

UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL DE HUAMANGA

Facultad de Ingeniería de Minas, Geología y Civil

Escuela de Formación Profesional de Ingeniería de Sistemas



MINERÍA DE DATOS PARA LA GESTIÓN DE RIESGO CREDITICIO DE LA
COOPERATIVA DE AHORRO Y CRÉDITO SAN CRISTÓBAL DE HUAMANGA DE
AYACUCHO, 2016

Tesis presentada por : Bach. Davis Fernando CALDERÓN
SARMIENTO

Para obtener el Título Profesional de: Ingeniero de Sistemas

Tipo de investigación : Aplicada

Área de Investigación : Inteligencia de negocios

Asesora : Ing. Elinar CARRILLO RIVEROS

Ayacucho – Perú

2017

A Dios, por acompañarme en mi vida.
A mis padres, por su apoyo permanente
en todos mis pasos.
A mis hermanas, por su cariño.
A mis amigos, por su aliento para seguir
superándome

AGRADECIMIENTOS

A la Universidad Nacional de San Cristóbal de Huamanga, en cuyas aulas aprendí nuevos conocimientos y logré mi formación profesional.

A los docentes de la Escuela Profesional de Ingeniería de Sistemas, por compartirme sus conocimientos.

A la cooperativa San Cristóbal de Huamanga, por permitirme realizar esta investigación.

A mis amigos y compañeros de trabajo por su apoyo y motivación.

A mi familia por su apoyo incondicional y comprensión durante mi formación profesional y durante el tiempo de la elaboración de la tesis.

A todas las personas por su apoyo en cada momento del proceso de elaboración de la tesis.

ÍNDICE

AGRADECIMIENTOS	ii
ÍNDICE	iii
RESUMEN	viii
INTRODUCCIÓN	ix

CAPÍTULO I

PLANTEAMIENTO DE LA INVESTIGACIÓN

1.1. DIAGNÓSTICO Y ENUNCIADO DEL PROBLEMA	1
1.2. FORMULACIÓN DEL PROBLEMA DE INVESTIGACIÓN	3
1.2.1. PROBLEMA PRINCIPAL	3
1.2.2. PROBLEMAS SECUNDARIOS	3
1.3. OBJETIVOS DE LA INVESTIGACIÓN	3
1.3.1. OBJETIVO GENERAL.....	3
1.3.2. OBJETIVOS ESPECÍFICOS.....	4
1.4. HIPÓTESIS	4
1.4.1. HIPÓTESIS GENERAL	4
1.4.2. HIPÓTESIS ESPECÍFICAS.....	4
1.5. JUSTIFICACIÓN Y DELIMITACIÓN DE LA INVESTIGACIÓN	4
1.5.1. IMPORTANCIA DEL TEMA.....	4
1.5.2. JUSTIFICACIÓN	5
1.5.3. DELIMITACIÓN	5

CAPÍTULO II

REVISIÓN DE LA LITERATURA

2.1.	ANTECEDENTES DE LA INVESTIGACIÓN	6
2.2.	MARCO CONCEPTUAL.....	7
2.2.1.	MINERÍA DE DATOS.....	7
2.2.2.	GESTIÓN DE RIESGO CREDITICIO	9
2.2.5.	EL LENGUAJE R.....	13
2.2.6.	CRISP-DM.....	14
2.2.7.	ÁRBOL DE CLASIFICACIÓN Y REGRESIÓN	26
2.2.8.	SISTEMA GESTOR DE BASE DE DATOS RELACIONAL	27

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

3.1.	TIPO DE INVESTIGACIÓN.....	28
3.2.	NIVEL DE INVESTIGACIÓN	28
3.3.	DISEÑO DE LA INVESTIGACIÓN	28
3.4.	POBLACIÓN Y MUESTRA.....	29
3.5.	VARIABLES E INDICADORES.....	29
3.5.1.	DEFINICIÓN CONCEPTUAL DE LAS VARIABLES.....	29
3.5.2.	DEFINICIÓN OPERACIONAL DE LAS VARIABLES	30
3.5.3.	TÉCNICAS E INSTRUMENTOS	30
3.6.	METODOLOGÍAS Y HERRAMIENTAS TECNOLÓGICAS	31
3.6.1.	METODOLOGÍA.....	31
3.6.2.	HERRAMIENTAS.....	31

CAPÍTULO IV

RESULTADOS DE LA INVESTIGACIÓN

4.1.	FASE I. COMPRENSIÓN DEL NEGOCIO	32
4.1.1.	DETERMINACIÓN DE LOS OBJETIVOS DEL NEGOCIO	32
4.1.2.	VALORACIÓN DE LA SITUACIÓN.....	35

4.1.3.	DETERMINACIÓN DE LOS OBJETIVOS DE MINERÍA DE DATOS.....	38
4.1.4.	PRODUCCIÓN DE UN PLAN DE PROYECTO	38
4.2.	FASE II. COMPRENSIÓN DE LOS DATOS	39
4.2.1.	RECOLECCIÓN DE DATOS INICIALES	39
4.3.2.	DESCRIPCIÓN DE LOS DATOS	42
4.3.3.	EXPLORACIÓN DE DATOS	47
4.3.4.	VERIFICACIÓN DE CALIDAD DE DATOS	52
4.4.	FASE III. PREPARACIÓN DE LOS DATOS	52
4.4.1.	SELECCIÓN DE DATOS.....	52
4.4.2.	LIMPIEZA DE DATOS.....	53
4.4.3.	ESTRUCTURAR LOS DATOS	54
4.4.4.	INTEGRACIÓN DE DATOS.....	54
4.5.	FASE IV: MODELADO	55
4.5.1.	SELECCIONAR LA TÉCNICA DE MODELADO	55
4.5.2.	GENERACIÓN EL PLAN DE PRUEBA	56
4.5.3.	CONSTRUIR EL MODELO	57
4.5.4.	EVALUACIÓN DEL MODELO	86
4.6.	FASE V: EVALUACIÓN	86
4.6.1.	EVALUAR LOS RESULTADOS.....	87
4.6.2.	REVISIÓN DEL PROYECTO.....	91

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

5.1.	CONCLUSIONES.....	92
5.2.	RECOMENDACIONES.....	93
	BIBLIOGRAFÍA.....	94
	ANEXOS	96

ÍNDICE DE FIGURAS

Figura 1.1: Mora en el tiempo	3
Figura 2.1: Ciclo de vida del Modelo CRISP-DM	14
Figura 2.2: Fase 1. Comprensión del negocio	16
Figura 2.3: Fase 2. Comprensión de los datos	18
Figura 2.4: Fase 3. Preparación de los datos	20
Figura 2.5: Fase 4. Modelado	23
Figura 2.6: Fase 5. Evaluación	25
Figura 4.1: Organigrama de la Cooperativa San Cristobal	34
Figura 4.2: Cantidad de créditos por cierre de año	47
Figura 4.3: Cantidad de créditos otorgados por categoría	48
Figura 4.4: Días de atraso por agencias	49
Figura 4.5: Cantidad de créditos por agencia	50
Figura 4.6: Cantidad de créditos por agencia	51
Figura 4.7: Carga de los datos a R Studio	53
Figura 4.8: Esquema de implementación	55
Figura 4.9: Gráfico de Árbol de clasificación	59
Figura 4.10: Gráfico del error	61
Figura 4.11: Gráfico de Árbol de clasificación	64
Figura 4.12: Gráfico del error	66
Figura 4.13: Gráfico del Árbol de clasificación	68
Figura 4.14: Gráfico del error	70
Figura 4.15: Gráfico del Árbol de clasificación	73
Figura 4.16: Gráfico del error	76
Figura 4.17: Gráfico del Árbol de clasificación	78

Figura 4.18: Gráfico del error	80
Figura 4.19: Gráfico de Árbol de clasificación	83
Figura 4.20: Gráfico del error	85

RESUMEN

La Gestión de Riesgo crediticio es un proceso importante en las entidades financieras, está diseñada para identificar eventos que pueda afectar a los movimientos de estas instituciones, gestionar de acuerdo con el riesgo y proveer una seguridad en el logro de sus objetivos.

Los indicadores de riesgo crediticio miden la clasificación crediticia, morosidad y cobertura de provisiones.

El objetivo de esta investigación es implementar un modelo de minería de datos que procese información eficaz para la gestión de riesgo crediticio en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga, 2016, mediante la metodología de minería de datos CRISP-DM y un gestor de base de datos, con la finalidad de obtener los indicadores para mejorar la gestión de riesgo crediticio.

La investigación se realizó tomando los datos de la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga. Es una investigación de tipo aplicada-tecnológica, nivel de investigación descriptiva. Se utilizó la metodología de minería de datos CRISP-DM para obtener la información necesaria de los indicadores del riesgo crediticio.

PALABRAS CLAVE: Minería de datos, Riesgo crediticio, CRISP-DM, Gestor de base de datos.

INTRODUCCIÓN

La minería de datos es el estudio y tratamiento de datos masivos para extraer conclusiones e información relevante de ellos. Aborda la solución de problemas de predicción, clasificación y segmentación, mediante la obtención de patrones o modelos a partir de un gran cúmulo de datos compilados y donde dicha información reside en forma oculta.

La motivación para implementar la minería de datos, es contar con un modelo que pueda procesar información eficaz para la gestión de riesgos, a partir de indicadores, por ejemplo la morosidad que permitirá pronosticar en el tiempo la calificación del crédito y poder regularla de acuerdo con la información.

Actualmente, existen reportes que acceden a la base de datos para consultar información, pero con diversas limitaciones, por ejemplo la complejidad de procesamiento, carga de memoria en servidor, errores en tiempo de ejecución, excesiva cantidad de tiempo, información poco elaborada o reportes realizados manualmente.

Los principales objetivos son: a) Generar indicadores de la minería de datos que procese información eficaz de la clasificación crediticia en relación a la morosidad en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga; b) Generar indicadores de la minería de datos que procese información eficaz de la clasificación crediticia en relación a la cobertura de provisiones en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga.

CAPÍTULO I

PLANTEAMIENTO DE LA INVESTIGACIÓN

1.1. DIAGNÓSTICO Y ENUNCIADO DEL PROBLEMA

Las bases de datos son el mayor motor donde se almacenan gran cantidad de datos, que al ser explotadas, usando la minería de datos, nos puede ayudar a generar información valiosa y la extracción de patrones de un archivo de datos previamente elaborado.

El otorgamiento de créditos es la principal actividad en el sistema financiero que genera beneficios, pero, a la vez, se les expone a diferentes riesgos, que puede incurrir en pérdidas, como consecuencia del incumplimiento de las obligaciones por parte del deudor.

Actualmente, en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga, para obtener información de la gestión de riesgo crediticio se realiza manualmente, es decir, la información de la base de datos se exporta a un archivo Excel y luego se procesa obteniendo así cuadros estadísticos; sin embargo, no cuenta con un modelo de minería de datos, que por medio de los datos históricos se pueda estudiar la calificación del crédito, el nivel de morosidad y la cobertura de provisiones que son indicadores importantes para la gestión de riesgo crediticio, de esta manera la cooperativa pueda proyectar su crecimiento.

Como se observa en Tabla N° 1.1, 1.2 y 1.3, la Gerencia de Riesgos utiliza estos indicadores para la gestión de riesgos y controlar la mora estableciendo metas mensuales.

AVANCE DE METAS EN NÚMERO DE SOCIOS						
Agencia	Cierre mes anterior	mes	Cierre mes actual	Variación	Meta Actual	Var. Meta
Totales	22,107		22,551	444	22,600	49

Tabla N° 1.1. Avance de metas en número de socios (Revisión documental, 31 de diciembre del 2016)

AVANCE DE METAS EN SALDO EN COLOCACIONES						
Agencia	Cierre mes anterior	mes	Cierre mes actual	Variación	Meta Actual	Variación a meta
Totales	214,675,181		219,544,412	4,869,231	220,000,000	455,588

Tabla N° 1.2. Avance de metas en saldo en colocaciones (Revisión documental, 31 de diciembre del 2016).

AVANCE DE METAS EN SALDO EN MORA						
Agencia	Cierre mes anterior	mes	Cierre mes actual	Variación	Meta Actual	Variación a meta
Totales	17,601,591		17,130,526	-471,064	17,000,000	-130,526

Tabla N° 1.3. Avance de metas en saldo en mora (Revisión documental, 21 de diciembre del 2016).

En la figura 1.1 se muestra el incremento de la mora en el tiempo, desde el año 2013 al 2016.

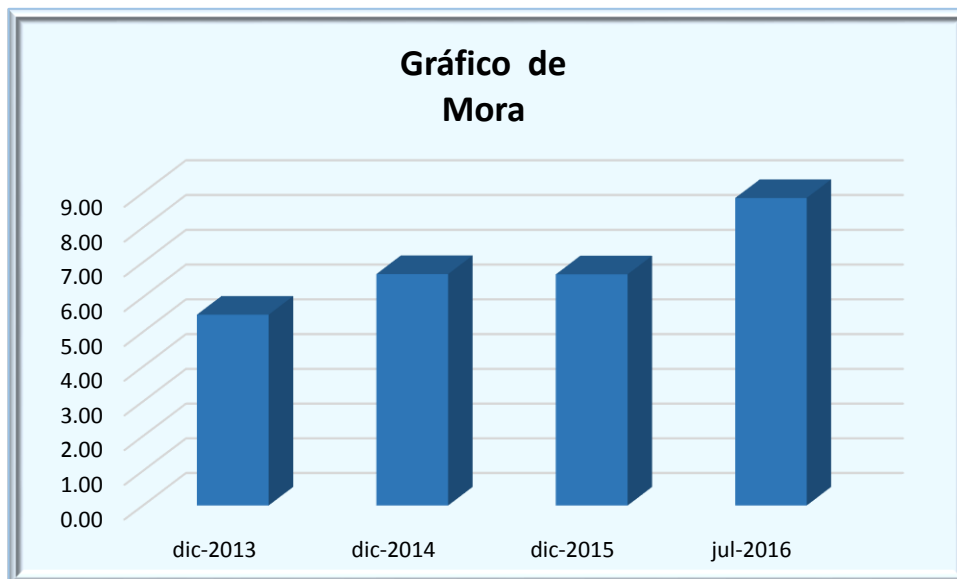


Figura 1.1. Mora en el tiempo (Elaboración propia, 2016)

1.2. FORMULACIÓN DEL PROBLEMA DE INVESTIGACIÓN

1.2.1. PROBLEMA PRINCIPAL

¿Cómo procesar la información para la gestión de riesgo crediticio en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga, 2016?

1.2.2. PROBLEMAS SECUNDARIOS

- a) ¿Cómo procesar información eficaz de la clasificación crediticia en relación a la morosidad en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga?
- b) ¿Cómo procesar información eficaz de la clasificación crediticia en relación a la cobertura de provisiones en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga?

1.3. OBJETIVOS DE LA INVESTIGACIÓN

1.3.1. OBJETIVO GENERAL

Implementar un modelo de minería de datos que procese información eficaz para la gestión de riesgo crediticio en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga, 2016, mediante la metodología de minería de datos

CRISP-DM, un gestor de base de datos, con la finalidad de obtener los indicadores para mejorar la gestión de riesgo crediticio.

1.3.2. OBJETIVOS ESPECÍFICOS

- a) Generar un modelo de minería de datos que procese información eficaz de la clasificación crediticia en relación a la morosidad en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga.
- b) Generar un modelo de minería de datos que procese información eficaz de la clasificación crediticia en relación a la cobertura de provisiones en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga.

1.4. HIPÓTESIS

1.4.1. HIPÓTESIS GENERAL

Si se implementa un modelo de minería de datos, entonces se logra apoyar a la gestión de riesgo crediticio en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga, 2016.

1.4.2. HIPÓTESIS ESPECÍFICAS

- a) Si se genera un modelo de minería de datos entonces se procesa una información eficaz de la clasificación crediticia en relación a la morosidad en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga.
- b) Si se genera un modelo de minería de datos entonces se procesa una información eficaz de la clasificación crediticia en relación a la cobertura de provisiones en la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga.

1.5. JUSTIFICACIÓN Y DELIMITACIÓN DE LA INVESTIGACIÓN

1.5.1. IMPORTANCIA DEL TEMA

En el aspecto técnico, al contar con un modelo de minería de datos, se

podrá tener una herramienta para procesar los indicadores de morosidad, cobertura de provisiones y la clasificación del crédito. De esta manera obtener información estratégica detallada para la toma de decisiones, simplificando el procesamiento con datos mejor elaborados. Los resultados obtenidos pueden ayudar a proyectarnos mejor en el futuro.

En el aspecto económico, la implementación de un modelo de minería de datos utilizando la herramienta propuesta, además de ser software libre, permitirá reducir el esfuerzo físico y costo en procesamiento de los indicadores.

1.5.2. JUSTIFICACIÓN

Monitorear el riesgo crediticio es fundamental para preservar la estabilidad de una entidad financiera, más aún cuando se observa en el tiempo que uno de sus principales indicadores es el índice de morosidad, en el caso de la Cooperativa San Cristóbal de Huamanga se está incrementando. Este fenómeno genera la motivación de emplear herramientas, como es la minería de datos para que ayude a obtener la información necesaria, previo análisis de la evolución de la cartera de créditos y medir este riesgo, a través de sus indicadores. Por consiguiente, implementando un modelo de minería de datos, a través de la información histórica se podrá pronosticar en el tiempo el comportamiento y los patrones de riesgo en la mencionada cooperativa.

1.5.3. DELIMITACIÓN

La investigación se realizó con la información de la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga de los años 2014, 2015 y 2016.

CAPÍTULO II

REVISIÓN DE LA LITERATURA

2.1. ANTECEDENTES DE LA INVESTIGACIÓN

Bravo (2012), en su tesis doctoral, menciona que medir la probabilidad de no pago de un gran número de solicitantes de crédito, llamado riesgo de crédito a consumidores, es un problema clásico de la gestión financiera. El problema requiere de una gran cantidad de herramientas estadísticas que lo hacen idóneo para su estudio por el área de Business Analytics. Esta tesis estudia este problema en base a la formalización de lo que se conoce sobre las determinantes del no pago y la transformación de ese conocimiento en herramientas estadísticas para la medición del riesgo. La finalidad del estudio fue mejorar la clasificación de créditos y la comprensión del default. Para ello se explotó inicialmente la regresión logística multinomial y luego se profundizó en el análisis de supervivencia, estudiando teóricamente los métodos de riesgos en competencia y los modelos mixtos, luego desarrollando herramientas computacionales liberadas públicamente para apoyar futuras aplicaciones.

Palacios (2012), en su tesis aplicado al banco de la nación sobre la morosidad de un microcrédito, señala que es conveniente, en una institución financiera, aumente radicalmente su competencia tecnológica a fin de adoptar decisiones óptimas. Una de las formas tecnológicas es adoptar un modelo de Red Neuronal Artificial (RNA) que fue implementado y probado. En el trabajo de tesis se realizó una propuesta sobre un nuevo servicio de Microcrédito, el cual se intenta vincular a un aspecto muy puntual de la computación moderna aplicada, donde se muestre que es posible predecir la morosidad de los clientes, planteando un modelo basado en RNA. Este acercamiento innovador también incluye la metodología de Minería de Datos para proyectos relacionados con redes neuronales artificiales.

Hernández (2015), en su investigación de la minería de datos aplicada al tema de la educación, hace un análisis de la aplicación de técnicas de minería de datos para identificar patrones de comportamiento con el fin de predecir el fracaso escolar y el abandono. Los experimentos se realizaron en una institución de nivel medio superior privada, donde se identificaron las variables que intervienen en el aprovechamiento académico indispensables para tomar decisiones y realizar acciones pertinentes, compara y se muestra los mejores modelos resultantes. Para la implementación utilizó la metodología CRISP-DM que se estructura el proceso de minería de datos en seis fases, que interactúan entre ellas de forma iterativa. Se aplicó los modelos de Redes Neuronales, Árboles de decisión y Clúster K-medias para analizar el comportamiento de los alumnos.

2.2. MARCO CONCEPTUAL

2.2.1. MINERÍA DE DATOS

Date y Ruiz (2001) afirman que la minería de datos puede describirse como "Análisis de datos exploratorios". El propósito es buscar patrones interesantes en los datos, patrones que puedan utilizarse para especificar la estrategia del negocio o para identificar comportamientos fuera de lo común. Las herramientas de minería de datos aplican técnicas de estadística, una gran cantidad de datos almacenados para buscar patrones. Las bases de datos para minería de datos frecuentemente son extremadamente grandes, es importante que los algoritmos sean escalables.

Para el análisis de datos se ha incorporado la utilización de herramientas de minería de datos por tres razones principales:

- a) Para Barreiro (Citado por Mena, 1999) la minería de datos ofrece la posibilidad de llevar a cabo un proceso de descubrimiento de información automático, es decir por un algoritmo.
- b) Para Barreiro (Citado por Berson y Smith, 1997) la minería de datos se ha diseñado para ser empleada directamente por decisores o

personal de negocio y no necesariamente por técnicos.

- c) Para Barreiro (Citado por Mena, 1999) las herramientas de minería de datos están preparadas para trabajar con grandes volúmenes de registros; mientras que, generalmente, los paquetes estadísticos clásicos se saturan cuando se introduce gran cantidad de datos para analizar.

Según Hernández, Ramírez y Ferri (2004) un proceso típico de minería de datos consta de los siguientes pasos generales:

- 1) Selección del conjunto de datos: Tanto en lo que se refiere a las variables objetivo (aquellas que se quieren predecir, calcular o inferir), como a las variables independientes (las que sirven para hacer el cálculo o proceso), como posiblemente al muestreo de los registros disponibles.
- 2) Análisis de las propiedades de los datos, en especial los histogramas, diagramas de dispersión, presencia de valores atípicos y ausencia de datos (valores nulos).
- 3) Transformación del conjunto de datos de entrada. Se realizará de diversas formas en función al análisis previo, con el objetivo de preparar y aplicar la técnica de minería de datos que mejor se adapte a los datos y al problema. A este paso también se le conoce como preprocesamiento de los datos.
- 4) Seleccionar y aplicar la técnica de minería de datos. Se construye el modelo predictivo, de clasificación o segmentación.
- 5) Extracción de conocimiento. Mediante una técnica de minería de datos, se obtiene un modelo de conocimiento que representa

patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables. También pueden usarse varias técnicas a la vez, para generar distintos modelos, aunque generalmente cada técnica obliga a un preprocesamiento diferente de los datos.

- 6) Interpretación y evaluación de datos. Una vez obtenido el modelo, se debe proceder a su validación comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, debe alterarse alguno de los pasos anteriores para generar nuevos modelos.

2.2.2. GESTIÓN DE RIESGO CREDITICIO

En la Memoria anual de Scotiabank (2011) se indica que el riesgo crediticio es el riesgo de pérdidas por el incumplimiento de un cliente o contraparte de sus obligaciones financieras o contractuales. Surge de las operaciones de préstamo directo y de las actividades de financiamiento, inversión y negociación en virtud de las cuales las contrapartes se comprometen a cumplir con reembolsos a la entidad financiera u otras obligaciones con este. La gestión eficaz del riesgo crediticio requiere el establecimiento de una cultura adecuada de riesgo crediticio. Las políticas de riesgo y las estrategias de gestión de riesgos clave son elementos importantes para la creación de esta cultura.

Para Aguilar, Camargo, y Morales (2006) el riesgo de crédito es el tipo de riesgo más importante al que debe hacer frente cualquier entidad financiera. Un indicador de este riesgo es el nivel de morosidad de la entidad; es decir, la proporción de su cartera que se encuentra en calidad de incumplimiento.

Estupiñán (2006) menciona que el riesgo crediticio es el factor más importante que se mide, es la frecuencia esperada de no pago. Se tiene dos teorías en el mundo financiero para analizarlas, mediante las cuales puede decidirse si se presta o no y la otra son las metodologías continuas, que tienen en cuenta la frecuencia esperada de no pago y otros factores adicionales, que monitorean el crédito durante toda su vida.

Si se trata de identificar las características de un deudor con un patrón de comportamiento, se debe tener una base de datos lo suficientemente grande como para establecer los diferentes patrones de comportamiento que existan en un mercado, no una base de datos solamente de clientes, sino un universo del mercado que se maneja. Con esa base de datos se identifica la probabilidad, que la persona pague o no pague y por ende, si se le presta o no, por tanto la entidad debe estar inscrita y de consulta permanente a las centrales de riesgos que se tienen en cada país.

Existen dos modalidades del riesgo de crédito: el directo y el contingente. El directo es aquel que ya está en los libros y ha sido desembolsado y el contingente se caracteriza, porque existen ciertas garantías, tales como cartas de crédito, están abiertas y sin utilizar. Los indicadores de riesgo miden la clasificación del crédito, su morosidad la cobertura de provisiones.

2.2.3. CLASIFICACIÓN DE LA CARTERA DE CRÉDITO

Para Lárraga (2008) la calificación crediticia es una opinión emitida por una entidad especializada sobre la posibilidad futura de que un deudor afronte sus compromisos de pagos a su debido tiempo y abone totalmente sus deudas, tanto por lo que se refiere al capital como con los intereses.

La SBS (2008) en la ley N° 11356 define la clasificación crediticia del deudor de la cartera de créditos para pequeñas empresas, microempresas y consumo no revolvente según los siguientes criterios:

- a) **CATEGORÍA NORMAL.** Son aquellos deudores que vienen cumpliendo con el pago de sus créditos de acuerdo con los convenido o con un atraso de hasta ocho días calendario.
- b) **CATEGORÍA CON PROBLEMAS POTENCIALES.** Son aquellos deudores que registran atraso en el pago de sus créditos de nueve a treinta días calendario.
- c) **CATEGORÍA DEFICIENTE.** Son aquellos deudores que registran atraso en el pago de sus créditos de treinta y uno a sesenta días calendario.
- d) **CATEGORÍA DUDOSO.** Son aquellos deudores que registran atraso en el pago de sus créditos de sesenta y uno a ciento veinte días calendario.
- e) **CATEGORÍA PÉRDIDA.** Son aquellos deudores que muestran atraso en el pago de sus créditos de más de ciento veinte días calendario.

2.2.4. MOROSIDAD

Para La Federación Nacional de Cooperativas de Ahorro y Crédito del Perú (2013) se usa la ratio de cartera atrasada y cartera bruta como un indicador de la morosidad.

Cóndor y Cajamarca (2013) describen el índice de la morosidad como el porcentaje de la cartera atrasada frente a la cartera total. Las ratios de la morosidad se calculan para el total de la cartera bruta y por línea de negocio. Cuando se incrementa la morosidad crediticia tiene un efecto negativo sobre la rentabilidad a la vez que se da una ruptura en la rotación de fondos. Lo que trae consigo que la entidad financiera incremente sus provisiones por los créditos

impagos, esto a su vez afecta inmediatamente a sus utilidades. Por lo que, un incremento importante en la morosidad, hace que el problema del incumplimiento se traduzca en rentabilidad, liquidez y finalmente en solvencia.

Para Brachfield (2003) el término morosidad es el retraso en el cumplimiento de una obligación de pago. Es decir, que a pesar de que el cliente no ha pagado la factura del vencimiento, acaba finalmente pagando la deuda en su totalidad. Por lo tanto moroso es aquel deudor que paga tarde, pero paga, o sea no se trata de un incumplimiento definitivo de la obligación, sino un cumplimiento tardío de la misma. Es importante distinguir si se trata de un simple retraso de algunos días o de una demora prolongada en el pago.

C. COBERTURA DE PROVISIONES

Para Estupiñán (2006) muestra los niveles de provisiones entre los rangos, sus garantías u otros problemas. Se establecen 3 indicadores, el de los niveles de provisión de la cartera de difícil cobro e irrecuperable, la de la totalidad de la provisión de cartera vencida y la de la cartera en judicial.

La SBS (2008) en la ley N° 11356 menciona las siguientes clases de provisiones:

- a) Provisión genérica. Son aquellas que se constituyen, de manera preventiva, sobre los créditos directos y la exposición equivalente a riesgo crediticio de los créditos indirectos de deudores clasificados en categoría normal.

TIPO DE CRÉDITO	TABLA DE PROVISIONES
Crédito a medianas empresas	1%
Crédito a pequeñas empresas	1%
Crédito a microempresas	1%
Crédito de consumo no-revolventes	1%

Créditos hipotecarios para vivienda	0.7%
-------------------------------------	------

Tabla 2.1. Tabla de provisiones genéricas (SBS, 2008)

- b) Provisión específica. Son aquellas que se constituyen sobre los créditos directos y la exposición equivalente a riesgo crediticio de los créditos indirectos y la exposición equivalente a riesgo crediticio de los créditos indirectos de deudores a los que se ha clasificado en una categoría de mayor riesgo que la categoría normal.

CATEGORÍA DEL RIESGO	TABLA 1	TABLA 2
Categoría con Problemas potenciales	5%	2.5%
Categoría Deficiente	25%	12.5%
Categoría Dudoso	60%	30%
Categoría Pérdida	100%	60%

Tabla 2.2. Tabla de provisiones específicas (SBS, 2008)

2.2.5. EL LENGUAJE R

Santana y Mateos (2014) definen que R es un software poderoso, que puede ser aplicado ampliamente en el procesamiento de datos en ciencias. R es un lenguaje de programación interpretado y se mantiene en un ambiente para el cómputo estadístico y gráfico.

ALGUNAS CARACTERÍSTICAS IMPORTANTES

Para Santana y Mateos (2014) R está dividido en dos partes conceptuales: 1) El sistema base de R; 2) en todo lo demás. La funcionalidad de R consta en paquetes modulares. El sistema base de R contiene el paquete básico que se requiere para su ejecución y la mayoría de sus funciones fundamentales.

2.2.6. CRISP-DM

Champan et. Al (2000), en la metodología CRISP-DM, estructuran el ciclo de vida de un proyecto de explotación de datos en seis fases cuya sucesión no es rígida e interactúan entre ellas de forma iterativa durante el desarrollo del proyecto. Las flechas indican la dependencia más importante y frecuente entre las fases. El círculo exterior simboliza la naturaleza cíclica de los proyectos de este tipo (Figura 1.1).

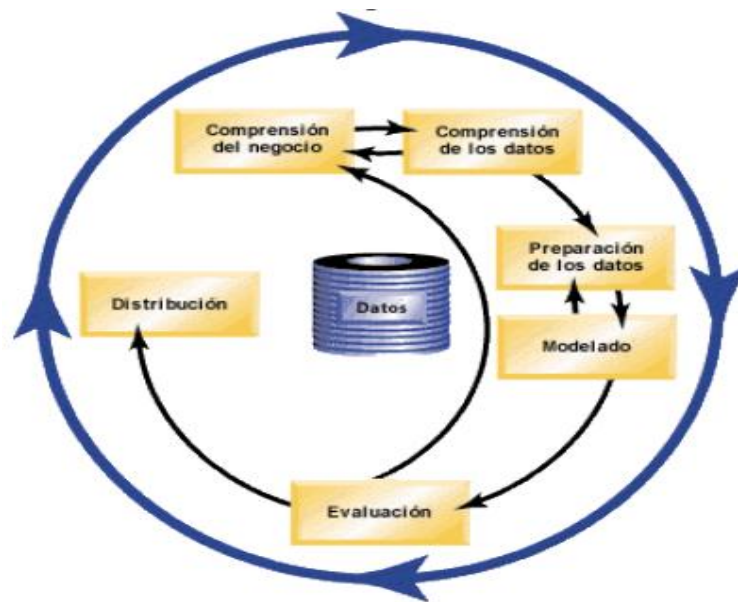


Figura 2.1. Ciclo de vida del Modelo CRISP-DM (Champan et. Al, 2000)

Para De Ville (2001) la metodología CRISP-DM es un enfoque basado en estándares multinacionales de describir, documentar y mejorar continuamente la minería de datos y negocio de la inteligencia. El marco de CRISP-DM identifica siete pasos en el proceso de minería de datos, para capturar la característica de bucle cerrado de un círculo virtuoso de mejora continua del proceso a través del plan sucesivo, analizar, implementar, medir iteraciones de proyectos de minería de datos.

1. Conocimiento del negocio
2. Comprensión de los datos
3. Preparación de datos

4. Modelado
5. Evaluación
6. Despliegue
7. Medición del desempeño

Las tareas que se realizan: elaborar un plan de implantación, elaborar un plan de monitoreo y mantenimiento, informar sobre el proyecto y la revisión del proyecto.

(Figura 1.2).

COMPRESIÓN DEL NEGOCIO	COMPRESIÓN DE LOS DATOS	PREPARACIÓN DE LOS DATOS	MODELADO	EVALUACION	DISTRIBUCIÓN
Determinar Objetivos del Negocio -Background del Negocio -Objetivos del Negocio -Criterios de éxito del negocio Valoración de la situación -Inventario de recursos -Requisitos, supuestos y restricciones -Riesgos y contingencias -Terminología -Costos y beneficios Determinar los objetivos de DM -Metas de Data Mining -Criterios de éxito de DM Realizar el plan del proyecto -Plan de proyecto -Valoración inicial	Recolectar los datos iniciales -Reporte de recolección de los datos. Descripción de los datos -Reporte de descripción de datos Exploración de los datos -Reporte de exploración de los datos Verificar la calidad de los datos -Reporte de calidad de los datos	Seleccionar los datos -Inclusión / exclusión de datos. Limpiar los datos -Reporte de calidad de los datos Estructurar los datos -Derivación de atributos -Generación de registros. Integrar los datos -Unificación de datos. Formateo de los datos -Reporte de calidad de datos.	Seleccionar técnica de modelado -Técnica seleccionada -Supuestos del modelo Generar el plan de prueba -Plan de pruebas Construir el modelo -Seteo de parámetros -Modelo -Descripción del modelo Evaluar el modelo -Evaluar el modelo -Revisión de los parámetros	Evaluación de los resultados -Valoración de los resultados. Modelos aprobados Revisión del proceso -Revisión del proceso Determinar próximos pasos -Técnica modelada -Listado de los posibles acciones.	Plan de implantación -Plan de implementación Plan de monitoreo y mantención -Plan de monitoriación y mantención Informe final -Informe final -Modelos aprobados Revisión del proyecto -Documentación de experiencias

Tabla 2.3. Tareas y salidas del modelo CRISP-DM (Champan et. Al, 2000)

A. FASE 1. COMPRENSIÓN DEL NEGOCIO O PROBLEMA

Champan et. Al (2000) mencionan que se centra en la comprensión de los objetivos y requisitos del proyecto desde una perspectiva de la empresa, para convertirlos en objetivos técnicos y en un plan de proyecto. El conocimiento adquirido del negocio se convierte en un problema de Data Mining (DM).

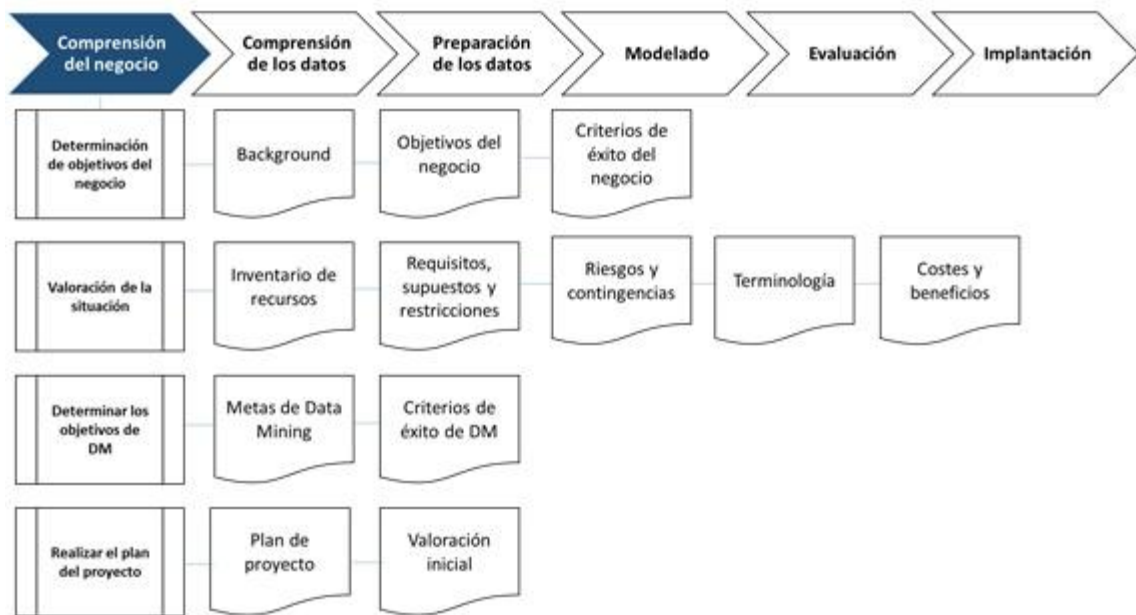


Figura 2.2. Fase1. Comprensión del negocio (Champan et. Al, 2000)

Las tareas que se realizan son: determinar los objetivos del negocio, evaluar la situación, determinar los objetivos del DM y elaborar un plan de proyecto.

a) DETERMINAR LOS OBJETIVOS DEL NEGOCIO

Para IBM (2012) la primera tarea es obtener la máxima información posible de los objetivos comerciales de la minería de datos. Es posible que esta tarea no sea tan fácil como parece, pero puede reducir los futuros riesgos clarificando los problemas, objetivos y recursos.

b) EVALUAR LA SITUACIÓN

Para Champan et Al. (2000) esta tarea implica una investigación detallada de todos los recursos, limitaciones, suposiciones y otros factores que deben ser considerados para determinar la meta de análisis de datos y el plan del proyecto. En esta se debe ampliar sobre los detalles.

Para la IBM (2012) este paso implica algunas cuestiones como:

¿Qué tipos de datos están disponibles para el análisis?

¿Dispone del personal necesario para completar el proyecto?

¿Cuáles son los principales factores de riesgo?

¿Dispone de planes de contingencia para cada factor de riesgo?

c) DETERMINAR LOS OBJETIVOS DE LA MINERÍA DE DATOS

Para Champan et Al. (2000) una meta comercial establece objetivos en la terminología comercial. Una meta de minería de datos establece los objetivos del proyecto en términos técnicos.

d) ELABORAR UN PLAN DE PROYECTO

IBM (2012) menciona que, en este punto, ya está listo producir un plan para el proyecto de minería de datos. Las cuestiones que hayan planteado hasta el momento los objetivos comerciales y de minería de datos que haya formulado constituirán la base de este plan.

B. FASE 2. COMPRESIÓN DE LOS DATOS

Para Champan et. Al (2000) comienza con una recopilación inicial de datos para establecer un primer contacto con el problema, familiarizarse con los datos, identificar la calidad de los datos y establecer relaciones evidentes que permitan definir la hipótesis de información oculta.

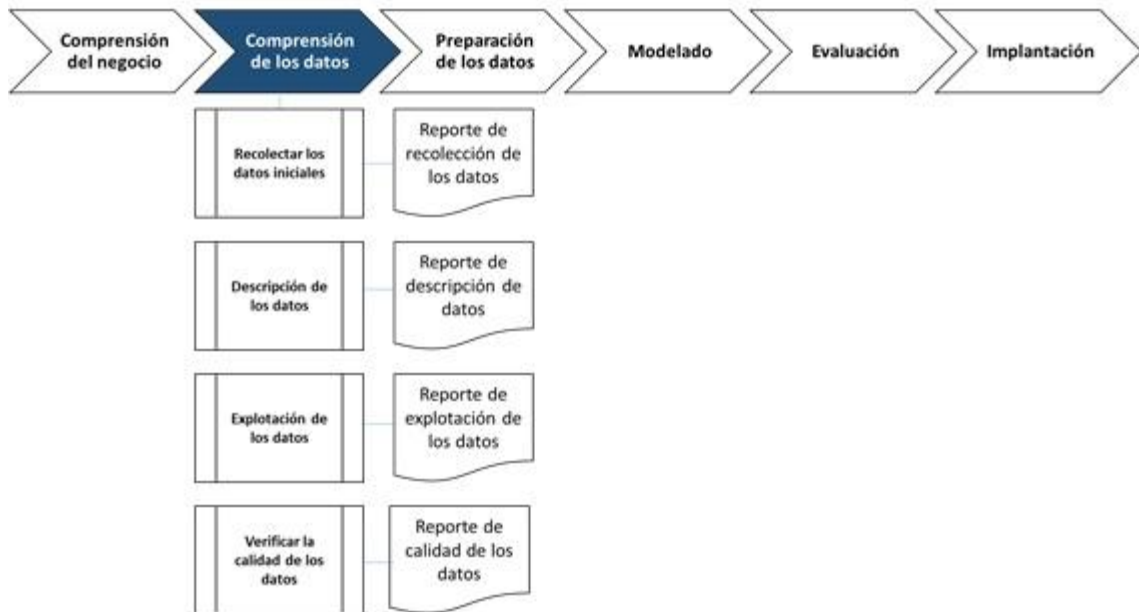


Figura 2.3. Fase 2. Comprensión de los datos (Champan et. Al, 2000)

Las tareas que se realizan: recolectar los datos iniciales, describir los datos, explorar los datos y validar la calidad.

1. RECOLECCIÓN DE DATOS INICIALES

Para IBM (2012) en este punto se puede acceder a los datos e incluirlos en un modelador. Las diferentes fuentes de los datos provienen de:

- Datos existentes. Incluye una amplia variedad de datos, como los datos transaccionales, datos de encuesta, registros Web, etc. Tenga en cuenta si los datos existentes son suficientes para adaptarse a sus necesidades.
- Datos adquiridos. Ver los datos adicionales si se considera necesario.
- Datos adicionales. Si las fuentes anteriores no satisfacen es posible realizar encuestas o complementar con los almacenes de datos actuales.

2. DESCRIPCIÓN DE LOS DATOS

Champan et. Al (2000) describen que después de adquiridos los datos iniciales, estos deben ser descritos. Este proceso involucra establecer volúmenes de datos (números de registros y campos por registro), su identificación, el significado de cada campo y la descripción del formato inicial.

3. EXPLORACIÓN DE DATOS

Según IBM (2012) se utiliza esta fase para explorar los datos con tablas, gráficos y otras herramientas de visualización. Estos análisis pueden ayudar a describir los objetivos de minería de datos generados durante la fase de comprensión comercial. También pueden ayudarle la hipótesis y dar forma a las tareas de transformación de datos que tiene lugar durante la preparación de los datos.

4. VERIFICAR LA CALIDAD DE LOS DATOS

Champan et. Al (2000) afirman que se debe examinar la calidad de los datos, abordando con preguntas si se completan los datos, si son correctos o contiene errores, y si contiene errores qué tan comunes son. Verificar si faltan valores en los datos, en caso afirmativo, cómo están representados y dónde se producen.

C. FASE 3. PREPARACIÓN DE LOS DATOS

Competen las actividades de preparación de los datos iniciales para adaptarlos a las técnicas de DM, tales como visualización, búsqueda de relaciones entre variables para la exploración.

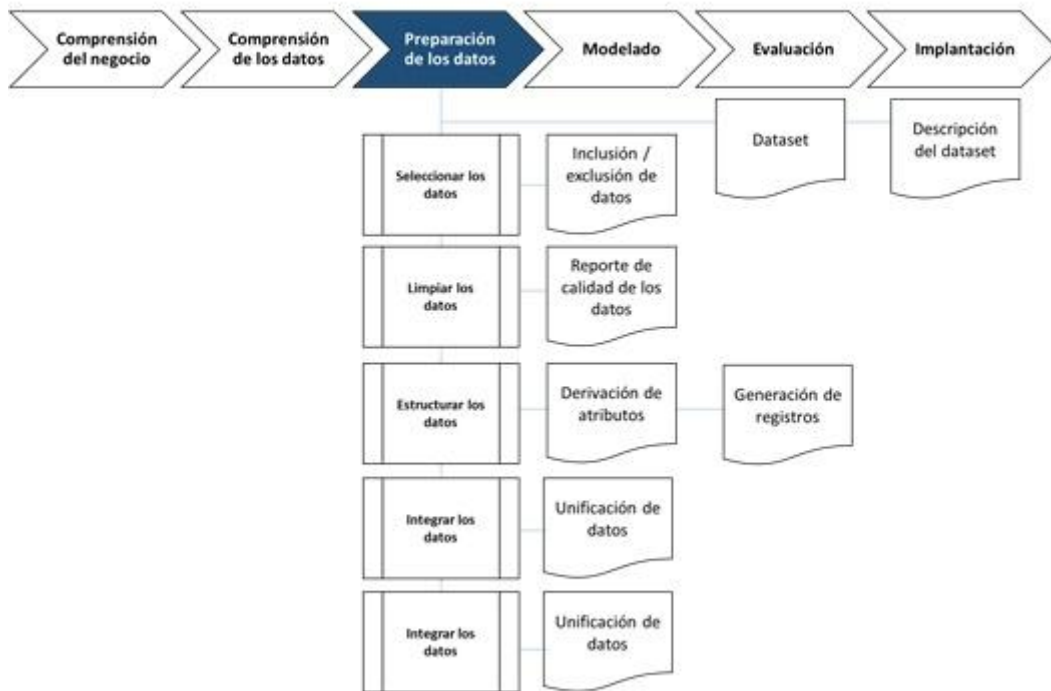


Figura 2.4. Fase 3. Preparación de los datos (Champan et. Al, 2000)

Las tareas que se realizan: selección de datos, limpieza de datos, estructurar los datos, integrar los datos y formatear los datos.

a) SELECCIONAR LOS DATOS

Para IBM (2012) ahora puede comenzar a seleccionar los datos relevantes a sus objetivos de minería de datos. De forma general, existen dos formas de seleccionar los datos:

- Selección de elementos (filas). Implica la toma de decisiones como las cuentas, productos o clientes que se van a incluir.
- Selección de atributos o características (columnas). Implica la toma de decisiones sobre el uso de características como la cantidad de las transacciones o los ingresos por hogar.

b) LIMPIAR LOS DATOS

Champan et. Al (2000) sostienen que se debe elevar la calidad de

los datos al nivel requerido por las técnicas de análisis seleccionadas. Esto puede implicar la selección de los subconjuntos limpios de los datos, la inserción de valores predeterminados adecuados o técnicas más ambiciosas como la estimación de datos faltantes mediante el modelamiento.

Problema de datos	Solución posible
Datos perdidos	Excluya las filas o características. O cumpliméntalas con un valor estimado.
Errores de datos	Utilice recursos lógicos para descubrir errores manuales y corríjalos. O, excluya las características.
Incoherencia de codificación	Decida un esquema de codificación simple y convierta y sustituya los valores.
Metadatos ausentes o erróneos	Examine manualmente los campos sospechosos y compruebe el significado correcto.

Tabla 2.4. Tipos de problemas concretos en los datos (IBM,2012)

c) ESTRUCTURAR LOS DATOS

Champan et. Al (2000) señalan que esta tarea incluye las operaciones de preparación de los datos tales como la generación de nuevos atributos a partir de atributos ya existentes, integración de nuevos registros o transformación de valores para atributos existentes.

Para IBM (2012) existen 2 formas de construir nuevos datos:

- Derivación de atributos (columnas o características)
- Generación de registros (filas)

d) INTEGRAR LOS DATOS

Champan et. Al (2000) mencionan que la integración de los datos

involucra la creación de nuevas estructuras a partir de los datos seleccionados, por ejemplo, generación de nuevos campos a partir de otros existentes, creación de nuevos registros, fusión de tablas o nuevas tablas donde se resumen características de múltiples registros o de otros campos en nuevas tablas de resumen.

Para IBM (2012) existen dos métodos básicos para integrar los datos:

- La fusión de datos. Implica unir dos conjuntos con registros similares, pero con atributos diferentes. Los datos se fusionan utilizando el mismo identificador clave en cada registro (ID). Los datos resultantes aumentan las columnas o las características.
- La adición de datos. Implica dos o más conjuntos de datos con atributos similares, pero con registros diferentes. Los datos se integran en función de los campos similares.

e) FORMATEO DE LOS DATOS

Gallardo (2009) afirma que esta tarea consiste principalmente, en la realización de transformaciones sintácticas de los datos sin modificar su significado, esto, con la idea de permitir o facilitar el empleo de alguna técnica de Data Mining en particular, por ejemplo la reordenación de los campos y/o registros de la tabla o el ajuste de los valores de los campos a las limitaciones de las herramientas de modelación (eliminar comas, tabulaciones, caracteres especiales, máximos y mínimos, para las cadenas de caracteres, etc.).

D. FASE 4. MODELADO DE DATOS

Diversas técnicas de modelado son seleccionadas y aplicadas y sus parámetros son calibrados a valores óptimos, por ejemplo, análisis de regresión, redes neuronales, razonamiento basado en casos (RBC), etc. Algunas técnicas tienen requerimientos específicos según las características de los datos y la

precisión que se requiera.

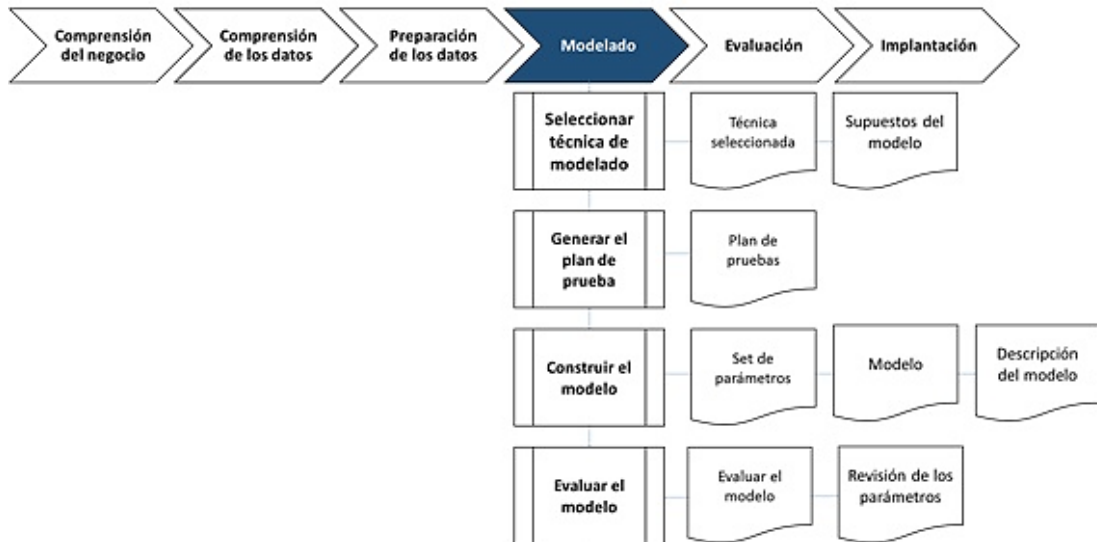


Figura 2.5. Fase 4. Modelado (Champan et. Al, 2000)

Las tareas que se realizan: selección de la técnica de modelado, generar el plan de pruebas, construir el modelo y evaluar el modelo.

a) SELECCIONAR LA TÉCNICA DE MODELADO

Para Champan et. Al (2000) como primer paso en el modelado, se selecciona la técnica de modelado real que se va a utilizar. Esta tarea se refiere a la Técnica de modelado, por ejemplo, construcción de árboles de decisión con 5.0, o generación de redes neuronales con respaldo propagación. Si se aplican múltiples técnicas, realice esta tarea por separado para cada técnica.

b) GENERAR EL PLAN DE PRUEBA

Champan et. Al (2000) dicen que una vez construido el modelo, se debe generar un procedimiento destinado a probar la calidad y validez del mismo. Típicamente se separan los datos en dos conjuntos, uno de entrenamiento y otro de prueba, para luego

construir el modelo basado en el conjunto de entrenamiento y medir la calidad del modelo generado con el conjunto de prueba.

c) CONSTRUIR EL MODELO

IBM (2012) sostiene que, en este punto, se debe tener la preparación suficiente para generar los modelos que haya considerado. Tómese el tiempo necesario para experimentar con diferentes modelos antes de llegar a conclusiones definitivas. Asegúrese de registrar los ajustes y datos utilizados para cada modelo. Al final se dispondrá de tres tipos de información que pueda utilizar en la toma de decisiones de minería de datos:

- Configuración de parámetros
- Los modelos reales producidos
- Descripciones de resultados de modelos

d) EVALUAR EL MODELO

Para Champan et. Al (2000) en esta tarea los ingenieros de DM interpretan los modelos de acuerdo con el conocimiento preexistente del dominio y los criterios de éxito preestablecidos. Expertos en el dominio del problema juzgan los modelos dentro del contexto del dominio y expertos en Data Mining aplican sus propios criterios.

E. FASE 5. EVALUACIÓN DEL MODELO

Una evaluación detallada del modelo y la revisión de los pasos ejecutados para construir el modelo y asegurar que se han alcanzado los objetivos de negocio. Un punto importante es determinar si hay algún objetivo que no ha sido considerado.

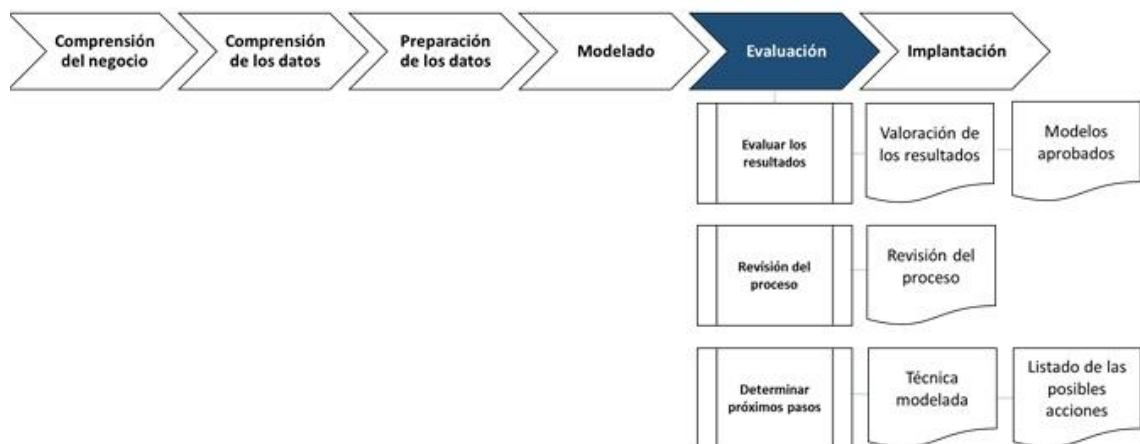


Figura 2.6. Fase 5. Evaluación (Champan et. Al, 2000)

Las tareas que se realizan: evaluar los resultados, revisar el proceso y determinar la implementación.

a) EVALUAR LOS RESULTADOS

IBM (2012) menciona que, en esta etapa, formalizará su evaluación en función de si los resultados del proyecto cumplen los criterios del rendimiento comercial. Este paso requiere una clara comprensión de los objetivos comerciales, por lo que debe estar seguro de incluir factores de toma de decisiones en la evaluación del proyecto.

Para Champan et. Al (2000) los resultados de la evaluación de los resultados son:

- Evaluación de los resultados de minería de datos con respecto a los criterios de éxito empresarial
- Modelos aprobados

b) REVISIÓN DEL PROCESO

Para Champan et. Al (2000) en este punto, los modelos resultantes parecen ser satisfactorios y satisfacer las necesidades del negocio.

Esto es ahora apropiado para hacer una revisión más completa del compromiso de minería de datos con el fin de determinar si hay un importante factor o tarea que de alguna manera se ha pasado por alto.

2.2.7. ÁRBOL DE CLASIFICACIÓN Y REGRESIÓN

Breiman (citado por Serna, 2009), desarrolló un algoritmo CART cuyo resultado genera un árbol de decisión, las ramas representan conjuntos de decisiones y cada decisión genera nuevas reglas sucesivas para continuar la clasificación (formando) así grupos homogéneos respecto a la variable que se desea discriminar. Las particiones se hacen de manera recursiva hasta que se alcanza un criterio de parada, el método utiliza datos históricos para construir el árbol de decisión, y este árbol se usa para clasificar los nuevos datos.

Timofeev (citado por Serna, 2009), plantea que el análisis de árboles de clasificación y regresión (CART) generalmente consiste en tres pasos:

1. Construcción del árbol máximo

Está construido utilizando un procedimiento de partición binaria, comenzando desde la raíz del árbol, este árbol es un modelo que describe el conjunto de entrenamiento y generalmente es sobreajustado.

2. Poda del árbol

El árbol es generalmente sobreajustado (podado), cortando sucesivamente ramas o nodos terminales hasta encontrar el tamaño adecuado del árbol.

3. Selección de árbol óptimo

Se requiere estimar el error de predicción con precisión y se hace con un procedimiento de validación cruzada, con el objetivo de

encontrar la proporción óptima entre la tasa de mala clasificación y complejidad del árbol (observaciones mal clasificadas/número de observación)

2.2.8. SISTEMA GESTOR DE BASE DE DATOS RELACIONAL

Nevado (2010) afirma que "El gestor de base de datos es un componente software encargado de garantizar el correcto, eficiente, íntegro y seguro acceso y almacenamiento de los datos. Además, proporciona una interfaz entre los datos almacenados y los programas de aplicación" (p.34).

Según Silberschatz (2002) un sistema gestor de bases de datos (SGBD) consiste en una colección de datos interrelacionados y un conjunto de programas para acceder a dichos datos. La colección de datos, normalmente denominada base de datos, contiene información relevante para una empresa. El objetivo principal de un SGBD es proporcionar una forma de almacenar y recuperar la información de una base de datos de manera que sea tanto práctica como eficiente. Los sistemas de bases de datos se diseñan para gestionar grandes cantidades de información. La gestión de los datos implica tanto la definición de estructuras para almacenar la información como la provisión de mecanismos para la manipulación de la información. Además, los sistemas de bases de datos deben proporcionar la fiabilidad de la información almacenada, a pesar de las caídas del sistema o los intentos de acceso sin autorización. Si los datos van a ser compartidos entre diversos usuarios, el sistema debe evitar posibles resultados anómalos.

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

3.1. TIPO DE INVESTIGACIÓN

La investigación es aplicada y tecnológica. Para Vara (2010), en una investigación aplicada, "el interés es práctico pues los resultados son utilizados inmediatamente en la solución de problemas de la realidad. La investigación aplicada normalmente identifica la situación problema y busca, dentro de las posibles soluciones, aquella que pueda ser la más adecuada".

3.2. NIVEL DE INVESTIGACIÓN

El nivel de investigación es descriptivo. Según Pérez (2012) en el nivel descriptivo se especifican los procedimientos involucrados en el procesamiento y generación de datos, y se refiere concretamente a los análisis macroscópicos y las tareas de experimentación que se efectuaron.

3.3. DISEÑO DE LA INVESTIGACIÓN

La presente investigación es no experimental, porque no se varían las variables de investigación. Según Hernández, Fernández y Baptista (2010), las investigaciones no experimentales son "aquellas que se realizan sin manipular deliberadamente las variables, es decir, no se varía intencionalmente la variable independiente, simplemente lo que se hace es observar las funciones tal y como se dan en su contexto natural para después analizarlo".

3.4. POBLACIÓN Y MUESTRA

POBLACIÓN

La población estuvo compuesta por la cartera de créditos de la Cooperativa de ahorro y crédito San Cristóbal de Huamanga.

MUESTRA

La muestra fue no probabilística y estuvo compuesta por la cartera de créditos de la Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga de los años 2014(8,889), 2015(10,603) y 2016(12,640).

3.5. VARIABLES E INDICADORES

3.5.1. DEFINICIÓN CONCEPTUAL DE LAS VARIABLES

VARIABLE

Gestión de riesgo crediticio. Es la posibilidad de que una empresa incurra en pérdidas que disminuyan el valor de sus activos, resultados y patrimonio, como consecuencia de que sus deudores o contrapartes fallen en el cumplimiento de los términos acordados en los contratos de crédito.

INDICADORES DE LA VARIABLE

Clasificación crediticia relacionada con morosidad. La clasificación crediticia está determinada por la capacidad de pago del deudor, el grado de incumplimiento de sus funciones y su historial crediticio. Las calificaciones son: A (Normal), Con problemas potenciales (B), Deficiente (C), Dudoso (D) y Pérdida (E). Se relaciona con la morosidad porque para la clasificación toma como variable los días de atraso, categoría y mora. La mora es falta de puntualidad o retraso. Existen mora temprana, son créditos vencidos con menos de 60 días, mora avanzada es desde mora temprana a un año de vencido y cartera castigada con créditos con más de un año de vencidos.

Clasificación crediticia relacionada con cobertura de provisiones. Este ratio, entendida como fórmula de protección ante problemas para el futuro, es uno de los indicadores más utilizados para evaluar la calidad crediticia de las empresas y, por ende, para analizar el estado financiero de la entidad en cuestión. Se relaciona con la morosidad porque se provisiona de acuerdo a la morosidad, días de atraso y categoría. La mora es la falta de puntualidad o retraso.

3.5.2. DEFINICIÓN OPERACIONAL DE LAS VARIABLES

VARIABLE

X: Gestión de riesgo crediticio

INDICADORES

X1: Clasificación Crediticia relacionado con morosidad

X2: Clasificación crediticia relacionada con cobertura de provisiones

3.5.3. TÉCNICAS E INSTRUMENTOS

Según Bernal (2006) la investigación descriptiva soporta técnicas como la encuesta, la entrevista, la observación y la revisión documental. Por lo tanto, en esta investigación se aplicó (Anexo 2):

A. TÉCNICAS

- Observación
- Análisis documental

B. INSTRUMENTOS

- Guía de observación
- Ficha Bibliográfica

3.6. METODOLOGÍAS Y HERRAMIENTAS TECNOLÓGICAS

Las siguientes metodologías y herramientas permitieron desarrollar la minería de datos en la cooperativa.

3.6.1. METODOLOGÍA

Metodología CRISP-DM

3.6.2. HERRAMIENTAS

- Base de datos Oracle
- Lenguaje R y R – Studio

CAPÍTULO IV

RESULTADOS DE LA INVESTIGACIÓN

Para la implementación de la solución se ha considerado la importancia de la utilización de metodologías para procesos de Minería de Datos, por este motivo, y teniendo en cuenta los avances tanto a nivel académico como industrial, se ha optado por utilizar CRISP-DM, por ello, este capítulo se estructura de acuerdo con las fases de la misma.

4.1. FASE I. COMPRENSIÓN DEL NEGOCIO

4.1.1. DETERMINACIÓN DE LOS OBJETIVOS DEL NEGOCIO

El objetivo del negocio es determinar comportamientos de los créditos otorgados en cuanto a morosidad, calificación y provisiones. Es de utilidad para la toma de decisiones el poder determinar características de los créditos donde están propensos a que un crédito caiga en mora.

A. CONTEXTO ACTUAL

La Cooperativa de Ahorro y Crédito San Cristóbal de Huamanga Ltda. N° 064 cuenta con más de 55 años de experiencia, crecimiento y solidez, dedicada a brindar productos financieros competitivos y servicios de calidad que buscan permanentemente satisfacer las necesidades de los socios, contribuyendo a su crecimiento, bienestar económico y social.

De acuerdo con la FENACREP la cooperativa obtuvo en las ratios de riesgo crediticio un porcentaje de morosidad, en diciembre de 2013 la morosidad fue de 5.49%, en noviembre de 2014, 6.66%, en diciembre de 2015, 6.65% y en noviembre de 2016, 8.20%. Como se muestra la morosidad ha ido aumentando, lo que preocupa al área de riesgos.

Dentro de la organización de la cooperativa los órganos de gobierno de la cooperativa están establecidas en concordancia con la Ley General de las Cooperativas y por el Estado (En la figura N° 4.1 se muestra el organigrama de la cooperativa):

1. ÓRGANO DE GOBIERNO. La Asamblea General de Delegados es la autoridad suprema de la cooperativa.

2. ÓRGANO DE DIRECCIÓN. El Consejo de Administración, es el órgano responsable del funcionamiento administrativo de la cooperativa.

3. ÓRGANOS DE CONTROL. El consejo de Vigilancia, es el órgano fiscalizador de la cooperativa para el cumplimiento de funciones, cuenta con el área de auditoría interna.

4. ÓRGANOS DE APOYO. Los órganos de apoyo reconocidos por el Estatuto son el Comité Electoral y el Comité de Educación.

5. OTROS ÓRGANOS DE APOYO. Los otros órganos de apoyo son los Comités y las Comisiones Especiales, El Comité de Créditos, Comité de Riesgos, Comité de Fondo de Previsión Social, Comité del Fondo de Desgravamen, Comisión de Adquisiciones y Contrataciones, Comisión de Selección de Personal, Comisión de Salud, Comisión de Riesgos de Lavado de Activos, Comisión de Altas y Bajas, Comisión de Desarrollo Institucional y Comisiones Especiales.

B. OBJETIVOS DEL NEGOCIO

El objetivo del negocio para el presente trabajo es detectar el comportamiento de los créditos otorgados en los últimos 3 años por la cooperativa.

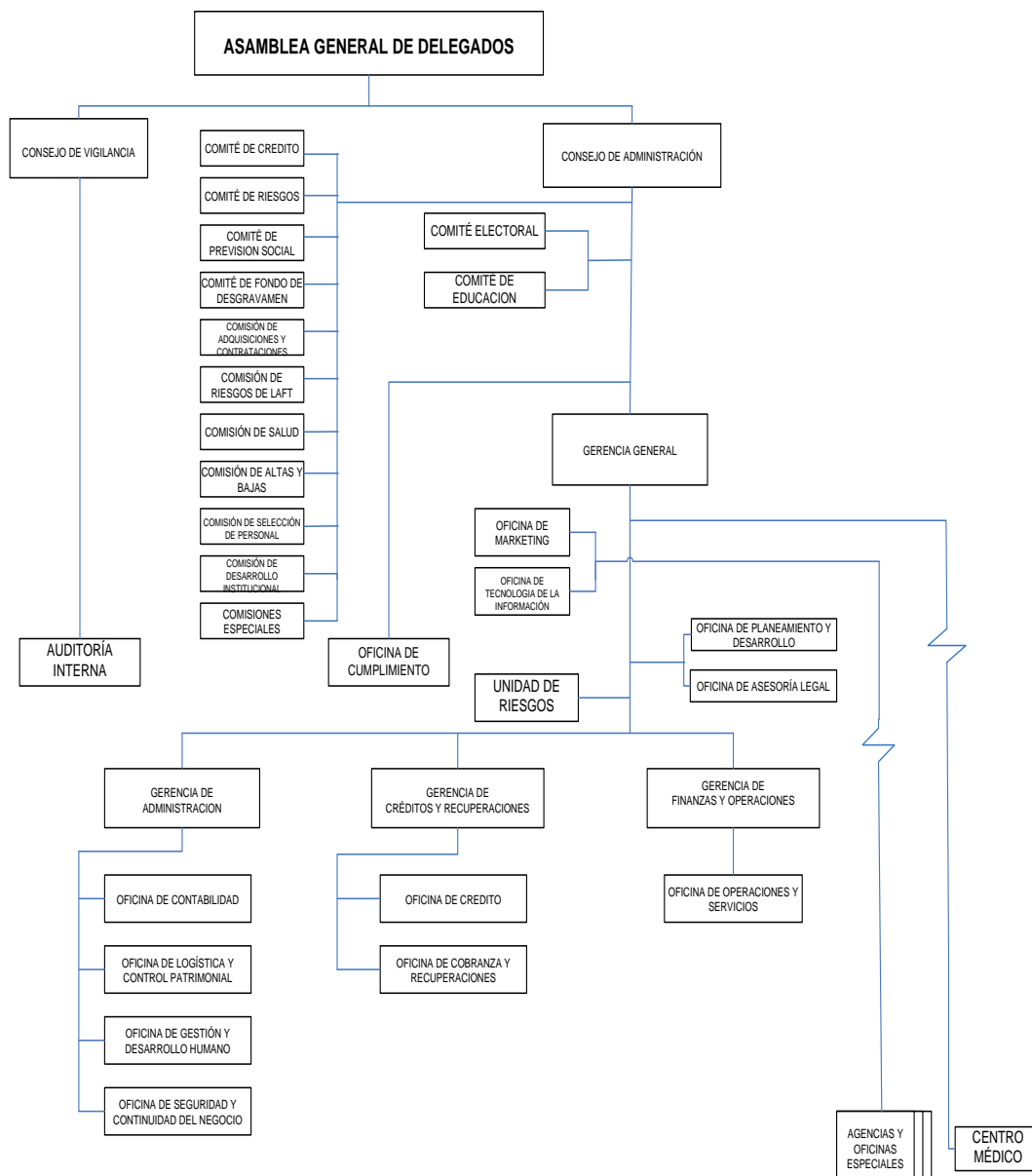


Figura 4.1. Organigrama de la Cooperativa San Cristóbal (Manual de organización y funciones, 2016)

C. SOLUCIÓN ACTUAL

La solución para resolver el problema de mora ha sido bonificar a los gestores de cobranzas por cada crédito vencido cobrado, también se da facilidades de pago condonando la mora a los socios que pidan una consideración en sus pagos. Otra medida que se tomó fue contratar asesores externos (abogados) en temas judiciales para cada agencia con

alta mora. También se tomó como solución la calificación de socios mediante un "scoring" donde, de acuerdo con una evaluación de parámetros, se obtiene la calificación de un socio.

4.1.2. VALORACIÓN DE LA SITUACIÓN

A. SITUACIÓN

Los sistemas de información de la Cooperativa San Cristóbal de Huamanga se encuentran desarrollados en varias plataformas, usando Oracle Builder y otros pequeños sistemas desarrollados en java, tanto en escritorio como en web, todos ellos utilizando como motor de base de datos Oracle 11G y para cada sistema utiliza varios esquemas de base de datos.

Se dispone en la fecha de una base de datos integrada, en la que todos los sistemas de información consumen los mismos datos evitando así la duplicidad de datos y la redundancia. El principal sistema de Core Financiero que cuenta la cooperativa es BESTERP. La observación a tener en cuenta en el Sistema Besterp es que tiene varios paquetes que no se utilizan y tablas que han sido creadas y no se utilizan. Desde que se empezó a utilizar el sistema BESTERP se realizó la migración de datos desde el año 2009, observándose que existen datos duplicados e inconsistentes. No existe un diccionario de datos que permita la comprensión del modelado.

También se observa que existen procesos en el sistema que no validan los registros de personas, lo que conlleva a tener los registros duplicados. Se encontró también que en recursos humanos tiene tablas donde se vuelven a registrar a los trabajadores generando redundancia de datos. En la data se observa que no se tiene un patrón de normalización de base de datos física.

B. FUENTES DE PERSONAL

La Cooperativa San Cristóbal de Huamanga cuenta con el área de sistemas, constituida por 6 profesionales de área de sistemas, un encargado de base de datos, uno que se encarga del desarrollo y mantenimiento de los sistemas de gestión, un encargado del soporte técnico de redes y servidores, un encargado de la seguridad y el jefe del área. En el equipo no existe personal dedicado a tareas relacionadas a inteligencia de negocios.

C. RECURSOS DE SOFTWARE

1. Sistema BESTERP

El Core Financiero de la Cooperativa se encuentra desarrollado con la tecnología "Oracle Forms" y "Reports Builder", utiliza como motor de base de datos a Oracle 11g. Entre los principales procesos que realiza se tiene:

- El proceso de admisión se encarga del registro de datos de personas, empresas, socios, trabajadores. También realiza cobros por inscripción.
- El proceso de Operaciones, se encarga de realizar las distintas operaciones de depósito y de retiros, así como la administración de la bóveda, cajas y transferencias a bancos.
- El proceso de créditos, donde se hacen la evaluación y aprobación de créditos, el seguimiento de las cobranzas y riesgo de créditos.
- El proceso de recursos humanos, apoya en la administración del personal, administración de los beneficios, pagos de sueldos, planilla, etc.
- El proceso de contabilidad, lleva a cabo las tareas de contabilidad como inventarios, balances, pagos y los asientos contables generados en caja.

2. Sistema Logística

El Sistema de Logística es un sistema web desarrollado en plataforma java, utiliza Jasper Reports para los reportes, utiliza como motor de base de datos Oracle 11g y tiene su propio esquema. Se encarga de gestionar los inventarios de activos de la cooperativa, los requerimientos de las áreas y el control del almacén.

3. Intranet Institucional

La Intranet es un sistema web desarrollado en plataforma java, utiliza Jasper Reports para los reportes, utiliza como motor de base de datos Oracle 11g y tiene su propio esquema. Se encarga de gestionar el directorio institucional, visualización de los documentos internos, los anuncios y las agendas de la cooperativa.

4. Sistema de Evaluaciones

Es una aplicación que comparte el mismo esquema que la Intranet, se encarga de gestionar los exámenes virtuales, a configurar las preguntas y a evaluar los resultados. Está desarrollado en plataforma java y usa como motor de base de datos el Oracle 11g.

5. Sistema de registro de marcado y asistencia

Es una aplicación de escritorio que permite el marcado o registro de la asistencia mediante el lector de huellas, comparte el mismo esquema del Sistema BESTERP.

6. Sistema Saldo mático

Es una aplicación que apoya a los socios a hacer consultas de su estado de cuenta previo ingreso de un usuario y una contraseña. Utiliza pantallas táctiles para la manipulación e impresoras térmicas para la impresión de recibos con la información de las cuentas. Está

desarrollado con las mismas características que el sistema logístico.

4.1.3. DETERMINACIÓN DE LOS OBJETIVOS DE MINERÍA DE DATOS

Los objetivos en términos de minería de datos son:

1. Determinar el comportamiento de los créditos, el monto vencido y su calificación.
2. Determinar el comportamiento de las provisiones de los créditos y las calificaciones crediticias.
3. Determinar el comportamiento de los productos otorgados y sus calificaciones obtenidas.
4. Determinar el comportamiento de las categorías de créditos y sus calificaciones crediticias.

ACTIVIDADES DE CLASIFICACIÓN			
ACTIVIDAD	2014	2016	2017
SalDOS vencidos	X	X	X
Categorías de créditos	X	X	X
Producto de crédito	X	X	X
Provisiones de crédito	X	X	X
Calificación del crédito	X	X	X
Tasa mensual	X	X	X
Monto de préstamo	X	X	X

Tabla 4.1. Actividades del Clasificación (Elaboración Propia)

4.1.4. PRODUCCIÓN DE UN PLAN DE PROYECTO

Se tomaron muestras de la base de datos en producción según los puntos descritos en el ítem anterior. A cada una de las muestras tomadas se analizó el contenido y se realizaron trabajos de normalización de información y adaptación de la información, para ello se ejecutaron las siguientes tareas:

- Analizar preliminarmente los datos.
- Seleccionar herramientas de software para el proceso.
- Implementar procesos de captura de datos.
- Implementar modelos de minería de datos
- Ejecutar pruebas sobre los sets de datos.

En las siguientes fases se detallaran las tareas.

4.2. FASE II. COMPRENSIÓN DE LOS DATOS

4.2.1. RECOLECCIÓN DE DATOS INICIALES

A. ADQUISICIÓN DE LOS DATOS DEL SISTEMA BESTERP

Los datos que se utilizan en el sistema BESTERP se encuentran en tablas relacionadas. Estas son: tablas de la evaluación, tablas de cartera de socios, tablas estado de cuenta y tablas de préstamos. La adquisición de los datos de los socios se registra al momento de la inscripción, se tiene el problema de registros duplicados y datos con errores de digitación, lo que hace el área de sistemas es unificar los datos uno por uno. Los registros no son completos antes de 2016, solo se encuentra información básica. A partir de julio de 2016, se estableció la política de actualización de la data de los socios al momento de realizar sus retiros. La tabla de préstamos contiene los diferentes registros de las evaluaciones de los préstamos que gestionan los analistas, se observa en algunas veces que los créditos otorgados han sido llenados con información inadecuada e incompleta, que genera un mayor riesgo que un crédito caiga en mora. La tabla de cartera es muy importante porque tiene toda la información resumida de los créditos y el historial de los créditos.

B. INFORME DE RECOLECCIÓN DE DATOS INICIALES

A continuación, se describen los primeros datos iniciales que nos servirán para armar nuestra tabla CARTERA_GERENCIAL, se ha utilizado para este resultado la guía de observación plasmada en el anexo 02 y a

continuación se describe los resultados:

COMPONENTES	DESCRIPCIÓN
Categoría del crédito	Se tomarán las categorías de los créditos que van desde los créditos consumo al hipotecario.
Calificación del crédito	Se tomará en rangos según días de atraso y categoría que van desde la A hasta el E.
Fecha desembolso	La fecha que se otorgó el crédito.
Tasa Mensual	Se procesará las tasas de interés otorgados en los créditos.
Días de atraso	Se tomarán los días de atraso de cada crédito en mora.
Monto vencido	Se tomarán el saldo vencido de los créditos que vencen de acuerdo con la categoría.
Producto	Los diferentes productos otorgados por la cooperativa.
Estado	Los estados que pasa el crédito desde el tiempo vigente al judicial.
Saldo Capital	Es la cantidad que le falta por pagar al crédito.
Provisión	El monto que se resguarda de acuerdo con el tipo de crédito.

Tabla 4.2. Componentes de CARTERA_GERENCIAL (Elaboración Propia)

Para la extracción se ha establecido el promedio de plazo en las cuotas hasta el 2016 y se ha obtenido que el promedio del plazo para otorgamiento de créditos es de 24 meses. Por lo tanto, para realizar el seguimiento del comportamiento de los créditos se tomarán 24 meses atrás, es decir, créditos desembolsados desde el 2014 hasta el 2016.

C. REQUERIMIENTOS DE INFORMACIÓN

La información necesaria para el logro de los objetivos del presente trabajo ha sido descrita en los puntos anteriores y se encuentran

disponibles en la base de datos.

D. CRITERIOS DE SELECCIÓN DE LOS DATOS

Los atributos necesarios para la minería de datos son:

Atributos	Monto de préstamo
	Fecha de desembolso
	Días de atraso
	Categoría del crédito
	Monto vencido
	Saldo Total
	Agencia
	Producto
	Estado
	Calificación del crédito

Tabla 4.3. Atributos de la minería de datos (Elaboración Propia)

Tablas seleccionadas para la extracción de datos:

Tablas	PR_PRESTAMO
	PR_CUOTA_PRESTAMO
	PR_PROD_PRESTAMO
	PR_CUOTA_DIARIO
	PR_PROVISION_DET
	PR_DET_SALDO
	PR_DET_EJEC_CARTERA
	CS_OFICINA_ORG
	PR_CFG_CATEG_PRES
	PR_VERSION_PRES

Tabla 4.4. Tablas seleccionadas (Elaboración Propia)

4.3.2. DESCRIPCIÓN DE LOS DATOS

La siguiente lista muestra los tipos de datos extraídos:

ATRIBUTO	TIPO DE DATO
Monto de préstamo	Numérico
Fecha de desembolso	Fecha
Días de atraso	Numérico
Categoría del crédito	Cadena
Monto vencido	Numérico
Saldo Total	Numérico
Agencia	Cadena
Estado	Cadena
Calificación del crédito	Cadena

Tabla 4.5. Tipo de dato de CARTERA_GERENCIAL (Elaboración Propia)

A continuación se describe detalladamente el contenido de las tablas utilizadas para la extracción de los datos que sirven para armar la tabla CARTERA_GERENCIAL. Se ha utilizado la guía de observación del anexo N° 2 para poder plasmar las siguientes tablas:

La tabla N° 4.6 muestra los datos almacenados en la tabla PR_PRESTAMO que contiene datos de las tasas, monto de préstamo y los estados de un préstamo, en 'VIG' es un préstamo vigente, 'REF' es un préstamo refinanciado, en 'CAN' es cancelado y 'JUD' es un crédito que está en estado judicial.

PR_PRESTAMO		
Nombre de Columna	Tipo de datos	Longitud
idPrestamo	NUMBER	18
idProdPres	Number	18
idMoneda	Number	18

tasaMensual	Number	14,2
mntoPrestamo	Number	14,2
idTercero	NUMBER	18
cantCuotas	Number	10
idOficOrg	Number	18

Tabla 4.6. Tabla PR_PRESTAMO (Elaboración Propia)

La tabla N° 4.7 muestra los datos almacenados en la tabla PR_CUOTA_PRESTAMO donde contiene el plan de pagos del crédito, el saldo capital, los intereses, las cuotas mensuales y la fecha de vencimiento de cada cuota.

PR_CUOTA_PRESTAMO		
Nombre de Columna	Tipo de datos	Longitud
IDCUOPRES	NUMBER	18
IDPRESTAMO	Number	18
SLDOCAPITAL	Number	14,2
INTERES	Number	14,2
FECVENC	Number	14,2
STSCUOTA	VARCHAR2	2
MTOTOTAL	Number	14,2

Tabla 4.7. Tabla PR_CUOTA_PRESTAMO (Elaboración Propia)

La tabla N° 4.8, muestra a los datos almacenados en la tabla PR_CUOTA_DIARIO, que actualiza diariamente la información de las cuotas del préstamo, la mora, los intereses, cargos, el monto vencido y judicial.

PR_CUOTA_DIARIO		
Nombre de Columna	Tipo de datos	Longitud
IDCUODIA	NUMBER	18

IDCUOPRES	Number	18
IDPRESTAMO	NUMBER	18
FECTRXX	DATE	11
SLDOCAP	Number	14,2
SLDOMORA	Number	14,2
SLDOINTERES	NUMBER	14,2
SLDOTOT	Number	14,2

Tabla 4.8. Tabla PR_CUOTA_DIARIO (Elaboración Propia)

La tabla N° 4.9 muestra la tabla PR_DET_EJEC_CARTERA, donde se guarda toda la información de la cartera de créditos por cada cierre de mes. Esta tabla proporciona un resumen de la situación del crédito y guarda el historial de los créditos, sus estados, días de atraso y las cuotas vencidas.

PR_DET_EJEC_CARTERA		
Nombre de Columna	Tipo de datos	Longitud
idDetEjecCar	Number	18
idCabEjeCar	Number	18
idPrestamo	Number	18
canCuoVen	Number	10
diasVenc	Number	10
idMoneda	Number	18
stsPrestamo	Varchar2	6
mtoDesemb	Number	14,2
idCfgCatPres	Number	18
idOficOrg	Number	18

Tabla 4.9. Tabla PR_DET_EJEC_CARTERA (Elaboración Propia)

La tabla N° 4.10 muestra los datos almacenados en la tabla CS_OFICINA_ORG

donde se almacena el nombre de las agencias de la cooperativa.

CS_OFICINA_ORG		
Nombre de Columna	Tipo de datos	Longitud
IDOFICORG	NUMBER	18
NOMOFI	Number	18
INICIAL	NUMBER	18
INDOFIMAIN	DATE	11
STSOFIORG	Number	14,2

Tabla 4.10. Tabla CS_OFICINA_ORG (Elaboración Propia)

La tabla N° 4.11 muestra la tabla PR_PROVISION_DET, en esta se guarda toda la información de las provisiones de los créditos. Esta tabla proporciona los datos de provisiones con garantía y provisión sin garantía, procesados diariamente y mensualmente.

PR_PROVISION_DET		
Nombre de Columna	Tipo de datos	Longitud
IDPROVDET	Number	18
IDPROVCAB	Number	18
DIASATRAZO	Number	6
PROVSINGARANT	Number	14,2
PROVNOGARANT	Number	14,2
IDRESTAMO	Number	18
IDTERCERO	NUMBER	18
MTOGARUSO	Number	14,2

Tabla 4.11. Tabla PR_PROVISION_DET (Elaboración Propia)

La tabla N° 4.12 muestra los datos almacenados en la tabla PR_CFG_CATEG_PRES donde se almacena las categorías de los créditos.

PR_CFG_CATEG_PRES		
Nombre de Columna	Tipo de datos	Longitud
IDCFGATPRES	NUMBER	18
CODCATEG	VARCHAR2	6
DESCCATEG	VARCHAR2	256
FECINI	DATE	11
FECFIN	DATE	11

Tabla 4.12. Tabla PR_CFG_CATEG_PRES (Elaboración Propia)

La tabla N° 4.13 muestra los datos almacenados en la tabla PR_VERSION_PRESTAMO, tiene los mismos atributos de la tabla PR_PRESTAMO, pero allí se guardan el historial de cambios de un préstamo, como por ejemplo los estados y el historial.

PR_VERSION_PRESTAMO		
Nombre de Columna	Tipo de datos	Longitud
IDVERSIONPRES	NUMBER	18
idPrestamo	NUMBER	18
idProdPres	Number	18
idMoneda	Number	18
tasaMensual	Number	14,2
mtoPrestamo	Number	14,2
idTercero	NUMBER	18
cantCuotas	Number	10
IdOficOrg	Number	18
NROVERSION	NUMBER	4
FECVERSION	DATE	11

Tabla 4.13. Tabla PR_VERSION_PRESTAMO (Elaboración Propia)

4.3.3. EXPLORACIÓN DE DATOS

Se realizó la exploración de los datos sobre la información extraída de los créditos desembolsados desde el año 2014 al 2016. En los resultados iniciales mostrará la información estadística que servirá principalmente para determinar la consistencia y completitud de datos.

De la exploración inicial de datos, se dispone para este estudio al cierre anual de diciembre 2014, diciembre 2015 y de diciembre 2016. Cada periodo será extraído de la base de datos y será almacenada en las variables como se muestra de la siguiente manera:

VARIABLE	AÑO	CRÉDITOS VIGENTES
datos_12_14	diciembre 2014	8,899
datos_12_15	diciembre 2015	10,603
datos_12_16	diciembre 2016	12,640

Tabla 4.14. Asignación de variables (Elaboración Propia)

Para el informe de exploración de datos se detallan las diferentes cantidades de datos a analizar.

INFORME DE EXPLOTACIÓN	CANTIDAD
Cantidad total de registros a explotar	32,142
Cantidad total de créditos atrasados que intervienen	1,595
Cantidad total de agencias que intervienen	19
Categoría de créditos	6

Tabla 4.15. Informe de explotación (Elaboración Propia)

En las siguientes figuras mostramos los resultados de la exploración inicial con la información que se tiene para este informe. Se muestran gráficos de cajas,

gráfico de barras y diagramas pastel para una mejor interpretación de los resultados.

El código que se muestra sirve para obtener la figura N° 4.2:

```
>datos_12<- rbind(datos_12_14,datos_12_15,datos_12_16);  
> anio<-  
table(format(as.Date(datos_12$FECDSEMBOLSO,format="%d/%m/%Y"),"%Y"))  
> bpplot <-plot(anio,col="blue", main="Cantidad por  
año",las=2,ylab="Cantidad")  
>text(bpplot,c(1,1), (anio),pos=3)
```

La figura N° 4.2, nos muestra la cantidad de créditos vigentes en cada periodo.

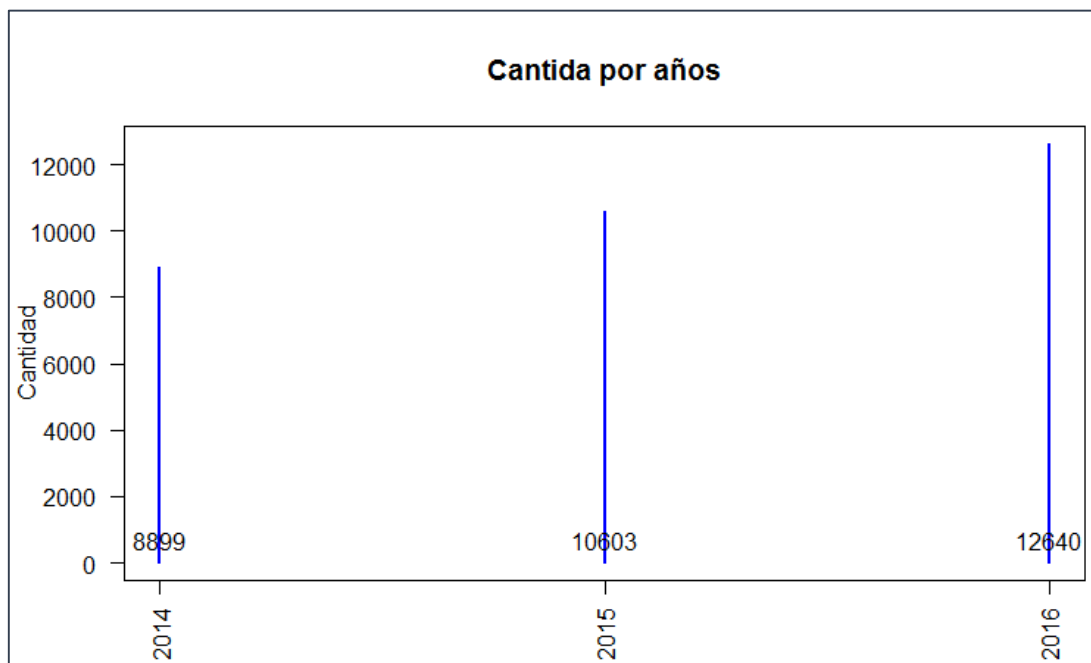


Figura 4.2. Cantidad de créditos por cierre de año (Elaboración propia, 2017)

El código que se muestra sirve para obtener la Figura 4.3:

```
> bpdatos<-  
barplot(table(datos_12$CATEGORIA),las=2,ylab="Cantidad",col=1:5,main="Creditos otorgados por Categoria",names.arg = c("Consumo Rev", "Grandes", "Medianas", "Micro", "Pequeñas", "Consumo no Rev", "Vivienda"),cex.names =0.8)  
>text(bpdatos,c(1,1),table(datos_12$CATEGORIA),pos=3)
```

La Tabla 4.16, nos muestra la cartera de créditos por categoría:

CATEGORÍA	CRÉDITOS VIGENTES
Créditos a Grandes empresas	2
Créditos a Medianas empresas	48
Créditos a Pequeñas empresas	4,611
Créditos a Microempresas	15,572
Créditos de Consumo no Revolvente	11,713
Créditos Hipotecarios para vivienda	205

Tabla 4.16. Créditos por categoría (Elaboración Propia)

La Figura 4.3, nos muestra la cartera de créditos por categoría en un gráfico de barras:

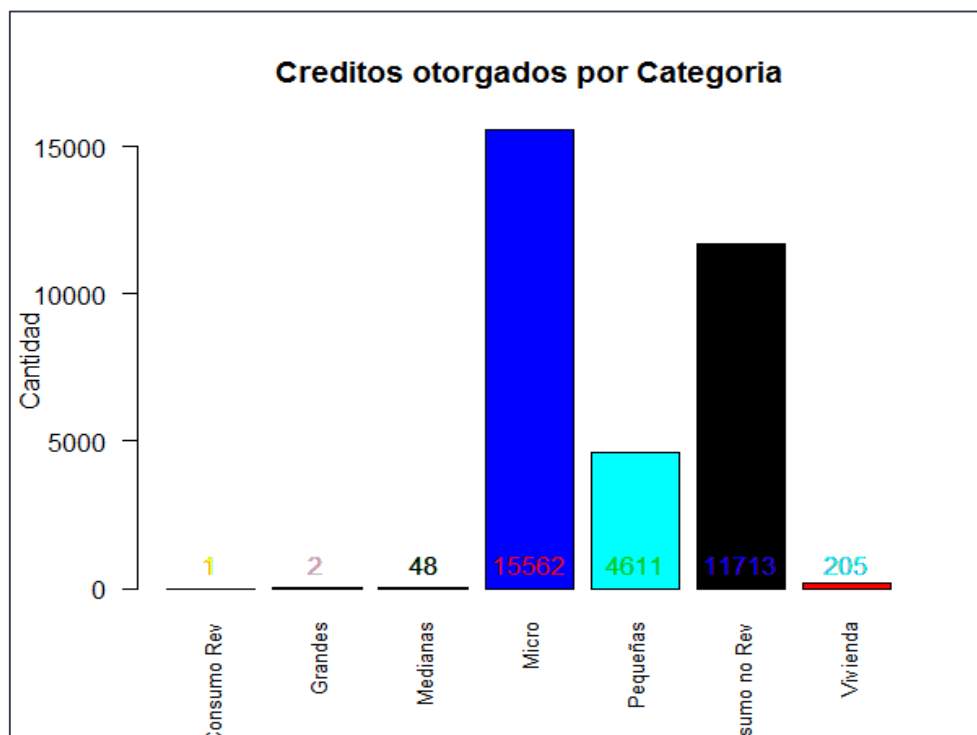


Figura 4.3. Cantidad de créditos otorgados por categoría (Elaboración propia, 2017)

El código que se muestra sirve para obtener la Figura 4.4:

```
> par(ps=10)
> boxplot(DIASATRASO ~
AGENCIA,data=datos_12,col=1:19,main="Box",las=2,ylab="Días de Atraso")
```

La figura 4.4, nos muestra los días de atraso por agencia:

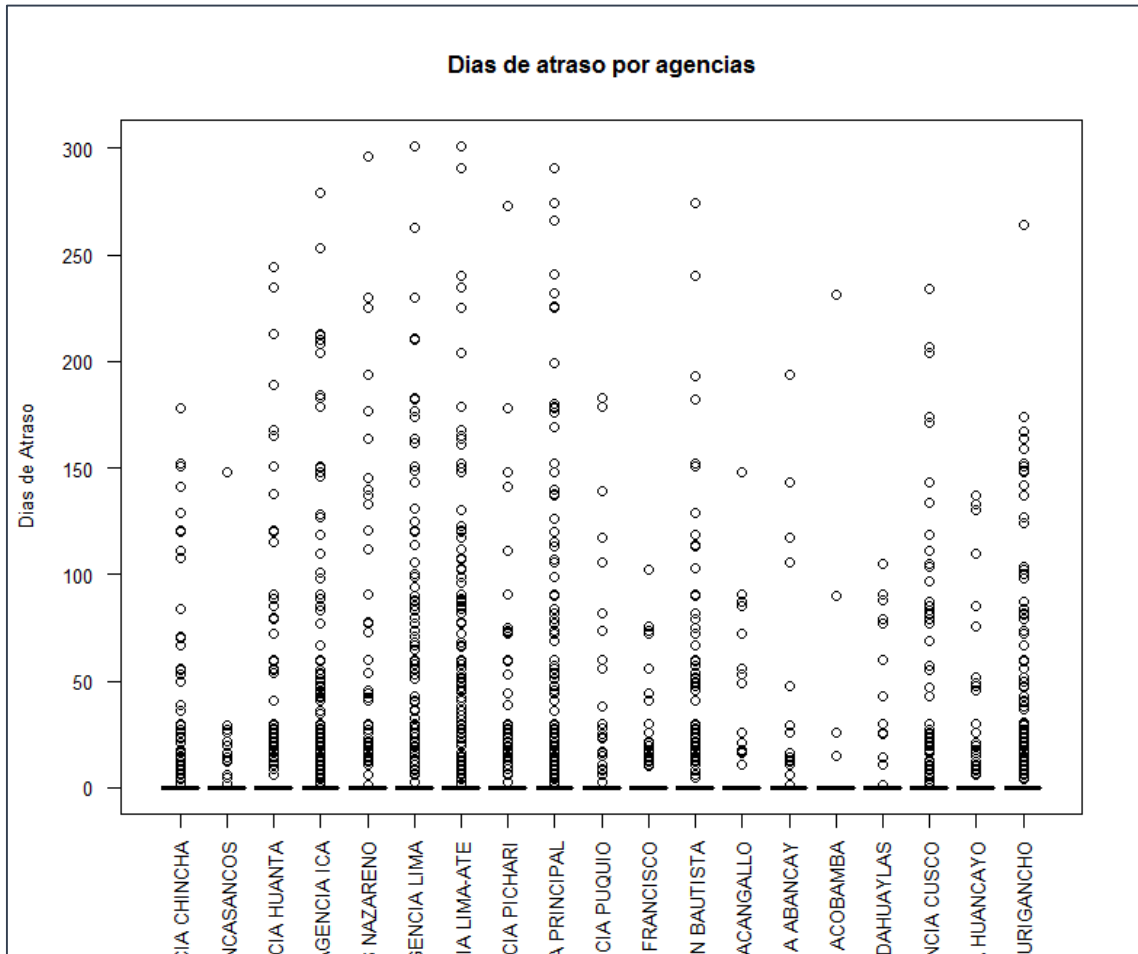


Figura 4.4. Días de atraso por agencias (Elaboración propia, 2017)

El código que se muestra sirve para obtener la figura 4.5:

```
> par(ps=10)
> bpdatos<-barplot(table(datos_12$AGENCIA),las=2,ylab="Cantidad de
créditos",col=1:19,main="Creditos otorgados por Agencia")
> text(bpdatos,c(6,2),table(datos_12$AGENCIA),srt=90,pos=3,adj=4,4:19,c
ol=21:40)
```

La figura 4.5, nos muestra la cantidad de créditos otorgados por agencia:

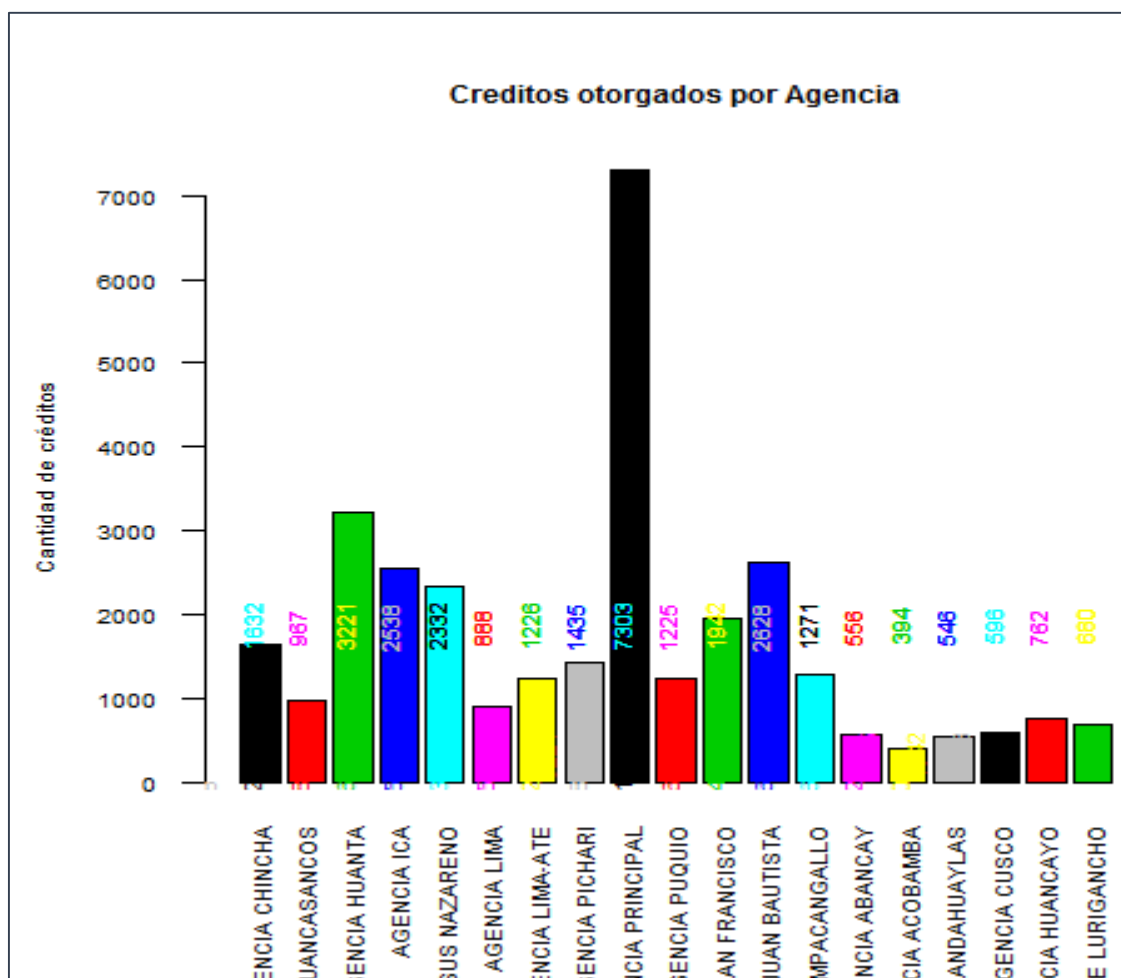


Figura 4.5. Cantidad de créditos por Agencia (Elaboración propia, 2017)

El código que se muestra sirve para obtener la figura 4.6:

```
>par(ps=10)
>porcentaje<-round(prop.table(table(datos_12$AGENCIA))*100,2)
>label<-paste(porcentaje,"%",sep=" ")
> color<-
c("white","antiquewhite1","antiquewhite4","aquamarine4","azure3","black",
"blue","blueviolet","brown1","burlywood4","chartreuse1","chocolate1",
"cornflowerblue","cyan","darkgoldenrod1","darkgreen","darkkhaki","deppink",
"yellow")
>pie(porcentaje,labels = label,clockwise = TRUE,main="Porcentaje de
Créditos por agencia",col= color)
> legend("left",rownames(table(datos_12$AGENCIA)),cex=0.8,fill= color)
```

La figura 4.6, nos muestra el porcentaje de créditos por agencia:

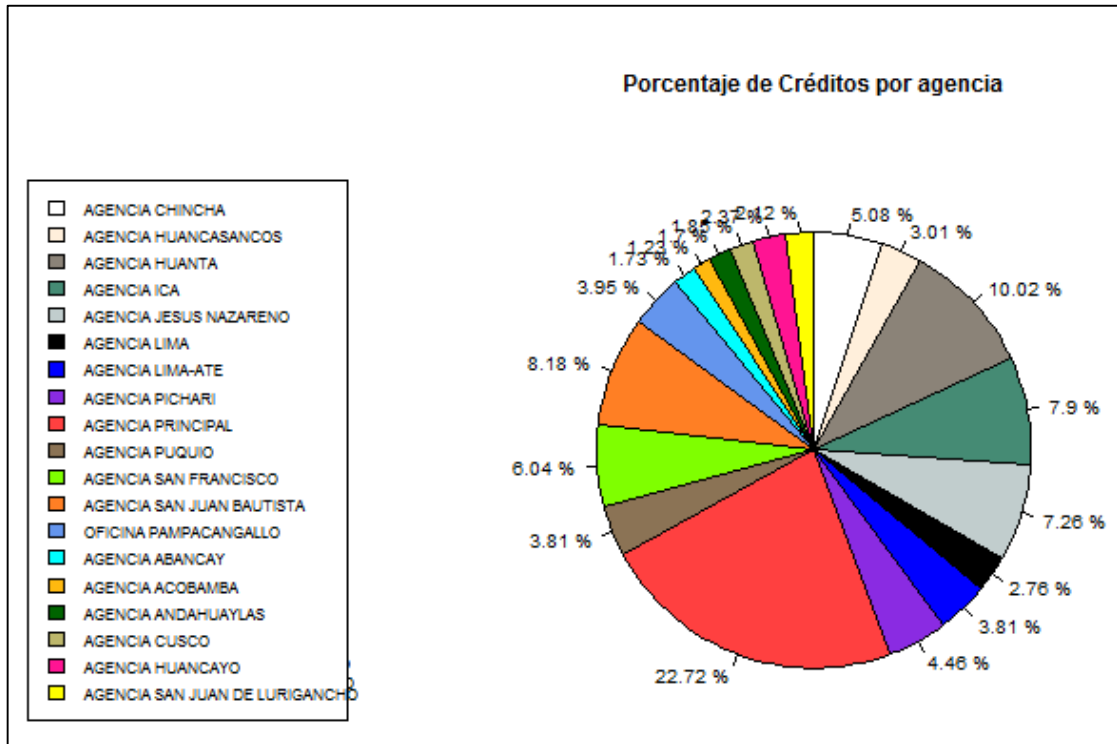


Figura 4.6. Cantidad de créditos por agencia (Elaboración propia, 2017)

4.3.4. VERIFICACIÓN DE CALIDAD DE DATOS

En función a la exploración inicial de datos se concluye que existe una categoría llamada Consumo Revolvente, como se muestra en el gráfico N° 4.4, por conocimiento obtenido en la investigación, la cooperativa no ofrece esa categoría de créditos, por lo tanto, no se tomó en cuenta en nuestro trabajo.

4.4. FASE III. PREPARACIÓN DE LOS DATOS

4.4.1. SELECCIÓN DE DATOS

Para seleccionar los datos de la base de datos y hacer la preparación se ha realizado una consulta a la base de datos de las tablas necesarias y los atributos necesarios para la explotación. Toda esta data final se guardará en una tabla llamada CARTERA_GERENCIAL. En el anexo N° 1 mostramos la consulta SQL de la selección de los datos primarios.

4.4.2. LIMPIEZA DE DATOS

Para realizar este paso se realizó la lectura de los datos de la tabla CARTERA_GERENCIAL, ha sido necesario el uso del software R Studio. A continuación se muestra la conexión, el uso de la librería 'RJDC' y la importación de la librería ojdbc7.

Conexión a la base de datos:

```
> install.packages("RJDBC");
> library(RJDBC);
> driver<-
  JDBC(driverClass="oracle.jdbc.OracleDriver", "C:/Users/davis/Documents/swirl_temp/ojdbc7.jar")
> conexion<-
  dbConnect(driver, "jdbc:oracle:thin:@192.168.1.1:puerto:orcl", "USUARIO", "CLAVE");
> options(java.parameters = "-Xmx2024m")
```

La base de datos contiene información relacionada con créditos de varios productos. Para esta investigación no se considera 3 productos (Cubierto M.N, Credimax M.N. y Credimax M.E.), porque estos productos no conllevan riesgo de recuperación, en caso de incumplimiento se descuentan de sus propias cuentas.

Los créditos cancelados serán excluidos, porque son créditos que cumplieron su ciclo y no genera ningún riesgo en su recuperación.

Cada dato se almacena en las variables datos_dec_16, datos_dec_15, datos_dec_14 extraída de los años 2016, 2015 y 2014. El código que se muestra sirve para obtener las variables:

```
> datos_dec_16<-sqlQuery(conexion," SELECT * FROM
CARTERA_GERENCIAL1 WHERE ANIO IN ('2016') AND MES IN ('12') AND
TO_CHAR(FECDESEM,'YYYY') = '2016' AND STSPRESTAMO IN
('VIG', 'REE', 'REF', 'JUD') AN CODPRODUCTO NOT IN
('201', '221', '247', '249') ");

> datos_dec_15<-sqlQuery(conexion," SELECT * FROM
CARTERA_GERENCIAL1 WHERE ANIO IN ('2015') AND MES IN ('12') AND
TO_CHAR(FECDESEM,'YYYY') = '2015' AND STSPRESTAMO IN
('VIG', 'REE', 'REF', 'JUD') AN CODPRODUCTO NOT IN
('201', '221', '247', '249')");

> datos_dec_14<-sqlQuery(conexion," SELECT * FROM
CARTERA_GERENCIAL1 WHERE ANIO IN ('2014') AND MES IN ('12') AND
TO_CHAR(FECDESEM,'YYYY') = '2014' AND STSPRESTAMO IN
```

```

('VIG', 'REE', 'REF', 'JUD') AN CODPRODUCTO NOT IN
('201', '221', '247', '249')");

> odbcClose(conexion);

```

En la figura 4.7., se detalla el resultado final de la limpieza y cómo se muestra en el software R studio.

	CODSOCIO	NOMBRES	NROPRESTAMO	CODUSR	MTOPRESTAMO	FECDSEMBOLSO	TASAMENSUAL
1	10316591	GARCIA MENDOZA DENYS ALEXANDER	2027522	GECI2	23500.00	04/01/2013	
2	10344926	GARCIA RAMOS GUILLERMO	2027580	RENF	12000.00	04/01/2013	
3	1051092326	QUISPE GUZMAN JULIANA BETSABE	2027655	VEVN	17000.00	05/01/2013	
4	1051269483	TASAYCO TASAYCO ABRAHAM EMILIANO	2028105	VEVN	20000.00	08/01/2013	
5	1011257976	TAPAHUASCO CALDERON WILBER	2027952	HUTM	13200.00	09/01/2013	
6	1031271240	ECHACCAVA PAHUARA ROSA ELVIRA	2028344	GECI2	15000.00	10/01/2013	
7	1031255688	HUAMANI MORALES ARMANDO OTTO	2028544	GECI	19000.00	11/01/2013	
8	10339958	SALAZAR HUARCAYA EDWIN OMAR	2028618	URQO	24520.00	11/01/2013	

Figura 4.7. Carga de los datos a R Studio (Elaboración Propia, 2017)

4.4.3. ESTRUCTURAR LOS DATOS

Para el proceso de transformación y generación de la CARTERA_GERENCIAL se ha realizado una programación de una consulta SQL con las sentencias correspondientes para la estructuración de los datos, como se muestra en el anexo N° 1 y que fueron llamados desde un formulario.

La explotación de los datos se realizará sobre la base de datos y se realizará con procedimientos para obtener los resultados procesados. Los productos Cubierto, Credimax y los créditos cancelados excluidos en el análisis fueron filtrados con el software de análisis R.

4.4.4. INTEGRACIÓN DE DATOS

Los datos son extraídos y combinados a partir de tablas relacionadas por sus respectivas llaves. Los atributos finales se obtuvieron de las siguientes tablas:

TABLA	ATRIBUTO
PR_PRESTAMO	Nro Préstamo
	Monto préstamo
	Fecha desembolso
	Tasa Mensual
	Días atraso
	Saldo Total
	Estado
PR_CFG_CATEG_PRES	Categoría
CS_OFICINA_ORG	Agencia
PR_PROD_PRESTAMO	Producto

Tabla 4.17. Atributos de las tablas extraídas (Elaboración Propia)

Las consultas SQL del anexo N° 1 son capaces de integrar toda la información requerida para el presente trabajo. Todas las líneas de código para la generación de archivos en función de los períodos establecidos quedan almacenadas en archivos con formato query.

4.5. FASE IV: MODELADO

4.5.1. SELECCIONAR LA TÉCNICA DE MODELADO

Para esta fase de la metodología se ha elegido la construcción de árboles de decisión. El lenguaje R tiene librerías para aplicar árboles de decisión, la que hemos utilizado es la librería Rpart. Los métodos utilizados se detallan en la siguiente tabla.

MÉTODO	DESCRIPCIÓN
Sample	Para la selección de la muestra con el set de entrenamiento y test.
Rpart	Para crear el árbol de decisión
Print	Permite imprimir la información del árbol.

Rpart.plot	Permite visualizar el árbol graficado.
Printcp	Permite identificar la información de probabilidad de error en la creación del árbol.
Plotcp	Muestra gráficamente el comportamiento del error.
Prune	Este método permite podar el árbol en caso de que haya algún error.
Predict	Este método permite ver la matriz de confusión.
Table	Muestra la cantidad de aciertos y desaciertos del árbol procesado en la matriz de confusión.

Tabla 4.18. Métodos para crear el árbol de decisión (Elaboración Propia)

En la figura 4.8 se muestra el esquema de implementación de la investigación.

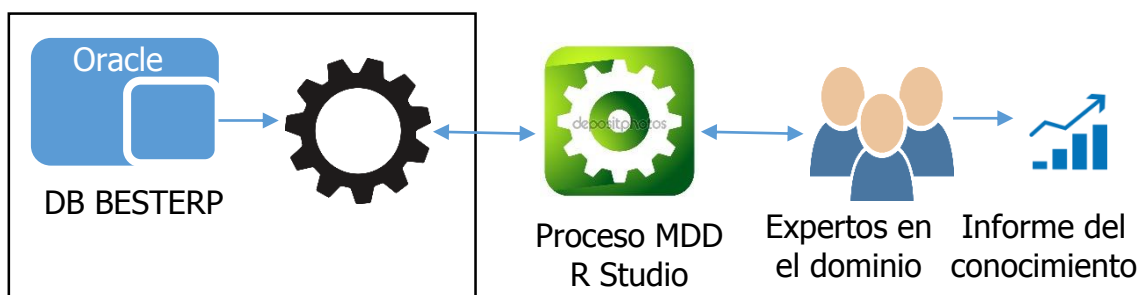


Figura 4.8. Esquema de implementación (Elaboración propia, 2017)

4.5.2. GENERACIÓN EL PLAN DE PRUEBA

En esta parte del modelado se eligió una muestra llamada el conjunto de entrenamiento para cada periodo, que es almacenada en la variable denominada "TrainData", en algunos casos se tomaron el 80% de la muestra y en otros el 60% de acuerdo con la mayor cantidad de aciertos obtenidos en la evaluación.

Para el conjunto de las pruebas fueron almacenados en la variable "Test Data", se han realizado muestras reducidas por cada periodo en algunos casos se tomó el 20% o el 40% de las muestras.

Durante las pruebas de diseño se desestimaron los atributos de agencia y días de atraso. La decisión fue que el resultado del árbol se hacía muy complejo y dificultaba su interpretación.

De las pruebas podemos concluir que en función de los resultados que se puedan llegar a obtener en el presente estudio, se podría tomar para futuras líneas de investigación otros datos para la obtención de patrones como edad, sexo y destino del crédito.

4.5.3. CONSTRUIR EL MODELO

A continuación, se describe la ejecución del modelo de clasificación en función de la técnica de árbol de decisiones utilizada en la investigación.

Para la construcción de un Modelo de Clasificación se analizó el comportamiento de los créditos desde su fecha de desembolso. Para la clasificación crediticia se usará la tabla 4.19.

VALOR	CALIFICACIÓN	DÍAS DE ATRASO
A	Normal	<0 a 8>
B	CPP	<9 a 30>
C	Deficiente	<31 a 60>
D	Dudoso	<61 a 120>
E	Pérdida	<120 a más>

Tabla 4.19. Calificación crediticia (SBS, 2008)

Para la evaluación de los resultados se mostrará paso a paso la elaboración del árbol de decisión para cada año.

a) MODELO DE MOROSIDAD CON CLASIFICACIÓN CREDITICIA

En esta parte del modelo se ha evaluado la morosidad y la clasificación

crediticia, en los resultados finales del árbol mostraremos cómo el comportamiento de la morosidad, la tasa y la categoría del crédito influyen en la clasificación final del crédito. A continuación, se muestra esta evaluación para los años 2014, 2015 y 2016.

PARA EL AÑO 2014

PASO 1. Los atributos utilizados para este análisis fueron el monto, la tasa, el monto vencido, la categoría y la clasificación crediticia que fueron elegidas de nuestra variable `datos_dec_14` que contiene la cartera a diciembre 2014.

El código que se muestra sirve para obtener las columnas del monto, tasa, monto vencido, categoría y clasificación crediticia del "dataframe":

```
> testTree<- datos_dec_14[,c(16,17,26,25,24)]
```

PASO 2. Seleccionaremos el número de muestras que necesitaremos para entrenamiento y para test. Con "sample" vamos a crear un vector que contendrá 2 grupos. El primer grupo `TrainData` tendrá el 80% de los elementos y el segundo grupo `TestData` tendrá el 20% de los elementos.

El código que se muestra sirve para obtener los sets de entrenamiento y test:

```
> ind <- sample(2,nrow(testTree),replace=TRUE,prob=c(0.8,0.2))
> trainData <- testTree[ind==1,] # entrenamiento
> testData <- testTree[ind==2,] # test
```

Paso 3. Creamos el árbol de clasificación con `Rpart`. Introducimos la variable dependiente `CLASIFICACIÓN` y explicamos con la selección de todas las demás variables. `Method="class"` significa que nuestra variable dependiente es cualitativa.

El código que se muestra sirve para calcular el árbol de clasificación:

```
> ArbolPart <- rpart(CALIFICACION ~.,
method="class",data=trainData)
```

Paso 4. Revisamos la información del gráfico. Donde n es el número de operaciones y Split es el número de cortes. Vemos que de 11,056 cortes la mayoría que es el 96% tiene clasificación A y 398 tendría clasificación B, C, D o E. El mejor corte que pudo hacer preguntarse es por el monto vencido si es menor o igual a 2.82 la mayoría 10,915 tendría clasificación A.

El código que se muestra sirve para mostrar la información del árbol de clasificación:

```
> print(ArbolPart)
```

```
1) root 11056 398 A (0.96 0.02 0.0052 0.0064 0.0049)
  2) MTOVENSOLES< 2.82 10915 257 A (0.98 0.02 0.0015
0.0016 0.00073) *
  3) MTOVENSOLES>=2.82 141 88 D (0 0.0071 0.29 0.38
0.33)
    6) MTOVENSOLES< 1171.74 45 20 C (0 0 0.56 0.38
0.067)
      12) TASAMENSUAL>=1.85 25 7 C (0 0 0.72 0.2 0.08)
      *
      13) TASAMENSUAL< 1.85 20 8 D (0 0 0.35 0.6 0.05)
      *
    7) MTOVENSOLES>=1171.74 96 53 E (0 0.01 0.17 0.38
0.45)
      14) CATEGORIA=CREDITOS A PEQUEÑAS
EMPRESAS,CREDITOS HIPOTECARIOS PARA VIVIENDA 31 19 D (0
0.032 0.32 0.39 0.26) *
      15) CATEGORIA=CREDITOS A MICROEMPRESAS,CREDITOS DE
CONSUMO NO REVOLVENTE 65 30 E (0 0 0.092 0.37 0.54) *
```

Analizando el árbol:

n= número de observaciones 11,056
Split= número de cortes

Nos indica que de un total de 11,056 observaciones la mayoría 96% tiene clasificación A. Si tiene un saldo vencido mayor a 2.82 la mayoría tendría clasificación D.

PASO 5. Gráfico del árbol de clasificación.

El código que se muestra sirve para obtener el gráfico del árbol de clasificación que se muestra en la figura 4.9:

> rpart.plot(ArbolPart)

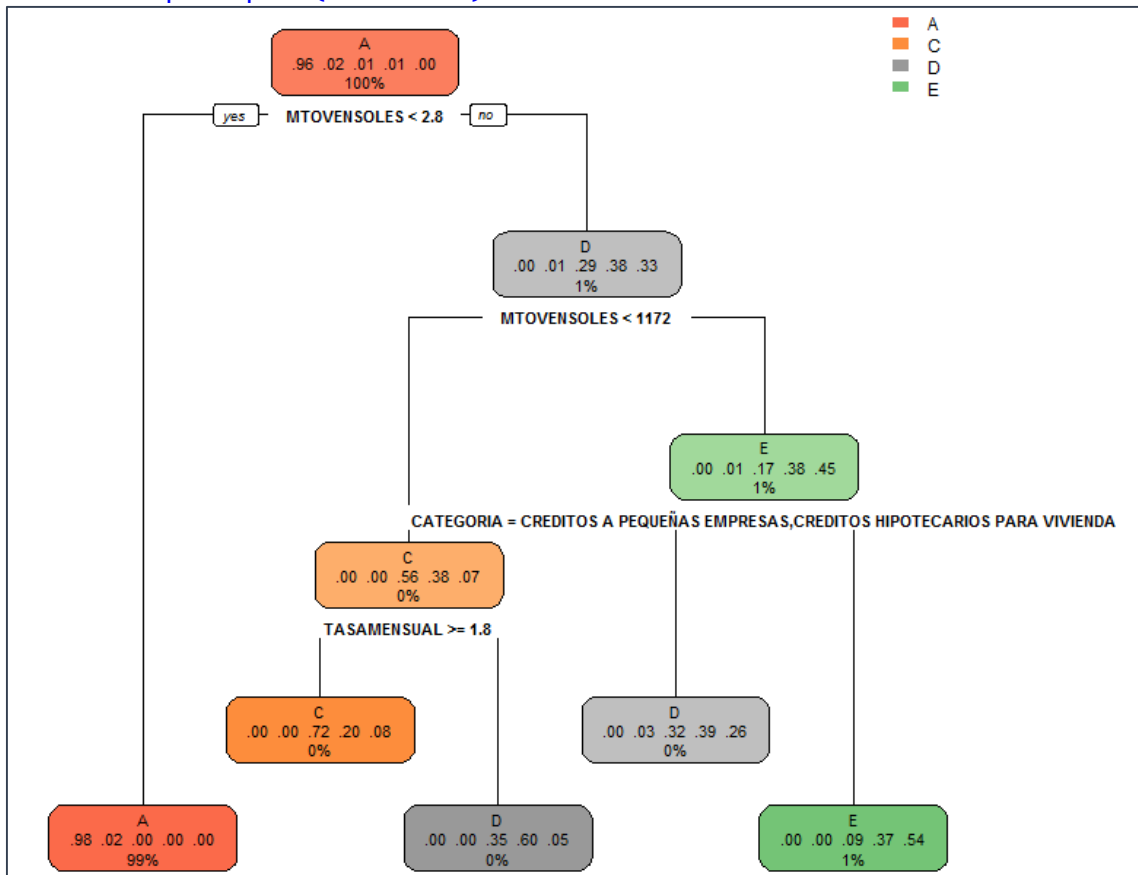


Figura 4.9. Gráfico del Árbol de decisión (Elaboración propia, 2017)

Conclusiones del árbol de la figura 4.9, son:

- La variable Monto vencido (MTOVENSOLES) es el mejor predictor para la Calificación crediticia con 5 clasificaciones: Normal (A), CPP (B), Deficiente(C), Dudoso (D) y Pérdida (E).
- La probabilidad más alta de tener una calificación Normal (96%) se da a los créditos que tengan un monto vencido (MTOVENSOLES) menor a 2.8.
- La probabilidad más alta de tener un calificación Deficiente (72%) se da entre los créditos que tengan un monto vencido (MTOVENSOLES) menor a 1,172 y una tasa mensual mayor a 1.8.
- La probabilidad más alta de tener una calificación Dudoso (35%) se da entre los créditos que tengan un monto vencido (MTOVENSOLES) menor a 1,172 y una tasa mensual menor a 1.8.

- La probabilidad más alta de tener una calificación Pérdida (54%) se da entre los créditos que tengan un monto vencido (MTOVENSOLES) mayor a 1,172 y no pertenezcan a las categorías Pequeña empresa e Hipotecarios para vivienda.

PASO 6. Visualizar las estadísticas de los resultados: Nos muestra que a partir de la cuarta fila incrementa el error de 0.86 a 0.87, es allí donde podaremos el árbol identificando el cp: 0.012563.

El código que se muestra sirve para mostrar las estadísticas de los resultados:

```
> printcp(ArbolPart)
```

```
Classification tree:
```

```
rpart(formula = CLASIF ~ ., data = trainData, method = "class")
```

```
Variables actually used in tree construction:
```

```
[1] CATEGORIA MTOVENSOLES TASAMENSUAL
```

```
Root node error: 398/11056 = 0.035999
```

```
n= 11056
```

	CP	nsplit	rel error	xerror	xstd
1	0.133166	0	1.00000	1.00000	0.049215
2	0.037688	1	0.86683	0.88693	0.046447
3	0.012563	2	0.82915	0.86683	0.045935
4	0.010050	3	0.81658	0.87437	0.046128
5	0.010000	4	0.80653	0.86935	0.045999

PASO 7. Gráfico de error.

El código que se muestra sirve para mostrar el gráfico del error como se muestra en la figura 4.10:

```
> plotcp(ArbolPart)
```

La figura 4.10, nos muestra el gráfico del error:

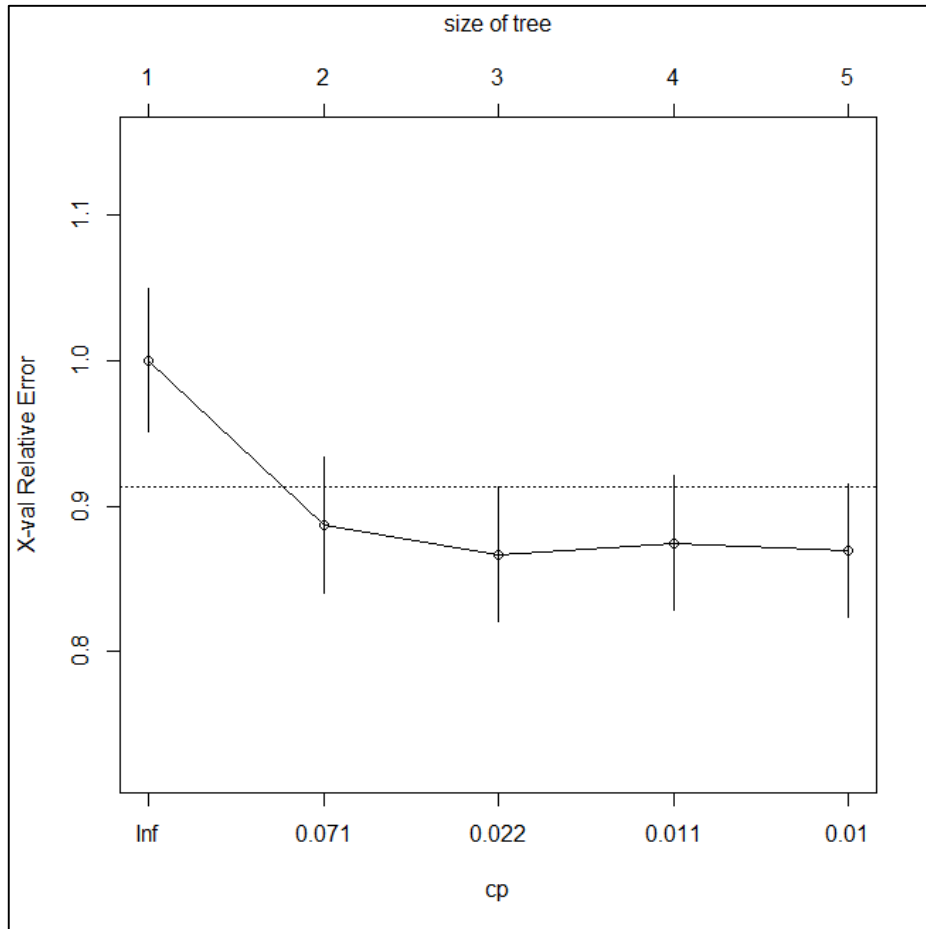


Figura 4.10. Gráfico del error (Elaboración propia, 2017)

PASO 8: Poda del árbol. Con el $cp=0.012563$ que es donde aumenta el error podaremos el árbol y mostraremos los resultados.

El código que se muestra sirve para podar el árbol:

```
> pArbolRpart<-prune(ArbolPart,cp=0.012563)
> printcp(pArbolRpart)
```

```
Classification tree:
rpart(formula = CLASIF ~ ., data = trainData, method = "class")
```

```
Variables actually used in tree construction:
[1] MTOVENSOLES
```

```
Root node error: 398/11056 = 0.035999
```

```
n= 11056
```

```
      CP nsplit rel error  xerror  xstd
1 0.133166      0  1.00000 1.00000 0.049215
```



```
2 0.037688      1  0.86683 0.88693 0.046447
4      0.012563      2  0.82915 0.86683 0.045935
```

PARA EL AÑO 2015

PASO 1. Los atributos utilizados para este análisis fueron el monto, tasa, monto vencido, la categoría y la clasificación crediticia que fueron elegidas de nuestro dataframe datos_dec_15 donde contiene la cartera a diciembre 2015.

El código que se muestra sirve para obtener las columnas del monto, tasa, monto vencido, categoría y clasificación crediticia de "dataframe":

```
> testTree<- datos_dec_15[,c(18,19,28,27,26)]
```

PASO 2. Seleccionaremos el número de muestras que necesitaremos para entrenamiento y para test. Con "sample" vamos a crear un vector que contendrá 2 grupos. El primer grupo TrainData tendrá el 80% de los elementos y el segundo grupo TestData tendrá el 20% de los elementos.

El código que se muestra sirve para obtener los sets de entrenamiento y test:

```
> ind <- sample(2,nrow(testTree),replace=TRUE, prob=c(0.8,0.2))
> trainData <- testTree[ind==1,] # entrenamiento
> testData <- testTree[ind==2,] # test
```

Paso 3. Creamos el árbol de clasificación con Rpart. Introducimos la variable dependiente CLASIFICACIÓN y explicamos con la selección de todas las demás variables. Method="class" significa que nuestra variable dependiente es cualitativa.

El código que se muestra sirve para calcular el árbol de clasificación:

```
> ArbolPart <- rpart(CALIFICACION ~.,
method="class",data=trainData)
```

Paso 4. Revisamos la información del gráfico. Donde n es el número de operaciones y Split es el número de cortes. Vemos que de 12,636 cortes la mayoría tiene clasificación A el 97% y 439 tendría clasificación B, C, D o E. El mejor corte que pudo hacer preguntarse por el monto vencido si es menor a 0.46 la mayoría 12,516 tendría clasificación A.

El código que se muestra sirve para mostrar la información del árbol de clasificación:

```
> print(ArbolPart)
n= 12636
node), split, n, loss, yval, (yprob)
      * denotes terminal node
  1) root 12636 439 A (0.97 0.023 0.004 0.0037 0.0044)
    2) MTOVENSOLES< 0.465 12516 319 A (0.97 0.023 0.00072 0.0014
0.00064) *
      3) MTOVENSOLES>=0.465 120 72 E (0 0.0083 0.35 0.24 0.4)
        6) MTOPRESTAMO>=266 112 70 C (0 0.0089 0.38 0.26 0.36)
          12) MTOVENSOLES< 1091.75 48 22 C (0 0 0.54 0.27 0.19)
            24) TASAMENSUAL< 4 37 12 C (0 0 0.68 0.3 0.027) *
              25) TASAMENSUAL>=4 11 3 E (0 0 0.091 0.18 0.73) *
                13) MTOVENSOLES>=1091.75 64 33 E (0 0.016 0.25 0.25
0.48) *
                  7) MTOPRESTAMO< 266 8 0 E (0 0 0 0 1) *
```

Analizando el árbol:

n= número de observaciones 12,636
Split= número de cortes

Nos indica que de un total de 12,636 observaciones la mayoría 96% tiene clasificación A. Si tiene un saldo vencido mayor a 0.46 la mayoría tiene clasificación E.

PASO 5. Graficando el árbol:

El código que se muestra sirve para obtener el gráfico del árbol de clasificación que se muestra en la figura 4.11.

```
> rpart.plot(ArbolPart)
```

En la figura 4.11, nos muestra si la condición es verdadera se irá por el

lado izquierdo y si es falsa por lado derecho.

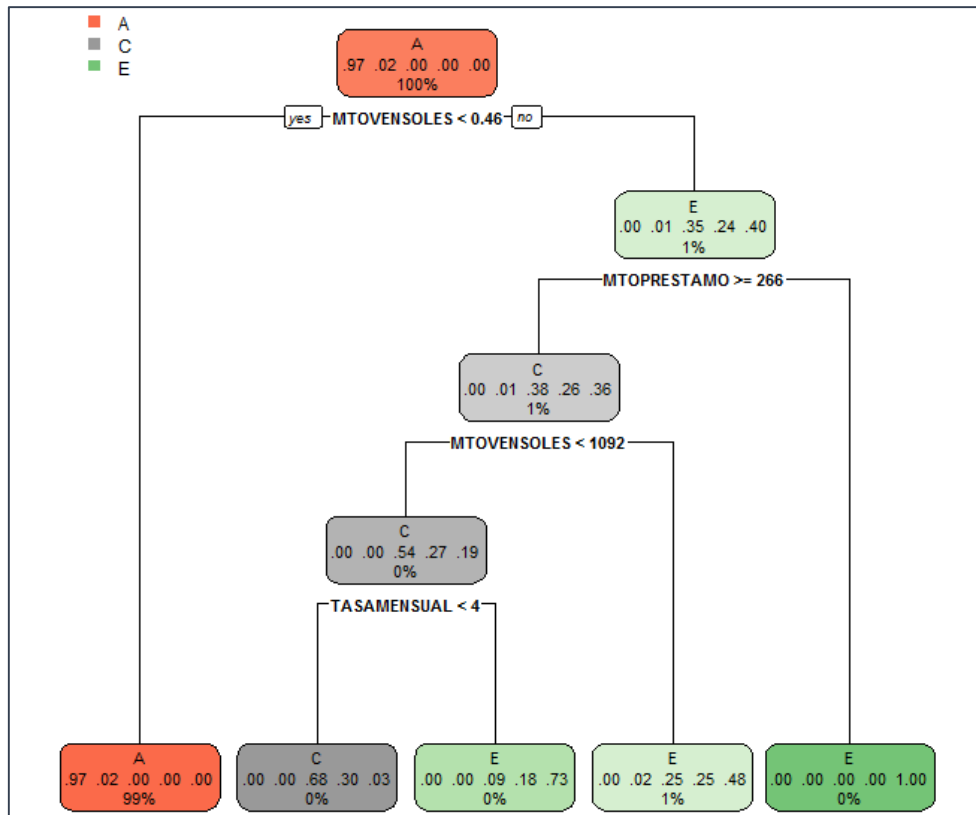


Figura 4.11. Gráfico del Árbol de clasificación (Elaboración propia, 2017)

Conclusiones del árbol de la figura 4.11, son:

- La variable Monto vencido (MTOVENSOLES) es el mejor predictor para la Calificación crediticia con 5 clasificaciones: Normal (A), CPP (B), Deficiente(C), Dudoso (D) y Pérdida (E).
- La probabilidad más alta de tener una calificación Normal (97%) se da a los créditos que tengan un monto vencido (MTOVENSOLES) menor a 0.46.
- La probabilidad más alta de tener un calificación Deficiente (68%) se da entre los créditos que tengan un monto vencido (MTOVENSOLES) menor a 1,092 y una tasa mensual mayor a 4.
- La probabilidad más alta de tener una calificación Pérdida (100%) se da entre los créditos que tengan un monto vencido (MTOVENSOLES) mayor a 0.46 y el monto de préstamo sea menor o igual a 266.

PASO 6. Visualizar las estadísticas de los resultados: Nos muestra que el error disminuye en el error.

El código que se muestra sirve para mostrar las estadísticas de los resultados:

```
> printcp(ArbolPart)
```

```
Classification tree:
rpart(formula = CLASIF ~ ., data = trainData, method = "class")
```

```
Variables actually used in tree construction:
[1] MTOPRESTAMO MTOVENSOLES TASAMENSUAL
```

```
Root node error: 439/12636 = 0.034742
```

```
n= 12636
```

	CP	nsplit	rel error	xerror	xstd
1	0.109339	0	1.00000	1.00000	0.046891
2	0.019362	1	0.89066	0.89294	0.044395
3	0.015945	3	0.85194	0.86788	0.043787
4	0.010000	4	0.83599	0.84282	0.043170

PASO 7. Gráfico de error.

El código que se muestra sirve para mostrar el gráfico del error que se muestra en la figura 4.12:

```
> plotcp(ArbolPart)
```

En la figura 4.12, nos muestra el error

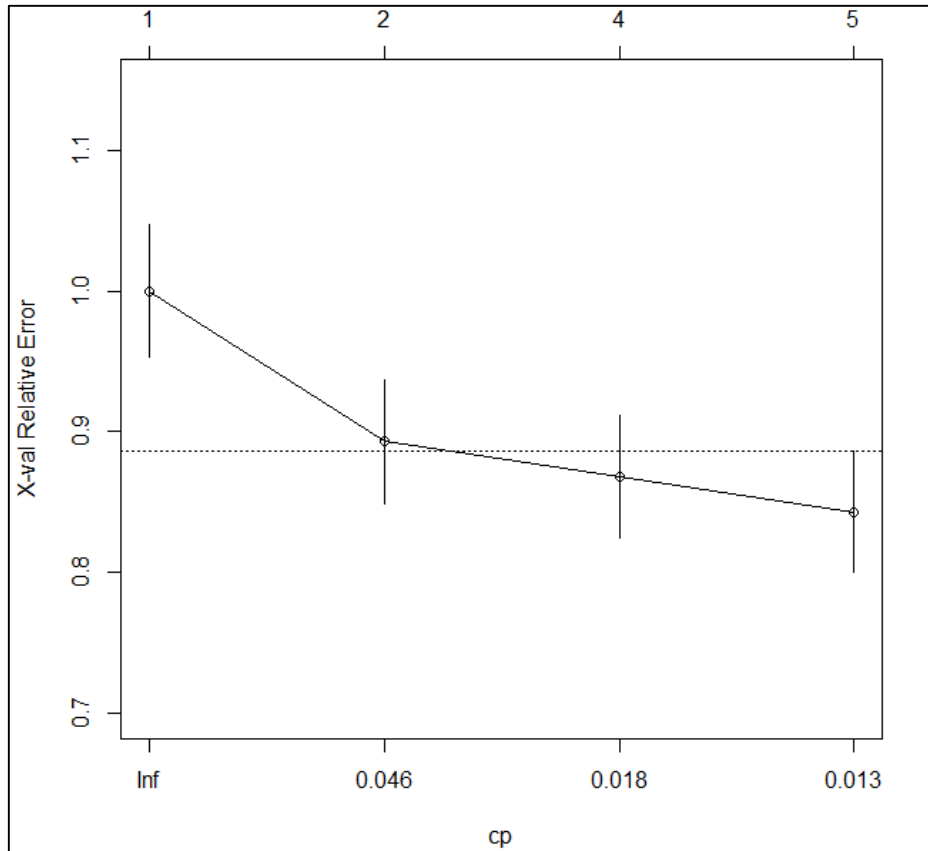


Figura 4.12. Gráfico del error (Elaboración propia, 2017)

PARA EL AÑO 2016

PASO 1. Los atributos utilizados para este análisis fueron el monto, tasa, monto vencido, categoría y clasificación crediticia del dataframe `datos_dec_2016` donde contiene la cartera gerencial al 31 diciembre 2016.

El código que se muestra sirve para obtener las columnas del monto, tasa, monto vencido, categoría y clasificación crediticia del "dataframe" `datos_dec_16`:

```
> testTree<- datos_dec_16[,c(18,19,28,27,26)]
```

PASO 2. Seleccionaremos el número de muestras que necesitaremos para entrenamiento y para test. Con "sample" vamos a crear un vector que contendrá 2 grupos. El primer grupo `TrainData` tendrá el 80% de los

elementos y el segundo grupo TestData tendrá el 20% de los elementos.

El código que se muestra sirve para obtener los sets de entrenamiento y test:

```
> ind <- sample(2,nrow(testTree),replace=TRUE,prob=c(0.8,0.2))
> trainData <- testTree[ind==1,] # entrenamiento
> testData <- testTree[ind==2,] # test
```

Paso 3. Creamos el árbol de clasificación con Rpart. Introducimos la variable dependiente CALIFICACION y explicamos con la selección de todas las demás variables. Method="class" significa que nuestra variable dependiente es cualitativa.

El código que se muestra sirve para calcular el árbol de clasificación:

```
> ArbolPart <- rpart(CALIFICACION ~.,
method="class",data=trainData)
```

Paso 4. Revisamos la información del gráfico. Donde n es el número de operaciones y Split es el número de cortes. Vemos que de 3,976 cortes la mayoría tiene clasificación A el 69% y 1419 tendría clasificación B, C, D o E. El mejor corte que pudo hacer preguntarse por el monto vencido si es menor a 0.01 la mayoría 3079 sería A.

El código que se muestra sirve para mostrar la información del árbol de clasificación:

```
> print(ArbolPart)
```

```
n= 3976
```

```
node), split, n, loss, yval, (yprob)
* denotes terminal node
```

```
1) root 3976 1252 A (0.69 0.084 0.025 0.034 0.17)
  2) MTOVENSOLES< 0.01 3079 362 A (0.88 0.11 0.0016 0.0049
0.0032) *
  3) MTOVENSOLES>=0.01 897 224 E (0.0078 0.0011 0.1 0.14 0.75)
    6) MTOVENSOLES< 3257.19 183 111 D (0.016 0.0055 0.32 0.39
0.27) *
    7) MTOVENSOLES>=3257.19 714 90 E (0.0056 0 0.05 0.07
0.87) *
```

Analizando el árbol:

n= número de observaciones 3,976

Split= número de cortes

PASO 5. Graficando el árbol se tiene:

El código que se muestra sirve para obtener el gráfico del árbol de clasificación que se muestra en la figura 4.13.

> `rpart.plot(ArbolPart)`

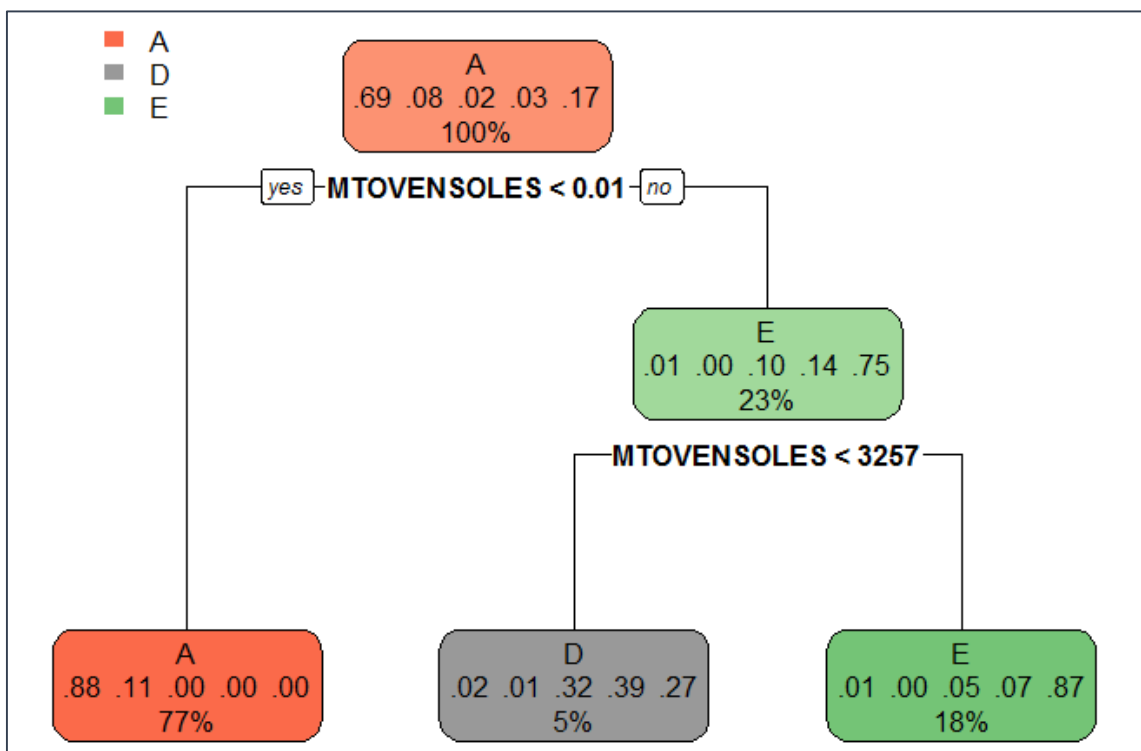


Figura 4.13. Gráfico del Árbol de clasificación (Elaboración propia, 2017)

Algunas conclusiones posibles del árbol de la figura 4.13, son:

- La variable Monto vencido (MTOVENSOLES) es el mejor predictor para la Calificación crediticia con 5 clasificaciones: Normal (A), CPP (B), Deficiente(C), Dudoso (D) y Pérdida (E).
- La probabilidad más alta de tener una calificación Normal (88%) se da a los créditos que tengan un monto vencido (MTOVENSOLES) menor a 0.01.
- La probabilidad más alta de tener una calificación Dudoso (32%) se

da entre los créditos que tengan un monto vencido (MTOVENSOLES) menor a 3,257.

- La probabilidad más alta de tener una calificación Pérdida (87%) se da entre los créditos que tengan un monto vencido (MTOVENSOLES) mayor a 3,257.

PASO 6. Visualizar las estadísticas de los resultados: Nos muestra que el error disminuye.

El código que se muestra sirve para mostrar las estadísticas de los resultados:

```
> printcp(ArbolPart)
```

```
Classification tree:  
rpart(formula = CLASIF ~ ., data = trainData, method = "class")
```

```
Variables actually used in tree construction:  
[1] MTOVENSOLES
```

```
Root node error: 1252/3976 = 0.31489
```

```
n= 3976
```

	CP	nsplit	rel error	xerror	xstd
1	0.531949	0	1.00000	1.00000	0.023393
2	0.018371	1	0.46805	0.46885	0.017866
3	0.010000	2	0.44968	0.46326	0.017777

PASO 7: Gráfico de error.

El código que se muestra sirve para obtener la figura 4.14:

```
> plotcp(ArbolPart)
```

En la figura 4.14, nos muestra el error

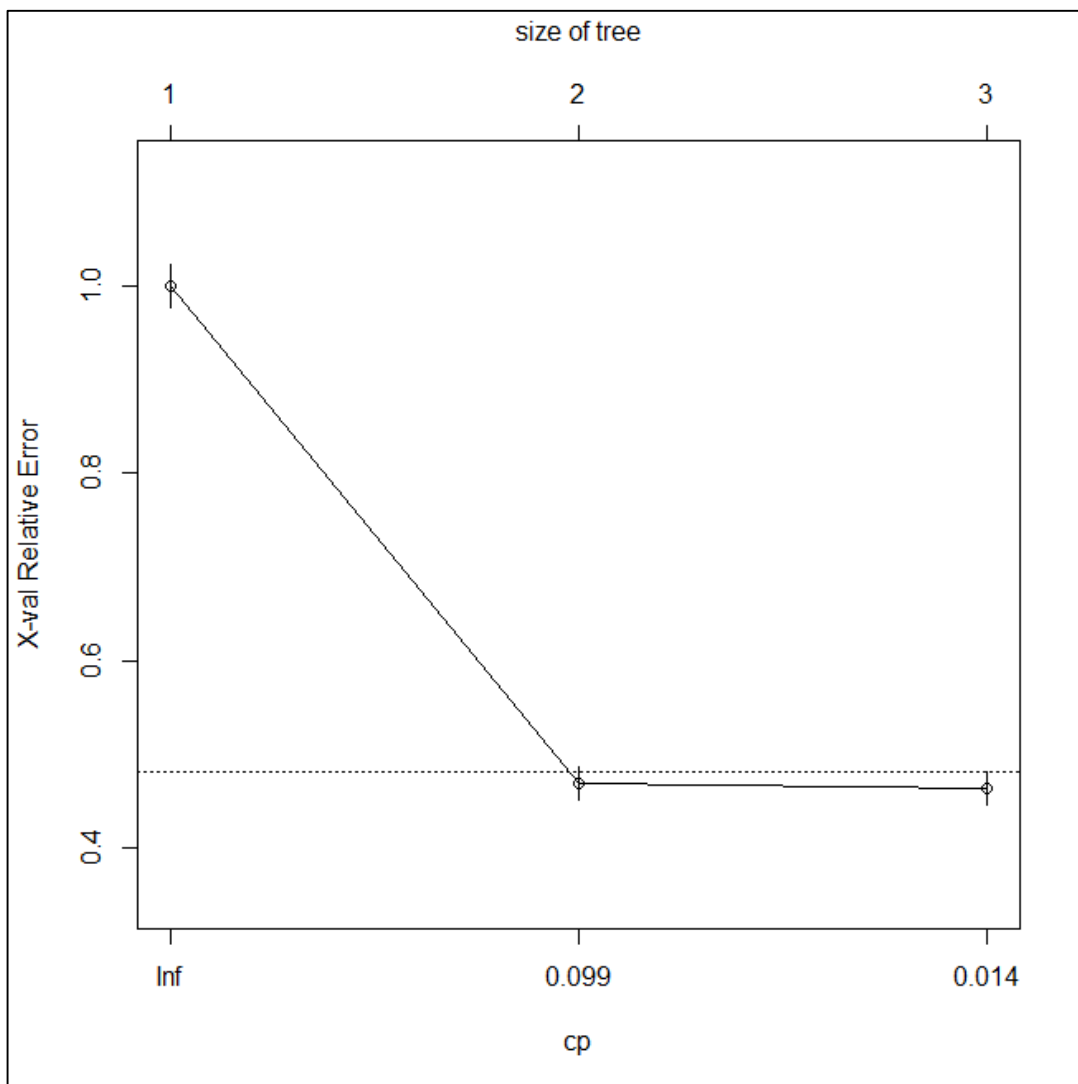


Figura 4.14. Gráfico del error (Elaboración propia, 2017)

b) MODELO DE PROVISIONES CON CLASIFICACIÓN CREDITICIA

En esta parte del modelo se ha evaluado las provisiones de cada crédito y la clasificación crediticia, en los resultados finales del árbol mostraremos como el comportamiento de la provisión de créditos, producto y la categoría del crédito, influye en la clasificación final del crédito.

PARA EL AÑO 2014

PASO1. Los atributos utilizados para este análisis fueron la categoría, las

provisiones, clasificación y producto como datos del dataframe `datos_dec_2014` que contiene la cartera gerencial al 31 de diciembre del 2014.

El código que se muestra sirve para obtener las columnas de categoría, provisiones, producto y clasificación crediticia del "dataframe":

```
> testTree<- datos_dec_14[,c(24,25,30,37)]
```

PASO 2. Seleccionaremos el número de muestras que necesitaremos para entrenamiento y para test. Con "sample" vamos a crear un vector que contendrá 2 grupos. El primer grupo `TrainData` tendrá el 80% de los elementos y el segundo grupo `TestData` tendrá el 20% de los elementos.

El código que se muestra sirve para obtener los sets de entrenamiento y test:

```
> ind <- sample(2,nrow(testTree),replace=TRUE,prob=c(0.8,0.2))
> trainData <- testTree[ind==1,] # entrenamiento
> testData <- testTree[ind==2,] # test
```

Paso 3. Creamos el árbol de clasificación con `Rpart`. Introducimos la variable dependiente `CALIFICACION` y explicamos con la selección de todas las demás variables. `Method="class"` significa que nuestra variable dependiente es cualitativa.

El código que se muestra sirve para calcular el árbol de clasificación:

```
> ArbolPart <- rpart(CLASIF ~., method="class",data=trainData)
```

Paso 4. Revisamos la información del gráfico. Donde `n` es el número de operaciones y `Split` es el número de cortes. Vemos que de 11,151 cortes la mayoría tiene clasificación A el 96% y 407 tendría clasificación B, C, D o E. El mejor corte que pudo hacer preguntarse por la provisión si es menor a 503.29 la mayoría 10,927 tendría clasificación A.

El código que se muestra sirve para mostrar la información del árbol de

clasificación:

```
> print(ArbolPart)
```

```
n= 11151
```

```
node), split, n, loss, yval, (yprob)
* denotes terminal node
```

```
1) root 11151 407 A (0.96 0.02 0.0054 0.0061 0.0054)
  2) PROVISION_CON_GARANTIA< 503.295 10927 185 A (0.98 0.013
0.002 0.0011 0.00037) *
  3) PROVISION_CON_GARANTIA>=503.295 224 152 B (0.0089 0.32
0.17 0.25 0.25)
    6) PROVISION_CON_GARANTIA< 2153.675 105 34 B (0.0095 0.68
0.15 0.12 0.038) *
    7) PROVISION_CON_GARANTIA>=2153.675 119 67 E (0.0084
0.0084 0.18 0.36 0.44)
      14) PRODUCTO.1=CREDICAMPAÑA CON P DE PATRIA,CREDICAMPAÑA
DIA DE LA MADRE,CREDICONVENIO M.N.,CREDINEGOCIO M.N.,INDIVIDUAL
- RURAL,PERSONAL M.N.,VIVIENDA M.N. 90 54 D (0.011 0.011 0.24
0.4 0.33)
        28) PROVISION_CON_GARANTIA< 14794.48 63 35 D (0.016
0.016 0.35 0.44 0.17)
          56) PROVISION_CON_GARANTIA< 5223.88 30 16 C (0.033
0.033 0.47 0.27 0.2) *
            57) PROVISION_CON_GARANTIA>=5223.88 33 13 D (0 0
0.24 0.61 0.15) *
              29) PROVISION_CON_GARANTIA>=14794.48 27 8 E (0 0 0
0.3 0.7) *
                15) PRODUCTO.1=ADMINISTRATIVO M.N.,CREDI CRECER
M.N.,CREDICAMPAÑA ESCOLAR,CREDICAMPAÑA NAVIDEÑA 2013 29 7 E
(0 0 0 0.24 0.76) *
```

PASO 5. Graficando el árbol:

El código que se muestra sirve para obtener el gráfico del árbol de clasificación que se muestra en la figura 4.15:

```
> rpart.plot(ArbolPart)
```

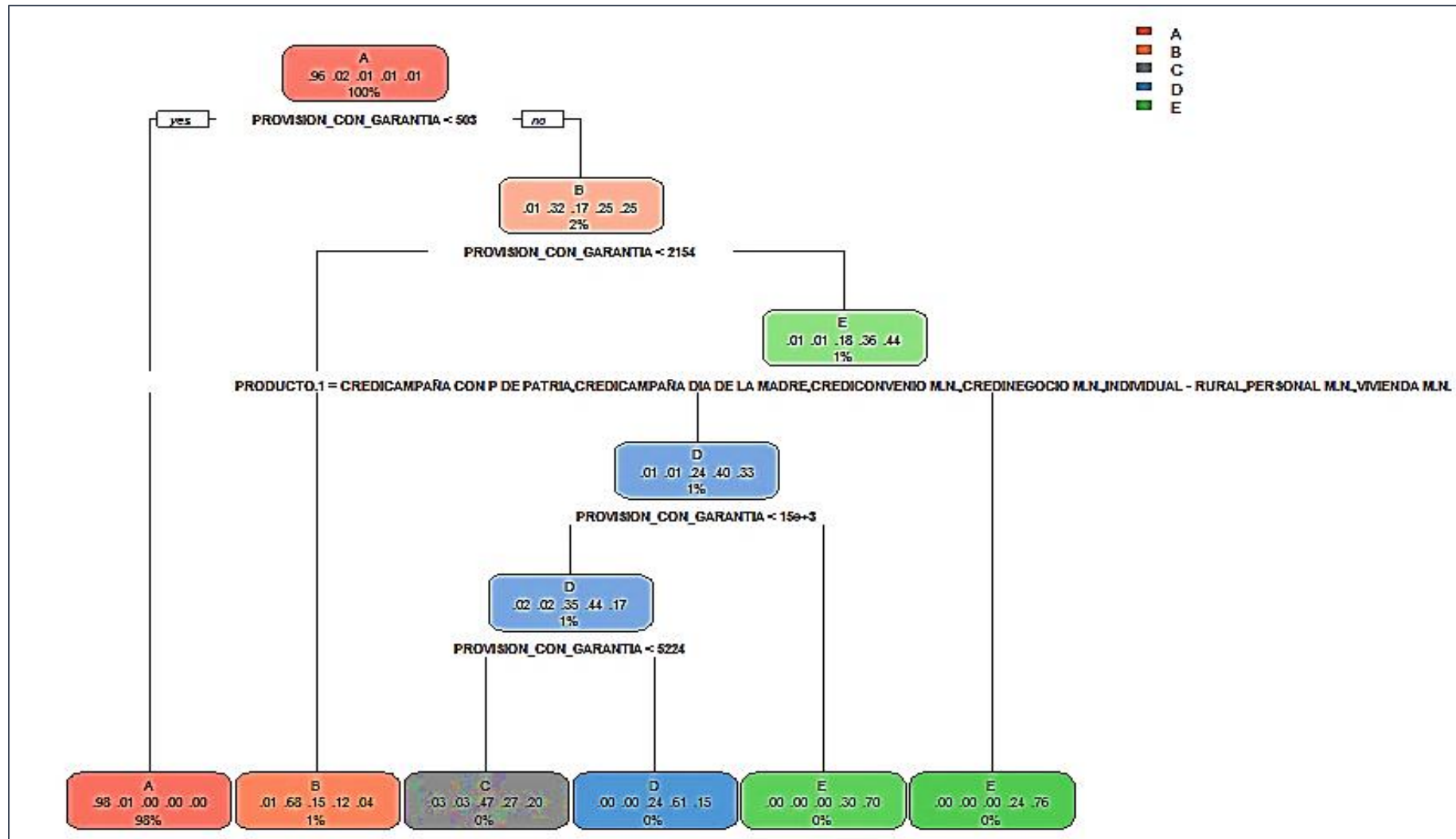


Figura 4.15. Gráfico de Árbol de clasificación (Elaboración propia, 2017)

Conclusiones del árbol de la figura 4.15, son:

- La variable Provisión (PROVISION_CON_GARANTÍA) es el mejor predictor para la Calificación crediticia con 5 clasificaciones: Normal (A), CPP (B), Deficiente(C), Dudoso (D) y Pérdida (E).
- La probabilidad más alta de tener una calificación Normal (98%) se da a los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) menor a 503.
- La probabilidad más alta de tener un calificación CPP (68%) se da entre los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) menor a 2,154.
- La probabilidad más alta de tener un calificación Deficiente (47%) se da a los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) menor a 5,224 y a los créditos que su producto sean Credicampaña con P de patria, Credicampaña día de la madre, Crediconvenio MN, Credinegocio MN, Individual Rural, Personal MN y Vivienda.
- La probabilidad más alta de tener un calificación Dudoso (61%) se da a los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) mayor a 5,224 y a los créditos que su producto sean Credicampaña con P de patria, Credicampaña día de la madre, Crediconvenio MN, Credinegocio MN, Individual Rural, Personal MN y Vivienda.
- La probabilidad más alta de tener un calificación Pérdida (76%) se da a los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) mayor a 2,154 y a los créditos que su producto no sean Credicampaña con P de patria, Credicampaña día de la madre, Crediconvenio MN, Credinegocio MN, Individual Rural, Personal MN y Vivienda.

PASO 6. Visualizar las estadísticas de los resultados: Nos muestra que a partir de 4 incrementa el error de 0.714 a 0.717, es allí donde podaremos

el árbol identificando el cp: 0.012563

El código que se muestra sirve para mostrar las estadísticas de los resultados:

```
> printcp(ArbolPart)
```

Classification tree:

```
rpart(formula = CLASIF ~ ., data = trainData, method = "class")
```

Variables actually used in tree construction:

```
[1] PRODUCTO.1          PROVISION_CON_GARANTIA
```

Root node error: 407/11151 = 0.036499

n= 11151

	CP	nsplit	rel error	xerror	xstd
1	0.171990	0	1.00000	1.00000	0.048655
2	0.125307	1	0.82801	0.83047	0.044482
3	0.020885	2	0.70270	0.71499	0.041363
4	0.014742	4	0.66093	0.71744	0.041432
5	0.010000	5	0.64619	0.71744	0.041432

PASO 7. Gráfico de error.

El código que se muestra sirve para obtener la figura 4.16:

```
> plotcp(ArbolPart)
```

En la figura 4.16, nos muestra el error

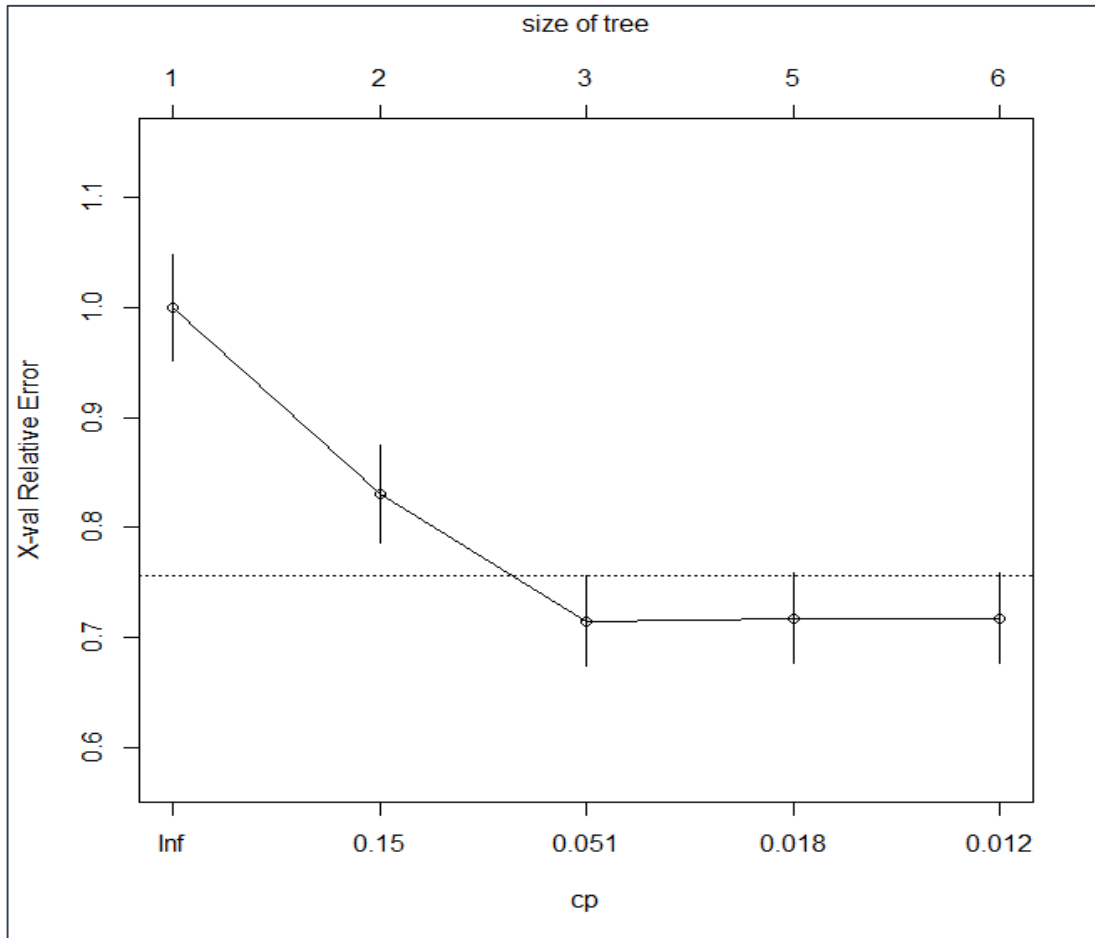


Figura 4.16. Gráfico del error (Elaboración propia, 2017)

PARA EL AÑO 2015

PASO1. Los atributos utilizados para este análisis fueron la categoría, las provisiones, clasificación y producto como datos del dataframe datos_dec_2015 que contiene la cartera gerencial al 31 de diciembre del 2015.

El código que se muestra sirve para obtener las columnas de categoría, provisiones, producto y clasificación crediticia del "dataframe":

```
> testTree<- datos_dec_15[,c(26,27,32,39)]
```

PASO 2. Seleccionaremos el número de muestras que necesitaremos para entrenamiento y para test. Con "sample" vamos a crear un vector que contendrá 2 grupos. El primer grupo TrainData tendrá el 60% de los

elementos y el segundo grupo TestData tendrá el 40% de los elementos.

El código que se muestra sirve para obtener los sets de entrenamiento y test:

```
> ind <- sample(2,nrow(testTree),replace=TRUE,prob=c(0.6,0.4))
> trainData <- testTree[ind==1,] # entrenamiento
> testData <- testTree[ind==2,] # test
```

Paso 3. Creamos el árbol de clasificación con Rpart. Introducimos la variable dependiente CALIFICACION y explicamos con la selección de todas las demás variables. Method="class" significa que nuestra variable dependiente es cualitativa.

El código que se muestra sirve para calcular el árbol de clasificación:

```
> ArbolPart <- rpart(CLASIF ~., method="class",data=trainData)
```

Paso 4. Revisamos la información del gráfico. Donde n es el número de operaciones y Split es el número de cortes. Vemos que de 12,748 cortes la mayoría tiene clasificación A el 96% y 454 tendría clasificación B, C, D o E. El mejor corte que pudo hacer preguntarse por la provisión si es menor a 503.29 la mayoría 12,556 tendría clasificación A.

El código que se muestra sirve para mostrar la información del árbol de clasificación:

```
> print(ArbolPart)
```

```
n= 12748
```

```
node), split, n, loss, yval, (yprob)
* denotes terminal node
```

```
1) root 12748 454 A (0.96 0.023 0.0042 0.0038 0.0043)
  2) PROVISION_CON_GARANTIA< 503.295 12556 265 A (0.98 0.018
0.0017 0.00064 0.001) *
  3) PROVISION_CON_GARANTIA>=503.295 192 117 B (0.016 0.39
0.17 0.21 0.22)
    6) PROVISION_CON_GARANTIA< 2123.085 115 40 B (0.026 0.65
0.14 0.087 0.096) *
    7) PROVISION_CON_GARANTIA>=2123.085 77 46 E (0 0 0.21
0.39 0.4)
      14) PROVISION_CON_GARANTIA< 4900.72 38 23 C (0 0 0.39
0.24 0.37) *
```


15) PROVISION_CON_GARANTIA >= 4900.72 39 18 D (0 0 0.026 0.54 0.44)

30) PRODUCTO.1 = CREDI CRECER M.N., CREDICONVENIO M.N., INDIVIDUAL - RURAL 16 4 D (0 0 0.062 0.75 0.19) *

31) PRODUCTO.1 = CAMPAÑA ESCOLAR 2015, CREDINEGOCIO M.N., PERSONAL M.N. 23 9 E (0 0 0 0.39 0.61) *

PASO 5. Graficando el árbol:

El código que se muestra sirve para obtener el gráfico del árbol de clasificación que se muestra en la figura 4.17:

```
> rpart.plot(ArbolPart)
```

En la figura 4.17, nos muestra si la condición es verdadera se irá por el lado izquierdo y si no por el lado derecho.

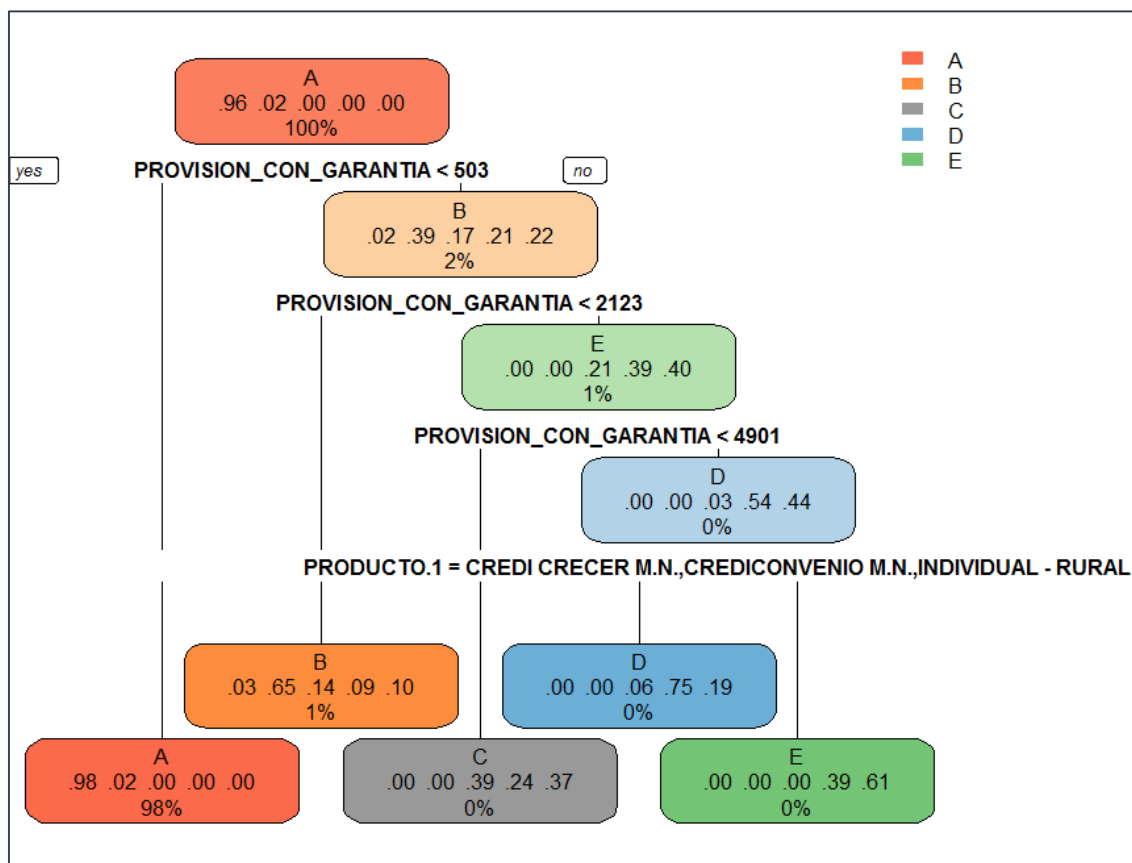


Figura 4.17. Gráfico del Árbol de clasificación (Elaboración propia, 2017)

Conclusiones del árbol de la figura 4.17, son:

- La variable Provisión (PROVISION_CON_GARANTÍA) es el mejor predictor para la Calificación crediticia, que se clasifica en 5: Normal

(A), CPP (B), Deficiente(C), Dudoso (D) y Pérdida (E).

- La probabilidad más alta de tener una calificación Normal (96%) se da a los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) menor a 503.
- La probabilidad más alta de tener un calificación CPP (65%) se da entre los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) menor a 2,123.
- La probabilidad más alta de tener un calificación Deficiente (39%) se da a los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) menor a 4,901.
- La probabilidad más alta de tener un calificación Dudoso (75%) se da a los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) mayor a 4,901 y a los créditos que su producto sean Credi Crecer, Crediconvenio MN, Individual Rural.
- La probabilidad más alta de tener un calificación Pérdida (61%) se da a los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) mayor a 4,901 y a los créditos que su producto no sean Credi Crecer, Crediconvenio MN, Individual Rural.

PASO 6. Visualizar las estadísticas de los resultados: Nos muestra que el error disminuye y no hay necesidad de podar el árbol.

El código que se muestra sirve para mostrar las estadísticas de los resultados:

```
> printcp(ArbolPart)
```

```
Classification tree:
rpart(formula = CLASIF ~ ., data = trainData, method =
"class")
```

```
Variables actually used in tree construction:
[1] PRODUCTO.1          PROVISION_CON_GARANTIA
```

```
Root node error: 454/12748 = 0.035613
```

```
n= 12748
```

```
CP nsplit rel error xerror xstd
```

1	0.158590	0	1.00000	1.00000	0.046089
2	0.068282	1	0.84141	0.84141	0.042400
3	0.011013	2	0.77313	0.80617	0.041530
4	0.010000	4	0.75110	0.79736	0.041309

PASO 7. Gráfico de error.

El código que se muestra sirve para obtener la figura 4.18:

> plotcp(ArbolPart)

En la figura 4.18, nos muestra el error

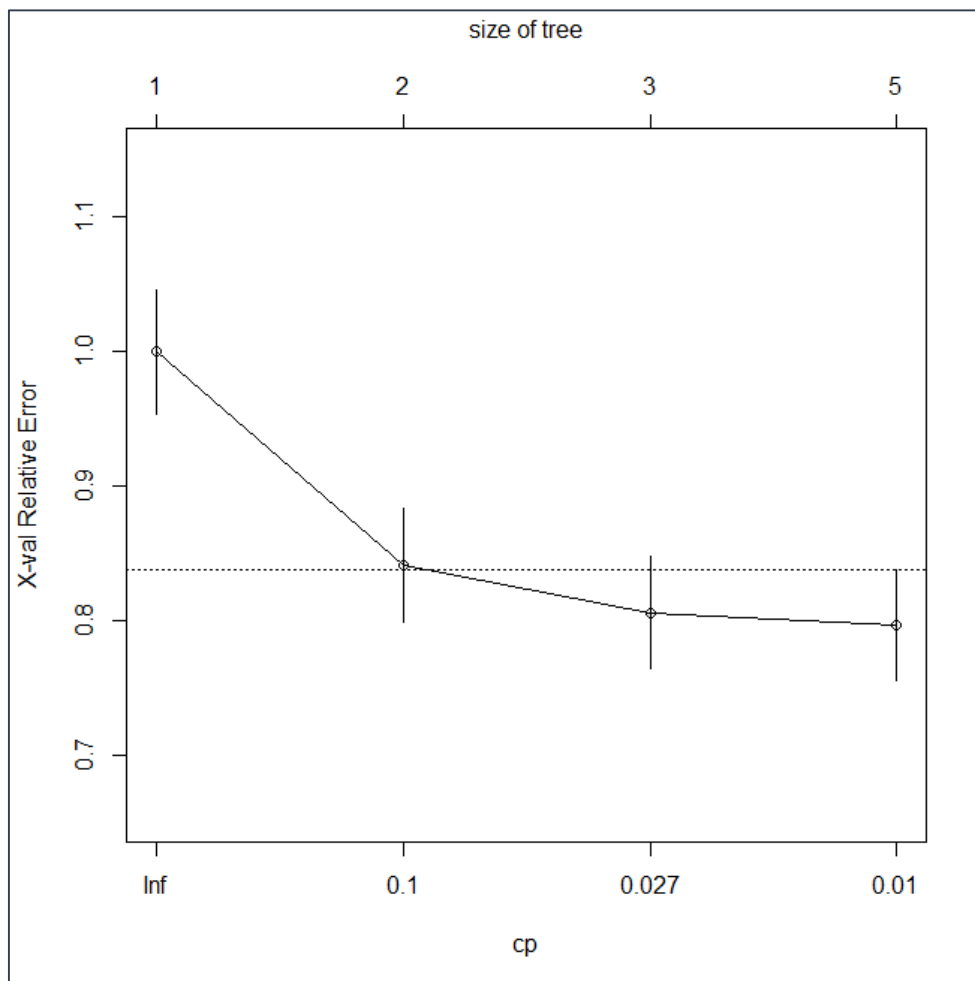


Figura 4.18: Gráfico del error (Elaboración propia, 2017)

PARA EL AÑO 2016

PASO1. Los atributos utilizados para este análisis fueron la categoría, las provisiones, clasificación y producto como datos del dataframe datos_dec_2016 que contiene la cartera gerencial al 31 de diciembre del

2016.

El código que se muestra sirve para obtener las columnas de la categoría, provisiones y clasificación crediticia del "dataframe":

```
> testTree<- datos_dec_16[,c(26,27,32)]
```

PASO 2. Seleccionaremos el número de muestras que necesitaremos para entrenamiento y para test. Con "sample" vamos a crear un vector que contendrá 2 grupos. El primer grupo TrainData tendrá el 80% de los elementos y el segundo grupo TestData tendrá el 20% de los elementos.

El código que se muestra sirve para obtener los sets de entrenamiento y test:

```
> ind <- sample(2,nrow(testTree),replace=TRUE,prob=c(0.8,0.2))
> trainData <- testTree[ind==1,] # entrenamiento
> testData <- testTree[ind==2,] # test
```

Paso 3. Creamos el árbol de clasificación con Rpart. Introducimos la variable dependiente CALIFICACION y explicamos con la selección de todas las demás variables. Method="class" significa que nuestra variable dependiente es cualitativa.

El código que se muestra sirve para calcular el árbol de clasificación:

```
> ArbolPart <- rpart(CLASIF ~., method="class",data=trainData)
```

Paso 4. Revisamos la información del gráfico. Donde n es el número de operaciones y Split es el número de cortes. Vemos que de 4,507 cortes la mayoría tiene clasificación A el 69% y 1,390 tendría clasificación B, C, D o E. El mejor corte que pudo hacer preguntarse por la provisión si es menor a 421.86 la mayoría 3,421 tendría clasificación A.

El código que se muestra sirve para mostrar la información del árbol de clasificación:

```
> print(ArbolPart)
```

```
n= 4513
```

```
node), split, n, loss, yval, (yprob)
* denotes terminal node
```

```
1) root 4513 1390 A (0.69 0.078 0.023 0.033 0.17)
  2) PROVISION_CON_GARANTIA< 421.86 3421 301 A (0.91 0.06
0.007 0.0058 0.015) *
  3) PROVISION_CON_GARANTIA>=421.86 1092 358 E (0.0027 0.13
0.075 0.12 0.67)
    6) PROVISION_CON_GARANTIA< 3160.815 340 195 B (0.0088
0.43 0.2 0.2 0.17)
      12) CATEGORIA=CREDITOS A PEQUEÑAS EMPRESAS,CREDITOS DE
CONSUMO NO REVOLVENTE 224 97 B (0.0045 0.57 0.21 0.11 0.11) *
      13) CATEGORIA=CREDITOS A MICROEMPRESAS,CREDITOS
HIPOTECARIOS PARA VIVIENDA 116 74 D (0.017 0.16 0.18 0.36
0.28) *
        7) PROVISION_CON_GARANTIA>=3160.815 752 75 E (0 0 0.019
0.081 0.9) *
```

PASO 5. Graficando el árbol:

El código que se muestra sirve para obtener el gráfico del árbol de clasificación que se muestra en la figura 4.19:

```
> rpart.plot(ArbolPart)
```

En la figura 4.19, nos muestra que si la condición es verdadera se irá por el lado izquierdo y si no por el lado derecho.

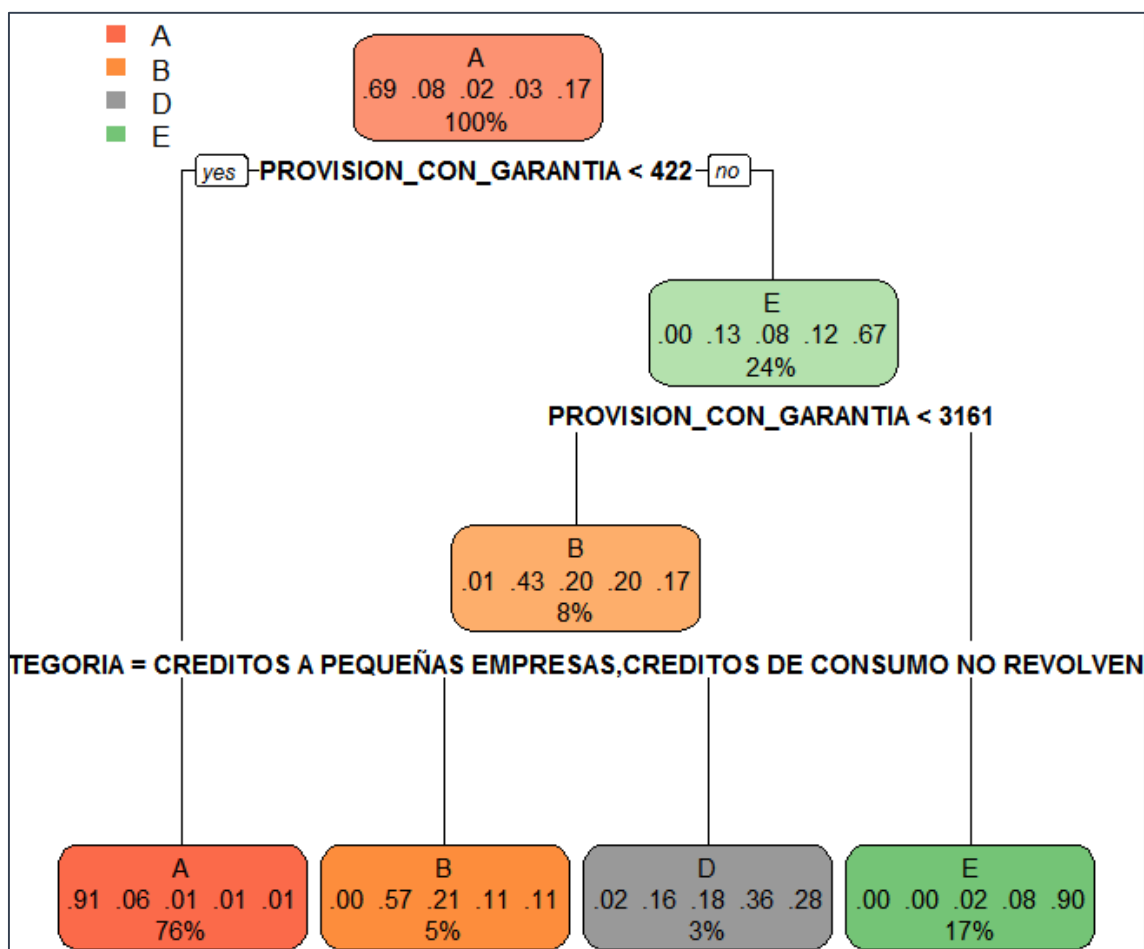


Figura 4.19. Gráfico del Árbol de clasificación (Elaboración propia, 2017)

Conclusiones del árbol de la figura 4.19, son:

- La variable Provisión (PROVISION_CON_GARANTÍA) es el mejor predictor para la Calificación crediticia, que se clasifica en 5: Normal (A), CPP (B), Deficiente(C), Dudoso (D) y Pérdida (E).
- La probabilidad más alta de tener una calificación Normal (91%) se da a los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) menor a 422.
- La probabilidad más alta de tener un calificación CPP (57%) se da entre los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) entre 422 y 3,161, además que sus categorías sean Crédito a pequeñas empresas y Créditos Consumo no Revolvente.
- La probabilidad más alta de tener un calificación Dudoso (36%) se da

a los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) entre 422 y 3,161, además que sus categorías no sean Crédito a pequeñas empresas y Créditos Consumo no Revolvente.

- La probabilidad más alta de tener un calificación Pérdida (61%) se da a los créditos que tengan una provisión (PROVISION_CON_GARANTÍA) mayor a 4,901 y a los créditos que su producto no sean Credi Crecer, Crediconvenio MN, Individual Rural.

PASO 6. Visualizar las estadísticas de los resultados: Nos muestra que el error disminuye y no hay necesidad de podar el árbol.

El código que se muestra sirve para mostrar las estadísticas de los resultados:

```
> printcp(ArbolPart)
```

```
Classification tree:
```

```
rpart(formula = CLASIF ~ ., data = trainData, method = "class")
```

```
Variables actually used in tree construction:
```

```
[1] CATEGORIA          PROVISION_CON_GARANTIA
```

```
Root node error: 1390/4513 = 0.308
```

```
n= 4513
```

	CP	nsplit	rel error	xerror	xstd
1	0.525899	0	1.00000	1.00000	0.022312
2	0.063309	1	0.47410	0.47554	0.017088
3	0.017266	2	0.41079	0.41367	0.016115
4	0.010000	3	0.39353	0.41079	0.016067

PASO 7. Gráfico de error.

El código que se muestra sirve para obtener la figura 4.20:

```
> plotcp(ArbolPart)
```

En la figura 4.20, nos muestra el error.

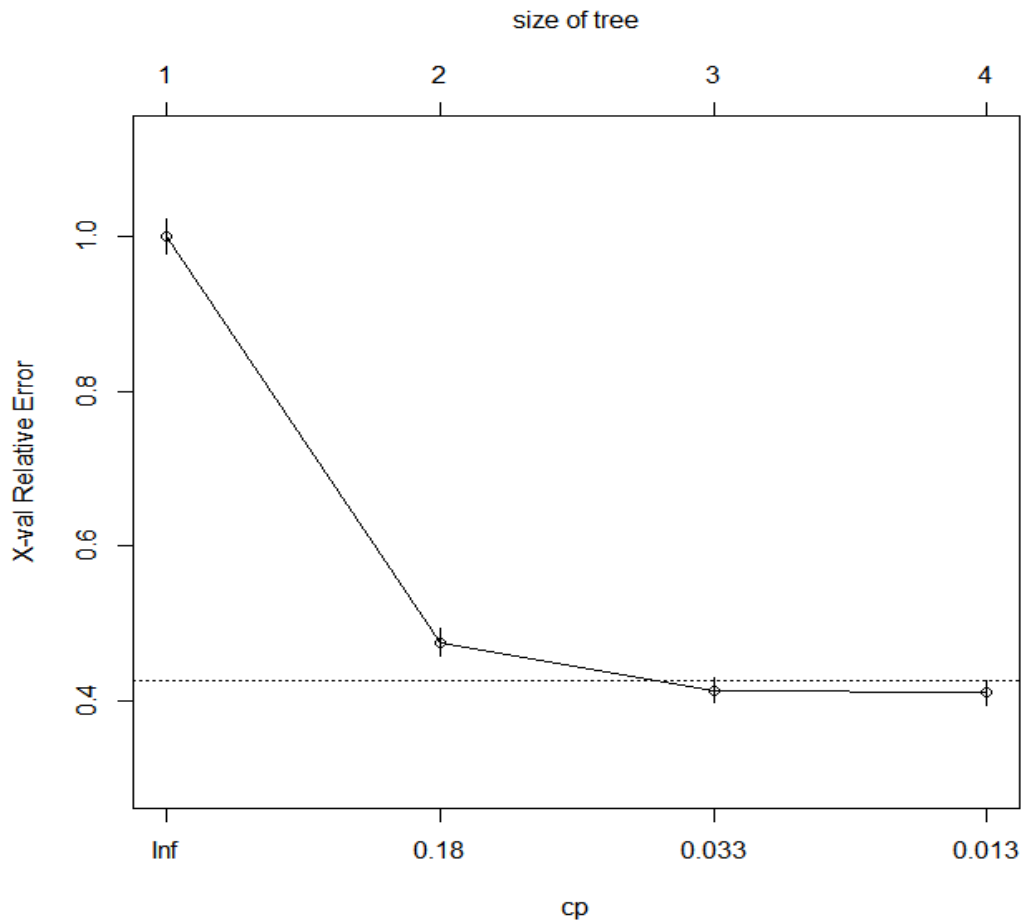


Figura 4.20. Gráfico del error (Elaboración propia, 2017)

4.5.4. EVALUACIÓN DEL MODELO

Las dificultades encontradas en el procesamiento del modelo desarrollado fue obtener los datos y dar el formato necesario para tener listo para la explotación en cada año correspondiente. Por lo tanto, en la ejecución del Modelo de Clasificación al contener una sola clasificación y evaluación de variable Calificación del crédito con algunos datos el modelo no resultaba lo esperado o se hacía muy complejo.

4.6. FASE V: EVALUACIÓN

El modelo generado y los resultados obtenidos para cada año han sido satisfactorios. La ejecución del modelo ha generado los datos esperados en función de los objetivos del negocio, establecidos previamente para el presente trabajo. No se descarta la posibilidad de tomar nuevas preguntas del negocio.

A continuación, se desarrolla un resumen de los resultados obtenidos del modelo aplicado y por cada año.

4.6.1. EVALUAR LOS RESULTADOS

a) EVALUACIÓN DEL MODELO DE MOROSIDAD CON CLASIFICACIÓN CREDITICIA.

PARA EL AÑO 2014

Para la evaluación de la predicción del modelo utilizamos la variable TestData. El código que se muestra sirve para obtener la tabla de predicción.

```
> testPredRpart<-  
predict(ArbolPart,newdata=testData,type="class")  
> table(testPredRpart,testData$CLASIF)
```

Se muestra que en el cálculo del modelo tienen calificación A para 2,667 casos que en realidad tienen calificación A; 51 casos que en realidad tenían calificación B, 8 casos que en realidad tenían calificación C, 4 casos que en realidad tenían calificación D y 4 casos que en realidad tenían calificación E.

testPredRpart	A	B	C	D	E
A	2667	51	8	4	4
B	0	0	0	0	0
C	0	0	4	2	0
D	0	0	2	6	5
E	0	0	2	4	7

Luego, evaluaremos la Estadística del modelo. Sumamos todos los aciertos y los dividimos entre el número de predicciones. El modelo arroja una efectividad de 97,03% de aciertos.

```
>  
sum(testPredRpart==testData$CLASIF)/length(testData$CLASIF)*100  
[1] 97.03543
```

PARA EL AÑO 2015

Evaluamos la predicción del modelo con TestData. El código que se muestra sirve para obtener la tabla de predicción.

```
> testPredRpart<-  
predict(ArbolPart,newdata=testData,type="class")  
> table(testPredRpart,testData$CLASIF)
```

Se muestra que el cálculo del modelo se tiene calificación A para 3,108 casos que tenían calificación A; para 87 casos que en realidad tenían calificación B, para 0 casos que en realidad tenían calificación C, 5 casos que en realidad tenían calificación D y 7 casos que en realidad tenían calificación E.

testPredRpart	A	B	C	D	E
A	3108	87	0	5	7
B	0	0	0	0	0
C	0	0	9	3	1
D	0	0	0	0	0
E	0	0	6	2	8

Luego evaluamos la Estadística del modelo. Sumamos todos los aciertos y los dividimos entre el número de predicciones. El modelo arroja una efectividad de 96,56% de aciertos.

```
>  
sum(testPredRpart==testData$CLASIF)/length(testData$CLASIF)*100  
[1] 96.56984
```

PARA EL AÑO 2016

Evaluamos la predicción del modelo con TestData. El código que se muestra sirve para obtener la tabla de predicción.

```
> testPredRpart<-  
predict(ArbolPart,newdata=testData,type="class")  
> table(testPredRpart,testData$CLASIF)
```

Se muestra que el cálculo del modelo se tiene calificación A para 1195 casos que tenían calificación A; para 118 casos que en realidad tenían calificación B, para 6 casos que en realidad tenían calificación C, 4 casos que en realidad tenían calificación D y 3 casos que en realidad tenían calificación E.

testPredRpart	A	B	C	D	E
---------------	---	---	---	---	---

A	1195	118	6	4	3
B	0	0	0	0	0
C	0	0	0	0	0
D	1	2	25	23	38
E	2	0	5	19	247

Luego evaluamos la Estadística del modelo. Sumamos todos los aciertos y los dividimos entre el número de predicciones. El modelo arroja una efectividad de 86,78% de aciertos.

```
>
sum(testPredRpart==testData$CLASIF)/length(testData$CLASIF)*100
[1] 86.7891
```

b) EVALUACIÓN DEL MODELO DE PROVISIONES CON CLASIFICACIÓN CREDITICIA PARA EL AÑO 2014

Evaluamos la predicción del modelo con TestData. Se muestra que en el cálculo del modelo se tiene calificación A para 2,580 casos que tenían calificación A. Calificación A para 32 casos que eran B, para 6 casos que eran C, 4 casos que eran D y 1 casos que eran E.

```
> testPredRpart<-
predict(ArbolPart,newdata=testData,type="class")
> table(testPredRpart,testData$CLASIF)
```

testPredRpart	A	B	C	D	E
A	2580	32	6	4	1
B	0	16	1	9	1
C	1	0	2	2	0
D	0	0	4	3	3
E	0	0	0	1	5

Luego evaluamos la Estadística del modelo. Sumamos todos los aciertos y los dividimos entre el número de predicciones. El modelo arroja una efectividad de 97,56% de aciertos.

```
>
sum(testPredRpart==testData$CLASIF)/length(testData$CLASIF)*100
[1] 97.56645
```

PARA EL AÑO 2015

Evaluamos la predicción del modelo con TestData. Se muestra que en el

cálculo del modelo se tiene calificación A para 3011 casos que tenían calificación A, y calificación A para 62 casos que eran B, para 6 casos que eran C, 3 casos que eran D y 4 casos que eran E.

```
> testPredRpart<-
predict(ArbolPart,newdata=testData,type="class")
> table(testPredRpart,testData$CLASIF)
```

testPredRpart	A	B	C	D	E
A	3011	62	6	3	4
B	0	11	3	3	3
C	0	1	3	1	3
D	0	0	0	1	2
E	0	0	1	1	5

Luego evaluamos la Estadística del modelo. Sumamos todos los aciertos y los dividimos entre el número de predicciones. El modelo arroja una efectividad de 97,02% de aciertos.

```
>
sum(testPredRpart==testData$CLASIF)/length(testData$CLASIF)*100
[1] 97.02305
```

PARA EL AÑO 2016

Evaluamos la predicción del modelo con TestData. Se muestra que en el cálculo del modelo se tiene calificación A para 797 casos que tenían calificación A, y calificación A para 63 casos que eran B, para 6 casos que eran C, 2 casos que eran D y 10 casos que eran E.

```
> testPredRpart<-
predict(ArbolPart,newdata=testData,type="class")
> table(testPredRpart,testData$CLASIF)
```

testPredRpart	A	B	C	D	E
A	797	63	6	2	10
B	2	33	8	5	4
C	0	0	0	0	0
D	0	6	4	5	5
E	0	0	11	23	167

Luego evaluamos la Estadística del modelo. Sumamos todos los aciertos y los dividimos entre el número de predicciones. El modelo arroja una efectividad de 87,05% de aciertos.

```
>
```

```
sum(testPredRpart==testData$CLASIF)/length(testData$CLASIF)*100
[1] [1] 87.05474
```

4.6.2. REVISIÓN DEL PROYECTO

Los resultados obtenidos del modelo aplicado tienen relación entre la realidad y en un primer análisis se puede deducir que los mismos han sido los esperados especialmente en los que respecta a cantidades y porcentajes.

El árbol de clasificación nos da el siguiente resultado que es ilustrado resumidamente en la tabla N° 4.20 y 4.21. Se puede ver claramente los resultados obtenidos sobre las cantidades de aciertos por tipos de clasificación crediticia por cada año.

Clasificación	2014	2015	2016
A	2667	3108	1195
B	0	0	0
C	4	9	0
D	6	0	23
E	7	8	247

Tabla 4.20. Cantidad de aciertos para el modelo de morosidad con clasificación crediticia.

Clasificación	2014	2015	2016
A	2580	3011	797
B	16	11	33
C	2	3	0
D	3	1	5
E	5	5	167

Tabla 4.21. Cantidad de aciertos para el modelo de provisiones con clasificación crediticia.

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

1. Se implementó de forma satisfactoria el modelo de minería de datos en la cooperativa, gracias a la metodología CRISP-DM y la técnica de árbol de decisión, se incluyó los 3 objetivos propuestos. Esta herramienta proporciona información útil para obtener un patrón de comportamiento y para el control de la gestión de riesgos de la Cooperativa de Ahorro y Crédito san Cristóbal de Huamanga, como se muestra en la tabla 4.20 y tabla 4.21 en relación a los años 2014, 2015 y 2016.

2. Se logró generar el modelo de minería de datos para procesar información de la calificación del crédito en relación a la morosidad, evaluado con cada indicador utilizando el árbol de decisión, que están orientados a apoyar gestión de riesgos de la cooperativa, como se muestra en la Figura 4.9 para el año 2014, Figura 4.10 para el año 2015 y Figura 4.13 para el año 2016.

2. Se logró generar el modelo de minería de datos para procesar información de la calificación del crédito en relación a la cobertura de provisiones, evaluado con cada indicador utilizando el árbol de decisión, que están orientados a apoyar gestión de riesgos de la cooperativa, como se muestra en la Figura 4.15 para el año 2014, Figura 4.17 para el año 2015 y Figura 4.19 para el año 2016.

5.2. RECOMENDACIONES

- A. Se recomienda incrementar al modelo de clasificación el número de variables como puede ser sexo, edad, producto, agencia. Ya que esto dará un panorama más preciso de la organización.

- B. Se recomienda utilizar otras técnicas de minería de datos como la regresión lineal o redes neuronales y comparar los resultados con el árbol de clasificación.

- C. Se recomienda realizar investigaciones adicionales para determinar que herramienta de Inteligencia de Negocios representa una mejor alternativa para el diseño y configuración del modelo de clasificación.

- D. La institución debe mejorar el uso de herramientas de inteligencia de negocios mediante el uso de técnicas de minería de datos, ya que estas herramientas proporcionan información útil, relevante y a tiempo.

- F. Se recomienda realizar investigaciones adicionales para apoyar otros procesos de gestión y administración de la institución.

BIBLIOGRAFÍA

1. Aguilar, G., Camargo, G., y Morales, R. (2006). *Economía y Sociedad*, (62), 1,2. Recuperado de http://cies.org.pe/sites/default/files/files/otros/economiaysociedad/10_aguilar.pdf
2. Barreiro, J. (2003). *Gestión científica empresarial*. Coruña: Netbiblo.
3. Brachfield, P. (2003). *Recobrar impagados y negociar con morosos*. Barcelona: Gestión 2000.
4. Bravo, C. (2012). *Métodos para estimar riesgo crediticio en base a minería de datos y teoría de juegos* (doctorado). Universidad de Chile.
5. Chapman P. et. Al (2000). *CRISP-DM 1.0: Guide, Step by Step DataMining* (Vol. 1) Londres, Nueva York y Bruselas: CRISP-DM Consortium.
6. Córdor, J. y Cajamarca, R. (2013). *Matrices de transición y análisis de cosechas en el contexto de riesgo de crédito*. Quito, Ecuador: SBS.
7. Date, C. & Ruiz, S. (2001). *Introducción a los sistemas de bases de datos* (7° Ed.). México: Pearson Educación.
8. De Ville, B. (2001). *Microsoft Data Mining: Integrated Business Intelligence for E-Commerce and Knowledge Management*. Digital Press.
9. Diaz, M. (2017). *Metodología CRISP-DM para minería de datps Parte 3 - Sección Final*. [online] SmartBase Group. Recuperado el 13 de Octubre del 2016, de <http://smartbasegroup.com/metodologia-crisp-dm-final/>
10. Estupiñan, R. (2006). *Administración o Gestión de riesgos E.R.M. y la Auditoría interna*. Bogotá, Colombia: Ecoe Ediciones.
11. FENACREP (2013). *Evolución y proyección de las principales cuentas e indicadores de las cooperativas de ahorro y crédito del Perú*. Presentación, Lima, Perú.
12. Flores, H. (2009). *Detección de patrones de daños y averías en la industria automotriz*. (Tesis de Maestría). Universidad Tecnológica Nacional Facultad Regional Buenos Aires.
13. Hernández, R., Fernández, C. y Baptista, P. (2010). *Metodología de la*

- investigación* (5ª Ed.). México, D.F., México: McGraw Hill Interamericana.
14. Hernández, J., Ramírez, M., y Ferri, C. (2004). *Introducción a la Minería de Datos*. Madrid: Pearson Educación S. A.
 15. Ibm.com. (2016). *IBM Knowledge Center*. Recuperado el 14-11-2016, de http://www.ibm.com/support/knowledgecenter/es/SS3RA7_17.0.0/modeler_crispdm_ddita-gentopic1.html.
 16. IBM (2012). *Manual de CRISP-DM de IBM SPSS Modeler*. Estados Unidos: IBM Corp.
 17. Larraga, P. (2008). *Mercado monetario y mercado de renta fija*. Barcelona, España: Bresca.
 18. MEMORIA ANUAL DE SCOTIABANK (2011). Media.scotiabank.com. Recuperado el 04-09-2016, de http://media.scotiabank.com/AR/2011_ARWeb/Spanish/22I2_mda_risk.php
 19. Nevado, V. (2010). *Introducción a las bases de datos relacionales*. Madrid, España: Visión Libros
 20. Palacios, E. (2012). *Modelo de RNA para predecir la morosidad de microcrédito en la Banca Estatal Peruana*. (Tesis de Maestría). Universidad Nacional Mayor de San Marcos.
 21. Pérez, S. (2012). *Tecnología lítica de la puna meridional Argentina*. Buenos Aires, Argentina: Dunken
 22. Gallardo, J. (2009). *Metodología ara la Definición de Requisitos en Proyectos de Data Mining (ER-DM)* (doctorado). Universidad Politécnica de Madrid, España.
 23. Santana, J., Mateos, E. (2014). *El arte de programar en R: un lenguaje para la estadística*. Instituto Mexicano de Tecnología del Agua. Morelos, México.
 24. SBS (2008). *Resolución SBS N° 11356-2008*. Lima, Peru.
 25. Serna, S. (2009). *Comparación de árbol de regresión y clasificación logística* (Tesis de Magíster), Universidad Nacional de Colombia, Colombia.

26. Silberschatz, A. (2002). Fundamentos de bases de datos (4ta. Ed.).77 Madrid, España: McGRAW Hill
27. Vara, A. (2010). *Desde la idea hasta la sustentación: 7 pasos para una tesis exitosa. Un método efectivo para las ciencias empresariales.* (2da Ed.). Universidad de San Martín de Porres. Lima, Perú.
28. Yatchesen, F. (2015). *Aplicación de técnicas de minería de datos para la detección de fraude tributario, caso de estudio Municipalidad de Apóstoles, provincia de Misiones* (Tesis de grado). Universidad Nacional de Misiones, Argentina.

ANEXOS

ANEXO 1. CONSULTA ESTRUCTURADA PARA BASE DE DATOS PARA GENERAR CARTERA

DECLARE

```

LC_CALFANT VARCHAR2 (16) ;
LN_TIPTER VARCHAR2 (2) ;
LC_CALFACT VARCHAR2 (16) ;
LN_DIASANT NUMBER;
LN_AUTORIZADOR VARCHAR2 (128) ;
LN_MOTIVO VARCHAR2 (255) ;
LN_GLOSA VARCHAR2 (255) ;
LN_SLDOINT NUMBER (14, 2) ;
LN_RPTA NUMBER (10) ;
PC_RPTA varchar2 (200) ;
LN_IDPROVCAB NUMBER;
LN_PROVI_CONSTI NUMBER (14, 2) ;
LN_PROVI_REQUERI NUMBER (14, 2) ;
LN_PROVI_CONSTI_D NUMBER (14, 2) ;
LN_PROVI_REQUERI_D NUMBER (14, 2) ;
LN_TASA_CAMBIO NUMBER (14, 6) ;
LN_ROWS NUMBER;
LN_IDSALDO NUMBER;
LC_MERCADO VARCHAR2 (512) ;
LC_GESTOR VARCHAR2 (32) ;
PC_DET_RPTA varchar2 (200) ;
LN_IDPERIODO NUMBER;
LC_TIPO VARCHAR2 (1) ;
DDATE DATE;
LC_FECHA VARCHAR2 (200) ;
OUT_FILE TEXT_IO.FILE_TYPE;
OUT_FILE1 TEXT_IO.FILE_TYPE;
LC_LINEBUF VARCHAR2 (6000) ;
LC_LINEBUF1 VARCHAR2 (6000) ;
LD_FECULTMOV DATE;
LN_IDSESIONUSR NUMBER (18) ;
LC_MACHINE VARCHAR2 (128) ;

```

```

LN_IDUSUARIO NUMBER(18);
LN_IDOFILOG NUMBER(18);
LN_IDOFICORG NUMBER(18);
LC_CODUSR VARCHAR2(32);
LC_NOMUSR VARCHAR(256);
LD_FINMES DATE;
LN_IDCABEJCAR NUMBER;
LN_PROV_SINGAR NUMBER(14,2);
LN_PROV_CONGAR NUMBER(14,2);
LN_INTNORMAL NUMBER(14,2);
LN_INTSUSPEN NUMBER(14,2);
LN_MTOGARANTIA NUMBER(14,2);
LC_TIPOGAR VARCHAR2(512);
LN_IDCASTIGO NUMBER(18);
DDATE1 DATE;
LC_ANALISTA VARCHAR2(32);
LC_ANALISTA1 VARCHAR2(128);
LC_EJECUTIVO VARCHAR2(32);
LC_ACTIVIDAD VARCHAR2(32);
LC_CATEGORIA VARCHAR2(32);
LC_INDRECVAL VARCHAR2(1) :='N';
LN_PROVI_CICLI NUMBER(14,2);
LN_PROCI_SALDO NUMBER(14,2);
LN_PROCI_REQUERIDA NUMBER(14,2);
LN_PROCI_CONSTIUIDA NUMBER(14,2);
LN_FLAG NUMBER;
LN_FLAG1 NUMBER;
LN_FLAG2 NUMBER;
LN_FLAG3 NUMBER;
LN_FLAG4 NUMBER;
LN_FLAG5 NUMBER;
LN_FLAG6 NUMBER;
LN_OFICINA VARCHAR2(128);
LC_CATEGORIA1 VARCHAR2(64);
LN_CAPRECUP NUMBER(18,2);
LN_INTRECUP NUMBER(18,2);
LN_OTRORECUP NUMBER(18,2);
LN_TOTALCUOTA NUMBER(18,2);
LN_MTODESMB NUMBER(18,2);
LN_CONDSOCIO VARCHAR2(2);
LD_feccastigo date;
---MIGRACION---
TEXTO VARCHAR2(2048);
TEX VARCHAR2(2048);
POS NUMBER(8);
DIRECCION1 VARCHAR2(1024);
UBIGEO VARCHAR2(16);
DEPARTAMENTO VARCHAR2(512);
PROVINCIA VARCHAR2(512);
DISTRITO VARCHAR2(512);
CONT NUMBER(8);

CURSOR C1 (N NUMBER,P_TC NUMBER) IS
select F.NOMOFI,REPLACE(C.NOMPROD,',',' ') NOMPROD,D.CANTCUOTAS
NROCUOTAS,
A.IDPRESTAMO,B.IDTERCERO, b.NUMTERREMOTO
codsocio,b.FULLNOMBRE,c.CODPRODMIG,d.VALKEYCARGA nroprestamo,
decode(d.IDMONEDA,1,'SOLES','DOLARES') MONEDA,NULL
DIRECCION,

```

E.CODUSR,F.CODAGENCIA,D.FECDESEMB,D.FECVENC,D.MTOPRESTAMO,
D.TASAMENSUAL,a.DIAS_ATRASO
,a.MTOSLDOVIG,a.MTOSLDOVEN,a.MTOSLDOJUD,a.MTOSLDOREE,a.MTOSLDOREF,a.Cl
asif,

g.DESCCATEG,
decode(d.idmoneda,1,a.MTOSLDOVEN+a.MTOSLDOJUD,(**NVL**(a.MTOSLDOVEN,0)+**NVL**
(a.MTOSLDOJUD,0))*P_TC) mtovensoles,

decode(d.idmoneda,1,a.MTOSLDOVIG+a.MTOSLDOVEN+a.MTOSLDOJUD+a.MTOSLDORE
E+a.MTOSLDOREF,(
NVL(a.MTOSLDOVIG,0)+**NVL**(a.MTOSLDOVEN,0)+**NVL**(a.MTOSLDOJUD,0)+**NVL**(a.MTOS
LDOREE,0)+**NVL**(a.MTOSLDOREF,0))*P_TC) saldototalsoles
,b.nrodocdef,
b.fecnac,

NVL(a.MTOSLDOVIG,0)+**NVL**(a.MTOSLDOVEN,0)+**NVL**(a.MTOSLDOJUD,0)+**NVL**(a.MTOS
LDOREE,0)+**NVL**(a.MTOSLDOREF,0) saldo,

A.STSPRESTAMO,A.IDMERCADO,B.SEXO,D.TASAANUAL,D.IDUSRANALISTA,

D.CATEGPRES,D.IDCFGACTECO,B.TIPOINTI,D.IDGESTOR,D.IDUSRASESOR
IDUSRASESOR,a.codprovi codprovi,a.codproviorig codproviorig
from pr_det_saldo a,pe_tercero b,pr_prod_prestamo c,pr_prestamo
d,CS_USUARIO E,CS_OFICINA_ORG F,pr_cfg_categ_pres g

where idcabslido **IN** (N,LN_IDCASTIGO)
and a.IDTERCERO=b.IDTERCERO
and c.IDPRODPRES=a.IDPRODPRES
and d.IDPRESTAMO=a.IDPRESTAMO
AND E.IDUSUARIO=A.IDUSRASESOR
AND F.IDOFICORG=A.IDOFIORG
and g.IDCFGATPRES=a.categpres
AND (:C04.IDOFISOCIO=0 **OR** A.IDOFIORG=:C04.IDOFISOCIO)
AND (:C04.IDUSRASESOR=0 **OR** D.IDUSRASESOR=:C04.IDUSRASESOR);
--and

a.MTOSLDOVIG+a.MTOSLDOVEN+a.MTOSLDOJUD+a.MTOSLDOREE+a.MTOSLDOREF>0

FUNCTION DAME_DIRECCION(CCODIGO **NUMBER**) **RETURN** **VARCHAR2** **IS**

N **VARCHAR2** (2000);
LC_UBIGEO **VARCHAR2** (32);
LN_IDDISTRITO **NUMBER** (18);
LC_DISTRITO **VARCHAR2** (512);
LN_IDESTADO **NUMBER** (18);
LN_IDPROV **NUMBER** (18);
LC_PROVINCIA **VARCHAR2** (512);
LC_ESTADO **VARCHAR2** (512);

BEGIN

SELECT ltrim(rtrim(replace(replace(replace(DIRECCION,chr(13),'
'),chr(10),' '),',','.')) ,CODUBIGEO,IDDISTRITO,IDESTADO,IDPROV

INTO N,LC_UBIGEO,LN_IDDISTRITO,LN_IDESTADO,LN_IDPROV

FROM PE_DIRECCION_TERCERO

WHERE IDTERCERO=CCODIGO **AND** ROWNUM=1;

IF LN_IDDISTRITO>0 **THEN**

BEGIN

SELECT REPLACE(DESCRIPCION,',',' '),CODUBIGEO

INTO LC_DISTRITO,LC_UBIGEO

FROM CS_DISTRITO

WHERE IDDISTRITO = LN_IDDISTRITO;

EXCEPTION

WHEN OTHERS **THEN**

```

        LC_DISTRITO:=NULL;
    END;
END IF;
IF LN_IDESTADO>0 THEN
    BEGIN
        SELECT REPLACE (DESCRIPCION,',' ,',' ' ')
        INTO LC_ESTADO
        FROM CS_ESTADO
        WHERE IDESTADO = LN_IDESTADO;
    EXCEPTION
        WHEN OTHERS THEN
            LC_ESTADO:=NULL;
    END;
END IF;
IF LN_IDPROV>0 THEN
    BEGIN
        SELECT REPLACE (DESCRIPCION,',' ,',' ' ')
        INTO LC_PROVINCIA
        FROM CS_PROVINCIA
        WHERE IDPROVINCIA = LN_IDPROV;
    EXCEPTION
        WHEN OTHERS THEN
            LC_PROVINCIA:=NULL;
    END;
END IF;
RETURN
N||',' ||LC_UBIGEO||',' ||LC_ESTADO||',' ||LC_PROVINCIA||',' ||LC_DISTRITO
;
    EXCEPTION WHEN OTHERS THEN
        RETURN ', , , ,';
    END;
    ---
FUNCTION DIRECCION(CCODIGO NUMBER) RETURN VARCHAR2 IS
    DIRECCION VARCHAR2 (2000);
BEGIN
    SELECT ltrim( rtrim(replace(replace(replace (DIRECCION,chr(13),'
'),chr(10),' '),',' ,',' ' ') ))
    INTO DIRECCION
    FROM PE_DIRECCION_TERCERO;
    RETURN DIRECCION;
    EXCEPTION WHEN OTHERS THEN
        RETURN '';
    END;

begin
    BEGIN

:P_TC:=PKG_GL_TASA_CAMBIO.OBTENER_TASA('3',:C04.FECFIN,2,1,:P_IDORG,'F
');
    EXCEPTION WHEN OTHERS THEN

:P_TC:=PKG_GL_TASA_CAMBIO.OBTENER_TASA('2',:C04.FECFIN,2,1,:P_IDORG,'F
');
    END;

    LD_FINMES:=ADD_MONTHS (LAST_DAY (:C04.FECFIN) ,-1);
    SELECT IDCABEJCAR,FECEVAL,IDPROVCAB
    INTO LN_IDCABEJCAR,LD_FINMES,LN_IDPROVCAB
    FROM PR_CAB_EJEC_CARTERA A
    WHERE A.IDORG = :P_IDORG

```

```

AND A.INDCIERRE = 'S'
AND A.STSEJEC = 'ACT'
AND A.FECEVAL = (SELECT MAX(FECEVAL)
                  FROM PR_CAB_EJEC_CARTERA B
                  WHERE B.IDORG = :P_IDORG
                        AND B.INDCIERRE = 'S'
                        AND B.STSEJEC = 'ACT'
                        AND B.FECEVAL<=:C04.FECFIN
                  );

--MESSAGE('SE GENERARA ARCHIVO');
SET_APPLICATION_PROPERTY(CURSOR_STYLE,'BUSY');
PKG_PR_CAB_SALDO.PROCESAR_SALDO(:C04.FECFIN,
                                :P_IDORG,
                                0,
                                0,
                                :C04.IDOFISOCIO,
                                1,
                                :C04.P_IDROW,
                                DDATE,
                                '0',
                                NULL);

:C04.IDCASTIGO:=0;
IF :C04.CASTIGO='S' THEN
    PKG_PR_CAB_SALDO.PROCESAR_SALDO_CASTIGO(:C04.FECFIN,
                                             :P_IDORG,
                                             0,
                                             0,
                                             1,
                                             LN_IDCASTIGO,
                                             DDATE1
                                             );

:C04.IDCASTIGO:=LN_IDCASTIGO;
END IF;
commit;

select COUNT(*) INTO LN_FLAG4 from cs_oficina_org where
idoficorg=:C04.IDOFISOCIO;
IF LN_FLAG4>0 THEN
    select replace(nomofi,' ','_') INTO LN_OFICINA from
cs_oficina_org where idoficorg=:C04.IDOFISOCIO;
ELSE
    LN_OFICINA:='GENERAL';
END IF;
SELECT COUNT(*) INTO CONT FROM PR_CARTERA_GERENCIAL WHERE
ANIO=TO_CHAR(:C04.FECFIN,'YYYY') AND MES=TO_CHAR(:C04.FECFIN,'MM');
LC_FECHA:=TO_CHAR(SYSDATE,'DDMMYYYY');
--RIESGOS
OUT_FILE :=
TEXT_IO.FOPEN('C:\CG_'||LN_OFICINA||'_':C04.FECFIN||'.CSV', 'W');
LC_LINEBUF
:='CODSOCIO, TIPO_PERSONA, NOMBRES, CONDIC_SOCIO, DIRECCION, UBIGEO, DEPTO, P
ROVINCIA, DISTRITO, PRODUCTO, NROPRESTAMO, MONEDA, CODUSR, USUARIO, CARGO, COD
AGENCIA, FECDESEMB, FECVENC, MTOPRESTAMO, MTONETODESEMB, TASAMENSUAL, DIASAT

```



```

SELECT COUNT(*) INTO LN_FLAG2 FROM CS_USUARIO WHERE
IDUSUARIO=(SELECT IDUSRACT FROM PR_PRESTAMO WHERE
IDPRESTAMO=r1.idprestamo);
IF LN_FLAG2 <> 0 THEN
BEGIN
--LN_AUTORIZADOR:=LN_FLAG2;
SELECT FULLNOMBRE into LN_AUTORIZADOR FROM PE_TERCERO WHERE
IDTERCERO=(SELECT IDTERCERO FROM CS_USUARIO WHERE IDUSUARIO=(SELECT
IDUSRACT FROM PR_PRESTAMO WHERE IDPRESTAMO=r1.idprestamo));
END;
ELSE
LN_AUTORIZADOR:='N/A';
END IF;

SELECT MAX(FECTR)
INTO LD_FECULTMOV
FROM PR_EC_CUOTAS
WHERE IDPRESTAMO=R1.IDPRESTAMO
AND NVL(STSREG, 'ACT')='ACT';
LC_MERCADO:=NULL;
LC_ANALISTA:=NULL;
LC_EJECUTIVO:=NULL;
LC_CATEGORIA:=NULL;
LC_ACTIVIDAD:=NULL;

LN_FLAG:=0;
SELECT IDEVALUACION INTO LN_FLAG FROM PR_PRESTAMO WHERE
IDPRESTAMO = R1.IDPRESTAMO;

IF LN_FLAG IS NOT NULL THEN
BEGIN
SELECT COUNT(IDEVALUACION) INTO LN_FLAG1 FROM
EC_DESTINO_EVAL WHERE IDEVALUACION=LN_FLAG;
SELECT COUNT(*) INTO LN_FLAG3 FROM CS_VAL_VALUE_SET CS
WHERE CS.CODVALOR = (SELECT IDMOTIVO FROM EC_DESTINO_EVAL WHERE
IDEVALUACION=LN_FLAG AND ROWNUM=1) AND CS.IDVALUESET=1002030;

IF LN_FLAG1 <> 0 THEN
BEGIN
SELECT NVL(replace(GLOSA, '-', ' '), ' ') INTO
LN_GLOSA FROM EC_DESTINO_EVAL WHERE IDEVALUACION=LN_FLAG AND
ROWNUM=1;

END;
ELSE
LN_GLOSA:='CREDITO CON EVALUACION
ELIMINADA';
END IF;

IF LN_FLAG3 <> 0 THEN
BEGIN
SELECT NVL(CS.DESCVLOR, '--') INTO LN_MOTIVO
FROM CS_VAL_VALUE_SET CS WHERE CS.CODVALOR = (SELECT IDMOTIVO FROM
EC_DESTINO_EVAL WHERE IDEVALUACION=LN_FLAG AND ROWNUM=1) AND
CS.IDVALUESET=1002030 ;

END;
ELSE
LN_MOTIVO:='CREDITO SIN DESTINO_1';

```



```

                                END IF;
        END;
ELSE
    BEGIN
        LN_MOTIVO:='CREDITO SIN EVALUACION CREADA';
        LN_GLOSA:='CREDITO SIN EVALUACION CREADA';
        END;
    END IF;

    IF R1.IDMERCADO>0 THEN
        BEGIN
            SELECT REPLACE(DESCMERCADO,',',' ')
            INTO    LC_MERCADO
            FROM    PR_MERCADO
            WHERE   IDMERCADO = R1.IDMERCADO;
            EXCEPTION
                WHEN OTHERS THEN
                    LC_MERCADO:=NULL;
            END;
        END IF;
        IF R1.IDUSRANALISTA>0 THEN
            BEGIN
                SELECT CODUSR
                INTO    LC_ANALISTA
                FROM    CS_USUARIO
                WHERE   IDUSUARIO = R1.IDUSRANALISTA;
                EXCEPTION
                    WHEN OTHERS THEN
                        LC_ANALISTA:=NULL;
                END;
            END IF;
            IF R1.IDUSRASESOR>0 THEN
                BEGIN
                    SELECT CODUSR
                    INTO    LC_EJECUTIVO
                    FROM    CS_USUARIO
                    WHERE   IDUSUARIO = R1.IDUSRASESOR;
                    EXCEPTION
                        WHEN OTHERS THEN
                            LC_EJECUTIVO:=NULL;
                    END;
                END IF;
                IF R1.IDCFGACTECO>0 THEN
                    BEGIN
                        SELECT CODACTIV
                        INTO    LC_ACTIVIDAD
                        FROM    PR_CFG_ACTIECON B
                        WHERE   B.IDCFGACTECO=R1.IDCFGACTECO;
                        EXCEPTION
                            WHEN OTHERS THEN
                                LC_ACTIVIDAD:=NULL;
                            END;
                    END IF;
                    IF R1.CATEGPRES>0 THEN
                        BEGIN
                            SELECT DESCcateg
                            INTO    LC_CATEGORIA

```

```

FROM PR_CFG_CATEG_PRES B
WHERE B.IDCFGCATPRES=R1.CATEGPRES;
EXCEPTION
WHEN OTHERS THEN
LC_CATEGORIA:=NULL;
END;
END IF;
IF R1.IDGESTOR>0 THEN
BEGIN
SELECT CODUSRGESTOR
INTO LC_GESTOR
FROM PR_GESTOR_PRES
WHERE IDGESTORPRES = R1.IDGESTOR;
EXCEPTION
WHEN OTHERS THEN
LC_GESTOR:=NULL;
END;
ELSE
LC_GESTOR:=NULL;
END IF;

SELECT COUNT(*) INTO LN_FLAG5 FROM PE_TERCERO WHERE
IDTERCERO=(SELECT IDTERCERO FROM CS_USUARIO WHERE codusr=LC_ANALISTA);

IF LN_FLAG5>0 THEN
SELECT FULLNOMBRE
INTO LC_ANALISTA1
FROM PE_TERCERO
WHERE IDTERCERO=(SELECT IDTERCERO FROM CS_USUARIO WHERE
codusr=LC_ANALISTA);
ELSE
LC_ANALISTA1:=' ';
END IF;

BEGIN
--messAge(ADD_MONTHS (:C04.FECFIN, -1));
SELECT
DECODE (IDMONEDA, '1', PROVSIGARANT+PROVNOGARANT+PROVSIGARANTPRO+PROVNOGA
RANTPRO, (PROVSIGARANT+PROVNOGARANT+PROVSIGARANTPRO+PROVNOGARANTPRO) *PK
G_GL_TASA_CAMBIO.OBTENER_TASA ('3', ADD_MONTHS (:C04.FECFIN, -
1), 2, 1, :P_IDORG, 'F')),
DECODE (IDMONEDA, '1', PROVSIGARACT+PROVNOGARACT+PROVSIGARACTPRO+PROVNOGA
RACTPRO, (PROVSIGARACT+PROVNOGARACT+PROVSIGARACTPRO+PROVNOGARACTPRO) *PK
G_GL_TASA_CAMBIO.OBTENER_TASA ('3', :C04.FECFIN, 2, 1, :P_IDORG, 'F')),
PROVSIGARACTPRO+PROVNOGARACTPRO
INTO LN_PROVI_CONSTI, LN_PROVI_REQUERI, LN_PROVI_CICLI
FROM PR_PROVISION_DET
WHERE IDPROVCAB = LN_IDPROVCAB
AND IDPRESTAMO = R1.IDPRESTAMO;
EXCEPTION
WHEN OTHERS THEN
LN_PROVI_CONSTI:=0;
LN_PROVI_REQUERI:=0;
LN_PROVI_CICLI:=0;
END;

--- SISTEMAS---
BEGIN

```

```

SELECT
DECODE(IDMONEDA,'2',PROVSIGARANT+PROVNOGARANT+PROVSIGARANTPRO+PROVNOGARANTPRO,0),

DECODE(IDMONEDA,'2',PROVSIGARACT+PROVNOGARACT+PROVSIGARACTPRO+PROVNOGARACTPRO,0)
INTO LN_PROVI_CONSTI_D,LN_PROVI_REQUERI_D
FROM PR_PROVISION_DET
WHERE IDPROVCAB = LN_IDPROVCAB
AND IDPRESTAMO = R1.IDPRESTAMO;
EXCEPTION
WHEN OTHERS THEN
LN_PROVI_CONSTI:=0;
LN_PROVI_REQUERI:=0;
LN_PROVI_CICLI:=0;
END;
--MONTO NETO DESEMBLSADO
SELECT PR.MTOPRESTAMO-PR.MTOPRESXCANC-PR.MTODISMIN INTO
LN_MTODESMB FROM PR_PRESTAMO PR WHERE IDPRESTAMO=R1.IDPRESTAMO;

--CONDICION DEL SOCIO
LN_CONDSOCIO:=PKG_PE_TERCERO.GET_TIPO_SOCIO(R1.IDTERCERO);

SELECT
SUM(AMORTIZACAP),SUM(INTERES),SUM(OTROMTO),SUM(AMORTIZACAP)+SUM(INTERES)+SUM(OTROMTO)
INTO LN_CAPRECUP,LN_INTRECUP,LN_OTRORECUP,LN_TOTALCUOTA
FROM PR_CUOTA_PRESTAMO
WHERE IDPRESTAMO=R1.IDPRESTAMO
AND FECPAGO BETWEEN TRUNC(:C04.FECFIN,'MM') AND
:C04.FECFIN;
--
BEGIN
SELECT SUM(SLDOINTE)
INTO LN_SLDOINT
FROM PR_DET_EJEC_CARTERA
WHERE IDCABEJCAR=LN_IDCABEJCAR
AND IDPRESTAMO = R1.IDPRESTAMO;
EXCEPTION WHEN OTHERS THEN
LN_SLDOINT:=0;
END;
BEGIN
SELECT DECODE(CODCATEGACT,'A',SLDOCREDITO,0), --SLDOCREDITO,
PROVSIGARACTPRO+PROVNOGARACTPRO,-- PROCI_REQUERIDA,
PROVNOGARANTPRO+PROVSIGARANTPRO-- PROCI_CONSTIUIDA,
INTO LN_PROCI_SALDO,LN_PROCI_REQUERIDA,LN_PROCI_CONSTIUIDA
FROM PR_PROVISION_DET
WHERE IDPROVCAB= LN_IDPROVCAB
AND IDPRESTAMO = R1.IDPRESTAMO;
EXCEPTION
WHEN OTHERS THEN
LN_PROCI_SALDO:=0;
LN_PROCI_REQUERIDA:=0;
LN_PROCI_CONSTIUIDA:=0;
END;
select feccastigo into LD_feccastigo from pr_prestamo where
idprestamo=R1.IDPRESTAMO;
--FN_IDCARGO_EMPLEADO(pr.IDUSRANALISTA)
begin

```

```

        select idusuario INTO LN_IDUSUARIO from cs_usuario where
codusr=R1.CODUSR;
        SELECT DESCARGO INTO LC_CATEGORIA1 FROM RH_CARGO_X_AREA
WHERE IDCARGOXAREA= FN_IDCARGO_EMPLEADO(LN_IDUSUARIO);
        EXCEPTION
        WHEN OTHERS THEN
            LC_CATEGORIA1:='POR ASISGNAR';
end;

        TEXTO:=DAME_DIRECCION(R1.IDTERCERO);

        POS:=instr(TEXTTO,' ');
        DIRECCION1:=SUBSTR(TEXTTO,1,POS-1);
        TEXTO:=SUBSTR(TEXTTO,POS+1);

        POS:=INSTR(TEXTTO,' ');
        UBIGEO:=SUBSTR(TEXTTO,1,POS-1);
        TEXTO:=SUBSTR(TEXTTO,POS+1);

        POS:=INSTR(TEXTTO,' ');
        DEPARTAMENTO:=SUBSTR(TEXTTO,1,POS-1);
        TEXTO:=SUBSTR(TEXTTO,POS+1);

        POS:=INSTR(TEXTTO,' ');
        PROVINCIA:=SUBSTR(TEXTTO,1,POS-1);
        TEXTO:=SUBSTR(TEXTTO,POS+1);

        DISTRITO:=SUBSTR(TEXTTO,1);

        LC_LINEBUF :=R1.CODSOCIO||','||'|'
                    LN_TIPTER||','||'|'
                    ltrim(
rtrim(replace(replace(replace(r1.FULLNOMBRE,chr(13),' '),chr(10),'
'),',',' ','.'))||','||'|'
                    LN_CONDSOCIO||','||'|'
                    DAME_DIRECCION(R1.IDTERCERO)||','||'|'
                    R1.CODPRODMIG||','||'|'
                    R1.NROPRESTAMO||','||'|'
                    R1.MONEDA||','||'|'
                    r1.CODUSR||','||'|'
                    LC_NOMUSR||','||'|'
                    LC_CATEGORIA1||','||'|'
                    R1.CODAGENCIA||','||'|'
                    to_char(R1.FECDESEMB,'dd/mm/yyyy')||','||'|'
                    to_char(R1.FECVENC,'dd/mm/yyyy')||','||'|'
                    R1.MTOPRESTAMO||','||'|'
                    LN_MTODESMB||','||'|'
                    R1.TASAMENSUAL||','||'|'
                    R1.DIAS_ATRASO||','||'|'
                    R1.MTOSLDOVIG||','||'|'
                    R1.MTOSLDOVEN||','||'|'
                    R1.MTOSLDOJUD||','||'|'
                    R1.MTOSLDOREE||','||'|'
                    R1.MTOSLDOREF||','||'|'
                    R1.CLASIF||','||'|'
                    R1.DESCCATEG||','||'|'
                    R1.MTOVENSOLES||','||'|'
                    R1.SALDOTOTALSOLES||','||'|'

```

```

to_char(LD_FINMES,'dd/mm/yyyy')||','||
LN_PROV_SINGAR||','||
LN_PROV_CONGAR||','||
LN_INTNORMAL||','||
LN_INTSUSPEN||','||
LN_MTOGARANTIA||','||
LC_TIPOGAR||','||
R1.NOMOFI||','||
LN_AUTORIZADOR||','||
R1.NOMPROD||','||
R1.NROCUOTAS||','||
to_char(LD_FECULTMOV,'dd/mm/yyyy')||','||
R1.nrodocdef||','||
to_char(R1.FECNAC,'dd/mm/yyyy')||','||
TRUNC(months_between(sysdate,R1.FECNAC)/12)||','||
R1.SALDO||','||
R1.STSPRESTAMO||','||
LD_feccastigo||','||
LC_MERCADO||','||
R1.SEXO||','||
R1.TASAANUAL||','||
LC_ANALISTA1||','||
LC_CATEGORIA||','||
LC_ACTIVIDAD||','||
R1.TIPOINSTI||','||
LC_GESTOR||','||
LN_PROVI_CONSTI||','||
LN_PROVI_REQUERI||','||
LN_PROVI_CONSTI_D||','||
LN_PROVI_REQUERI_D||','||
LN_PROVI_CICLI||','||
LC_EJECUTIVO||','||
LN_PROCI_SALDO||','||
LN_PROCI_REQUERIDA||','||
LN_PROCI_CONSTIUIDA||','||
R1.codprovi||','||
R1.codproviorig||','||
LN_SLDOINT||','||
LN_MOTIVO||','||
LN_GLOSA||','||
LN_CAPRECUP||','||
LN_INTRECUP||','||
LN_OTRORECUP||','||
LN_TOTALCUOTA;

BEGIN
IF LD_FINMES=:C04.FECFIN AND CONT=0 THEN
INSERT INTO PR_CARTERA_GERENCIAL
(
CODSOCIO,
TIPTER,
FULLNOMBRE,
DIRECCION,
UBIGEO,
DEPARTAMENTO,
PROVINCIA ,
DISTRITO ,
CODPRODUCTO ,
NROPRESTAMO ,
MONEDA ,

```

CODUSR ,
NOMUSR ,
CARGO ,
CODAGENCIA ,
FECDESEMB ,
FECVENC ,
MTOPRESTAMO ,
TASAMENSUAL ,
DIAS_ATRASO ,
MTOSLDOVIG ,
MTOSLDOVEN ,
MTOSLDOJUD ,
MTOSLDOREE ,
MTOSLDOREF ,
CLASIF , --1
DESCCATEG ,
MTOVENSOLES ,
SALDOTOTALSOLES ,
FECPROVIC ,
PROV_SINGAR ,
PROV_CONGAR ,
INTNORMAL ,
INTSUSPEN ,
MTOGARANTIA ,
TIPOGAR ,
NOMOFI ,
AUTORIZADOR ,
NOMPROD ,
NROCUOTAS ,
FECULTMOV ,
nrodocdef ,
FECNAC ,
EDAD ,
SALDO ,
STSPRESTAMO ,
feccastigo ,
MERCADO ,
SEXO ,
TASAANUAL ,
ANALISTA1 ,
PLANINVER ,
ACTIVIDAD ,
TIPOINSTI ,
GESTOR ,
PROVI_CONSTI ,
PROVI_REQUERI ,
PROVI_CONSTI_D ,
PROVI_REQUERI_D ,
PROVI_CICLI ,
EJECUTIVO ,
PROCI_SALDO ,
PROCI_REQUERIDA ,
PROCI_CONSTIUIDA ,
CALFCALIN ,
CALFSALIN ,
SLDOINT ,
DESTINO1 ,
DESTINO2 ,
DIASANT ,
CALFANT ,

```

CALFACT ,
LN_CAPRECUP ,
LN_INTRECUP ,
LN_OTRORECUP ,
ANIO ,
MES ,
IDCARTGER
)
VALUES (
R1.CODSOCIO ,
LN_TIPTER ,
ltrim( rtrim(replace(replace(replace(replace(r1.FULLNOMBRE,chr(13),'
'),chr(10),' '),',' ','.')) ,
DIRECCION1 ,
UBIGEO ,
DEPARTAMENTO ,
PROVINCIA ,
DISTRITO ,
R1.CODPRODMIG ,
R1.NROPRESTAMO ,
R1.MONEDA ,
r1.CODUSR ,
LC_NOMUSR ,
replace(LC_CATEGORIAL,' ',''),
R1.CODAGENCIA ,
R1.FECDESEMB,--22
R1.FECVENC ,
R1.MTOPRESTAMO ,
R1.TASAMENSUAL ,
R1.DIAS_ATRASO ,
R1.MTOSLDOVIG ,
R1.MTOSLDOVEN ,
R1.MTOSLDOJUD ,
R1.MTOSLDOREE ,
R1.MTOSLDOREF ,
R1.CLASIF ,
R1.DESCCATEG ,
R1.MTOVENSOLES ,
R1.SALDOTOTALSOLES ,
LD_FINMES ,
LN_PROV_SINGAR ,
LN_PROV_CONGAR ,
LN_INTNORMAL ,
LN_INTSUSPEN ,
LN_MTOGARANTIA ,
LC_TIPOGAR ,
R1.NOMOFI ,
LN_AUTORIZADOR ,
R1.NOMPROD ,
R1.NROCUOTAS ,
LD_FECULTMOV ,
R1.nrodocdef ,
R1.FECNAC ,
TRUNC(months_between(sysdate,R1.FECNAC)/12) ,
R1.SALDO ,
R1.STSPRESTAMO ,
LD_feccastigo ,
LC_MERCADO ,
R1.SEXO ,
R1.TASAANUAL ,

```

```

LC_ANALISTA1,
LC_CATEGORIA,
LC_ACTIVIDAD,
R1.TIPOINSTI,
LC_GESTOR,
LN_PROVI_CONSTI,
LN_PROVI_REQUERI,
LN_PROVI_CONSTI_D,
LN_PROVI_REQUERI_D,
LN_PROVI_CICLI,
LC_EJECUTIVO,
LN_PROCI_SALDO,
LN_PROCI_REQUERIDA,
LN_PROCI_CONSTIUIDA,
R1.codprovi,
R1.codproviorig,
LN_SLDOINT,
LN_MOTIVO,
LN_GLOSA,
LN_DIASANT,
LC_CALFANT,
LC_CALFACT,
LN_CAPRECUP,
LN_INTRECUP,
LN_OTRORECUP,
TO_CHAR(:C04.FECFIN,'YYYY'),
TO_CHAR(:C04.FECFIN,'MM'),
SQ_CARTGER.NEXTVAL
);
END IF;
EXCEPTION WHEN OTHERS THEN
    MESSAGE (R1.NROPRESTAMO);
END;

LC_LINEBUF1 :=R1.CODSOCIO||','||
ltrim(
rtrim(replace(replace(replace(r1.FULLNOMBRE,chr(13),' '),chr(10),'
'),',' ','.'))||','||

DIRECCION(R1.IDTERCERO)||','||

R1.NROPRESTAMO||','||
r1.CODUSR||','||
R1.CODAGENCIA||','||
R1.MTOPRESTAMO||','||

to_char(R1.FECDESEMB,'dd/mm/yyyy')||','||

R1.TASAMENSUAL||','||
R1.DIAS_ATRASO||','||
R1.DESCCATEG||','||
R1.MTOVENSOLES||','||
R1.SALDOTOTALSOLES||','||
R1.NOMOFI||','||
R1.NOMPROD||','||
R1.STSPRESTAMO||','||
LC_GESTOR||','||
LN_PROVI_REQUERI||','||
LC_EJECUTIVO||','||
LC_ANALISTA1

;
TEXT_IO.PUT_LINE (OUT_FILE,LC_LINEBUF);
TEXT_IO.PUT_LINE (OUT_FILE1,LC_LINEBUF1);

```



```

END LOOP;
Text_IO.Fclose(OUT_FILE);
Text_IO.Fclose(OUT_FILE1);

LN_IDSESIONUSR:=PKG_CS_SESION_USR.GET_IDSESIONUSR;

SELECT MACHINE, IDUSUARIO, IDOFILOG
INTO LC_MACHINE, LN_IDUSUARIO, LN_IDOFILOG
FROM CS_SESION_USR
WHERE IDSESIONUSR=LN_IDSESIONUSR;

IF :C04.INDRIESGO='S' THEN
  IF :C04.FECFIN=LAST_DAY(:C04.FECFIN) THEN
    SELECT MAX(IDCABEJCAR), MAX(IDSLDOPRES)
    INTO LN_IDCABEJCAR, :C04.IDSLDOPRES
    FROM PR_CAB_EJEC_CARTERA A
    WHERE A.IDORG = :P_IDORG
      AND A.INDCIERRE = 'S'
      AND A.STSEJEC = 'ACT'
      AND A.FECEVAL =:C04.FECFIN;
    :C04.IDCABEJCAR:=LN_IDCABEJCAR;
  ELSE
    SELECT SQ_IDCABEJCAR.NEXTVAL INTO LN_IDCABEJCAR FROM DUAL;
  END IF;
END IF;

LN_TASA_CAMBIO:=PKG_GL_TASA_CAMBIO.OBTENER_TASA('2',:C04.FECFIN,2,1,:P_IDORG,'F');
INSERT INTO PR_CAB_EJEC_CARTERA
(IDCABEJCAR, IDPERIODO, FECCREA,
FECEVAL, IDUSUARIO, IDORG, STSEJEC, INDCIERRE, MTOTC,
FECINIPROC, IDOFICORG, TIPOFILTRO)
VALUES
(LN_IDCABEJCAR, 0, SYSDATE,
:C04.FECFIN, LN_IDUSUARIO, :P_IDORG, 'REG', 'N', LN_TASA_CAMBIO,
SYSDATE, 0, '0'
);
PKG_PR_CAB_EJEC_CARTERA.ACTIVAR(LN_IDCABEJCAR,
'N'
);
LN_IDSALDO:=PKG_PR_SALDOS_PRES.CREAR_DATOS(LN_IDCABEJCAR);
COMMIT;
:C04.IDSLDOPRES:=LN_IDSALDO;
:C04.IDCABEJCAR:=LN_IDCABEJCAR;
END IF;
END IF;

INSERT INTO AU_EJEC_PANTALLA
(IDAUEJECPAN, IDUSUARIO, FECHA,
HORA, IDOBJMOD, CODSOCIO,
IDOFICORG, IDORG, MACHINE, IDROW, IDSLDOPRES, IDCABEJCAR
)
VALUES
(SQ_IDAUEJECPAN.NEXTVAL, LN_IDUSUARIO, TRUNC(SYSDATE),
TO_CHAR(SYSDATE, 'HH24:MI:SS'), 4581, null,
LN_IDOFILOG, :P_IDORG, LC_MACHINE, :C04.P_IDROW, :C04.IDSLDOPRES, :C04.IDCABEJCAR
);
:SYSTEM.MESSAGE_LEVEL:=25;

COMMIT;

```

```

SET_APPLICATION_PROPERTY(CURSOR_STYLE,'DEFAULT');

LC_INDRECVAL := 'S';
IF LC_INDRECVAL = 'S' THEN
    LN_RPTA := ALERTA('NOTIFICA','Archivo Generado ','Mensaje');
END IF;

EXCEPTION WHEN OTHERS THEN

MESSAGE('ERROR ' || SQLERRM);
Text_IO.Fclose(OUT_FILE);
Text_IO.Fclose(OUT_FILE1);

SET_APPLICATION_PROPERTY(CURSOR_STYLE,'DEFAULT');
--- COMMIT;
END;

```

ANEXO 2: GUÍA DE OBSERVACIÓN

GUÍA DE OBSERVACIÓN	
RESPONSABLE:	
FECHA OBSERVACIÓN	
DESCRIPCIÓN	

ANEXO 3: FICHA BIBLIOGRÁFICA

FICHA BIBLIOGRÁFICA	
AUTOR	
TITULO	
PAGINA	
RESUMEN DEL CONTENIDO	