexo X: Correlación de variables

	Antiguedad	Clasificación_6m	Clasificación_12m	Clasificación_18m	Clasificación_24m	Clasificación_30m	condicion_6m	condicion_12m	condicion_18m	condicion_24m	condicion_30m	condicion_24m_m	condicion_30m_m	D_Displacement_6m	N_Displacement_6m
Antiguedad	1	.052	.051	.048	.067	.085	.060	.062	.055	.060	.092	.064	.090	.208	.215
Clasificación_6m	.052	1	.363	.206	.121	.096	.835	.355	.213	.150	.132	.038	.040	.073	.075
Clasificación_12m	.051	.363	1	.391	.233	.157	.313	.832	.375	.246	.182	.087	.061	.025	.022
Clasificación_18m	.048	.206	.391	1	.445	.265	.183	.332	.826	.417	.270	.149	.111	.000	-.009
Clasificación_24m	.067	.121	.233	.445	1	.556	.099	.176	.343	.767	.442	.253	.216	-.027	-.034
Clasificación_30m	.085	.096	.157	.265	.556	1	.079	.117	.188	.337	.734	.176	.255	-.021	-.028
condicion_6m	.060	.835	.313	.183	.099	.079	1	.382	.242	.163	.153	.040	.045	.098	.099
condicion_12m	.062	.355	.832	.332	.176	.117	.382	1	.402	.247	.197	.069	.055	.050	.047
condicion_18m	.055	.213	.375	.826	.343	.188	.242	.402	1	.423	.281	.126	.095	.027	.017
condicion_24m	.060	.150	.246	.417	.767	.337	.163	.247	.423	1	.423	.193	.140	.010	.002
condicion_30m	.092	.132	.182	.270	.442	.734	.153	.197	.281	.423	1	.147	.191	.025	.017
condicion_24m_m	.064	.038	.087	.149	.253	.176	.040	.069	.126	.193	.147	1	.415	-.013	-.016
condicion_30m_m	.090	.040	.061	.111	.216	.255	.045	.055	.095	.140	.191	.415	1	-.008	-.011
D_Displacement_6m	.208	.073	.025	.000	-.027	-.021	.098	.050	.027	.010	.025	-.013	-.008	1	.940
N_Displacement_6m	.215	.075	.022	-.009	-.034	-.028	.099	.047	.017	.002	.017	-.016	-.011	.940	1
N_Displacement_12m	.228	.100	.069	.003	-.043	-.042	.119	.096	.039	.004	.017	-.023	-.019	.742	.800
Nro_emp_6m_m	.165	-.033	-.040	-.042	-.041	-.035	-.029	-.032	-.034	-.036	-.024	.046	.030	.005	.001
Nro_ENT_1m	.343	.077	.008	-.024	-.063	-.060	.115	.054	.025	.010	.025	-.032	-.032	.423	.428
Nro_Normal_6m	.271	-.516	-.332	-.212	-.163	-.139	-.443	-.313	-.207	-.168	-.145	-.116	-.099	.101	.103
Nro_Normal_12m	.289	-.175	-.469	-.304	-.197	-.151	-.143	-.390	-.286	-.199	-.153	-.119	-.095	.110	.114
Nro_Normal_18m	.206	-.098	-.199	-.502	-.317	-.194	-.087	-.167	-.420	-.300	-.191	-.175	-.117	.092	.100
Nro_Normal_24m	.226	-.067	-.120	-.238	-.546	-.301	-.060	-.099	-.195	-.449	-.289	-.184	-.192	.070	.079
Nro_Normal_30m	.285	-.057	-.086	-.132	-.267	-.543	-.055	-.072	-.111	-.200	-.426	-.069	-.188	.060	.070
NroNromales	.565	-.084	-.182	-.289	-.395	-.416	-.054	-.128	-.221	-.314	-.336	-.203	-.259	.232	.249
Peor_12m	.080	.354	.907	.357	.215	.149	.313	.830	.341	.232	.174	.116	.079	.071	.066
Peor_18m	.076	.207	.381	.904	.394	.244	.190	.330	.816	.380	.252	.244	.150	.050	.040
Peor_24m	.092	.117	.226	.437	.977	.543	.097	.172	.337	.748	.431	.490	.328	.013	.005
Peor_30m	.108	.093	.152	.259	.547	.977	.080	.116	.185	.332	.717	.268	.523	.017	.009

	N_Displ_1 2m	DoP	Nro_emp_6m_m	Nro_ENT_1m	Nro_Norm al_6m	Nro_Norm al_12m	Nro_Norm al_18m	Nro_Norm al_24m	Nro_Norm al_30m	NroNromales	Peor_12m	Peor_18m	Peor_24m	Peor_30m
Antiguedad	.228	.209	.165	.343	.271	.289	.206	.226	.285	.565	.080	.076	.092	.108
Clasificacion_6m	.100	.128	-.033	.077	-.516	-.175	-.098	-.067	-.057	-.084	.354	.207	.117	.093
Clasificacion_12m	.069	.211	-.040	.008	-.332	-.469	-.199	-.120	-.086	-.182	.907	.381	.226	.152
Clasificacion_18m	.003	.288	-.042	-.024	-.212	-.304	-.502	-.238	-.132	-.289	.357	.904	.437	.259
Clasificacion_24m	-.043	.464	-.041	-.063	-.163	-.197	-.317	-.546	-.267	-.395	.215	.394	.977	.547
Clasificacion_30m	-.042	.510	-.035	-.060	-.139	-.151	-.194	-.301	-.543	-.416	.149	.244	.543	.977
condicion_6m	.119	.102	-.029	.115	-.443	-.143	-.087	-.060	-.055	-.054	.313	.190	.097	.080
condicion_12m	.096	.171	-.032	.054	-.313	-.390	-.167	-.099	-.072	-.128	.830	.330	.172	.116
condicion_18m	.039	.235	-.034	.025	-.207	-.286	-.420	-.195	-.111	-.221	.341	.816	.337	.185
condicion_24m	.004	.290	-.036	.010	-.168	-.199	-.300	-.449	-.200	-.314	.232	.380	.748	.332
condicion_30m	.017	.318	-.024	.025	-.145	-.153	-.191	-.289	-.426	-.336	.174	.252	.431	.717
condicion_24m_m	-.023	.207	.046	-.032	-.116	-.119	-.175	-.184	-.069	-.203	.116	.244	.490	.268
condicion_30m_m	-.019	.252	.030	-.032	-.099	-.095	-.117	-.192	-.188	-.259	.079	.150	.328	.523
D_Displ_6m	.742	.022	.005	.423	.101	.110	.092	.070	.060	.232	.071	.050	.013	.017
N_Displ_6m	.800	.012	.001	.428	.103	.114	.100	.079	.070	.249	.066	.040	.005	.009
N_Displ_12m	1	-.006	.003	.447	.091	.107	.104	.093	.081	.269	.106	.054	-.004	-.005
Nro_emp_6m_m	.003	.001	1	.466	.157	.140	.100	.085	.070	.235	.000	-.020	-.027	-.027
Nro_ENT_1m	.447	-.040	.466	1	.194	.210	.176	.154	.134	.400	.056	.028	-.021	-.020
Nro_Normal_6m	.091	-.123	.157	.194	1	.256	.186	.152	.134	.437	-.315	-.205	-.152	-.130
Nro_Normal_12m	.107	-.143	.140	.210	.256	1	.262	.188	.143	.507	-.376	-.277	-.175	-.135
Nro_Normal_18m	.104	-.177	.100	.176	.186	.262	1	.260	.183	.527	-.151	-.411	-.285	-.175
Nro_Normal_24m	.093	-.205	.085	.154	.152	.188	.260	1	.247	.532	-.097	-.189	-.465	-.279
Nro_Normal_30m	.081	-.206	.070	.134	.134	.143	.183	.247	1	.532	-.067	-.106	-.229	-.454
NroNromales	.269	-.228	.235	.400	.437	.507	.527	.532	.532	1	-.139	-.237	-.336	-.360
Peor_12m	.106	.201	.000	.056	-.315	-.376	-.151	-.097	-.067	-.139	1	.385	.223	.157
Peor_18m	.054	.272	-.020	.028	-.205	-.277	-.411	-.189	-.106	-.237	.385	1	.412	.256
Peor_24m	-.004	.470	-.027	-.021	-.152	-.175	-.285	-.465	-.229	-.336	.223	.412	1	.567
Peor_30m	-.005	.519	-.027	-.020	-.130	-.135	-.175	-.279	-.454	-.360	.157	.256	.567	1

Anexo XI: Resultados de la correlación de variables

Variables	Correlación	IV	Gini	Resultado
condicion_6m	.835	0.152	0.182	Vence
Clasificacion_6m		0.112	0.106	Eliminar

Variables	Correlación	IV	Gini	Resultado
condicion_12m	.832	0.089	0.190	Vence
Clasificacion_12m		0.069	0.127	Se va

Variables	Correlación	IV	Gini	Resultado
condicion_18m	.826	0.065	0.193	Se va
Clasificacion_18m		0.056	0.139	Vence ²⁹

Variables	Correlación	IV	Gini	Resultado
condicion_24m	.767	0.066	0.193	Se va
Clasificacion_24m		0.052	0.883	Vence

Variables	Correlación	IV	Gini	Resultado
condicion_30m	.734	0.063	0.280	Se va
Clasificacion_30m		0.057	0.880	Vence

Variables	Correlación	IV	Gini	Resultado
Peor_12m	.907	0.069	0.095	Se va
Clasificacion_12m		0.069	0.127	Vence

Variables	Correlación	IV	Gini	Resultado
Peor_18m	.904	0.054	0.102	Se va
Clasificacion_18m		0.056	0.139	Vence

Variables	Correlación	IV	Gini	Resultado
Peor_24m	.977	0.051	0.896	Vence
Clasificacion_24m		0.052	0.883	Se va

²⁹ Se eligió Clasificacion_18m ya que es la que mejor GINI tiene luego de la Clasificacion_24m y Clasificacion_30m que fueron descartadas al tener fuerte correlación con Peor_24m y Peor_30m.

Variables	Correlación	IV	Gini	Resultado
Peor_30m	.977	0.057	0.892	Vence
Clasificacion_30m		0.057	0.880	Se va

Variables	Correlación	IV	Gini	Resultado
Peor_12m	.830	0.069	0.095	Se va
condicion_12m		0.089	0.190	Vence

Variables	Correlación	IV	Gini	Resultado
Peor_18m	.816	0.054	0.102	Se va
condicion_18m		0.065	0.193	Vence

Variables	Correlación	IV	Gini	Resultado
Peor_24m	.748	0.051	0.896	Vence
condicion_24m		0.066	0.193	Se va

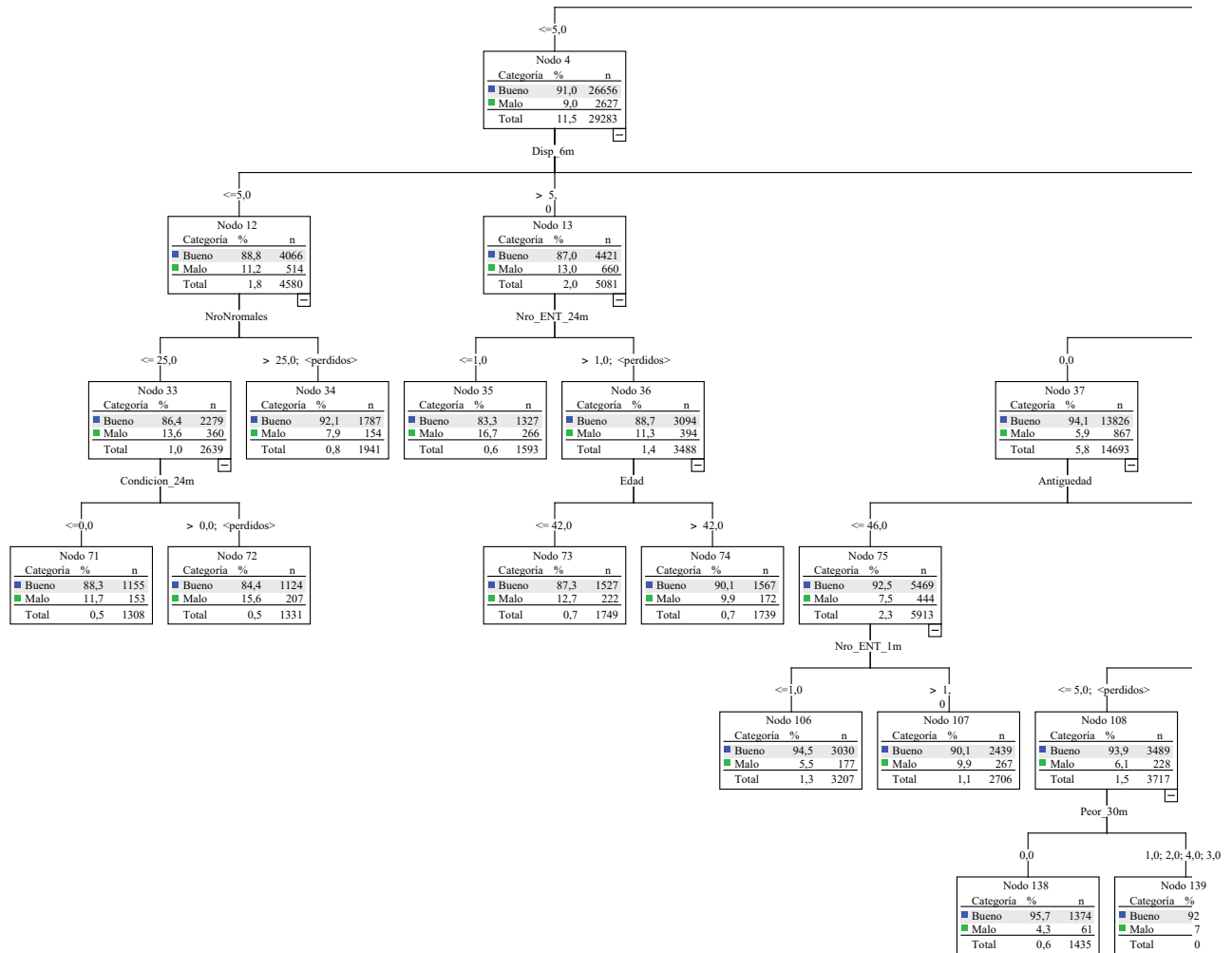
Variables	Correlación	IV	Gini	Resultado
Peor_30m	.717	0.057	0.892	Vence
condicion_30m		0.063	0.280	Se va

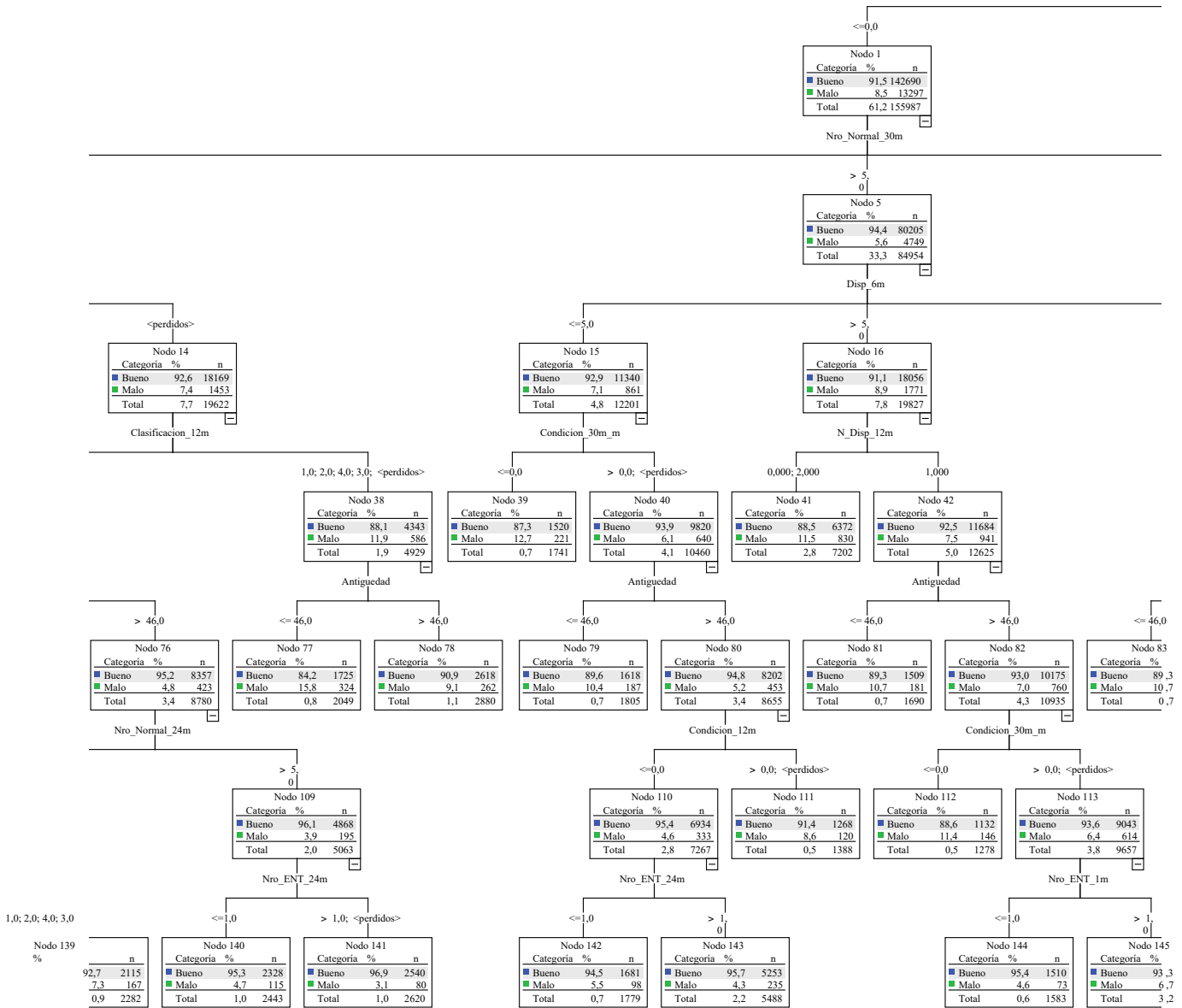
Variables	Correlación	IV	Gini	Resultado
N_Displ_6m	.940	0.038	0.221	Vence
D_Displ_6m		0.037	0.221	Se va

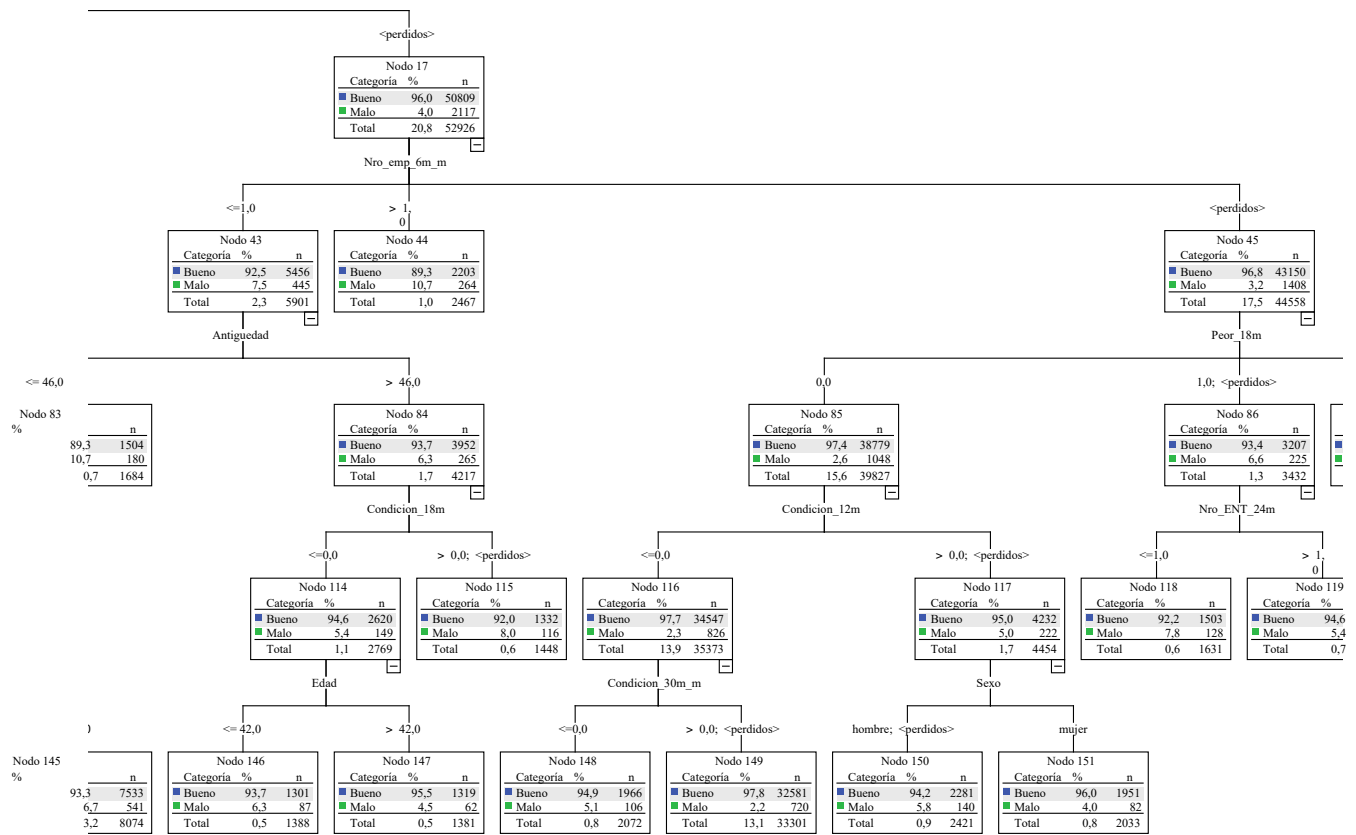
Variables	Correlación	IV	Gini	Resultado
N_Displ_12m	.742	0.021	0.173	Se va
D_Displ_6m		0.037	0.221	Vence

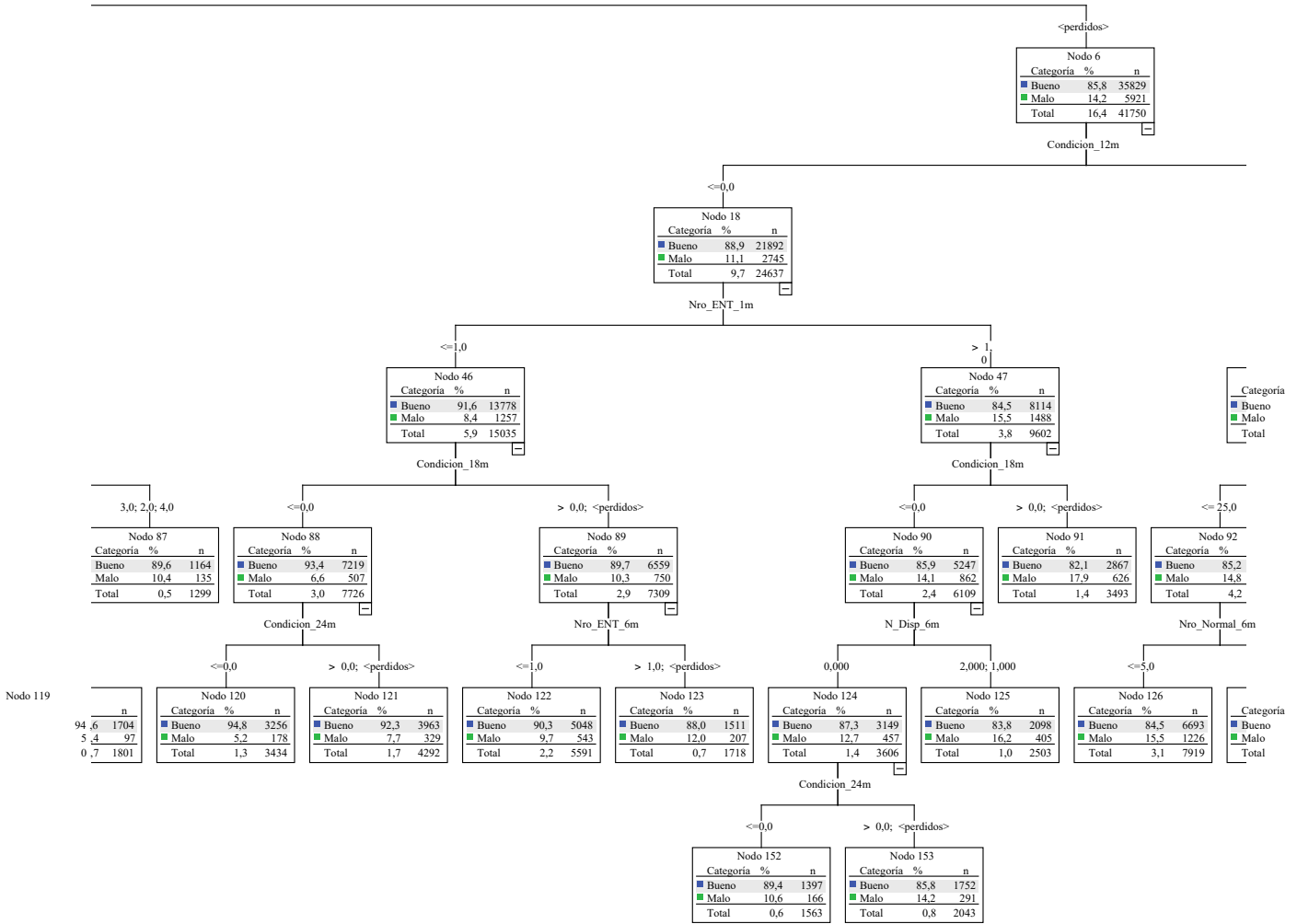
Variables	Correlación	IV	Gini	Resultado
N_Displ_12m	.800	0.021	0.173	Se va
N_Displ_6m		0.038	0.221	Vence

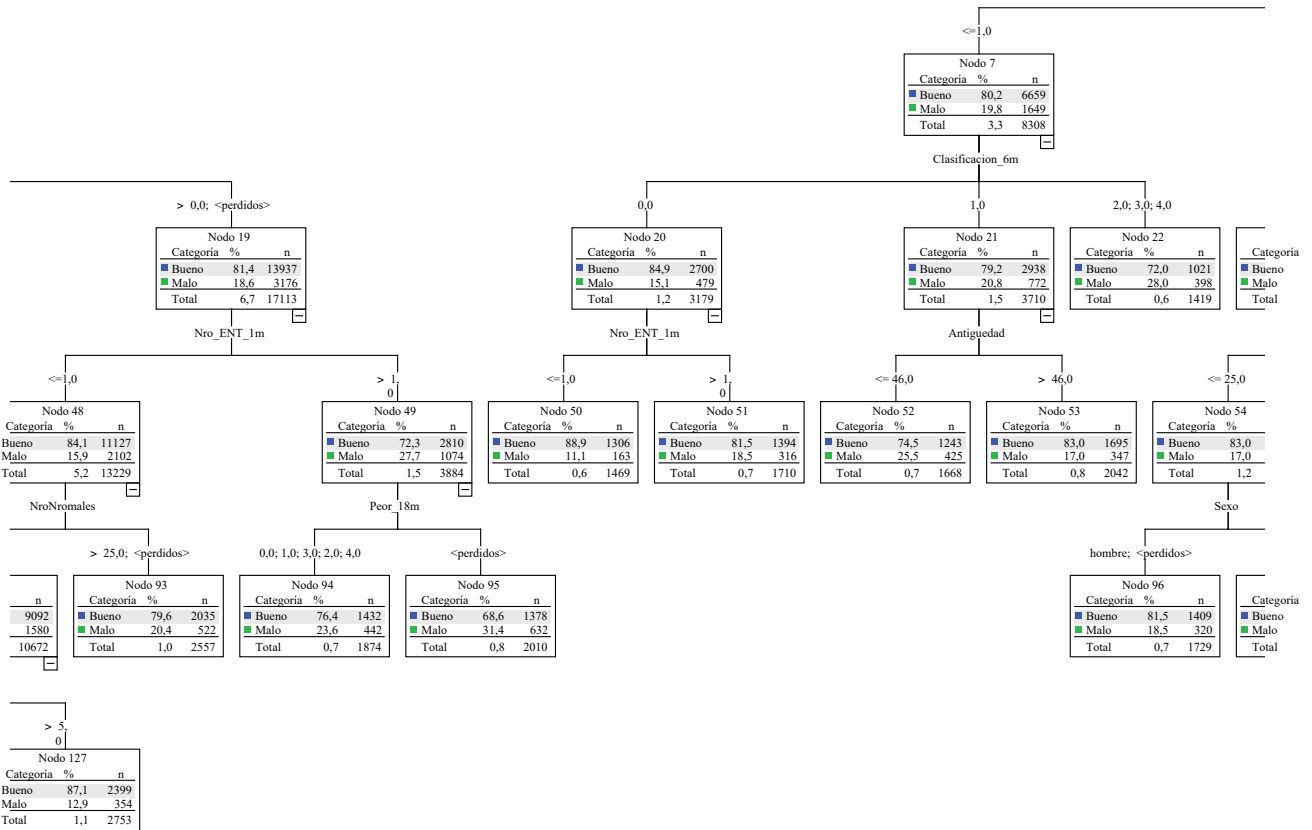
Anexo XII: Resultados del Árbol de Decisión



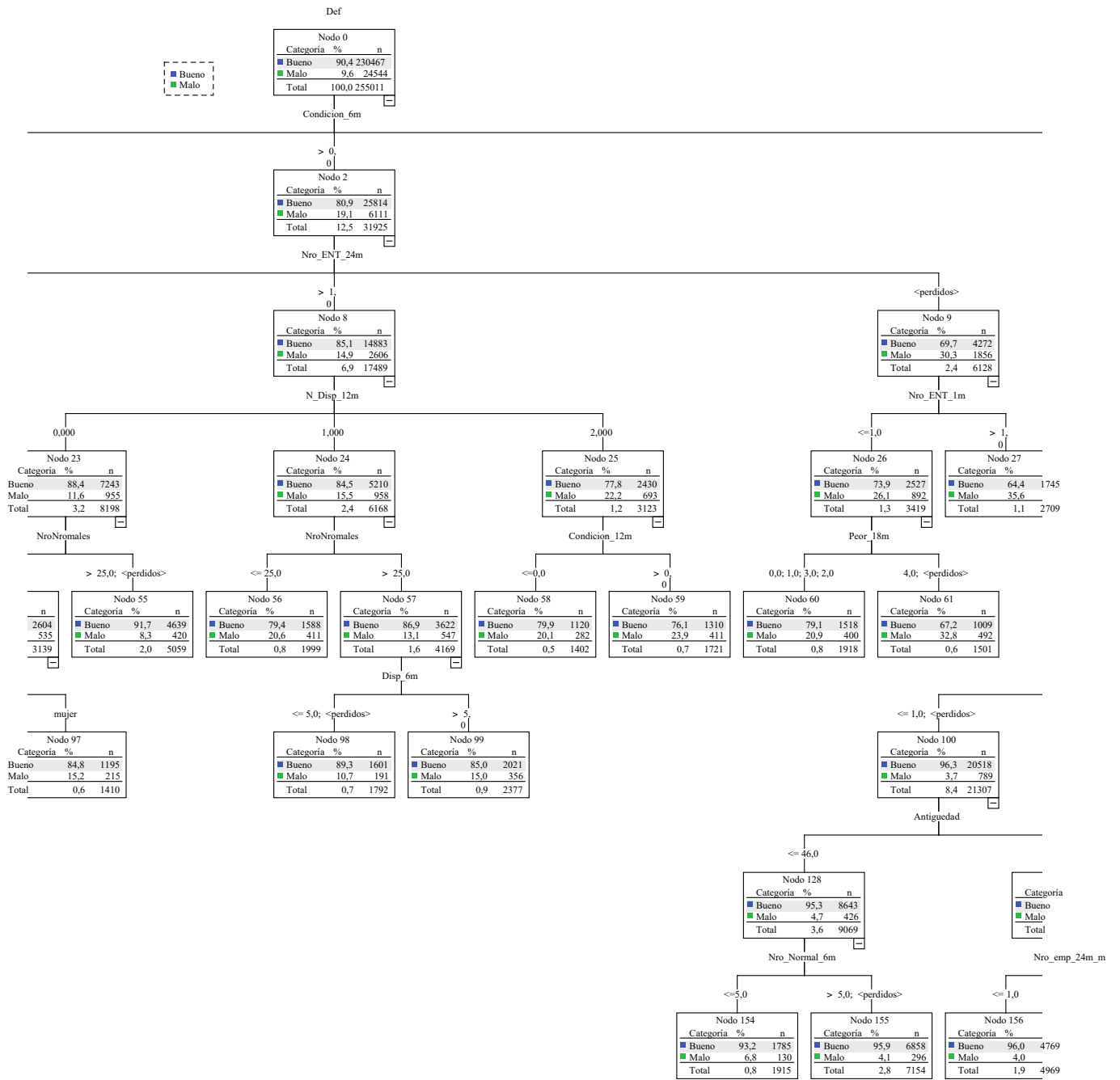


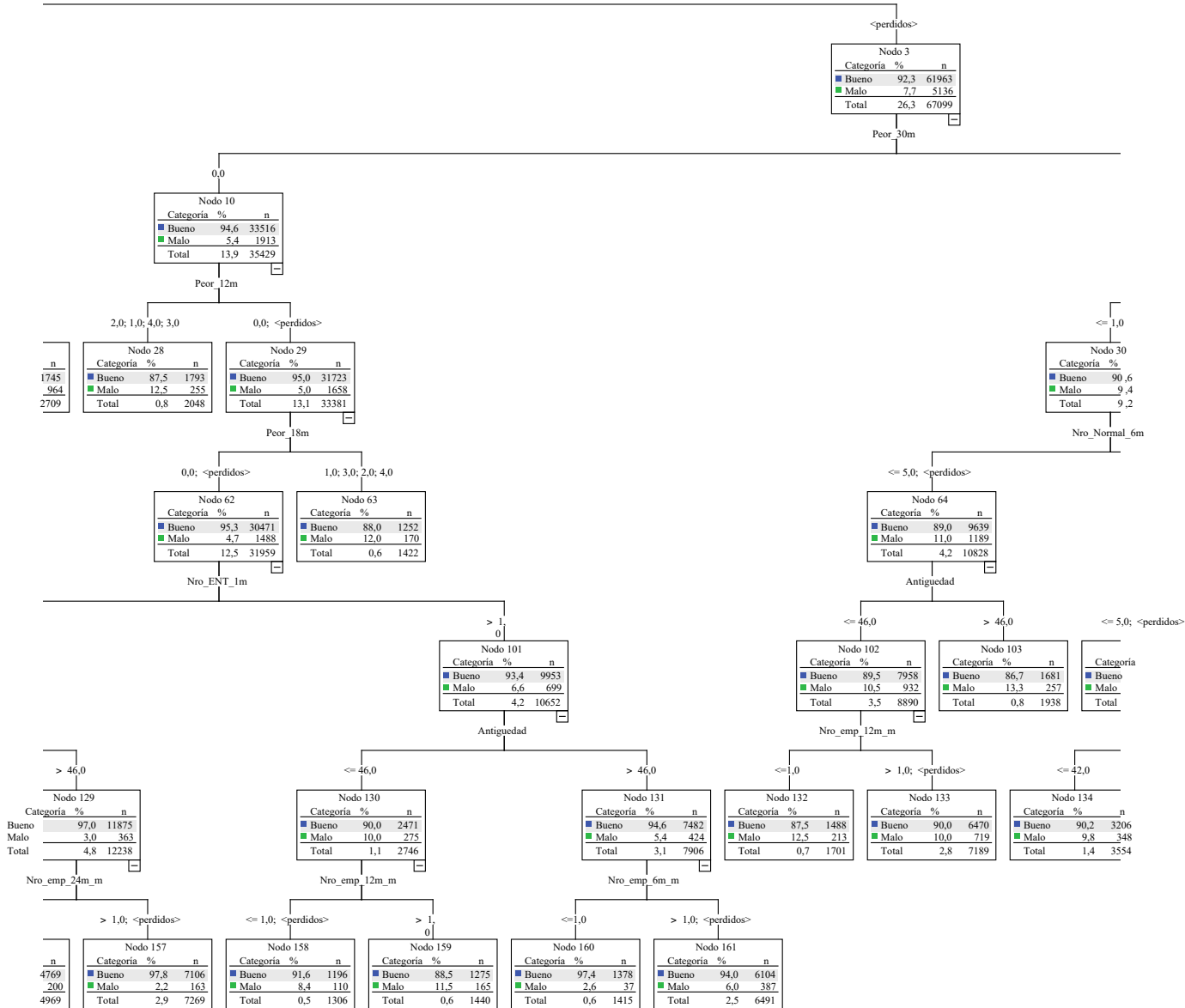


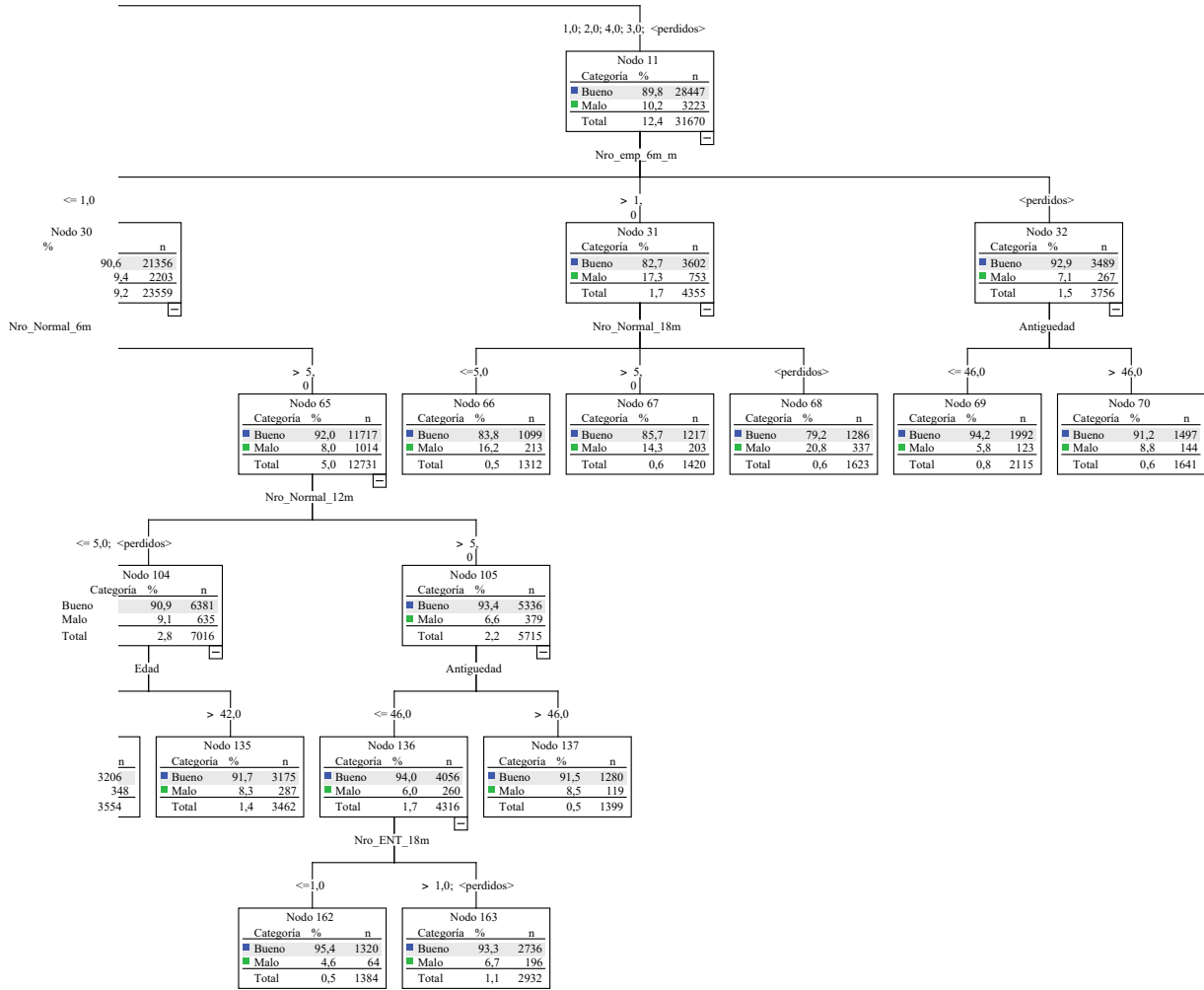




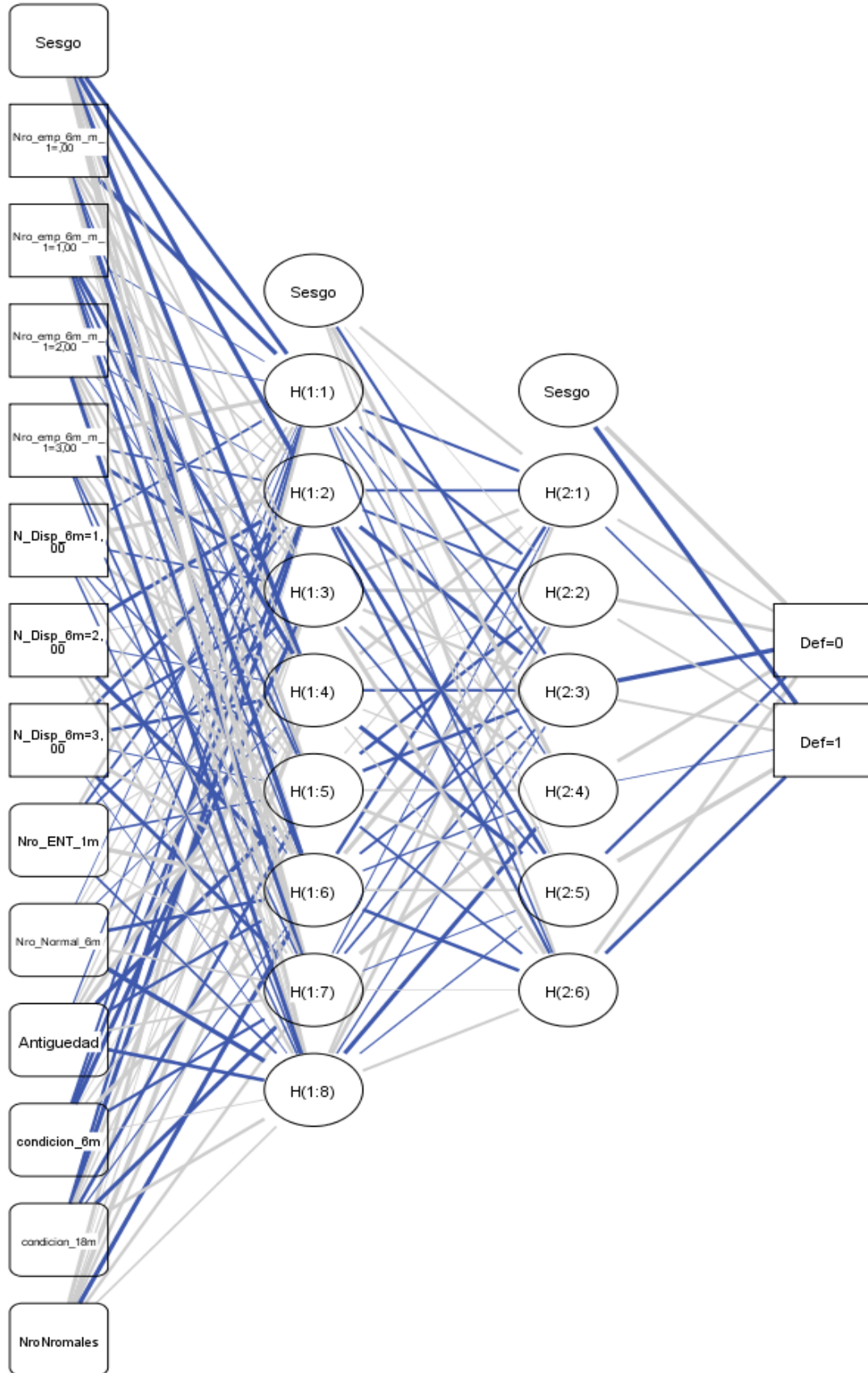
Nro_Normal_6m







Anexo XIII: Modelo de Redes Neuronales



Anexo XIV: Pesos Sinápticos del Modelo de Redes Neuronales

Predictor		Estimaciones de parámetro							
		Capa oculta 1							
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)
Capa de entrada	(Sesgo)	-.057	.034	-.421	-.099	-.215	-.272	.170	-.303
	Nro_emp_6m_m_1=0	.501	.065	-.294	.033	.409	-.245	.076	-.190
	Nro_emp_6m_m_1=1	-.011	.314	.248	.463	-.067	-.188	.392	-.398
	Nro_emp_6m_m_1=2	-.081	-.052	-.392	-.101	-.571	.403	-.364	.108
	Nro_emp_6m_m_1=3	.288	.335	-.148	.143	.308	-.098	-.024	-.326
	N_Dis_6m=1	-.350	.274	.049	-.252	.186	-.193	.349	.361
	N_Dis_6m=2	-.526	.344	.167	.522	.094	-.375	-.005	.187
	N_Dis_6m=3	-.330	.328	.464	.469	-.304	.433	-.268	.237
	Nro_ENT_1m	.040	-.287	.004	.311	-.429	.358	.040	.432
	condicion_6m	-.754	.013	-.124	.053	.042	-.283	.282	.576
	Nro_Normal_6m	.046	-.051	-.270	.108	-.124	-.292	-.457	-.161
	Antigüedad	.059	-.656	-.225	-.196	.295	.302	-.483	.172
	NroNromales	-.205	-.442	-.532	.013	.214	-.210	-.031	-.384
	condicion_18m	-.315	-.036	.302	.461	.178	-.605	.121	-.218

Predictor		Estimaciones de parámetro					
		Capa oculta 2					
		H(2:1)	H(2:2)	H(2:3)	H(2:4)	H(2:5)	H(2:6)
Capa oculta 1	(Sesgo)	-.332	.407	-.073	-.215	-.056	-.691
	H(1:1)	-.223	.293	-.363	.148	.359	-.320
	H(1:2)	-.235	.261	.830	.371	.020	-.028
	H(1:3)	.529	.196	-.437	.537	.058	.300
	H(1:4)	.133	.031	.492	.108	-.340	-.241
	H(1:5)	.134	.516	-.135	.016	.274	.029
	H(1:6)	.143	-.452	.442	-.181	.406	-.055
	H(1:7)	-.242	-.098	-.075	-.030	.307	-.226
	H(1:8)	.026	.457	0.00	.448	-.096	.182

Predictor		Estimaciones de parámetro	
		Capa de salida	
		[Def=0]	[Def=1]
Capa oculta 2	(Sesgo)	.157	-1.022
	H(2:1)	-.141	.281
	H(2:2)	.488	-.119
	H(2:3)	-.687	.464
	H(2:4)	-.525	.252
	H(2:5)	-.084	-.424
	H(2:6)	-.459	.548