

**UNIVERSIDAD NACIONAL DE CAJAMARCA**

**ESCUELA DE POSGRADO**



**MAESTRÍA EN CIENCIAS**

**MENCIÓN: ADMINISTRACIÓN Y GERENCIA EMPRESARIAL**

**TESIS**

**EFFECTO DE LA IMPLEMENTACIÓN DE MINERÍA DE DATOS EN LA  
CALIFICACIÓN DE CRÉDITO A LOS CLIENTES DE LA MICRO  
FINANCIERA PROGRESO DE LA CIUDAD DE CAJAMARCA**

Para optar el Grado Académico de

**MAESTRO EN CIENCIAS**

Presentada por:

**JAIME AMADOR MEZA HUAMÁN**

Asesor:

**Dr. HÉCTOR DIÓMEDES VILLEGAS CHÁVEZ**

CAJAMARCA - PERÚ

2018

COPYRIGHT © 2018 by  
**JAIME AMADOR MEZA HUAMÁN**  
Todos los derechos reservados

# **UNIVERSIDAD NACIONAL DE CAJAMARCA**

## **ESCUELA DE POSGRADO**



## **MAESTRÍA EN CIENCIAS**

### **MENCIÓN: ADMINISTRACIÓN Y GERENCIA EMPRESARIAL**

#### **TESIS APROBADA**

#### **EFFECTO DE LA IMPLEMENTACIÓN DE MINERÍA DE DATOS EN LA CALIFICACIÓN DE CRÉDITO A LOS CLIENTES DE LA MICRO FINANCIERA PROGRESO DE LA CIUDAD DE CAJAMARCA**

Para optar el Grado Académico de

#### **MAESTRO EN CIENCIAS**

Presentada por:

**JAIME AMADOR MEZA HUAMÁN**

#### **Comité Científico**

Dr. Héctor Diómedes Villegas Chávez  
Asesor

Dr. Alejandro Vásquez Ruiz  
Jurado Evaluador

Dr. Julio Sánchez De La Puente  
Jurado Evaluador

Mg. Carlos Aparicio Arteaga  
Jurado Evaluador

CAJAMARCA, PERÚ

2018



# Universidad Nacional de Cajamarca

"NORTE DE LA UNIVERSIDAD PERUANA"

## Escuela de Posgrado

CAJAMARCA - PERÚ

### ACTA DE SUSTENTACIÓN PÚBLICA DE TESIS

Siendo las 11:00 am. de la mañana del día 27 de marzo de 2018, reunidos en el Auditorio de la Escuela de Posgrado de la Universidad Nacional de Cajamarca, los miembros del Jurado Evaluador presidido por el **Dr. ALEJANDRO VÁSQUEZ RUIZ**, y como Miembros del Jurado Evaluador, **Dr. JULIO SÁNCHEZ DE LA PUENTE** y **M.Cs. CARLOS APARICIO ARTEAGA**, en calidad de Asesor **Dr. HÉCTOR VILLEGAS CHÁVEZ**; actuando de conformidad con el Reglamento de la Escuela de Posgrado, se dio inicio a la **SUSTENTACIÓN PÚBLICA** de la tesis titulada **EFECTO DE LA IMPLEMENTACIÓN DE MINERÍA DE DATOS EN LA CALIFICACIÓN DE CRÉDITO A LOS CLIENTES DE LA MICRO FINANCIERA PROGRESO DE LA CIUDAD DE CAJAMARCA**; presentada por el **Bach. en Ingeniería de Sistemas JAIME AMADOR MEZA HUAMÁN**, con la finalidad de optar el Grado Académico de **MAESTRO EN CIENCIAS**, en la Unidad de Posgrado de la Facultad de Ciencias Económicas, Contables y Administrativas, con Mención en **ADMINISTRACIÓN Y GERENCIA EMPRESARIAL**.

Realizada la exposición de la Tesis y absueltas las preguntas formuladas por el Jurado Evaluador, y luego de la deliberación, se acordó... Aprobada..... la mencionada Tesis con la calificación de 18 (Dieciocho)..... Excelente.....; en tal virtud el alumno **Bach. en Ingeniería de Sistemas JAIME AMADOR MEZA HUAMÁN**, está apto para recibir en ceremonia especial el Diploma que la acredita como **MAESTRO EN CIENCIAS**, en la Unidad de Posgrado de la Facultad de Ciencias Económicas, Contables y Administrativas, con Mención en **ADMINISTRACIÓN Y GERENCIA EMPRESARIAL**.

Siendo las 11:50..... horas del mismo día, se dio por concluido el acto.

Dr. Alejandro Vásquez Ruiz  
JURADO EVALUADOR

Dr. Héctor Villegas Chávez  
ASESOR

Dr. Julio Sánchez De La Puente  
JURADO EVALUADOR

M.Cs. Carlos Aparicio Arteaga  
JURADO EVALUADOR

## DEDICATORIA

A:

A mi familia Meza Espinoza integrada por mi esposa Sofía, mis hijos Eduardo, Fátima y Vicente, así como también mis padres Celia y Julián y hermanos Edita, Liliana y Jorge por su apoyo moral, paciencia, comprensión y consejos de superación. Así como también a mi hermano C.P.C. Rafael que siempre me protegió desde mi niñez hasta donde estuvo en vida y sé que ahora me proteges espiritualmente.

## **AGRADECIMIENTO**

Al asesor de la tesis Dr. Héctor Diómedes Villegas Chávez por la dedicación y esfuerzo en la orientación académica brindada durante el desarrollo de la presente tesis.

A los Ingenieros Walter Pérez y Gilmer Cayotopa Jara pertenecientes al área de desarrollo y soporte del sistema de información de créditos de la Institución Micro Financiera Progreso por el apoyo constante en brindar las facilidades para obtener la información necesaria para el presente desarrollo de la tesis.

Al Director de la Escuela Académica Profesional de Ingeniería de Sistemas MSc. Ing. Carlos Aparicio Arteaga y compañeros del área de Sistemas pertenecientes al Departamento de Sistemas Estadística e Informática da la Universidad Nacional de Cajamarca, por sus buenos consejos brindados durante el desarrollo de esta investigación.

Los datos recogen un conjunto de hechos (una base de datos) y los patrones son expresiones que describen un subconjunto de los datos (un modelo aplicable a ese subconjunto).

José M. Molina y Jesús García

## ÍNDICE GENERAL

<b>Ítem</b>	<b>Página</b>
<b>DEDICATORIA .....</b>	<b>v</b>
<b>AGRADECIMIENTO.....</b>	<b>vi</b>
<b>ÍNDICE GENERAL.....</b>	<b>viii</b>
<b>LISTA DE TABLAS.....</b>	<b>xiii</b>
<b>LISTA DE CUADROS.....</b>	<b>xv</b>
<b>LISTA DE FIGURAS.....</b>	<b>xvi</b>
<b>LISTA DE SIGLAS.....</b>	<b>xviii</b>
<b>GLOSARIO.....</b>	<b>xix</b>
<b>RESUMEN .....</b>	<b>xxi</b>
<b>ABSTRACT .....</b>	<b>xxii</b>
 <b>CAPÍTULO I: INTRODUCCIÓN</b>	
<b>1.1. Planteamiento del problema .....</b>	<b>1</b>
<b>1.1.1. Contextualización.....</b>	<b>1</b>
<b>1.1.2. Descripción del problema.....</b>	<b>2</b>
<b>1.1.3. Formulación del problema .....</b>	<b>3</b>
<b>1.2. Justificación e importancia .....</b>	<b>4</b>
<b>1.2.1. Justificación científica.....</b>	<b>4</b>
<b>1.2.2. Justificación técnica-práctica .....</b>	<b>5</b>
<b>1.2.3. Justificación institucional y personal .....</b>	<b>6</b>
<b>1.3. Delimitación de la investigación .....</b>	<b>6</b>
<b>1.4. Objetivos.....</b>	<b>7</b>
<b>1.4.1. Objetivo General .....</b>	<b>7</b>
<b>1.4.2. Objetivos Específicos.....</b>	<b>7</b>

## CAPÍTULO II: MARCO TEÓRICO

<b>2.1. Antecedentes teóricos de la investigación.....</b>	<b>8</b>
2.1.1. Antecedentes Internacionales.....	8
2.1.2. Antecedentes Nacionales.....	10
<b>2.2. Bases teóricas .....</b>	<b>11</b>
2.2.1. Teoría del Crédito.....	11
<b>2.3. Base conceptuales .....</b>	<b>13</b>
2.3.1. Minería de datos .....	13
2.3.2. Técnicas de minerías de datos.....	13
2.3.3. Análisis y modelamiento multidimensional.....	14
2.3.4. Sistemas de información .....	15
2.3.5. Aplicaciones de la minería de datos en la medición de software.....	15
2.3.6. Metodologías de minería de datos.....	17
2.3.7. Factores BioSocio culturales.....	26
2.3.8. Factores económicos.....	26
2.3.9. Calificación de crédito .....	27
2.3.10. Calidad crediticia.....	27
2.3.11. Clasificación de la calidad crediticia.....	27
2.3.12. Riesgo crediticio.....	29
2.3.13. Enfoque cuantitativo de la investigación.....	29
2.3.14. Estudio de alcance descriptivo.....	29
2.3.15. Estudio de alcance correlativo.....	29
2.3.16. Data Mart.....	30
2.3.17. Modelo de regresión logística .....	30
<b>2.4. Definición de términos básicos .....</b>	<b>31</b>

## **CAPÍTULO III: PLANTEAMIENTO DE LAS HIPÓTESIS Y VARIABLES**

<b>3.1. Hipótesis .....</b>	<b>35</b>
3.1.1. Hipótesis General .....	35
<b>3.2. Variables.....</b>	<b>35</b>
3.2.1. Variable X .....	35
3.2.2. Variable Y .....	35
<b>3.3. Operacionalización de los componentes de las hipótesis.....</b>	<b>36</b>
<b>4.1. Ubicación geográfica .....</b>	<b>37</b>
<b>4.2. Diseño de la investigación .....</b>	<b>38</b>
4.2.1. Tipo de Investigación .....	38
4.2.2. Diseño de contrastación de la hipótesis.....	38
<b>4.3. Método de investigación.....</b>	<b>38</b>
<b>4.4. Población, muestra y unidad de análisis.....</b>	<b>39</b>
<b>4.5. Técnicas e instrumentos de recopilación de información .....</b>	<b>41</b>
<b>4.6. Técnicas para el procesamiento y análisis de la información.....</b>	<b>41</b>
<b>4.7. Equipos, materiales e insumos.....</b>	<b>42</b>
<b>4.8. Matriz de consistencia metodológica .....</b>	<b>43</b>

## **CAPÍTULO V: RESULTADOS Y DISCUSIÓN**

<b>5.1. Presentación de resultados.....</b>	<b>44</b>
5.1.1. Compresión del Negocio .....	44
5.1.1.1. Objetivos.....	44
5.1.1.2. Restricciones.....	44
5.1.1.3. Tipo de resultados.....	45
5.1.1.4. Requerimientos del Negocio .....	45

5.1.2.	Comprensión de los Datos.....	48
5.1.3.	Preparación de los Datos .....	51
5.1.3.1.	Revisión de la información.....	51
5.1.3.2.	Diseño de la Ventana de Tiempo.....	53
5.1.3.3.	Modelado dimensional .....	54
5.1.3.4.	Diseño e implementación del subsistema de ETL.....	55
5.1.3.5.	Muestreo .....	62
5.1.4.	Modelo .....	64
5.1.4.1.	Análisis Univariante .....	64
5.1.4.2.	Desarrollo del modelo .....	83
5.1.5.	Evaluación.....	88
<b>5.2.</b>	<b>Análisis, interpretación y discusión de resultados.....</b>	<b>91</b>
5.2.1.	Indicadores de la variable Independiente .....	91
5.2.1.1.	Nivel de respuesta al usuario .....	91
5.2.1.2.	Nivel de conformidad .....	92
5.2.2.	Indicadores de la variable dependiente .....	95
5.2.2.1.	Riesgo crediticio .....	95
5.2.2.2.	Recuperación de capital.....	98
<b>5.3.</b>	<b>Contrastación de hipótesis .....</b>	<b>99</b>
5.3.1.	Formulación de la Hipótesis.....	99
5.3.2.	Ubicación de la región crítica.....	100
5.3.3.	Determinación y ubicación de valor esperado en la región crítica .....	100
5.3.4.	Aceptación o rechazo de la hipótesis .....	101
5.3.5.	Discusión de resultados .....	101
	<b>CONCLUSIONES .....</b>	<b>104</b>
	<b>RECOMENDACIONES Y/O SUGERENCIAS .....</b>	<b>106</b>
	<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>107</b>

<b>APÉNDICES .....</b>	<b>109</b>
APÉNDICE A. Implementación de la base de datos temporal.....	110
APÉNDICE B. Tablas complementarias de la categorización de variables .....	115
APÉNDICE C. Corridas para encontrar las variables en la ecuación del modelo....	126
APÉNDICE C. Algoritmo en SPSS para la construcción del modelo.....	128
APÉNDICE E. Validación y llenado del cuestionario.....	138

## LISTA DE TABLAS

	<b>Pág.</b>
Tabla 1. Montos por proyecto en el año 2015.....	5
Tabla 2. Muestra de la cartera de clientes Buenos y Malos .....	63
Tabla 3. Muestra de la cartera de clientes seleccionados para la construcción del modelo.....	63
Tabla 4. Muestra de la cartera de clientes seleccionados para la validación del modelo .....	63
Tabla 5. Contingencia de la variable Edad (agrupado) por tipo de cliente .....	67
Tabla 6. Medidas direccionales de la variable Edad .....	67
Tabla 7: Medidas direccionales de la variable Edad_Cat .....	69
Tabla 8. Medidas direccionales de la variable Sexo_Cat.....	70
Tabla 9. Medidas direccionales de la variable EstadoCivil_Cat.....	71
Tabla 10. Medidas direccionales de la variable Distrito_Cat .....	72
Tabla 11. Medidas direccionales de la variable Zona_Cat .....	73
Tabla 12. Medidas direccionales de la variable Grado de Instrucción .....	74
Tabla 13. Medidas direccionales de la variable Profesión_Cat .....	74
Tabla 14. Medidas direccionales de la variable EstadoCredito_Cat.....	75
Tabla 15. Medidas direccionales de la variable AntigüedadMeses_Cat.....	76
Tabla 16. Medidas direccionales de la variable RangoMoraMaxima.....	77
Tabla 17. Medidas direccionales de la variable PeorHbitoPago_6meses .....	78
Tabla 18. Porcentaje de clientes malos vs categoría hábito de pago promedio .....	79
Tabla 19. Medidas direccionales de la variable Refinanciamiento.....	80
Tabla 20. Medidas direccionales de la variable SaldoTotal_Cat .....	81
Tabla 21. Medidas direccionales de la variable MontoInteres_Cat .....	82

Tabla 22. Medidas direccionales de la variable Promedio de Pagos_Cat.....	83
Tabla 23. Prueba K-S Aplicado a la muestra de validación.....	89
Tabla 24. Tiempo de requerido por requerimiento .....	91
Tabla 25. Resultado del Pre – Test .....	93
Tabla 26. Resultado del Post – Test.....	94
Tabla 27. Diferencia de medias del Pre y Post – Test.....	94
Tabla 28. Poder Discriminatorio Univariado .....	96
Tabla 29. Contingencia del puntaje agrupado por Saldo Total.....	98

## LISTA DE CUADROS

	<b>Pág.</b>
Cuadro 1. Operacionalización de variables .....	36
Cuadro 2. Primera población de estudio.....	39
Cuadro 3. Técnicas e instrumentos de recopilación de información.....	41
Cuadro 4. Equipos, materiales e insumos.....	42
Cuadro 5. Matriz de consistencia metodológica.....	43
Cuadro 6. Requerimiento R01 .....	46
Cuadro 7. Requerimiento R02.....	46
Cuadro 8. Requerimiento R03 .....	47
Cuadro 9. Requerimiento R04 .....	47
Cuadro 10. Requerimiento R05 .....	47
Cuadro 11. Diccionario de datos a nivel general de la base de datos transaccional.....	50
Cuadro 12. Variables explicativas del modelo a incluir.....	64
Cuadro 13. Variables seleccionadas para el modelo .....	65
Cuadro 14. Categorización de la variable Edad .....	68
Cuadro 15. Ecuación del score en base a dos escenarios .....	84
Cuadro 16. Variables en la ecuación del modelo .....	85
Cuadro 17. Niveles de score para KS.....	90
Cuadro 18. Criterios de evaluación del cuestionario del Pre y Post –Test.....	92

## LISTA DE FIGURAS

	<b>Pág.</b>
Figura 1. Metodología de Ralph Kimball.....	17
Figura 2. Metodología Inmon – Data Warehouse Corporativo.....	22
Figura 3. Ciclo de CRISP.....	23
Figura 4. Mapa de ubicación de la Micro Financiera Progreso.....	37
Figura 5. Modelo relacional de la base de datos del Sistema de Crédito.....	49
Figura 6. Esquema de las tablas de la base de datos temporal.....	52
Figura 7. Diseño de la Ventana de Tiempo.....	53
Figura 8. Diseño físico del Data Mart.....	55
Figura 9. Diseño ETL.....	56
Figura 10. Flujo de datos DIM_Cliente con atributos categóricas.....	57
Figura 11. Flujo de datos DIM_Cliente con atributos escalares.....	58
Figura 12. Flujo de datos DIM_Cliente con atributos escalares - continuación.....	59
Figura 13. Flujo de datos DIM_Solicitud.....	60
Figura 14. Flujo de datos DIM_Credito.....	61
Figura 15. Flujo de datos DIM_FACT.....	62
Figura 16. Porcentaje de clientes malos por categoría edad.....	68
Figura 17. Porcentaje de clientes malos vs categoría Sexo.....	70
Figura 18. Porcentaje de clientes malos vs categoría EstadoCivil_Cat.....	71
Figura 19. Porcentaje de clientes malos vs categoría Distrito_Cat.....	72
Figura 20. Porcentaje de clientes malos vs categoría GradoInstruccion_Cat.....	73
Figura 21. Porcentaje de clientes malos vs categoría estado de Crédito_Cat.....	75
Figura 22. Porcentaje de clientes malos vs categoría AntigüedadMeses_Cast.....	76
Figura 23. Porcentaje de clientes malos vs categoría RangoMoraMaxima.....	77

Figura 24. Porcentaje de clientes malos vs categoría Peor hábito de pago .....	78
Figura 25. Porcentaje de clientes malos vs categoría Hábito de pago promedio .....	79
Figura 26. Porcentaje de clientes malos vs variable Refinanciamiento .....	80
Figura 27. Porcentaje de clientes malos vs categoría Saldo Total .....	81
Figura 28. Porcentaje de clientes malos vs categoría Monto de Interés.....	82
Figura 29. Porcentaje de clientes malos vs categoría Promedio de Pagos_Cat.....	83
Figura 30. Prueba K-S .....	90
Figura 31. Comportamiento de los bueno y malos clientes.....	97
Figura 32. Ubicación del valor esperado en la región crítica .....	100

**LISTA DE SIGLAS**

AFP: Administradoras de Fondos de Pensiones.

B2B: Negocio entre empresas.

B2C: Negocio entre empresa y cliente.

CIF: Fábrica de Información Corporativa.

CRISP-DM: Modelo de proceso de minería de datos que describe los enfoques comunes que utilizan los expertos en minería de datos.

EDPYME: Entidades de Desarrollo de la Pequeña y Micro Empresa.

ETL: Proceso de extracción, transformación y carga.

IMFs: Instituciones de Micro Finanzas.

IRB: Métodos Basados en Calificaciones Internas.

SBS: Superintendencia de Banca y Seguros.

OLAP: Procesamiento analítico en línea.

OLTP: Procesamiento Transaccional en línea.

## GLOSARIO

**Credit Scorig:** Consiste en asignarle un puntaje al potencial deudor que, a su vez, representa una estimación del desempeño del crédito para el banco.

**Clustering:** Conocido como Análisis de Conglomerados, Taxonomía Numérica o Reconocimiento de Patrones, es una técnica estadística multivariante cuya finalidad es dividir un conjunto de objetos en grupos de forma que los perfiles de los objetos en un mismo grupo sean muy similares entre sí.

**Data Mart:** Almacén de datos a los sistemas de extracción y almacenamiento de datos de diversas fuentes, con vistas multidimensionales de un área específica de la organización.

**Data Warehouse:** Es una de las tecnologías más importantes para la obtención de información, ya que lo hace sin sacrificar el rendimiento de las aplicaciones operacionales. Su manera de generar bases tangibles mediante la fusión de datos de múltiples fuentes, permite que éstos se mantengan actualizados sin que cambien la velocidad de los sistemas transaccionales. Además, se convierte en un almacén de datos transaccionales para consultas operativas, con la información necesaria para realizar un análisis estratégico y decisorio de toda la organización.

**Fact Table:** Traducido al español como una tabla de hechos y es la tabla central de un esquema dimensional (en estrella o en copo de nieve) y contiene los valores de las medidas de negocio o dicho de otra forma los indicadores de negocio.

**Framework:** Es un esquema (un esqueleto, un patrón) para el desarrollo y/o la implementación de una aplicación.

**Lenguaje SQL:** se considera, por un lado, un lenguaje diseñado específicamente para la comunicación entre usuarios y, por otro con la base de datos para realizar todas las tareas requeridas para resolver los requerimientos como obtener información almacenada, realizar cálculos, modificar lo existente y agregar nuevas filas que contengan información de clientes, productos y transacciones.

**Metadata:** Se los denomina así porque son datos acerca de los datos.

## RESUMEN

La minería de datos en las instituciones se considera como una herramienta tecnológica orientada a la extracción de datos y el análisis, se desarrolla en base a técnicas estadísticas en las bases de datos y de esta forma se pueden encontrar cualidades ocultas de los clientes refiriéndose a los objetivos del negocio. En tal sentido, esta investigación tiene como objetivo implementar minería de datos en la calificación de crédito a los clientes de la Micro Financiera Progreso, con el propósito de determinar los clientes de alto riesgo y buenos clientes. La hipótesis planteada es: “La implementación de minería de datos a través del modelo planteado mejorará la calificación de crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca”, para contrastar esta hipótesis, se aplicó un diseño no experimental, descriptiva y para medir el efecto producido por la Implementación de Minería de Datos, se aplicó una Pre – prueba y Post – prueba. En la recolección y el procesamiento de datos, se utilizó la metodología de minería de datos CRISP-DM y el cuestionario para saber el nivel de conformidad del área usuaria, se procesa a través del estadístico t Student; después del análisis, interpretación y discusión de los resultados, se concluye que la hipótesis tiene una aceptación positiva porque los usuarios evidenciaron que la implementación de minería de datos: está programado para que se ejecute en un tiempo aproximado 31.083 minutos y la calificación de los clientes se calculó en función del modelo de regresión logística binaria. Con dicha calificación, se determinaron clientes con alto nivel de riesgo que representan el 20.16%, por lo tanto, se puede asegurar el retorno del capital en 80.15%.

**Palabras Clave:** Minería de Datos, Calificación de Crédito, Modelo de Regresión Logística.

## ABSTRACT

The mining of data in institutions is considered as a technological tool that is oriented to data extraction and analysis, it is developed based on statistical techniques on the databases and thus being able to find hidden qualities of the clients referring to the objectives of the business. In this regard, this research aims to implement data mining in the credit rating to the customers of the Micro Financiera Progreso, for the purpose of determining high risk clients and good clients. The hypothesis raised is: "The implementation of the Data Mining through the raised model will improve the credit rating to the customers of the Micro Financiera Progreso of the city of Cajamarca", To contrast this hypothesis, a non-experimental, descriptive design was applied and to measure the effect produced by the Implementation of Data Mining, a Pre - test and Post - test was applied. In data collection and processing, the CRISP-DM data mining methodology was used and the questionnaire to know the level of conformity of the user area is processed through the statistician t Student; after the analysis, interpretation and discussion of the results, it is concluded that the hypothesis has a positive acceptance because the users evidenced that the implementation of data mining: is programmed to run in an approximate time of 31.083 minutes and the rating score of the clients was calculated based on the binary logistic regression model. With this qualification, customers with a high level of risk representing 20.16% were determined, therefore, it can ensure return of capital by 80.15%

**Keywords:** Data Mining, Credit Rating, Logistic Regression Model.

# CAPÍTULO I

## INTRODUCCIÓN

### 1.1. Planteamiento del problema

#### 1.1.1. Contextualización

Actualmente las organizaciones están comenzando a darse cuenta a valorar los datos que van generando día a día. En este contexto se estudia uno de los campos como es la extracción de conocimiento a partir de fuentes masivas de datos. Para ello se emplean las denominadas técnicas de minería de datos, que son algoritmos capaces de obtener relaciones entre distintos atributos o conceptos para ayudar, por ejemplo, a la toma de decisiones (García Herrero & Molina López, 2012).

La minería de datos, sea B2C o B2B, genera tras el análisis exhaustivo de datos unos modelos ya sean ocultos, que se han ido generando a través de la interacción con los clientes, pero no se han formalizado, y unos modelos predictivos basados en el análisis del comportamiento de los consumidores. Estos modelos son una de las variables fundamentales a tener en cuenta en la elaboración de estrategias; también la minería de datos ha pasado de convertirse de un diferencial a una necesidad para la creación de estrategias de mercado, al generar modelos que permitan conocer a los clientes e incluso predecir sus comportamientos (Empresas, 2015).

La presente investigación utilizará una metodología para implementar Minería de Datos y el método de Credit Scoring para elaborar un modelo que le permita medir la probabilidad de impago del crédito que se otorga (Rayo Cantón, Lara rubio, & Camino Blasco, 2010).

### 1.1.2. Descripción del problema

Los sistemas de información transaccionales, las tecnologías de la información y las comunicaciones están cambiando la forma de actuar y trabajar los procesos de negocio. Este cambio ha sido paulatino y continuo que en la actualidad las micro financieras, como parte de su estrategia política, han ido adoptando tecnologías crediticias en el proceso de evaluación y otorgamiento de crédito, según Valencia Ramírez (2009, pág. 57) se analiza varios factores como perfil de cliente, Capital Proceso de acumulación y entorno de la actividad económica, verificación in situ de la empresa, Comprobación de la información del prestatario, Riesgo de Crédito y Scoring o Sistema de Calificación de Crédito. Progreso es una Micro financiera que tiene su misión de “Contribuir a mejorar la calidad de vida de la población cajamarquina menos favorecida a través de servicios financieros y no financieros, siendo socialmente responsables” y en los últimos años los datos generados por concepto de cada transacción de otorgamiento de los diferentes tipos de préstamos a los clientes han ido incrementándose año tras años y se van almacenando en un Sistema de Gestión de Base de Datos y estos datos son utilizados para la micro financiera para automatizar tareas operativas de cada unidad de la institución y sirve para apoyar a disminuir los tiempos en su labores cotidianos de los trabajadores y tomar decisiones a las jefaturas(mandos intermedios), pero no apoyan a la gerencia ni al directorio en la toma de decisiones. Entonces se puede evidenciar

que cuenta con datos históricos, pero hasta la fecha no tiene un sistema que pueda generar información tal valiosa como es la predecir el riesgo de morosidad de sus clientes, basados en sus atributos socioeconómicos versus su comportamiento de pagos.

La micro financiera Progreso solicita una serie de requisitos para el otorgamiento de un préstamo de dinero, tales como Recibo de servicios (luz, agua) del último mes cancelado, declaración jurada de ingresos hasta un monto máximo de S/. 1,500, copia de 3 últimas Boletas de Pago o último reporte del estado de AFP, Recibos por Honorarios Profesionales, entre otros requerimientos para clientes internos y externo; también hacen consultas a las centrales de riesgo como al SBS si las personas está con una calificación 100% normal. Entonces la persona una vez que cumple con todos los requisitos anteriormente descritos, recién le aprueba su solicitud de crédito y le otorgan el dinero con el cronograma de pago que tiene que cumplir, por lo que se puede decir que ya es un cliente de la Micro Financiera Progreso. Llegado el tiempo de sus compromisos pago de los clientes, muchos de ellos no cumplen, por lo que se puede evidenciar que hay morosidad como se muestra en la tabla 1. En este ámbito se plantea la siguiente formulación del problema.

### 1.1.3. Formulación del problema

#### **Problema General:**

¿Cuál es el efecto de la implementación de minería de datos en la calificación de crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca?

## 1.2. Justificación e importancia

### 1.2.1. Justificación científica

La Micro Financiera Progreso actualmente tiene un área de crédito que se encarga de otorgar los préstamos a sus clientes que lo solicitan, para ello cuenta con un sistema información de crédito automatizado que permite el ingreso de datos concerniente al proceso de otorgamiento de préstamos y las salidas de los diferentes reportes que necesitan para toma de decisiones, pero ninguno de sus reportes se orientan fundamentalmente a dar información de cómo garantizar sus capitales de la Micro Financiera. Por ello nace las interrogantes ¿Cómo será el comportamiento del cliente a partir de que se le otorga un préstamo? y ¿cancelará su deuda?, actualmente consultan al sistema bancario, pero a pesar de eso no es suficiente debido a que hay muchos rasgos o atributos social, cultural, geográfico y económico del cliente y como se ha comportado en cuanto a los pagos y que nos permitirá predecir el comportamiento en cuanto a la deuda.

Se puede observar en la tabla 1 de los diferentes proyectos de crédito que se han prestado haciendo un Total general de S/. 3417928.44 de los cuales solo en ese año se han recuperado 1247765.08 el cual representa el 36.51% y la diferencia es 63.49% de lo cual no se sabe cuánto de ello será recuperado el siguiente año por compromiso de pago o por deuda, también se observa monto por mora pero no se detalla cuanto de ello es por préstamos realizados el año anterior o del mismo año, esto evidencia que el área de crédito de la micro financiera no cuenta con técnicas estadísticas de análisis de datos con respecto a

la morosidad de sus clientes, por lo tanto no se puede tomar decisiones con respecto a este tema.

Tabla 1. Montos por proyecto en el año 2015

Proyecto	Monto Crédito	Monto Capital Recuperado	Monto Mora	Monto Cobranza	Monto Interés
Crédito Progreso	50331.80	2489.74	45.22	45.00	867.13
Grupo Autogestionarios de Crédito	13282.00	13838.24	289.96	350.00	2543.94
Proyecto de Desarrollo Sostenible	998389.24	340851.95	6192.39	7010.00	127179.50
Proyecto de Micro Financiamiento de Vivienda Yanacocha	2355925.40	885311.67	12979.45	10220.00	285220.56
Proyecto Fondo Empleo	0.00	5273.48	348.06	55.00	612.76
Total General	3417928.44	1247765.08	19855.08	17680.00	416423.89

Fuente: Micro Financiera Progreso (2015)

Esta investigación es importante porque permitirá determinar el comportamiento a futuro de sus clientes por concepto de préstamo en relación de los datos históricos que se han generado año tras año. Por lo tanto, estamos frente a un problema que las técnicas de minería de datos persiguen para solucionar, con la finalidad de ayudar al área de crédito a analizar los hábitos de pagos de los clientes y su respectiva ponderación de calificación para poder predecir frente a una solicitud de crédito si se le puede conceder o no dicho préstamo.

### 1.2.2. Justificación técnica-práctica

La referida investigación permitirá conocer el proceso de crédito que se otorga a sus los clientes, identificar a los atributos que tiene el sistema de

información, y posteriormente proponer un modelo para poder observar el comportamiento de pago en el tiempo e identificar los diferentes atributos que los caracterizan y/o influyen en ser un buen pagador o un cliente con riesgo de adeudar, por tal razón es necesario la utilización de métodos analíticos avanzados, como es la minería de datos para la explotación de datos con la finalidad de ayudar al área de crédito a la tomar de decisiones desde el punto de vista de riesgo, convirtiéndose así en el fundamento teórico para contribuir a la mejora de la efectividad en la calificación de Crédito y con ello asegurar a que los capitales de la Micro financiera estén bien invertidos en clientes que puedan pagar su deudas.

### **1.2.3. Justificación institucional y personal**

Al diseñar un modelo de Calificación de Crédito, se contribuirá en minimizar el riesgo de morosidad por otorgamiento de préstamos a los clientes de la Micro financiera Progreso, a través de un puntaje de calificación que el modelo le asignará a cada uno de los clientes.

### **1.3. Delimitación de la investigación**

El presente proyecto de tesis, abarca el estudio de las transacciones realizadas por el concepto de otorgamiento de préstamos en el área de Créditos de la Micro financiera Progreso que se realizan en el periodo de tiempo del 2016 hasta la agosto 2017 y esos datos que se han ingresado a través del sistema de información transaccional serán analizados por una adecuada técnica de minería de datos.

## **1.4. Objetivos**

### **1.4.1. Objetivo General**

Implementar minería de datos en la calificación de crédito a los clientes de la Micro Financiera Progreso.

### **1.4.2. Objetivos Específicos**

- Analizar la validez de la información de la base de datos transaccional que involucra el proceso de otorgamiento de préstamo al cliente.
- Diseñar e implementar el Data Marts a partir de la base de datos transaccional del proceso de otorgamiento de préstamo al cliente.
- Diseñar un modelo de calificación que identifique las variables socio culturales y económicas que inciden en el cumplimiento o morosidad de préstamos de los clientes.
- Asignar puntaje de riesgo por morosidad de préstamos a los clientes en base al modelo de calificación planteado.

## CAPÍTULO II

### MARCO TEÓRICO

#### 2.1. Antecedentes teóricos de la investigación

##### 2.1.1. Antecedentes Internacionales

- Figueroa (2006), tesis de maestría “Minería de Datos Aplicada a Credit Scoring”, indica que las técnicas de clasificación frecuentemente utilizadas en credit scoring son, análisis discriminante, regresión logística, árboles de clasificación, redes neuronales, y una que considera en su tesis que no es muy frecuente denominada support vector machines. La presentación de cada una de estas técnicas es totalmente independiente, lo que permite concentrarse exclusivamente en la que sea de interés. Luego de haber descrito cada uno de estas técnicas resaltó la capacidad de reconocer los supuestos, las debilidades y fortalezas de cada una de las técnicas para credit scoring. Teniendo como objetivo Desarrollar un modelo de credit scoring que sirva como soporte del proceso de decisión si otorgar o rechazar una solicitud de crédito de consumo utilizando características propias del solicitante. Para el desarrollo del modelo de credit scoring se dispuso de 6,497 observaciones de 12 características correspondientes a individuos que han sido beneficiarios de crédito en la institución financiera en el año 2002. Los individuos considerados son ecuatorianos que han recibido un crédito en la institución involucrada que hasta el corte no tienen cuotas pendientes Después de desarrollar su modelo concluyó que la piedra angular de los modelos de credit scoring es la noción que una persona o institución tiene sobre el significado de lo que es un cliente

riesgoso. Por tal motivo es de vital importancia en el desarrollo de modelos de credit scoring definir el significado de cliente riesgoso. Otra conclusión que resaltó de vital importancia en el desarrollo de modelos de credit scoring es la selección del conjunto de entrenamiento.

- Peralta (2009) “Proyecto de minería de datos para el análisis del comportamiento de los clientes de telecomunicaciones”, determina: que la Empresa Pública Municipal de Telecomunicaciones, Agua Potable, Alcantarillado y Saneamiento de Cuenca E.T.A.P.A. no cuenta con procesos que le permitan predecir el hecho de que un cliente incumpla en sus pagos por consumo o como se conoce como “caer en mora”, por lo que se procede a la toma acciones correctivas consistentes en suspender el servicio hasta que el cliente cumpla con sus obligaciones. Por ende la autora pretende realizar estudios donde se prevea con cierto grado de precisión el incumplimiento de pago en las cuentas referentes al servicio de Telefonía Fija basado en el comportamiento de pago de los clientes y comportamiento de consumo de forma que la información que se obtenga se entregará a la Gerencia Comercial de la Empresa E.T.A.P.A. Después de desarrollar el estudio de investigación llegó a las conclusiones importantes tales como: “Con la realización del presente proyecto la Empresa cuenta ya con un plan que le permita predecir los clientes insolventes con un porcentaje de precisión del 72% en su primer entrenamiento mediante la aplicación de técnicas de Minería de Datos” y “Las técnicas de Minería de datos se pueden aplicar en cualquier área donde se quiera usar de la manera más óptima la información que se genera diariamente además aplicando la Inteligencia de Negocios se puede obtener resultados favorables para la Empresa.”

### 2.1.2. Antecedentes Nacionales

- Lara y Camino (2010) “Modelo de Credit scoring para instituciones de Microfinanzas en el marco de Basilea II Rayo”, mencionan que en el Perú, el reglamento para el requerimiento de Patrimonio Efectivo por Riesgo de Crédito de la Superintendencia de Banca y Seguros (SBS) del 2009 constituye un notable avance para la implementación de Basilea II entre las entidades sometidas a supervisión, entre ellas las Entidades de Desarrollo de la Pequeña y Micro Empresa (Edpyme). Si bien la norma establece un modelo estándar para la gestión del riesgo de crédito, también ofrece la posibilidad de que las IMFs puedan implementar modelos internos. Es decir, la nueva normativa derivada de este reglamento aporta como principal novedad la posibilidad de desarrollar Métodos Basados en Calificaciones Internas (IRB) que deberán estar homologados por la Superintendencia de Banca y Seguros. Por tanto, es necesario utilizar métodos estadísticos más precisos para la estimación de la probabilidad de insolvencia y de la severidad. Ambas estimaciones son fundamentales para calcular el requerimiento de capital, la pérdida esperada y la pérdida inesperada, aspectos claves para establecer la política de tasas de interés. En este nuevo contexto se requiere contar no solo con nuevos sistemas y metodologías estadísticas y econométricas, sino también, con profesionales con la formación adecuada para desarrollar e implantar en la institución financiera los modelos adecuados de estimación de la insolvencia y la severidad. En definitiva, Basilea II constituye un nuevo reto para las entidades de micro finanzas y para los profesionales que laboran en ellas. El objetivo de esta investigación es diseñar un modelo de credit scoring para una institución

sometida a supervisión y especializada en microcréditos, como es la Entidad de Desarrollo de la Pequeña y Micro Empresa (Edpyme) del sistema financiero del Perú. El resultado de la investigación muestra la metodología y Fases necesarias para diseñar el modelo, así como el proceso de valoración y validación para que pueda ser aplicado en el área de negocio, especialmente para establecer la política de tasas de interés con clientes. La estimación del modelo de credit scoring se realizó mediante el método de introducción por pasos, y aplicando la técnica paramétrica de regresión logística de las variables explicativas sobre la base de las fases y estadios obtenidos en el proceso de concesión de un microcrédito. De esta forma, la investigación realizada diseña un modelo de calificación estadística capaz de predecir correctamente en 78.3% de los créditos de la cartera de la Edpyme Proempresa, corroborado por un porcentaje similar en el proceso de validación del modelo. A este respecto, las medidas de valoración del modelo globalmente indican un ajuste aceptable en regresión logística. Por último, los investigadores, también muestran cómo puede utilizarse el modelo para desarrollar una gestión del riesgo de crédito.

## **2.2. Bases teóricas**

### **2.2.1. Teoría del Crédito**

Los hermanos Morales Castro (2014, pág. 18) analizan y describen que la historia del crédito inicia en la región de la Mesopotamia en la década 1792 – 1750 a.C. donde se desarrolla el comercio de la banca en toda Babilonia y utilizan como moneda la cebada o la plata, también tienen un reglamento de crédito denominado código de Hamurabi que consiste en compilación de

decisiones judiciales en litigios entre deudores y acreedores, donde se conceden ciertas facilidades a los campesinos para pagar sus deudas.

Plantean las 5 C del crédito:

La Primera C: conducta, evaluar la conducta es determinar la calidad moral y capacidad administrativa de los clientes.

Segunda C: capacidad de pago histórica, analizar la capacidad de pago histórica es evaluar la habilidad del cliente de haber generado, en el pasado, los recursos suficientes para cumplir con sus compromisos financieros a través de un análisis cuantitativo de su riesgo financiero.

Tercera C: capacidad de endeudamiento, medir la solidez de la estructura financiera de la empresa.

Cuarta C: condiciones macroeconómicas, determinar el comportamiento de la industria en su conjunto, para determinar la influencia que tiene en la capacidad y fortaleza financiera del deudor.

Quinta C: capacidad de pago proyectada, analizar la capacidad que tiene un cliente para generar efectivo suficiente en el futuro, y cumplir sus compromisos financieros.

La definición de crédito en el campo de los negocios se define según Del Valle (1998, pág. 7) es la confianza dada o tomada a cambio de dinero, bienes o servicios. También se debe tener en cuenta la definición de otorgar un crédito “Otorgar un crédito significa adelantar fondos, o bien, conceder un plazo para recibir un pago exigible” (Morales Castro & Morales Castro, 2014). Por lo tanto se puede decir que el crédito y el otorgamiento de crédito son teorías similares.

## **2.3. Base conceptuales**

### **2.3.1. Minería de datos**

La minería de datos está más orientada al descubrimiento, ya que provee perspectivas hacia los datos corporativos que no se pueden obtener mediante OLAP, al encontrar patrones y relaciones ocultas en las bases de datos grandes e inferir reglas a partir de estos patrones y relaciones, para predecir el comportamiento a futuro. Los patrones y reglas se utilizan para guiar la toma de decisiones y pronosticar el efecto de esas decisiones. Los tipos de información que se pueden obtener de la minería de datos son: asociaciones, secuencias, clasificaciones, agrupamientos y pronóstico. (Laudon & Laudon, 2012, págs. 224,255)

Consiste en la extracción no trivial de información que reside de manera implícita en los datos (Chambi Canahuire & Arias Huamani, 2014).

La minería de datos prepara, sondea y explora los datos para extraer la información oculta en ellos. (Chambi Canahuire & Arias Huamani, 2014)

### **2.3.2. Técnicas de minerías de datos.**

La minería de datos ha dado lugar a una paulatina sustitución del análisis de datos, la principal diferencia entre ambos se encuentra en que en el último se descubre información sin necesidad de formular previamente una hipótesis. La aplicación automatizada de algoritmos de minería de datos permite detectar fácilmente patrones en los datos, razón por la cual esta técnica es mucho más eficiente que el análisis dirigido a la verificación cuando se intenta explorar

datos procedentes de repositorios de gran tamaño y complejidad elevada. Dichas técnicas emergentes se encuentran en continua evolución como resultado de la colaboración entre campos de investigación tales como bases de datos, reconocimiento de patrones, inteligencia artificial, sistemas expertos, estadística, visualización, recuperación de información, y computación de altas prestaciones. (Moreno García, Miguel Quintales, García Peñalvo, & Polo Martín, 2009)

### **2.3.3. Análisis y modelamiento multidimensional**

En la actualidad, las empresas se encuentran constantemente condicionadas por factores del medio o negocio en el que estén involucradas, lo que implica una toma de decisiones que pueden ser del tipo tácticas (corto plazo) o estratégicas (mediano a largo plazo). En general, estas decisiones se basan en información histórica, proveniente de distintas fuentes de datos relacionadas directamente con el caso en estudio (ventas históricas, estudio de clientes, análisis de mercado, encuestas, etc.).

Por lo anterior, muchas empresas realizan un almacenamiento de la información de sus transacciones comerciales, ya sea para contar con un registro de su operación o en otros casos con un fin estratégico. Sin embargo, esta recopilación de datos y su análisis posterior puede tornarse un problema cuando aumenta el volumen y la variedad de éstos. Es aquí donde se aplica el uso de metodologías y de herramientas como los sistemas de información y de apoyo a la toma de decisiones.

El análisis multidimensional busca entregar al usuario final una manera fácil de representar la información, con varios componentes dimensionales o atributos, en una estructura común con la cual poder tomar decisiones (Velasquez & Pelade, 2008).

#### **2.3.4. Sistemas de información**

Un conjunto de componentes interrelacionados que recolectan (o recuperan), procesan, almacenan y distribuyen información para apoyar los procesos de toma de decisiones y de control en una organización. Además de apoyar la toma de decisiones, la coordinación y el control, los sistemas de información también pueden ayudar a los gerentes y trabajadores del conocimiento a analizar problemas, visualizar temas complejos y crear nuevos productos. (Laudon & Laudon, 2012, págs. 15-17).

Conjunto de elementos o componentes interrelacionados que recaban (entrada), manipulan (proceso), almacenan y distribuyen (salida) datos e información y proporciona una reacción correctiva (mecanismo de retroalimentación) si no se ha logrado cumplir un objetivo. El mecanismo de retroalimentación es el componente que ayuda a las organizaciones a cumplir sus objetivos, tales como incrementar sus ganancias o mejorar sus servicios al cliente (Stair & Reynolds, 2010, pág. 10).

#### **2.3.5. Aplicaciones de la minería de datos en la medición de software.**

De acuerdo a lo que mencionan Moreno García, Miguel Quintales, García Peñalvo y Polo Martín (2009) en su trabajo de investigación:

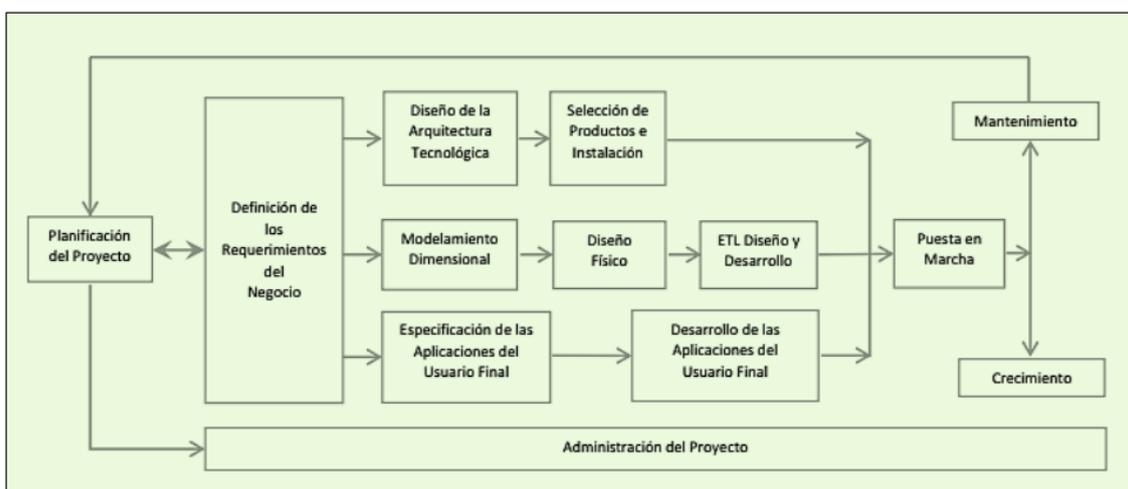
Invocando Khoshgoftaar y Allen, 1999; Porter y Selby, 1990 Tian y Palma, 1989, indica que Las técnicas de minería de datos se están utilizando desde hace varios años para la obtención de patrones en los datos y para la extracción de información valiosa en el campo de la Ingeniería del Software. Entre estas aplicaciones podemos citar la utilización de árboles de decisión en la construcción de modelos de clasificación de diferentes características del desarrollo de software. Por otro lado, refiere que Krohn y Boldyreff, 1999 expresa que la aplicación de técnicas de “Clustering” en la planificación del mantenimiento y en la estimación de la fiabilidad del software para Podgurski et al., 1999 o el uso de redes neuronales en la predicción de riesgos de mantenimiento en módulos de programa, de acuerdo a lo que sustenta Khoshgoftaar y Lanning, 1995. La mayor parte de los trabajos realizados están dirigidos a la obtención de modelos de estimación de esfuerzo de desarrollo, según lo indican Srinivasan y Fisher, 1995; mientras que Khoshgoftaar et al., 1997 expresan los modelos de predicción de diferentes aspectos de la calidad del software. En ambos casos, las métricas tanto de productos como de procesos juegan un papel importante, constituyendo la base para la construcción de los modelos y posterior validación de los mismos. En publicaciones recientes aparece la introducción de algoritmos de minería en la realización de validaciones de modelos obtenidos mediante otras técnicas. Citando a Dolado 2000, se dice que estos trabajos se comprueban la validez de modelos de estimación mediante métodos de regresión, redes neuronales, algoritmos genéticos, etc., y para Mendonca y Brasili 2000, se validan métricas, e incluso “frameworks” de medición.

## 2.3.6. Metodologías de minería de datos

### A. Metodología de Ralph Kimball

El punto central de la metodología de Kimball es el modelado dimensional que propone el enfoque bottom-up es una metodología rápida que se basa en experimentos y prototipos. La idea es construir Data Marts independientes para evaluar las ventajas del nuevo sistema a medida que avanzamos. (Group, 2015)

Figura 1. Metodología de Ralph Kimball



Fuente: Group (2015)

#### a. Planeación y administración del proyecto

Existen varios escenarios posibles en los que surge un proyecto de bodega de datos para una empresa. Es importante identificar el escenario para determinar el alcance y definición del proyecto. Los escenarios, originados por una demanda del proyecto en una empresa son los siguientes:

- ✓ Demanda de un sector del negocio: En este escenario, un ejecutivo del negocio tiene el propósito de obtener mejor información con un mejor acceso para tomar mejores decisiones.

✓ Demasiada demanda de información: En este escenario, existen múltiples ejecutivos del negocio buscando mejor información.

✓ En busca de demanda. En este escenario usualmente está involucrado el presidente de una empresa, quien no identifica necesidades de una bodega de datos para su negocio, pero desea incorporar este sistema por razones diferentes a requerimientos o necesidades del negocio.

#### b. Análisis de requerimientos

Para entender mejor los requerimientos se debe empezar por hablar con los usuarios del negocio. No se debe preguntar a estos usuarios, qué datos quieren que aparezcan en el Data Mart, sino hablar con ellos sobre sus trabajos, objetivos y retos e intentar conocer cómo toman decisiones, actualmente y en el futuro. Se debe considerar lo que requiere el negocio comparando estos requerimientos con los datos disponibles en las bases de datos que servirán como fuente, para lograr el soporte de estos requerimientos.

#### c. Diseño de la arquitectura tecnológica

Los ambientes del Data Mart requieren la integración de muchas tecnologías. El diseño de la arquitectura tecnológica establece el marco arquitectónico y la visión. Hay que considerar tres factores: los requerimientos del negocio, el ambiente técnico actual y las direcciones técnicas estratégicas planificadas.

#### d. Selección e instalación de productos

Utilizando el Diseño de la Arquitectura Tecnológica, como marco, se especifican los componentes estructurales tales como: plataforma de hardware, administrador de base de datos, herramientas de preparación.

#### e. Modelamiento dimensional

El modelo dimensional es una técnica de diseño lógico que busca presentar los datos de una forma intuitiva y que proporcione acceso de alto desempeño. Cada modelo dimensional se compone de una tabla con múltiples llaves foráneas, llamada tabla de hechos (Fact Table), y un conjunto de tablas más pequeñas, llamadas tablas de dimensión.

Existen dos modelos dimensionales que predominan en las soluciones de bodegas de datos: el modelo estrella y el modelo copo de nieve. En el modelo estrella, se tiene una tabla de hechos y en ella llaves foráneas a cada una de las tablas de dimensión que tiene el modelo. Es decir, cada tabla dimensional está directamente relacionada a la tabla de hechos.

#### f. Diseño físico

El diseño físico de la base de datos consiste en definir las estructuras físicas necesarias para soportar el diseño lógico, incluye definir los nombres estándares y establecer el ambiente de bases de datos, así como la indexación preliminar y las estrategias de partición.

#### g. Diseño y desarrollo de ETL

El proceso de preparación de datos tiene tres pasos principales: extracción, transformación y carga. El proceso de extracción casi siempre muestra problemas de calidad en los datos que han estado ocultos en los sistemas operacionales. Ya que la calidad de los datos impacta significativamente a la credibilidad del Data Mart, hay que resolver aquí este problema, tanto para la carga inicial como para la carga regular incremental posterior.

#### h. Especificaciones de la aplicación del usuario final

Se define un conjunto de aplicaciones estándares ya que no todos los usuarios requieren acceso ad hoc al Data Mart. Se describen los universos (templates), los parámetros a ser utilizados por los usuarios, los cálculos requeridos. Estas especificaciones aseguran que los usuarios y el grupo de desarrollo tengan un entendimiento común de las aplicaciones que se entregarán.

#### i. Desarrollo aplicación usuario final

De acuerdo a las especificaciones de la aplicación se desarrollan las aplicaciones de usuario final, lo que involucra configurar la metadata y construir los informes especificados.

#### j. Puesta en marcha

La puesta en marcha representa la convergencia de tecnología, datos y aplicaciones de usuario final accesibles desde el desktop del usuario. Se requiere una planificación detallada para asegurar que todas las piezas del rompecabezas se ajusten apropiadamente. Se capacita a los usuarios en todos los aspectos de la convergencia. Adicionalmente se define el soporte a los usuarios, así como las comunicaciones y estrategias de retroalimentación antes de dar acceso a los usuarios el Data Mart.

#### k. Mantenimiento y crecimiento

Se da soporte y capacitación continua a los usuarios, asegurándose que los procesos y procedimientos aseguren una operación efectiva del Data Mart.

## **B. Metodología de BILL INMON**

Es una metodología descendente top-down (hacia abajo) donde los Data Marts se crearán después de haber terminado el Data Warehouse completo de la organización. (Inmon, 1998)

Bill Inmon ve la necesidad de transferir la información de los diferentes OLTP (Sistemas Transaccionales) de las organizaciones a un lugar centralizado donde los datos puedan ser utilizados para el análisis a la Fábrica de Información Corporativa (CIF o Corporate Information Factory). Insiste además en que ha de tener las siguientes características:

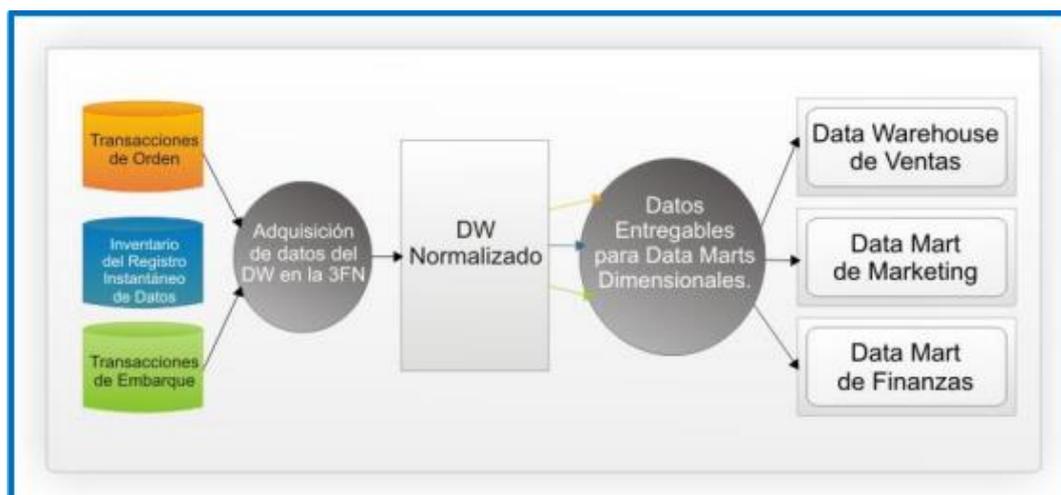
Orientado a temas: Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.

Integrado: La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.

No volátil: La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas.

Variante en el tiempo: Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones. La información debe estar a los máximos niveles de detalle. Los Data Warehouses departamentales o Data Marts son tratados como sub conjuntos de este Data Warehouse corporativo, son construidos para cubrir las necesidades individuales de análisis de cada departamento, y siempre a partir del Data Warehouse.

Figura 2. Metodología Inmon – Data Warehouse Corporativo

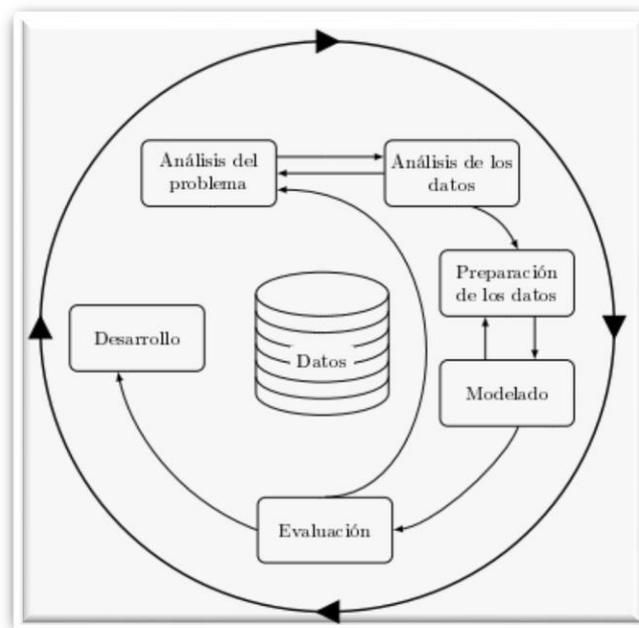


Fuente: Inmon (1998)

### C. Metodología CRISP

CRISP-DM (Cross- Industry Standard Processfor Data Mining). La metodología CRISP es una de las principales metodologías por seguir por los analistas en la inteligencia de negocios, donde se puede rescatar primordialmente Data Warehouse y Data Mining. La metodología CRISP está sustentada en estándares internacionales que reflejan la robustez de sus procesos y que facilitan la unificación de sus fases en una estructura confiable y amigable para el usuario. Además de ello, esta tecnología interrelaciona las diferentes fases del proceso entre sí, de tal manera que se consolida un proceso iterativo y recíproco. (Chapman, y otros, 2000) Propone uno de los modelos del proceso de minería de datos más avanzado es el CRISP-DM.

Figura 3. Ciclo de CRISP



Fuente: Chapman et al. (2000)

El ciclo de vida del proyecto según la metodología CRISP está basado en seis fases cambiantes entre sí y nunca terminantes, lo cual lo postula como un ciclo en constante movimiento.

#### a. Comprensión del negocio

Se trata de entender claramente los requerimientos y objetivos del proyecto siempre desde una visión de negocio. Esta fase se subdivide a su vez en las siguientes categorías:

Definición de los objetivos de negocio (inicial, objetivos de negocio y criterios de éxito del negocio).

Evaluación de la situación (inventario de recursos, requisitos supuestos y requerimientos, riesgos y contingencias, terminología y costes y beneficios).

Definición de los objetivos del Data Warehouse (objetivos y criterios de éxito).

Realización del plan del proyecto (plan del proyecto y valoración inicial de herramientas y técnicas).

b. Comprensión de los datos

Conseguir y habituarse con los datos, reconocer las dificultades en la calidad de los datos y reconocer también las fortalezas de estos mismos que pueden servir en el proceso de análisis. Sus subdivisiones son:

- ✓ Recolección inicial de datos (informe de recolección).
- ✓ Descubrimiento de los datos (informe descriptivo de los datos).
- ✓ Exploración de los datos (informe de exploración de los datos).
- ✓ Verificación de calidad de los datos (informe de calidad).

c. Preparación de los datos

Analizar los datos realmente importantes en el proceso de selección, depuración y transformación. Sus subdivisiones son:

- ✓ Selección de los datos (motivos para incluirlos o excluirlos).
- ✓ Depuración de los datos (reporte de depuración).
- ✓ Estructuración de los datos (generación de atributos y registros)
- ✓ Integración de los datos (agrupar los datos).
- ✓ Formateo de datos (informe de la calidad de datos formateados).

d. Modelado

Consiste en aplicar las técnicas de modelado o de Data Warehouse. Sus subdivisiones son:

- ✓ Selección de la técnica de modelado
- ✓ Generar el plan de pruebas

- ✓ Construcción del modelo (parámetros escogidos, modelos, descripción de los modelos).
- ✓ Evaluación del modelo (revisión de los parámetros elegidos).

#### e. Evaluación

Esta fase es muy importante y decisiva, pues corresponde a la evaluación de la escogencia de los modelos anteriores y la toma de decisión respecto a si realmente son útiles en el proceso. Sus subdivisiones son:

- ✓ Evaluar resultados (valoración de los resultados respecto al éxito del negocio, modelos aprobados).
- ✓ Proceso de revisión.
- ✓ Determinación de los pasos siguientes (listado de posibles acciones, técnica modelada).

#### f. Despliegue o divulgación

Es la fase de implementación o de divulgación de los modelos anteriormente escogidos y evaluados. Sus subdivisiones son:

- ✓ Plan de divulgación o implementación.
- ✓ Plan de monitoreo y mantenimiento.
- ✓ Presentación del informe final.
- ✓ Revisión del proyecto (documentación de la experiencia).

### **D. Elección de la Metodología de Minería de Datos**

Al hacer un análisis teórico de las tres metodologías citadas anteriormente **Ralph Kimball**, **BILL INMON** y **CRISP -DM** destaco a esta última por dos razones:

Interrelaciona las diferentes fases del proceso entre sí, de tal manera que se consolida un proceso iterativo y recíproco.

Tiene una fase de Modelado que consiste en aplicar las técnicas de modelado y tiene sus fases Selección de la técnica de Modelado, Generar el plan de pruebas, Construcción del modelo y Evaluación del modelo.

### **2.3.7. Factores BioSocio culturales**

La palabra Bio-socio-cultural refiere a la conjugación de 3 términos fundamentales e inherentes al ser humano.

**Biológico:** Para el presente análisis se delimitan dentro del rubro de marcadores biológicos las siguientes características físicas y psicológicas de un sujeto: grupo étnico, edad, sexo, peso, talla, estado general de salud, estado de nutrición, actividad física.

**Sociedad:** Es la organización cultural de los seres humanos. Es el agregado organizado de los individuos que siguen un mismo modo de vida.

La Sociedad como tal no existe porque es una abstracción, la sociedad sólo existe en lo concreto.

**Cultura:** Es el modo de vida de un pueblo. Una sociedad está compuesta de gentes, el modo en que se comporta su cultura. (Duran Lopez., 2012)

### **2.3.8. Factores económicos.**

Los factores económicos son actividades que tienden a incrementar la capacidad productiva de bienes y servicios de una economía, para satisfacer las necesidades socialmente humanas. Existen muchos factores económicos, estos están clasificados por su potencialidad en la economía de un país. (Gutiérrez Montes, 2012)

### **2.3.9. Calificación de crédito**

La calificación de crédito es una puntuación que mide la calidad crediticia de un país, empresa o persona. Se determina mediante una evaluación del historial de crédito y capacidad para pagar la deuda. Esta capacidad se mide teniendo en cuenta tanto los activos como los pasivos. (Forex, s.f.)

### **2.3.10. Calidad crediticia**

La calidad crediticia es la capacidad que posee una entidad emisora de deuda para hacer frente a sus compromisos de pago futuros, tanto en tiempo como en forma.

La calidad crediticia evalúa, por tanto, la probabilidad de incumplimiento de una determinada obligación financiera específica por parte del emisor de la deuda. Por un lado, una buena calidad crediticia nos indica que la obligación financiera en cuestión tiene poco riesgo de impago y, por otro, una muy mala calidad crediticia significa que las probabilidades de que la entidad emisora consiga pagar su deuda conforme a las condiciones pactadas son bastante bajas. (Juste, 2015)

### **2.3.11. Clasificación de la calidad crediticia**

Si atendemos al concepto estricto de calidad crediticia, lo cierto es que no podemos hablar de “tipos” o “variedades” de calidad crediticia puesto que ésta es buena o mala, mejor o peor. No obstante, sí que podemos establecer un orden de prelación de la deuda emitida por los organismos emisores, asociando el concepto de calidad crediticia al nivel de riesgo de impago en caso de liquidación.

En función del emisor de la deuda podemos distinguir:

**Deuda Pública (soberana):** Generalmente, este tipo de activos financieros poseen la mejor calidad crediticia del mercado puesto que, en escenarios de normalidad, los Estados tienen una mayor capacidad para hacer frente a sus compromisos de pago en comparación con entidades privadas. Las Letra del tesoro, los Bono del estado y las obligaciones del Estado son ejemplos de productos financieros de Deuda Pública.

**Deuda interbancaria:** Las entidades financieras se prestan continuamente dinero con un horizonte temporal de muy corto plazo y una elevada liquidez. Por ello, también suelen poseer una buena calidad crediticia. El producto típico que encontraríamos en esta categoría son los depósitos bancario.

**Deuda Privada (corporativa):** Se trata de la deuda asumida por cualquier persona, física o jurídica, que no sea una Administración Pública. Suele poseer una calidad crediticia menor. Dentro de la deuda corporativa podemos distinguir:

- ✓ Deuda Senior Secured
- ✓ Deuda senior
- ✓ Deuda subordinada
- ✓ Deuda Híbrida
- ✓ Acciones: En este caso, no estaríamos hablando propiamente de deuda, sino de inversión en capital.

**2.3.12. Riesgo crediticio**

Es la posibilidad de que una Institución Financiera sufra pérdidas no controladas por la falta de pago del deudor. Un ejemplo de riesgo crediticio constituye cuando se otorga un crédito de consumo y existe la probabilidad de que el socio en cuestión incumpla y entre en morosidad, y es precisamente esto lo que busca medir el riesgo de crédito bajo diferentes técnicas o modelos. (Córdova, 2009)

**2.3.13. Enfoque cuantitativo de la investigación.**

Enfoque cuantitativo usa la recolección de datos para probar hipótesis, con base en la medición numérica y el Análisis estadístico, para establecer patrones de comportamiento y probar teorías. (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2010, pág. 5)

**2.3.14. Estudio de alcance descriptivo.**

Investigación descriptiva busca especificar propiedades, características y rasgos importantes de cualquier fenómeno que se analice. Describe tendencias de un grupo o población. (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2010, pág. 80)

**2.3.15. Estudio de alcance correlativo.**

Este tipo de estudios tiene como finalidad conocer la relación o grado de asociación que exista entre dos o más conceptos, categorías o variables en un contexto en particular. (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2010, pág. 81)

### 2.3.16. Data Mart

Un Data Mart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un Data Mart puede ser alimentado desde los datos de un data warehouse, o integrar por sí mismo un compendio de distintas fuentes de información. (Sinergia e Inteligencia de Negocio, 2007-2016)

Un Data mart es un gran almacén de datos de una organización referente a un área, un tema o una función específica de una organización de la cual se pueden hacer consultas rápidas en un nivel más pequeño que una data warehouse, es decir, es un repositorio menos ambicioso que un DWH.

### 2.3.17. Modelo de regresión logística

Para modelos de regresión logística, los parámetros se estiman a través de los métodos de máxima Verosimilitud. Así, los coeficientes que estima el modelo hacen nuestros datos “más verosímiles”, puesto que el modelo es no lineal, se necesita un algoritmo iterativo para esta estimación. El método iterativo que se aplica es el método de Newton-Raphson. (Chambi Canahuire & Arias Huamani, 2014)

La representación matemática del modelo es la siguiente:

$$g_i = \log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + K + \beta_p x_{ip}$$

$g_i$  : Variable dependiente del modelo: “Bueno” y “Malo”

$P_i$  : Probabilidad de que el cliente sea bueno en los próximos meses

$\beta_i$  : Coeficientes del modelo (Parámetros a estimar)

$x_i$  : Variables explicativas del modelo

$$g_i = \log(ODDS) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

  $\frac{p_i}{1 - p_i}$  : ODDs Ratio, es una razón de probabilidades. Indica cuánto más probable es ser un cliente bueno que malo.

#### 2.4. Definición de términos básicos

**Aval:** Garantía de pago que otorga un tercero extraño en una letra de cambio o pagaré. El avalista generalmente responde del pago en las mismas condiciones que el librador y los endosantes en la letra y del deudor principal en el pagaré. (Micro Financiera Progreso, 2016)

**Bases de datos:** Los la de una colección de datos organizados para dar servicio a muchas aplicaciones de manera eficiente, al centralizar los datos y controlar los que son redundantes. En vez de guardar los datos en archivos separados para cada aplicación, se almacenan de modo que los usuarios creen que están en una sola ubicación. (Laudon & Laudon, 2012, pág. 212)

**Capital:** Es la suma de todos los recursos, bienes y valores movilizadas para la constitución y puesta en marcha de una empresa. Es su razón económica. Cantidad invertida en una empresa por los propietarios, socios o accionistas. (Micro Financiera Progreso, 2016)

**Compra venta:** Contrato por el cual una parte (vendedor) se obliga a entregar a la otra parte (comprador) un bien y transmitirle su dominio, y el adquiriente a su vez obliga a pagar cierto precio en dinero. (Micro Financiera Progreso, 2016)

**Contrato:** Acuerdo de voluntades entre dos o más personas que se obligan a dar y hacer algunas cosas o realizaciones conjuntas. Lo normal es que, mediante Procedimientos Legales, las partes que han asumido una obligación contractual pueden ser forzados a su cumplimiento. Desde un punto de vista económico, el contrato establece pautas reguladoras de conducta para los agentes económicos, reduciendo los riesgos de las actividades. (Micro Financiera Progreso, 2016)

**Cliente:** Persona que contrata un producto o servicio con una entidad financiera. (Asociación de Supervisores Bancarios de las Américas, s.f.)

**Crédito:** Obtención de recursos en el presente sin efectuar un pago inmediato, bajo la promesa de restituirlos en el futuro en condiciones previamente establecidas. Pueden ser Recursos Financieros o referirse a bienes y servicios. El crédito es fundamental en una economía moderna y reviste diversas formas entre las que destacan el crédito de consumo extendido a los individuos para financiar su consumo de bienes; el crédito comercial extendido por los oferentes de materias primas a las empresas o por éstas a los vendedores mayoristas; y el crédito bancario que consiste en préstamos a diversos agentes económicos. (Micro Financiera Progreso, 2016)

**Datos:** Los datos son flujos de elementos en bruto que representan los eventos que ocurren en las organizaciones o en el entorno físico antes de ordenarlos e interpretarlos en una forma que las personas puedan comprender y usar. (Laudon & Laudon, 2012, pág. 15)

**Desembolsos:** Es el monto total o parcial de un crédito que la Institución otorga al Prestatario o Deudor. (Micro Financiera Progreso, 2016)

**Endeudamiento:** Utilización de recursos de terceros obtenidos vía deuda para financiar una actividad y aumentar la capacidad operativa de la empresa. (Micro Financiera Progreso, 2016)

**Hardware de computadora:** El hardware de computadora es el equipo físico que se utiliza para las actividades de entrada, procesamiento y salida en un sistema de información. Consiste en lo siguiente: computadoras de diversos tamaños y formas (incluyendo los dispositivos móviles de bolsillo). (Laudon & Laudon, 2012, pág. 20)

**Información:** Información nos referimos a los datos que se han modelado en una forma significativa y útil para los seres humanos. (Laudon & Laudon, 2012, pág. 15)

**Micro créditos:** Son pequeños préstamos realizados a Prestatarios demasiado pobres como para que les concedan un Préstamo en un Banco tradicional. Los micro créditos posibilitan, especialmente en países en vías de desarrollo que muchas personas sin recursos puedan financiar proyectos laborales por su cuenta que les genere ingresos. (Micro Financiera Progreso, 2016)

**Planificación:** La planificación consiste en definir las metas de la organización, establecer una estrategia general para alcanzarlas y trazar planes exhaustivos para integrar y coordinar el trabajo de la organización. Se ocupa tanto de fines (qué hay que hacer) como los medios (cómo hay que hacerlo). (Robbins & Coulter, 2005, pág. 185)

**Producto:** Un producto es cualquier objeto que puede ser ofrecido a un mercado que pueda satisfacer un deseo o una necesidad. Sin embargo, es mucho más que un objeto físico ya que es un conjunto de beneficios o satisfacciones que los consumidores perciben al obtenerlo cuando lo compran, es la suma de los atributos físicos, psicológicos, simbólicos y de servicio. (Micro Financiera Progreso, 2015)

**Proceso de negocios:** Los procesos de negocios se refieren al conjunto de tareas y comportamientos relacionados en forma lógica, que las organizaciones desarrollan con el tiempo para producir resultados de negocios específicos y la forma única en que se organizan y coordinan estas actividades. (Laudon & Laudon, 2012, pág. 11)

**Software de computadora:** El software de computadora consiste en las instrucciones detalladas y pre-programadas que controlan y coordinan los componentes de hardware de computadora en un sistema de información. (Laudon & Laudon, 2012, pág. 20)

**Tecnología de información (TI):** Las TI consiste en todo el hardware y software que necesita usar una empresa para poder cumplir con sus objetivos de negocios. Esto incluye no sólo a los equipos de cómputo, los dispositivos de almacenamiento y los dispositivos móviles de bolsillo, sino también a los componentes de software, como los sistemas operativos Windows o Linux, la suite de productividad de escritorio Microsoft Office y los muchos miles de programas de computadora que se encuentran en la típica empresa de gran tamaño. (Laudon & Laudon, 2012, pág. 15)

## **CAPÍTULO III**

### **PLANTEAMIENTO DE LAS HIPÓTESIS Y VARIABLES**

#### **3.1. Hipótesis**

##### **3.1.1. Hipótesis General**

La implementación de minería de datos a través del modelo planteado mejorará la calificación de crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca.

#### **3.2. Variables**

##### **3.2.1. Variable X**

Variable independiente X: Implementación de Minería de Datos

##### **3.2.2. Variable Y**

Variable dependiente Y: Calificación de Crédito.

### 3.3. Operacionalización de los componentes de las hipótesis

Cuadro 1. Operacionalización de variables

$$Y \cong F(X)$$

<b>Variabes</b>	<b>Dimensiones</b>	<b>Indicadores</b>	<b>Ítems</b>	<b>Instrumento</b>
<b>Variable X</b> <b>Implementación de Minería de Datos:</b> La minería de datos prepara, sondea y explora los datos para extraer la información oculta en ellos (Chambi Canahuire & Arias Huamani, 2014)	Operatividad	Nivel de respuesta al usuario	Tiempo requerido para el análisis de datos.	Procedimiento / Observación
	Funcionalidad	Nivel de conformidad	Determinar la satisfacción del usuario, con respecto a la información procesada por minería de datos.	Observación/ Escala Likert
<b>Variable Y</b> <b>Calificación de Crédito:</b> La calificación de crédito es una puntuación que mide la calidad crediticia de un país, empresa o persona. (Forex, s.f.)	Calidad Crediticia	Riesgo Crediticio	N° de Clientes alto riesgo de Pago*100/N° total de Cliente	Procedimiento /Razón
	Comercio y Finanzas	Recuperación de Capital	Porcentaje de Recuperación del Saldo Total	Procedimiento / Razón

Fuente: Propia

## CAPÍTULO IV

### MARCO METODOLÓGICO

#### 4.1. Ubicación geográfica

Región : Cajamarca

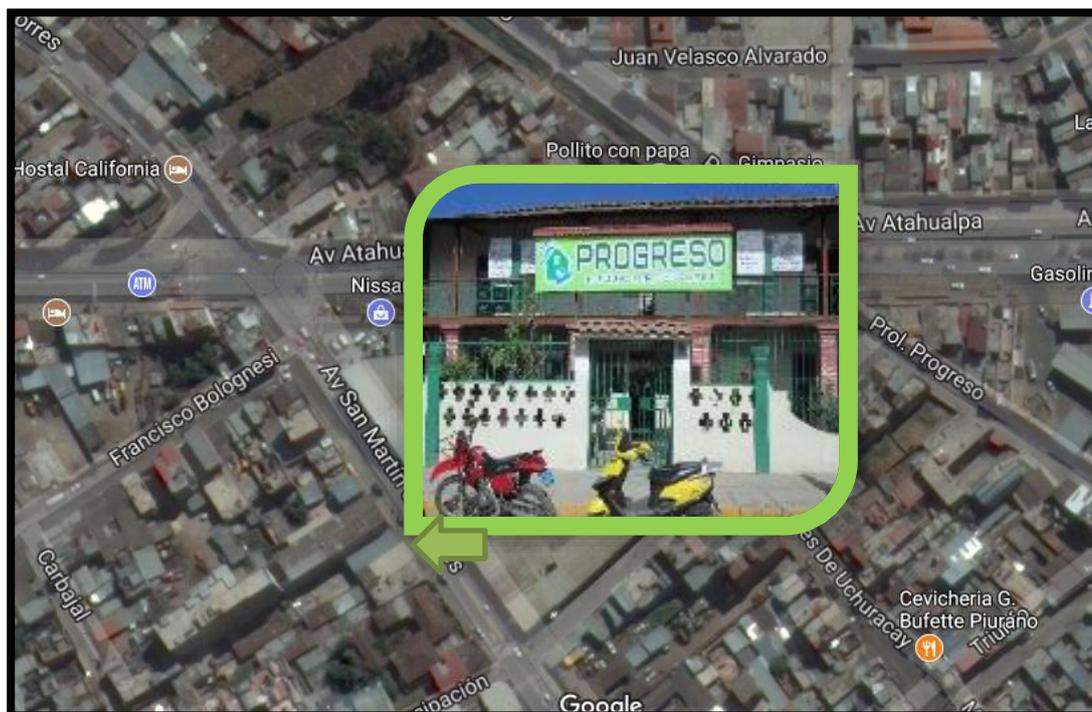
Provincia : Cajamarca

Distrito : Cajamarca

Localización : Micro Financiera Progreso

Av. San Martín de Porres N°556, del Barrio Aranjuez

Figura 4. Mapa de ubicación de la Micro Financiera Progreso



Fuente: Google Maps

## 4.2. Diseño de la investigación

### 4.2.1. Tipo de Investigación

La presente investigación es Aplicada de tipo Descriptiva – Correlacional (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2010, págs. 80,81), en razón a que se implementará una técnica de minería de datos, para construir un modelo que influyen en la calificación de riesgo por préstamos a los clientes de la Micro Financiera Progreso.

### 4.2.2. Diseño de contrastación de la hipótesis

Esta investigación, por el método de contrastación es no experimental y para medir el efecto que produce la Implementación de Minería de Datos en la Calificación de Crédito de los Clientes de la Micro financiera, se aplicarán una Pre – prueba y Post – prueba, de acuerdo al siguiente diseño:

    M    E1    X    E2    

E1: Pre – Prueba, antes de aplicar el Modelo.

E2: Post – Prueba, después de aplicar el Modelo.

M: Muestra de datos del historial de crédito de los clientes, que serán manipulados con el Modelo.

X: Implementación de Minería de Datos.

## 4.3. Método de investigación

Se aplica el *método científico*, ya que se parte de la observación directa de los hechos que están registrados en base de datos, luego se va a utilizar los métodos generales compuestos; el método deductivo permite caracterizar de forma general las teorías de la implementación de minería de datos y los

procesos administrativos relacionados a la calificación de crédito; y el método inductivo nos permite describir e identificar de forma particular las dimensiones e indicadores tanto de la variable independiente como dependiente.

El método analítico nos permite desagregar las variables para poder formular el problema, objetivos e hipótesis y el método sintético en base a los resultados de la investigación llevará a presentar las conclusiones del trabajo.

#### 4.4. Población, muestra y unidad de análisis

##### **Población de estudio:**

Para dicha investigación se tiene en cuenta 2 poblaciones de estudio:

- a. Primera población de estudio: Personal involucrado en el desarrollo y mantenimiento del sistema de crédito.

Cuadro 2. Primera población de estudio

N°	Cargo	Área	Cantidad
<b>Personal de Soporte del Área de Soporte del Sistema de Crédito</b>			
1	Administrador de la Base de datos del Sistema de Crédito.	Área de créditos y proyectos	1
2	Desarrollador del Sistema de Crédito.	Área de créditos y proyectos	1

Fuente: Propia

- b. Segunda población de estudio: está representada por 1030 clientes en la cartera activa y su respectivo historial de Crédito de cada uno de los clientes de la Micro Financiera Progreso.

### Muestra de estudio:

- a. Para la primera población, se consideró una *muestra por conveniencia*, de naturaleza censal al 100%, pues el tamaño de la población es pequeño; la disponibilidad del personal y fácil acceso a la información apoyó a obtener datos reales y exactos con respecto a la Base de Datos de los Clientes y sus respectivos registros de créditos. Posteriormente serán los usuarios que validarán de dicha investigación.
- b. En la segunda población, se obtuvo todos los datos válidos (la minería de datos procesa todos los datos) entonces para efectos de esta investigación la muestra para construir y validar el modelo se utilizará una representación de población finita de la cartera de clientes:

Dado  $N=1030$  y  $N \geq 30$  entonces  $n = \frac{N * p * q * Z^2}{e^2 * (N - 1) + p * q * Z^2}$

$N$ =es el tamaño de la población o universo.

$Z$ =es una constante que depende del nivel de confianza al 95% igual a 1.96; indica la probabilidad que los resultados de nuestra investigación sean ciertos.

$e$ =Es el error de muestreo deseado, este error muestral es la diferencia que puede haber entre un resultado que se obtenga preguntándole a una muestra de la población y que se obtendría si se procesara al total de ella; se consideró un error 7.0%.

$p$ =es la proporción de individuos que poseen en la población la característica del estudio. Este dato es generalmente desconocido y se suele suponer que  $p=q=0.5$ .

$q$ = es la proporción de individuos que no poseen esa característica, es decir es igual a  $1 - p$ .

Altos niveles de confianza y bajos márgenes de error no significan que la encuesta sea mayor confianza o esté libre de error necesariamente.

Aplicando la formula el total de encuestas serian:

$$n = \frac{1.96^2 \times 1030 \times 0.5 \times 0.5}{(0.07^2 \times (1030 - 1)) + (1.96^2 \times 0.5 \times 0.5)}$$

$$n=165$$

#### Unidad de análisis:

Área de créditos y proyectos y Soporte de Sistemas.

#### 4.5. Técnicas e instrumentos de recopilación de información

Cuadro 3. Técnicas e instrumentos de recopilación de información

Técnica	Instrumento	Fuente	Objetivo
Encuesta	Cuestionario	Trabajadores de la Micro financiera Progreso	Determinar la satisfacción del usuario, con respecto a la información procesada por minería de datos.
Observación Directa	Procedimientos	Fuentes de información (datos) del proceso de otorgamiento de crédito a los clientes de la Micro financiera Progreso.	Calcular el riesgo crediticio y recuperación de capital

Fuente: Propia

#### 4.6. Técnicas para el procesamiento y análisis de la información

La información recolectada fue procesada mediante software Servidor de Base de Datos SQL Server y el programa estadístico IBM SPSS para evaluar resultados de las técnicas de minería de datos que generan tablas y gráficos en Microsoft Excel. También se hizo uso técnicas estadísticas como la media y desviación estándar.

#### 4.7. Equipos, materiales e insumos

Está compuesto por los recursos descritos en el cuadro 4

Cuadro 4. Equipos, materiales e insumos

Recursos Humanos	
Recurso	Unidad de Medida
Jaime Amador Meza Huamán	Tesista
Héctor Diómedes Villegas Chávez	Asesor
Materiales e Insumos	
Recurso	Unidad de Medida
Lapiceros	Unidad
Papel bond A4	Millar
Cartucho Negro y Color	Unidad
Folder Manila	Unidad
Papel A3	Millar
DVD	Unidad
Hardware y Software	
Clasificación	Recurso
Hardware	Laptop
	Impresora
Software	ReportingServices
	SQL Server versión 2014
	Microsoft Excel Versión 2016
	IBM SPSS versión 22
Servicios	
Servicios	Unidad de Medida
Internet	Horas
Consumo Eléctrico	Kw/hora x 5 meses
Movilidad	Días
Anillado	Unidad
Empastado	Unidad

Fuente: Propia

#### 4.8. Matriz de consistencia metodológica

Cuadro 5. Matriz de consistencia metodológica

Efecto de la implementación de minería de datos en la calificación de Crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca.						
Formulación del Problema	Objetivos	Hipótesis	Variables	Dimensión	Indicador	Instrumento
¿Cuál es el efecto de la implementación de minería de datos en la calificación de crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca?	<u>Objetivo General</u> Implementar minería de datos en la calificación de crédito a los clientes de la Micro financiera Progreso.	La implementación de minería de datos a través del modelo planteado mejorará la calificación de crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca.	<b>Variable X</b> Implementación de Minería de Datos: La minería de datos prepara, sondea y explora los datos para extraer la información oculta en ellos (Chambi Canahuire & Arias Huamani, 2014)	Operatividad	Nivel de respuesta al usuario.	Procedimiento/ Razón
	<u>Objetivos Específicos</u> 1. Analizar la validez de la información de la base de datos transaccional que involucra el proceso de otorgamiento de préstamo al cliente. 2. Diseñar e implementar el data marts a partir de la base de datos transaccional del proceso de otorgamiento de préstamo al cliente. 3. Diseñar un modelo de calificación que identifique las variables socio culturales y económicas que inciden en el cumplimiento o morosidad de préstamos de los clientes. 4. Asignar puntaje de riesgo por morosidad de préstamos a los clientes en base al modelo de calificación planteado.			Funcionalidad	Nivel de conformidad	Observación/ Escala Likert
			Calidad Crediticia	Riesgo Crediticio	Procedimiento/ Razón	
			Comercio y Finanzas	Recuperación de Capital	Procedimiento / Razón	

Fuente: Propia

## CAPÍTULO V

### RESULTADOS Y DISCUSIÓN

#### 5.1. Presentación de resultados

Los resultados se presentan a través de cuadros, tablas y figuras en base al procedimiento de la implementación de la minería de datos siguiendo la metodología de CRISP-DM, justificada en el punto de Bases Teóricas, el cual se detalla el desarrollo paso a paso:

##### 5.1.1. Compresión del Negocio

###### 5.1.1.1. Objetivos

La Micro Financiera Progreso desea implementar minería de datos en la calificación de crédito a los clientes que con lleven a pronosticar si un cliente va ser buen pagador o mal pagador durante un periodo de tiempo, por concepto de otorgamiento de préstamo de un determinado producto y así disminuir la morosidad.

###### 5.1.1.2. Restricciones

- El proceso de otorgamiento al crédito se inicia cuando el cliente solicita adquirir producto como: Vivienda progresiva, Credicasa, Créditos microempresa, Credinegocio o crédito Progreso, cada una con sus respectivas restricciones a un determinado tipo de cliente. Posteriormente se compromete en pagar en un determinado periodo.
- Los clientes cuentan con límite de crédito impuesto por el área de crédito.
- Los datos a simular corresponderán desde el año 2016 hasta agosto del 2017.

- Si un cliente desea obtener un nuevo crédito primero se evalúa si tiene deuda pendiente en un determinado periodo, de ser el caso puede realizar un refinanciamiento de crédito.
- Se cuenta con datos de la apertura del crédito del producto que se ha otorgado a cada uno de clientes, así como los pagos detallados que van amortizando mes a mes.

#### **5.1.1.3. Tipo de resultados**

Para el modelo de puntuación en calificación de cumplimiento se desea obtener y aproximar la función que permita ajustar con la máxima exactitud las observaciones de la muestra, de manera que el error incurrido en la predicción sea mínimo. En consecuencia, se analizará las variables candidatas o explicativas para poder encontrar un patrón estándar.

#### **5.1.1.4. Requerimientos del Negocio**

Se analizó la base de datos relacional del sistema de créditos y según los Factores BioSocio Culturales y Factores Económicos, como los hábitos de pago y días de mora que van a incidir en el desarrollo del modelo de puntuación crediticia, por lo que se necesitan los siguientes requerimientos funcionales (ver cuadro 6 hasta cuadro 10).

Cuadro 6. Requerimiento R01

<b>Identificador</b>	R01	<b>Nombre</b>	Por cada cliente se debe tener las variables cualitativas Sexo, Estado Civil, Distrito, Zona de Residencia, Grado de Instrucción, Profesión; Estado de Crédito
<b>Tipo</b>	Funcional		
<b>Descripción</b>	El Data Mart nos mostrará los valores de las variables cualitativos que posteriormente serán categorizados.		

Fuente: Propia

Cuadro 7. Requerimiento R02

<b>Identificador</b>	R02	<b>Nombre</b>	Por cada cliente se debe tener las variables de tipo cuantitativas discreta y continuas Edad, Antigüedad en Días, Antigüedad en Meses, Mora de Pagos por cada Mes, Hábito de Pagos por cada Mes, Refinanciamiento Pago de algunos Clientes, Saldo Total, Monto de Interés, Promedio de Pagos, Máximo de Días de Mora en periodo de Tiempo, Veces de Mora mayor a 0 y Veces de Mora mayor a 30 por periodos de tiempo.
<b>Tipo</b>	Funcional		
<b>Descripción</b>	El Data Mart mostrará los valores de las variables cualitativos que posteriormente algunos de estas variables serán categorizados según sea el caso.		

Fuente: Propia

Cuadro 8. Requerimiento R03

<b>Identificador</b>	R03	<b>Nombre</b>	Generar un Flag de Tipo de cliente Bueno o Malo
<b>Tipo</b>	Funcional		
<b>Descripción</b>	El Data Mart mostrará el número de clientes buenos y malos, en base a los días de mora.		

Fuente: Propia

Cuadro 9. Requerimiento R04

<b>Identificador</b>	R04	<b>Nombre</b>	Generar un Flag de Muestra
<b>Tipo</b>	Funcional		
<b>Descripción</b>	El Data Mart nos permitirá saber que clientes (atributos cualitativos y cuantitativos) tiene un historial de crédito en un periodo de tiempo y si todavía está pendiente de pagar para que pueda estar incluido en la muestra de la construcción y validación del modelo de puntuación de cumplimiento.		

Fuente: Propia

Cuadro 10. Requerimiento R05

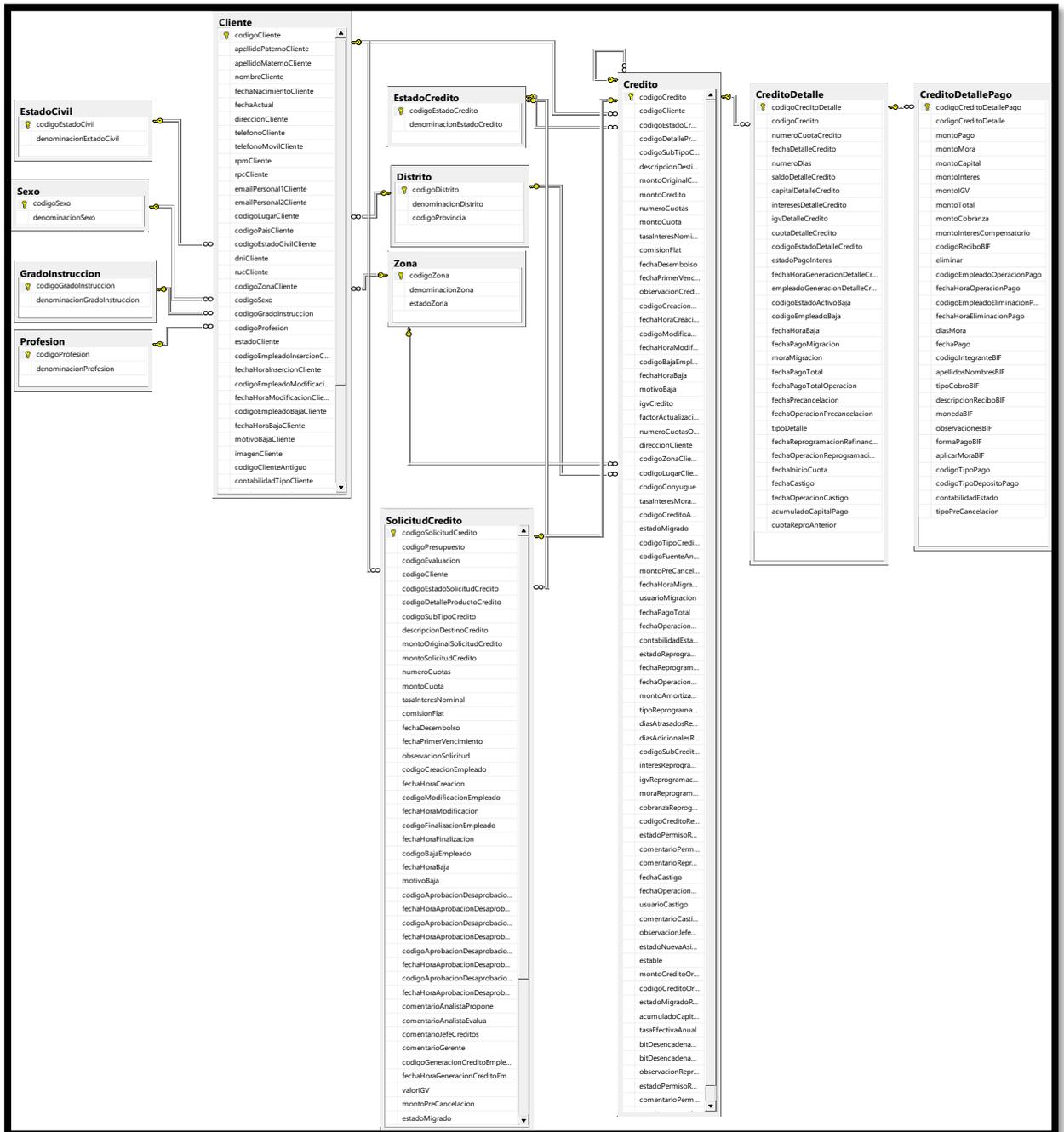
<b>Identificador</b>	R05	<b>Nombre</b>	Puntuación (Scoring Final) de Cumplimiento de Pago.
<b>Tipo</b>	Funcional		
<b>Descripción</b>	El Data Mart y la generación del modelo de puntuación, a través de la aplicación de la minería de datos nos permitirán calcular el puntaje de cumplimiento de pago a Futuro.		

Fuente: Propia

### 5.1.2. Comprensión de los Datos

Para poder comprender como está estructurado cada uno de los datos que requerimos para lograr el objetivo propuesto, se analizó las tablas de la base de datos transaccional que van a necesitar, para ello se logró realizar un modelo relacional de base de datos como se puede observar en la figura 5.

Figura 5. Modelo Relacional de la Base de Datos del Sistema de Crédito



Fuente: Propia

Se conoce la población histórica de los clientes con sus respectivos créditos y amortización de las cuotas de pagos como se detallan:

- Registro de Clientes categorizados por estado civil, sexo, grado de instrucción, profesión, distrito y zona de ubicación.
- Registros de Solicitudes de Crédito que realiza los clientes
- Una vez aprobado las solicitudes los créditos se registran los créditos, estos créditos se pueden realizar refinanciamiento por eso se considera una relación recursiva en la misma entidad.
- Cada registro de crédito puede genera uno o más registros de Crédito detalle dependiendo el número de cuotas que van a realizar.
- Por cada registro de crédito detalle se genera solo un registro de Crédito Detalle de Pago, en este caso esta entidad se registra el pago de la cuota que le toca realizar en el periodo indicado.

Cuadro 11. Diccionario de datos a nivel general de la base de datos transaccional

<b>Tabla</b>	<b>Descripción</b>
Cliente	Se registrar todas las altas de los nuevos clientes ingresantes a la cartera de clientes de la Micro Financiera Progreso.
Crédito	Se registran todos los créditos que se otorga a cada cliente.
Crédito Detalle	Se registran todos los compromiso de deuda por un determinado crédito
Crédito Detalle Pago	Se registran los pagos de cada una de las cuotas pactadas, que le pertenece únicamente a un crédito Detalle en un determinado mes.
Distrito	Se registran todas los distritos por cada provincia del Perú
Estado Crédito	Se registran todos los estados de crédito que maneja la Micro Financiera Progreso.
Grado Instrucción	Se registran todos los grados de instrucción de sus clientes.
Profesión	Se registran las profesiones que tiene cada uno de los clientes.
Sexo	Se registra el género de cada cliente.
Solicitud Crédito	Se registran las solicitudes de crédito que realiza cada cliente.

Fuente: Propia

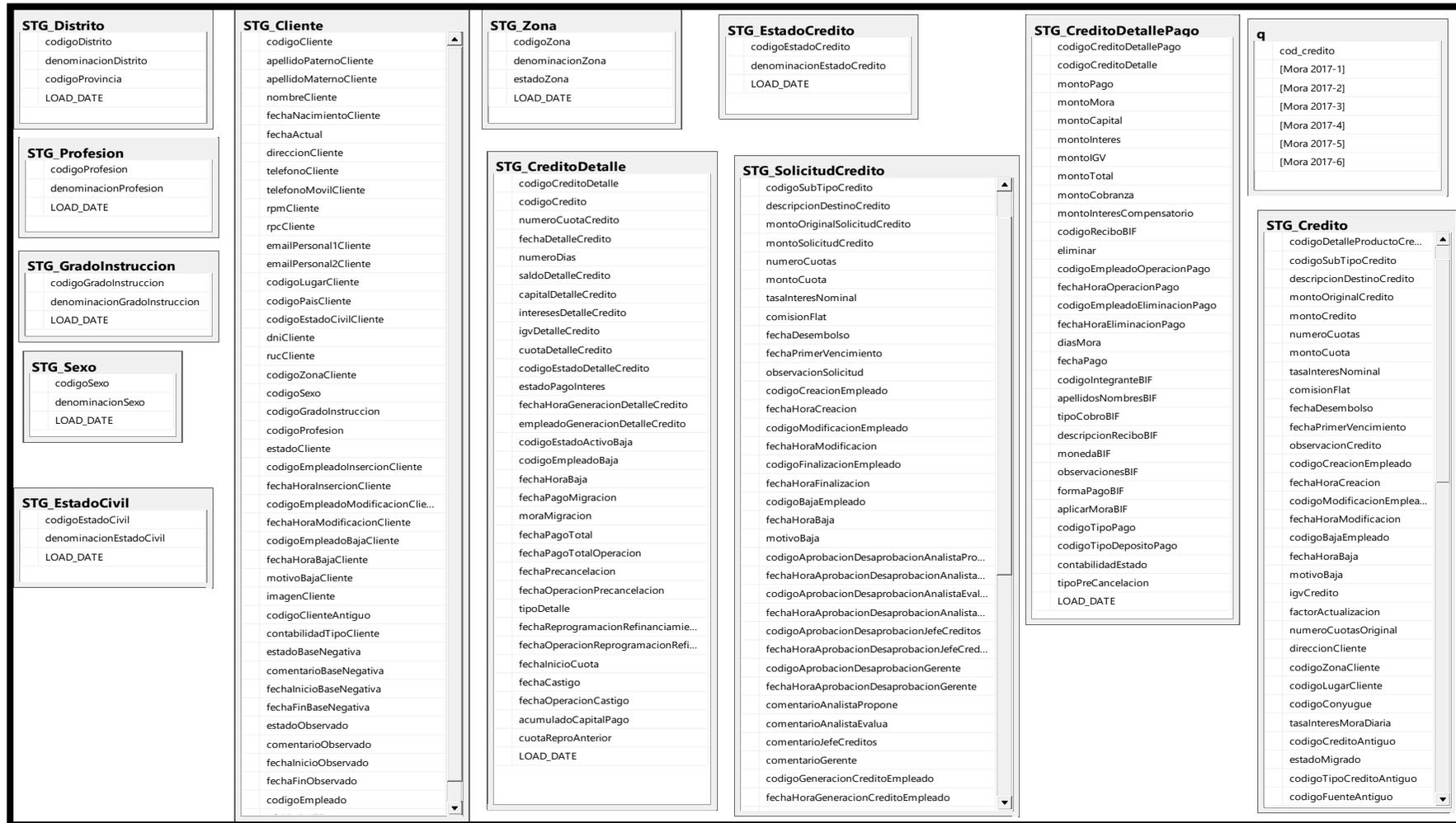
### 5.1.3. Preparación de los Datos

#### 5.1.3.1. Revisión de la información

La información de clientes (registros en tablas en la base de datos relacional) es la materia prima para llevar a cabo la elaboración del modelo y por lo tanto debe ser válida cada uno de los datos que se guardan.

- En primer lugar, se realizó una limpieza de la data para obtener resultados confiables, esto se llevó a cabo en coordinación con el responsable de la Administración de la Base de Datos de la Micro Financiera Progreso. En este caso se ha generado una base de datos temporal (Staging Area) con data sin errores es decir datos de calidad que se representa en la figura 6. Esta base de datos se crea a partir de sentencias sql como se detallan en el apéndice A

Figura 6. Esquema de las tablas de la base de datos temporal



Fuente: Propia

### 5.1.3.2. Diseño de la Ventana de Tiempo

#### a. Definición de clientes

Según los días de morosidad que el cliente registre, la cartera de clientes es segmentada entre Malos y Buenos, Según las Política de Provisiones para cubrir deudas incobrables, establecido en el manual de políticas para el otorgamiento y recuperación de créditos, un cliente es categorizado como deficiente si tiene al menos tiene un pago atrasado más de 30 días, por lo tanto, he tomado el siguiente criterio de clasificación:

Días de mora  $\leq 30$  días, entonces el cliente es **Bueno**.

Días de mora  $> 30$  días, entonces el cliente es **Malo**.

#### b. Diseño de la Ventana de Tiempo

La ventana de tiempo es el periodo de predicción del modelo, donde se define a los clientes malos y buenos (Chambi Canahuire & Arias Huamani, 2014). El periodo mínimo de antigüedad que se ha tenido en cuenta es 12 meses de actividad crediticia para cada uno de los clientes. Según la información con que se cuente para tal objetivo del modelo es 6 meses a partir del 01/01/2017 como se muestra en la figura 7 tal fecha es el punto de corte en el tiempo, lo más próximo posible a la fecha actual.

Figura 7. Diseño de la ventana de tiempo



Fuente: Propia

### 5.1.3.3. Modelado dimensional

En base al análisis de la data temporal se identifica las medidas y dimensiones orientadas a analizar los datos relacionados en sus diferentes niveles.

#### a. Elección de las dimensiones

Se identificó las variables de análisis por las cuales el modelo de calificación necesitará para procesar y encontrar los patrones respectivos.

Resalto las siguientes dimensiones:

- Clientes
- Créditos
- Sexo
- Zonas
- Distritos
- Profesión
- Grado de Instrucción

Se agrupan las variables en una sola entidad, comúnmente cada una es atributo de alguna entidad o tabla importante que podría ser una dimensión.

#### b. Dimensiones encontradas

Después de un análisis según a los requerimientos planteados se determina que las dimensiones que conforman el Data Mart son:

- OSD\_DIM\_CREDITO
- OSD\_DIM\_CLIENTE
- OSD\_DIM\_SOLICITUD

#### c. Medidas encontradas

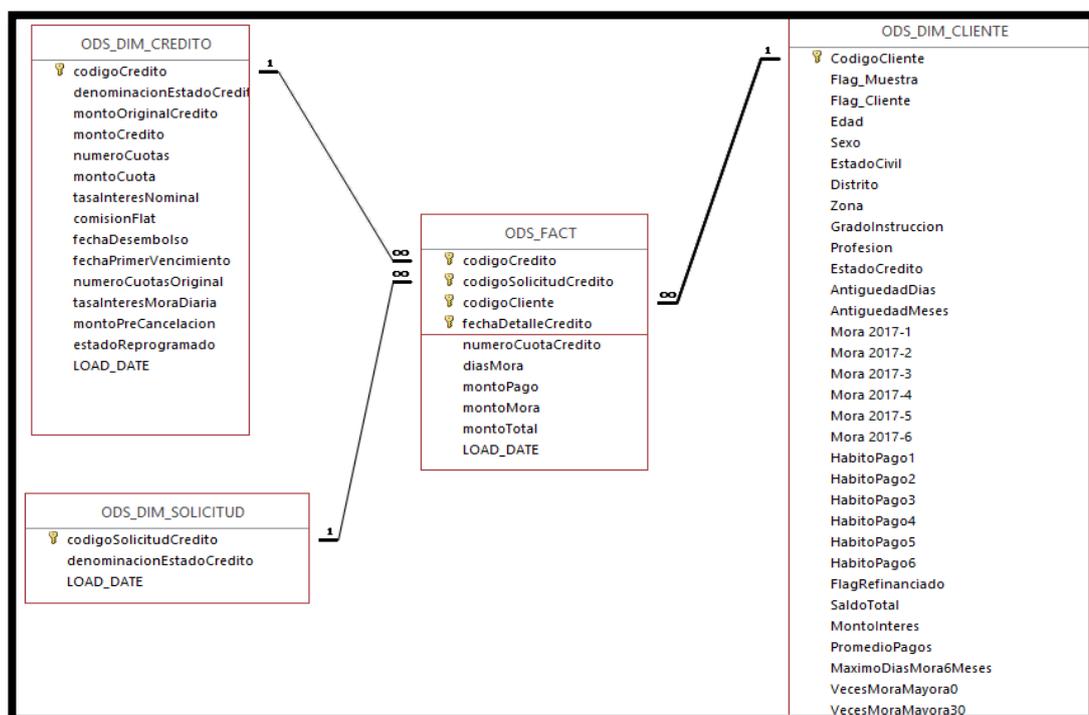
Las medidas encontradas son:

- Número de días de mora de una cuota a pagar en un mes.
- Número de veces que el cliente demoró en pagar en determinados meses.
- Antigüedad de días y meses que es cliente.
- Saldo total.
- Monto de interés.

#### d. Diseño del modelo dimensional

Está representado por un diseño físico del Data Mart como se ve en la figura 8.

Figura 8. Diseño físico del Data Mart

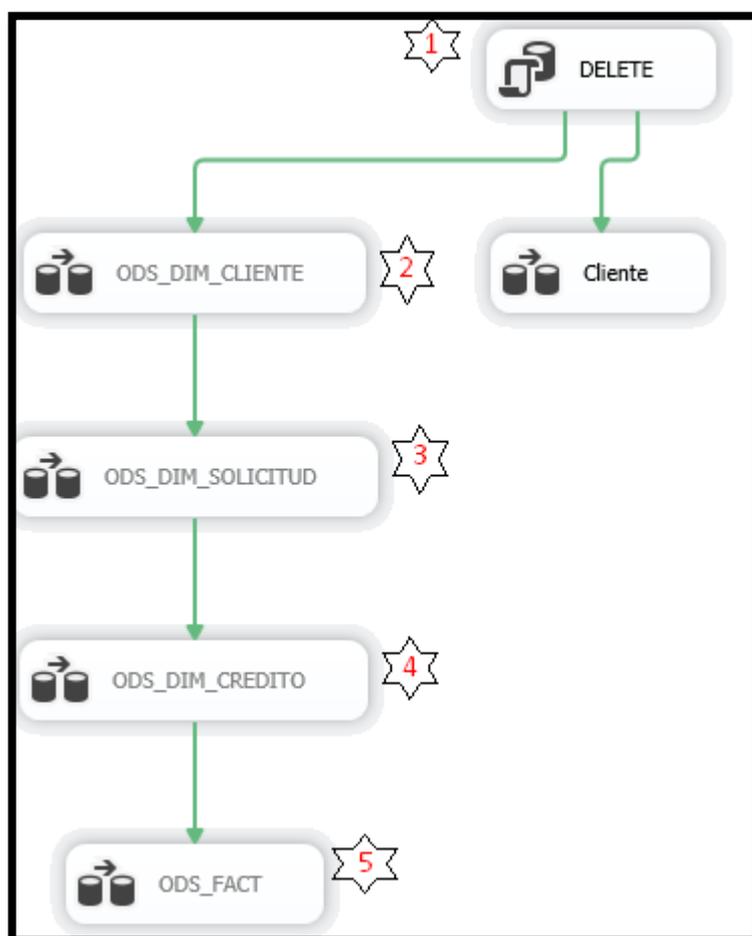


Fuente: Propia

#### 5.1.3.4. Diseño e implementación del subsistema de ETL

Para el desarrollo del proceso ETL de las dimensiones y fact table, se utilizó la herramienta INTEGRATION SERVICES que nos brinda SQL SERVER en su versión 2014 y su flujo de implementación se basa en un orden mostrado en la figura 9.

Figura 9. Diseño ETL

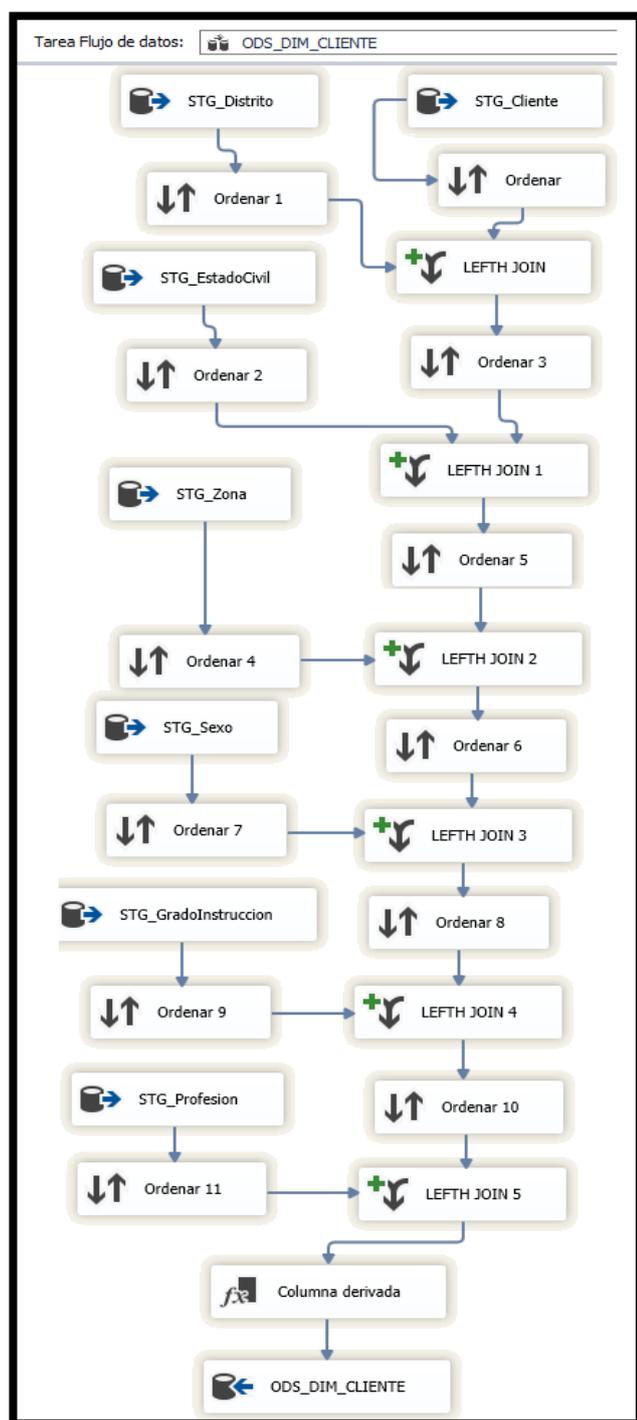


Fuente: Propia

#### a. Flujo de datos DIM\_Cliente

Para obtener la dimensión del cliente se realizó un flujo de datos, ver figura N° 10, este se inicia con las lecturas de las tablas origen ya validadas de Distrito y Clientes, luego ambas tablas se ordenan y se unen para poder añadir el atributo denominación distrito en la tabla Clientes. Estos mismo paso se repiten para poder llenar las denominaciones de Estado civil, Zona, Sexo, Grado de Instrucción y la Fecha Actual (columna derivada) en la tabla de tipo de dimensión de Cliente.

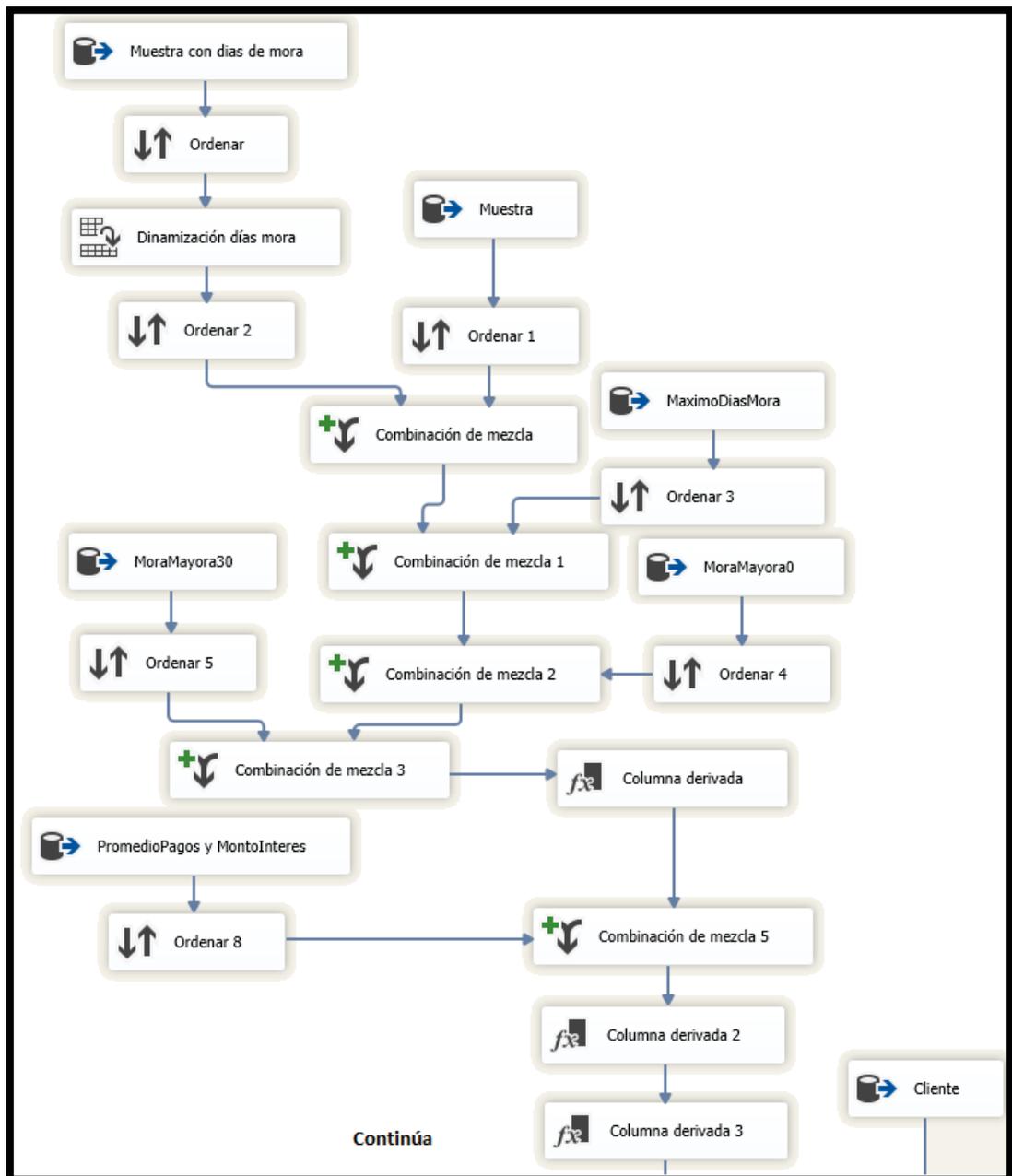
Figura 10. Flujo de datos DIM\_Cliente con atributos categóricas



Fuente: Propia

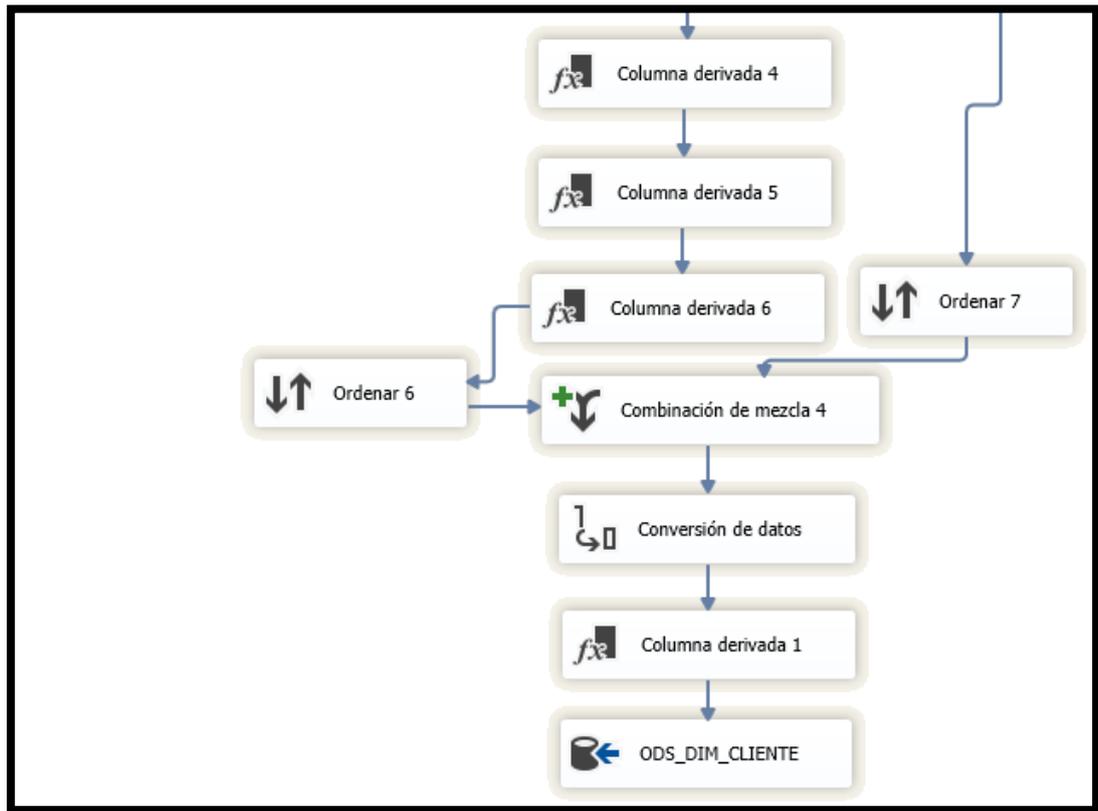
También se calcula por cada cliente los días de morosidad, número de veces que se demora más de cero días y 30 días, promedio de pagos y montos de interés, y los hábitos de pago como se puede apreciar en las figuras 11 y 12.

Figura 11. Flujo de datos DIM\_Cliente con atributos escalares.



Fuente: Propia

Figura 12. Flujo de datos DIM\_Cliente con atributos escalares - continuación

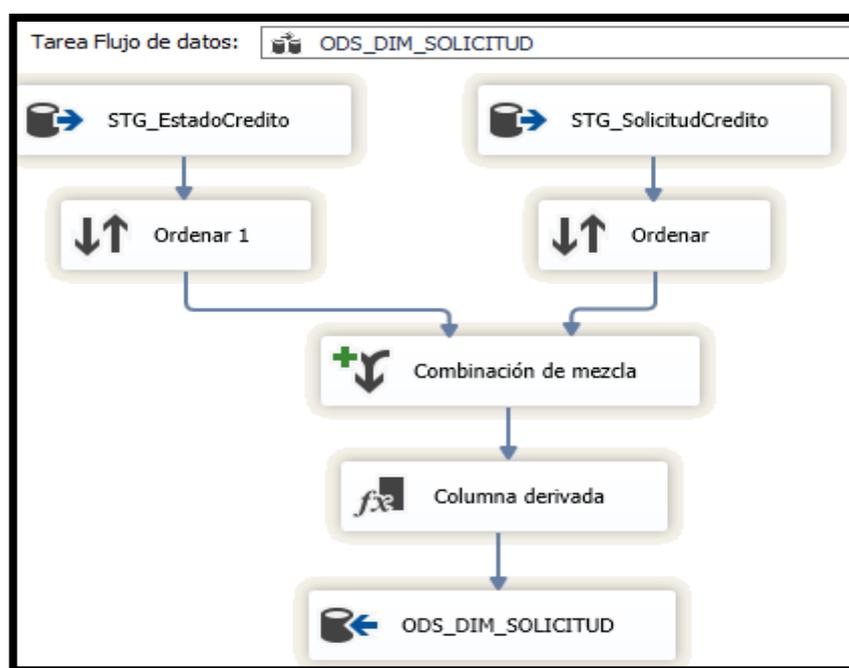


Fuente: Propia

### b. Flujo de datos DIM\_Solicitud

El flujo de datos para poblar a la dimensión solicitud, inicia con la lectura de las tablas orígenes y validadas de Solicitud de Crédito y Estado de Crédito, tal como se aprecia en la figura 13, ambas tablas se ordenan por el código de estado de solicitud de crédito, posteriormente las referidas tablas se combinan en una sola tabla, se agrega la fecha actual y finalmente se puebla la dimensión solicitud.

Figura 13. Flujo de datos DIM\_Solicitud

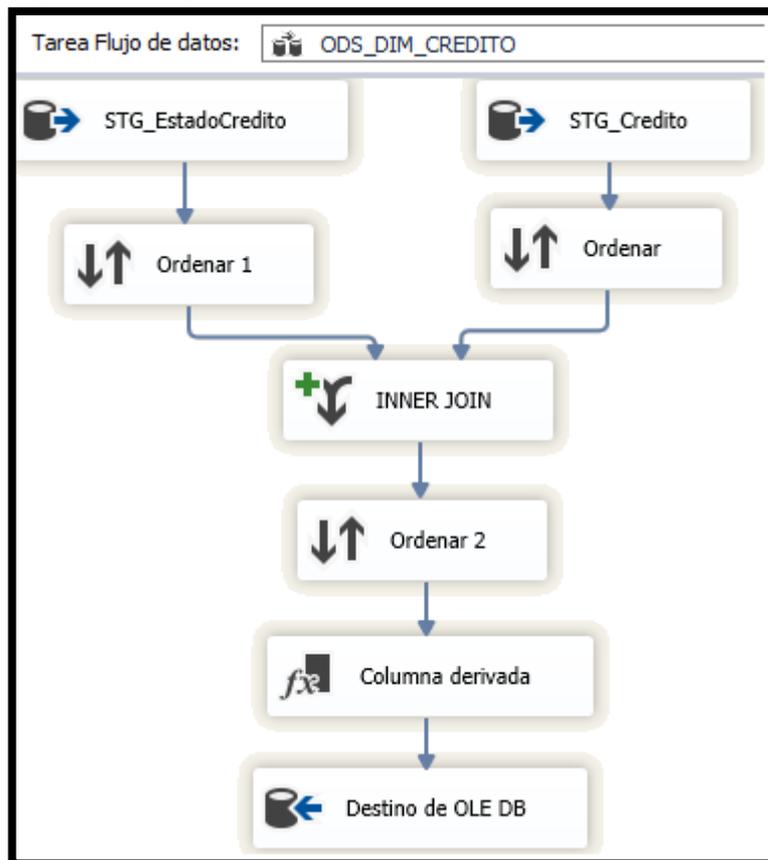


Fuente: Propia

### c. Flujo de datos DIM\_Credito

Este flujo comienza con la lectura de las tablas orígenes y validadas de Crédito y Estado de Crédito, ver figura 14. Ambas tablas se ordenan por el código de estado de crédito, luego dichas tablas se combinan en la tabla Crédito y se ordena por el código de Crédito, se agrega la fecha actual y finalmente se puebla la dimensión crédito.

Figura 14. Flujo de datos DIM\_Credito

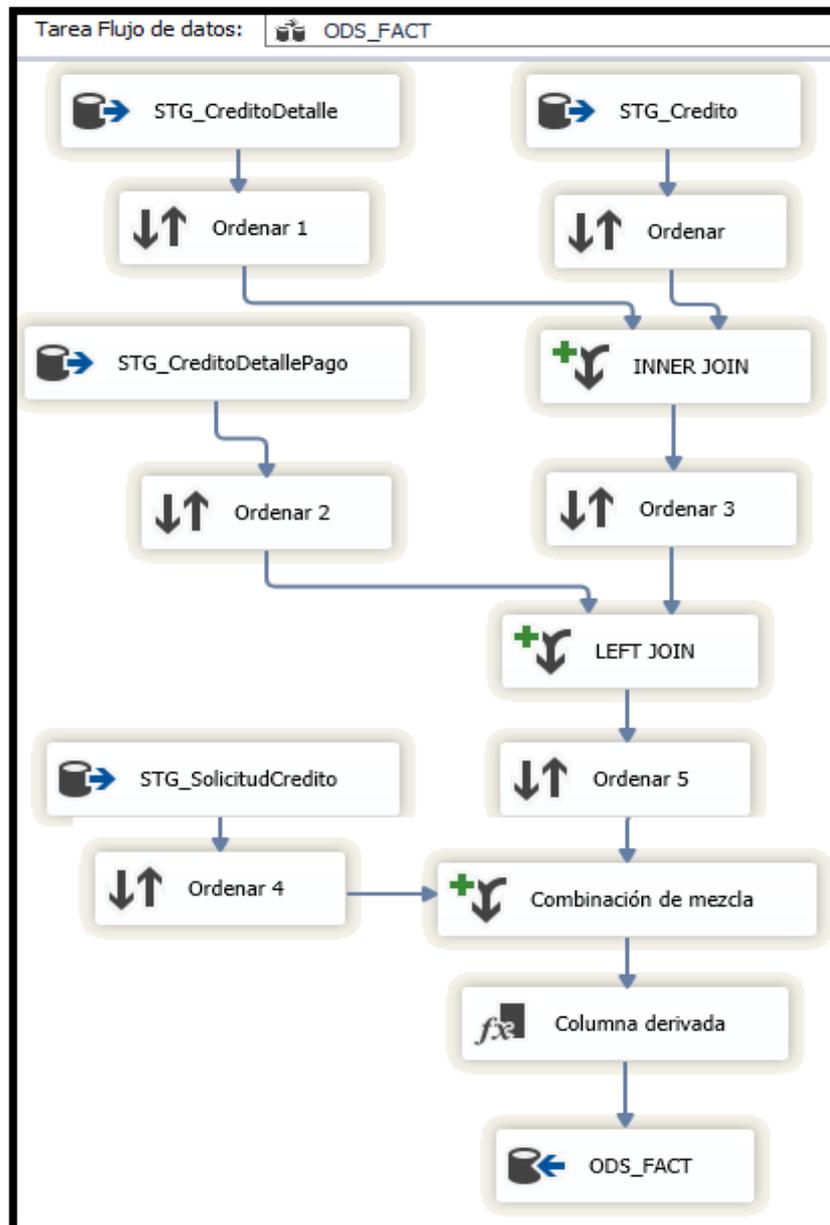


Fuente: Propia

#### d. Flujo de datos DIM\_FACT

Esta dimensión tiene como objetivo obtener los datos código de crédito, código de solicitud de crédito, código cliente, fecha de detalle de crédito, número de cuota de crédito, días de mora, monto de pago, monto de mora, monto total y la fecha actual, tal como se observa en la figura 15.

Figura 15. Flujo de datos DIM\_FACT



Fuente: Propia

### 5.1.3.5. Muestreo

El muestreo para la construcción del modelo de minería de datos con la que se va a trabajar está contemplado por la definición de la ventana de tiempo y del tipo de cliente (bueno y/o malo) y está representado por 253 clientes de los cuales el 81% son considerados como clientes buenos y el 19% como clientes malos, el cual se observa en la Tabla 2.

Tabla 2. Muestra de la cartera de clientes buenos y malos

	Frecuencia	Porcentaje %	Porcentaje válido	Porcentaje acumulado
Buenos	205	81.0	81.0	81.0
Malos	48	19.0	19.0	100.0
Total	253	100.0	100.0	

Fuente: Propia

Posteriormente se procedió a dividir la muestra de la Tabla 2 en dos submuestras, una para la construcción del modelo (ver tabla 3) que 227 clientes representan el 90% del total de clientes y la otra muestra servirá para la validación del modelo (ver tabla 4) que 26 clientes representa el 10% de los 253 clientes. Teniendo en cuenta que la proporción entre buenos y malos es la misma en ambas muestras; y para la elección de los clientes en dichas muestras se eligió de forma aleatoria.

Tabla 3. Muestra de la cartera de clientes seleccionados para la construcción del modelo

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Buenos	184	81%	81.0	81.0
Malos	43	19%	19.0	100.0
Total	227	100.0	100.0	

Fuente: Propia

Tabla 4. Muestra de la cartera de clientes seleccionados para la validación del modelo

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Buenos	21	81%	81.0	81.0
Malos	5	19%	19.0	100.0
Total	26	100.0	100.0	

Fuente: Propia

## 5.1.4. Modelo

### 5.1.4.1. Análisis Univariante

#### a. Selección de variable candidatas

Para identificar las variables candidatas para el modelo, se procedió en primer lugar a tener conocimiento de la información que nos proporciona el Data Mart., que son un total 29 de variables potenciales que a la vez son variables explicativas del modelo, ver cuadro 12.

Cuadro 12. Variables explicativas del modelo a incluir

Item	Nombre de la variable	Tipo de la variable
1	Edad	Escalar
2	Sexo	Nominal
3	Estado Civil	Nominal
4	Distrito	Nominal
5	Zona	Nominal
6	Grado Instrucción	Nominal
7	Profesión	Nominal
8	Estado de Crédito	Nominal
9	Antigüedad de la Cuenta Días	Escalar
10	Antigüedad de la Cuenta Meses	Escalar
11	Días de Mora 2017-1	Escalar
12	Días de Mora 2017-2	Escalar
13	Días de Mora 2017-3	Escalar
14	Días de Mora 2017-4	Escalar
15	Días de Mora 2017-5	Escalar
16	Días de Mora 2017-6	Escalar
17	Hábito de Pago 1(Enero)	Ordinal
18	Hábito de Pago 2(Febrero)	Ordinal
19	Hábito de Pago 3(Marzo)	Ordinal
20	Hábito de Pago 4(Abril)	Ordinal
21	Hábito de Pago 5(Mayo)	Ordinal
22	Hábito de Pago 6(Junio)	Ordinal
23	Refinanciamiento	Nominal
24	Saldo Total	Escalar
25	Promedio de Pagos	Escalar

26	Monto de Interés	Escalar
27	Máximo Días de Mora de los 6 meses	Ordinal
28	Número de Veces de Mora mayor a 0	Escalar
29	Número de Veces de Mora mayor a 30	Escalar

Fuente: Propia

Las variables explicativas que entran al modelo algunas fueron transformadas en categóricas (Rangos) y otras se quedaron como escalar, como se puede apreciar en el cuadro 13.

Cuadro 13. Variables seleccionadas para el modelo

<b>Variables</b>	<b>Tipo</b>
Edad	Categorías
Sexo	Categorías
Estado Civil	Categorías
Distrito	Categorías
Zona	Categorías
Grado Instrucción	Categorías
Profesión	Categorías
Estado de Crédito	Categorías
MoraMax_6meses	Categorías
Peor Hábito	Categorías
Hábito Promedio	Categorías
Refinanciación	Categorías
SaldoNormal	Categorías
MontoInteres	Categorías
VecesMoraMayor_0	Escalar
VecesMoraMayor_30	Escalar
Hábito de Pago 1(Enero)	Escalar
Hábito de Pago 2(Febrero)	Escalar
Hábito de Pago 3(Marzo)	Escalar
Hábito de Pago 4(Abril)	Escalar
Hábito de Pago 5(Mayo)	Escalar
Hábito de Pago 6(Junio)	Escalar

Fuente: Propia

## **b. Análisis exploratorio de datos**

El objetivo del análisis exploratorio es conocer de forma a priori la relación que existe entre las variables explicativas o independientes y la variable de respuesta o dependiente (posibles valores: Malo o Bueno).

Teniendo una serie de variables explicativas, donde varias de ellas son convertidas a categorizadas para ser considerado en el modelo. Por lo tanto, para lograr el objetivo de categorizarlo el primer paso que se realizó es el uso de las medidas descriptivas de posición cuantiles de tipo deciles, con el propósito de observar un adecuado número de rangos y el comportamiento de cada rango con respecto a la variable dependiente. El segundo paso se obtuvo el porcentaje de asociación entre la variable independiente con la dependiente. A continuación, se detalla la conversión a variables categorías:

### **➤ Categorización de la variable Edad**

Paso 1: Agrupar la variable edad en deciles (ver tabla 5)

Paso 2: Realizar el cruce de variable edad agrupado deciles con el objetivo de la variable tipo de cliente (ver tabla 5)

Tabla 5. Contingencia de la variable Edad (agrupado) por tipo de cliente

		Tipo de Cliente		Total	Tasa de Malos	Crear Categoría
		Bueno	Malo			
Edad (agrupado)	<= 29	30	3	33	9%	25%
	30 – 32	8	5	13	38%	
	33 – 38	16	10	26	38%	
	39 – 42	16	5	21	24%	22%
	43 – 45	22	6	28	21%	
	46 – 49	16	5	21	24%	
	50 – 52	15	2	17	12%	18%
	53 – 57	28	1	29	3%	
	58 – 63	16	1	17	6%	
	64+	17	5	22	23%	
Total		184	43	227		

Fuente: Propia

**Paso 3:** Se calcula y analiza la asociación de la dependencia o no entre las variables tipo de cliente y edad en base al valor de medidas direccionales (ver tabla 6)

Tabla 6. Medidas direccionales de la variable Edad

			Valor	Error típ. asint. <sup>a</sup>	T aproximadab	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	-.075	.047	-1.584	.113
		Edad (agrupado) Independiente	-.147	.091	-1.584	.113
		Flag_Cliente dependiente	<b>-.050</b>	.032	-1.584	.113
a. Asumiendo la hipótesis alternativa.						
b. Empleando el error típico asintótico basado en la hipótesis nula.						

Fuente: Propia

Interpretación: “Considerando a la variable Edad como independiente y al Tipo de Cliente como dependiente se obtiene  $d = -5.04\%$  (Tabla 6) lo cual indica no hay una asociación entre ambas variables”.

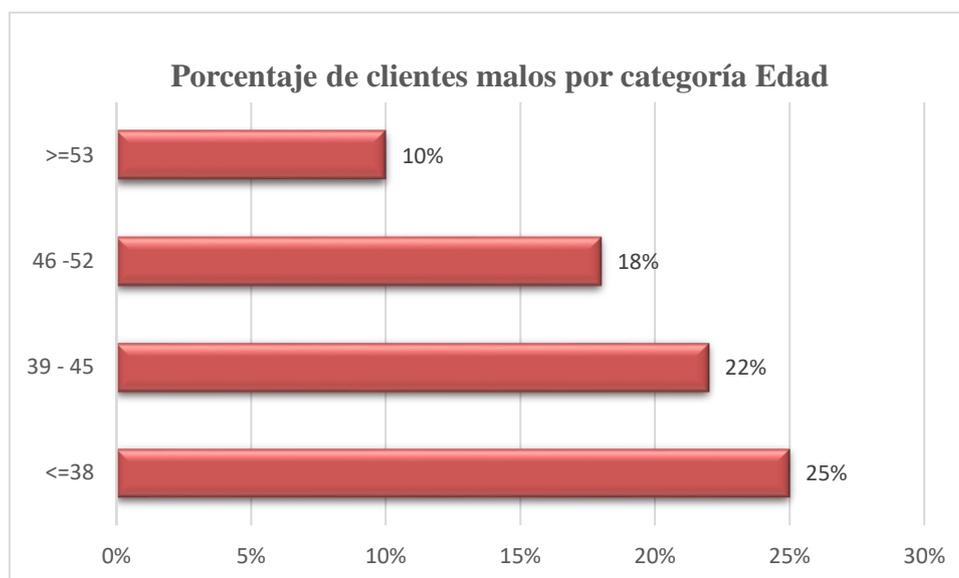
Paso 4: Se crea una nueva variable Edad\_Cat(categorización de la variable Edad) y se genera sus respectivos etiquetas y pesos, observar cuadro 14.

Cuadro 14. Categorización de la variable Edad

Nivel de Riesgo	Edad_Cat	Porcentaje de Clientes Malos	Peso de Categoría
+ Riesgo	<=38	25%	3
	39 - 45	22%	2
	46 -52	18%	1
-Riesgo	>=53	10%	0

Fuente: Propia

Figura 16. Porcentaje de clientes malos por categoría Edad



Fuente: Propia

De la figura 16, se observa que a medida que la edad de los clientes avanza se tiene menos riesgos de pago.

Paso 5: Se calcula y analiza la asociación de la dependencia o no entre las variables tipo de cliente y la Edad\_Cat en base al valor de medidas direccionales (ver tabla 7)

Tabla 7. Medidas direccionales de la variable Edad\_Cat

			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	.125	.052	2.349	.019
		Edad_Cat Independiente	.212	.088	2.349	.019
		TipoCliente dependiente	<b>.089</b>	.038	2.349	.019
a. Asumiendo la hipótesis alternativa.						
b. Empleando el error típico asintótico basado en la hipótesis nula.						

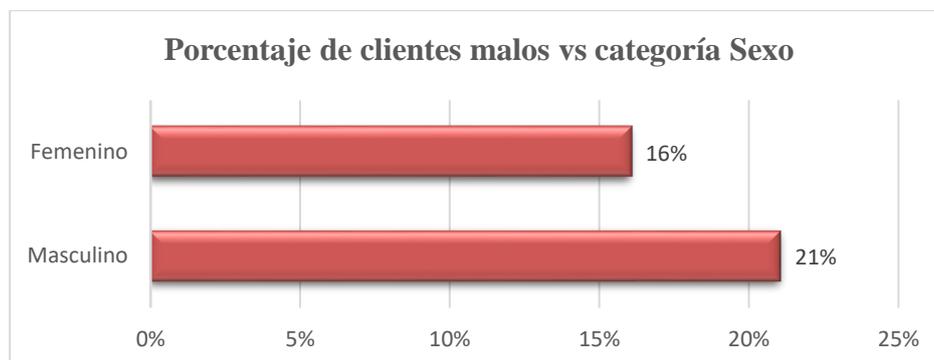
Fuente: Propia

Interpretación: “Considerando a la variable Categoría Edad como independiente y a Tipo de Cliente como dependiente se obtiene  $d=8.9\%$ . Lo cual indica una asociación positiva (aunque débil) entre ambas variables. En otras palabras, a mayor edad entonces crece la tendencia del cliente a ser bueno”.

Para la categorización de las demás variables sólo se mostrará la tabla de Medidas direccionales, el gráfico de porcentaje de clientes malos por categoría respectiva y la interpretación, las demás tablas se encuentran en apéndice B.

➤ **Categorización de la variable Sexo**

Figura 17. Porcentaje de clientes malos vs categoría Sexo



Fuente: Propia

Tabla 8. Medidas direccionales de la variable Sexo\_Cat

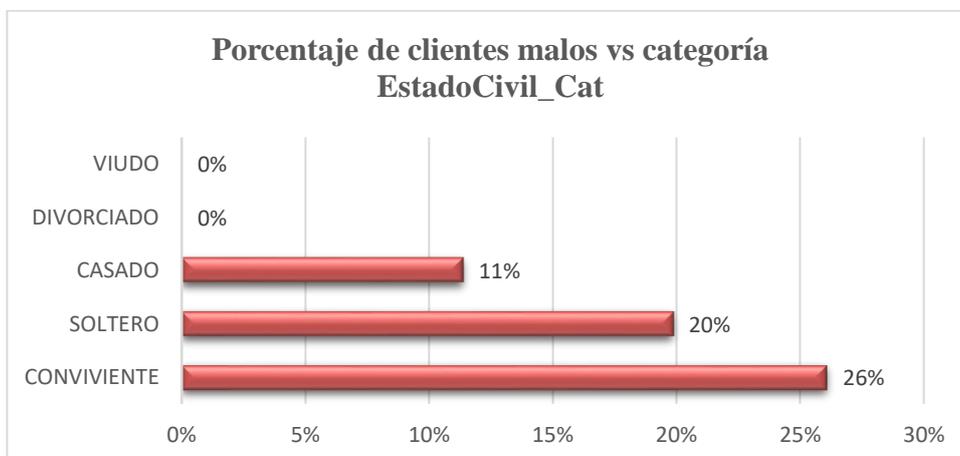
			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	.060	.063	.959	.337
		Sexo_Cat Independiente	.074	.077	.959	.337
		Flag_Cliente dependiente	<b>.051</b>	.053	.959	.337
a. Asumiendo la hipótesis alternativa.						
b. Empleando el error típico asintótico basado en la hipótesis nula.						

Fuente: Propia

Interpretación: “Considerando a la variable Sexo\_Cat como independiente y a tipo de cliente como dependiente se obtiene  $d=5.1\%$ . Lo cual indica una asociación positiva (aunque débil) entre ambas variables”. En otras palabras, las mujeres tienden a ser clientes buenos.

➤ **Categorización de la variable Estado civil**

Figura 18. Porcentaje de clientes malos vs categoría EstadoCivil\_Cat



Fuente: Propia

Tabla 9. Medidas direccionales de la variable EstadoCivil\_Cat

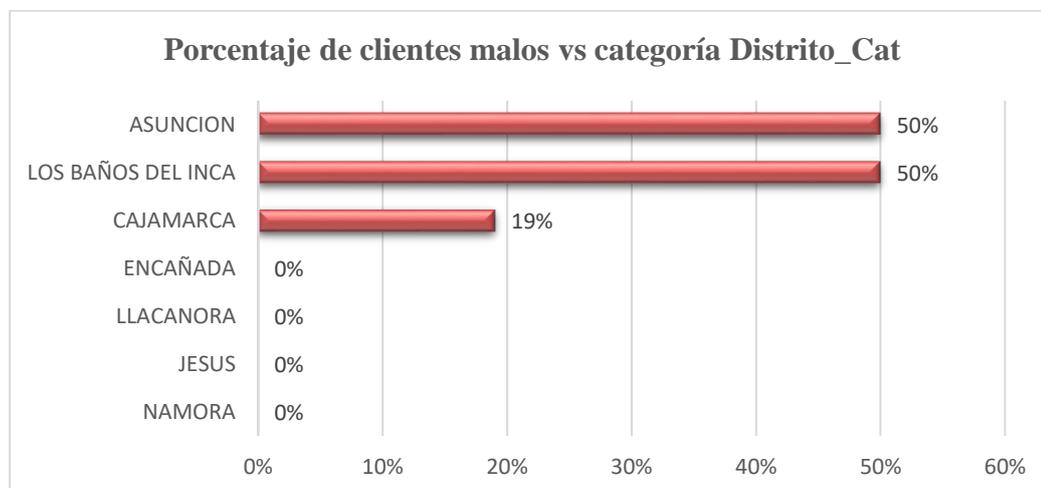
			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	.123	.056	2.164	.030
		EstadoCivil_Cat Independiente	.180	.081	2.164	.030
		TipoCliente dependiente	<b>.094</b>	.043	2.164	.030
a. Asumiendo la hipótesis alternativa.						
b. Empleando el error típico asintótico basado en la hipótesis nula.						

Fuente: Propia

Interpretación: “Considerando a la variable EstadoCivil\_Cat como independiente y la variable Tipo de Cliente como dependiente se obtiene  $d = 9.4\%$ . Lo cual indica una asociación positiva (aunque débil) entre ambas variables”. Los Casados, Divorciado y Viudo tienden hacer buenos clientes.

➤ **Categorización de la variable Distrito de Residencia**

Figura 19. Porcentaje de clientes malos vs categoría Distrito\_Cat.



Fuente: Propia

Tabla 10. Medidas direccionales de la variable Distrito\_Cat

Medidas direccionales de la variable Distrito_Cat						
			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	.149	.049	2.552	.011
		Distrito_Cat Independiente	.109	.041	2.552	.011
		TipoCliente dependiente	<b>.234</b>	.075	2.552	.011

Fuente: Propia

Interpretación: “Considerando a la variable Distrito\_Cat de Residencia como independiente y la variable Tipo de Cliente como dependiente se obtiene  $d = 23.4\%$ . Lo cual indica que hay una asociación positiva entre las mencionadas variables”. Los clientes de Cajamarca, Encañada, LLacanora, Jesús y Namora tienden hacer buenos clientes.

➤ **Categorización de la variable Zona Residencia**

Tabla 11. Medidas direccionales de la variable Zona\_Cat

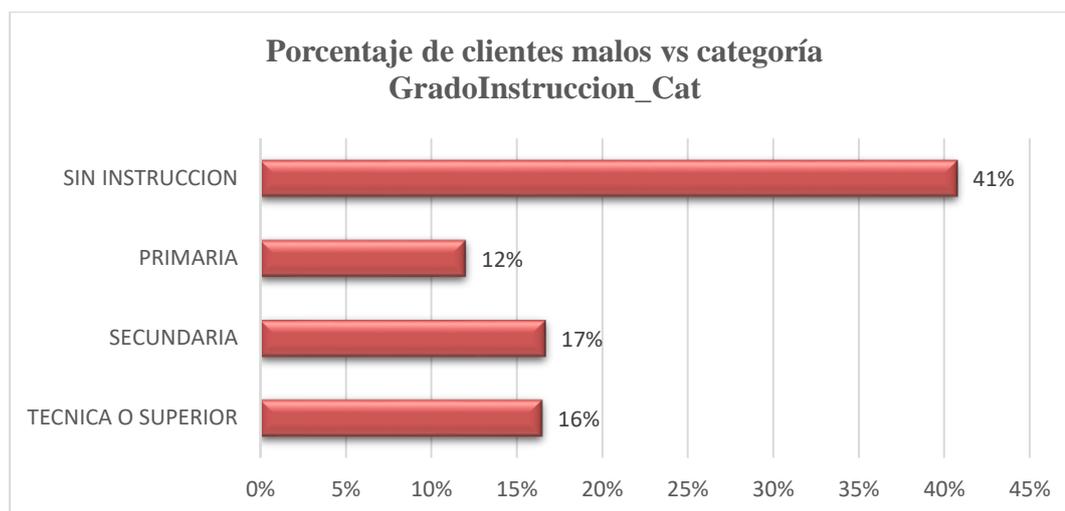
			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	-.009	.061	-.144	.885
		Zona_Cat independiente	-.013	.089	-.144	.885
		Flag_Cliente dependiente	<b>-.007</b>	.046	-.144	.885
a. Asumiendo la hipótesis alternativa.						
b. Empleando el error típico asintótico basado en la hipótesis nula.						

Fuente: Propia

Interpretación: “Considerando a la variable Zona\_Cat de Residencia como independiente y la variable Tipo de Cliente como dependiente se obtiene  $d = -0.7\%$ ”. Indica que no hay una asociación entre las mencionadas variables.

➤ **Categorización de la variable de Grado instrucción**

Figura 20. Porcentaje de clientes malos vs categoría GradoInstruccion\_Cat



Fuente: Propia

Tabla 12. Medidas direccionales de la variable Grado de instrucción

Medidas direccionales de la Variable Grado de Instrucción						
			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	.091	.061	1.466	.143
		GradoInstruccion_Cat Independiente	.145	.098	1.466	.143
		Tipo Cliente dependiente	<b>.066</b>	.045	1.466	.143

a. Asumiendo la hipótesis alternativa.

b. Empleando el error típico asintótico basado en la hipótesis nula.

Fuente: Propia

Interpretación: “Considerando a la variable Grado Instrucción como independiente y la variable Tipo de Cliente como dependiente se obtiene  $d = 6.6\%$ . Indica que hay una asociación positiva entre las mencionadas variables”. Los clientes que tiene más grado de instrucción tienden hacer buenos clientes.

### ➤ Categorización de la variable Profesión

Tabla 13. Medidas direccionales de la variable Profesión\_Cat

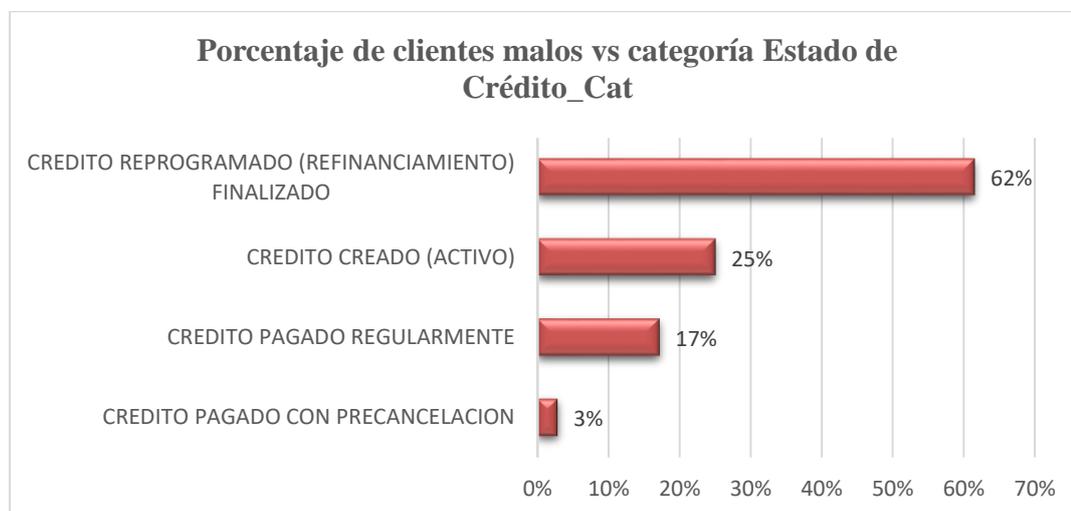
			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	-.023	.056	-.403	.687
		Profesion_Cat Independiente	-.044	.109	-.403	.687
		TipoCliente dependiente	<b>-.015</b>	.038	-.403	.687
a. Asumiendo la hipótesis alternativa.						
b. Empleando el error típico asintótico basado en la hipótesis nula.						

Fuente: Propia

Interpretación: “Considerando a la variable Profesión\_Cat como independiente y la variable Tipo de Cliente como dependiente se obtiene  $d = -1.5\%$ ”. Indica que no hay una asociación entre las mencionadas variables.

➤ **Categorización de la variable Estado de crédito**

Figura 21. Porcentaje de clientes malos vs categoría Estado de Crédito\_Cat



Fuente: Propia

Tabla 14. Medidas direccionales de la variable EstadoCredito\_Cat

			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. Aproximada
Ordinal por ordinal	d de Somers	Simétrica	.293	.046	5.432	.000
		EstadoCredito_Cat Independiente	.456	.069	5.432	.000
		TipoCliente dependiente	<b>.216</b>	.039	5.432	.000

a. Asumiendo la hipótesis alternativa.

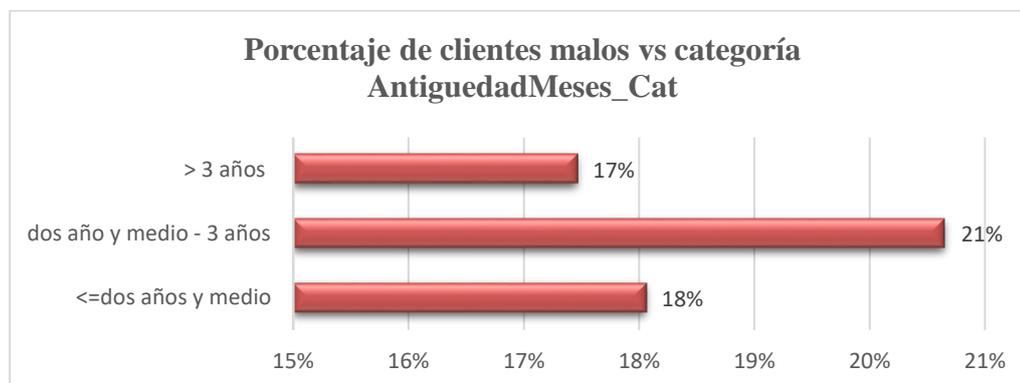
b. Empleando el error típico asintótico basado en la hipótesis nula.

Fuente: Propia

Interpretación: “Considerando a la variable Estado de Crédito\_Cat como independiente y la variable Tipo de Cliente como dependiente se obtiene  $d=21.6\%$ . Lo cual indica que hay una asociación positiva entre las mencionadas variables”. Los clientes que se encuentra en el estado Crédito Pagado con Precancelado y Regularmente tienden hacer buenos clientes.

➤ **Categorización de la variable Antigüedad de la cuenta en meses**

Figura 22. Porcentaje de clientes malos vs categoría AntigüedadMeses\_Cat.



Fuente: Propia

Tabla 15. Medidas direccionales de la variable AntigüedadMeses\_Cat

			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	.031	.058	.528	.597
		AntigüedadMeses_Cat Independiente	.049	.092	.528	.597
		TipoCliente dependiente	<b>.023</b>	.043	.528	.597

a. Asumiendo la hipótesis alternativa.

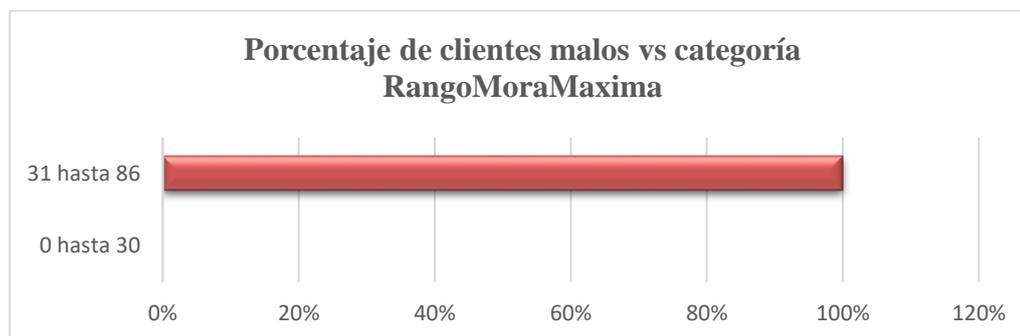
b. Empleando el error típico asintótico basado en la hipótesis nula.

Fuente: Propia

Interpretación: “Considerando a la variable AntigüedadMeses\_Cat como independiente y a Tipo de Cliente como dependiente se obtiene  $d=2.3\%$ . Lo cual indica una asociación positiva (aunque débil) entre ambas variables”. En otras palabras, hay una tendencia en los extremos de los clientes a ser bueno entre una antigüedad menor igual a dos años y medio y mayor a 3 años.

➤ **Categorización de la variable Rango de mora máxima en 6 meses**

Figura 23. Porcentaje de clientes malos vs categoría RangoMoraMaxima



Fuente: Propia

Tabla 16. Medidas direccionales de la variable RangoMoraMaxima

	Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal d de Simétrica	1.000	0.000	9.505	.000
por Somers RangoMoraMaxima	1.000	0.000	9.505	.000
ordinal Independiente				
Tipo Cliente dependiente	<b>1.000</b>	0.000	9.505	.000

a. Asumiendo la hipótesis alternativa.

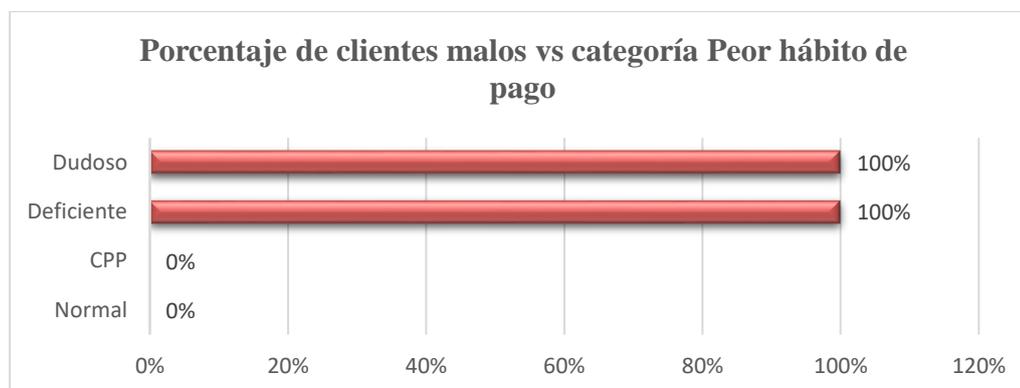
b. Empleando el error típico asintótico basado en la hipótesis nula.

Fuente: Propia

Interpretación: “Considerando a la variable Máximo Días de Mora en 6 Meses como independiente y la variable Tipo de Cliente como dependiente se obtiene  $d = 100\%$ , indica que hay una asociación positiva fuerte entre las mencionadas variables”. Los clientes que están dentro del rango 0 hasta 30 días tienden hacer buenos clientes.

➤ **Categorización de la variable Peor hábito de pago en 6 meses**

Figura 24. Porcentaje de clientes malos vs categoría Peor hábito de pago



Fuente: Propia

Tabla 17. Medidas direccionales de la variable PeorHabitopago\_6meses

			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	.644	.027	9.505	.000
		PeorHabitopago_6meses Independiente	1.000	0.000	9.505	.000
		Tipo Cliente dependiente	<b>.475</b>	.039	9.505	.000

a. Asumiendo la hipótesis alternativa.

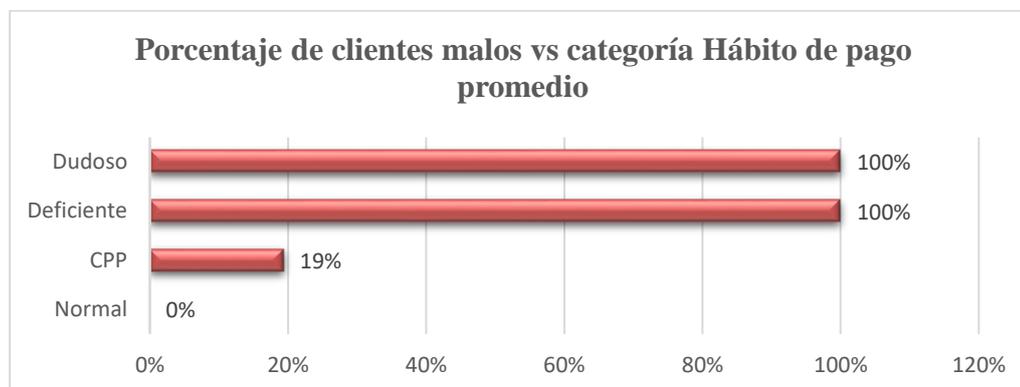
b. Empleando el error típico asintótico basado en la hipótesis nula.

Fuente: Propia

Interpretación: “Considerando a la variable Peor hábito de Pago en 6 Meses como independiente y la variable Tipo de Cliente como dependiente se obtiene  $d = 47,5\%$ , indica que hay una asociación positiva fuerte entre las mencionadas variables”. Los clientes que están dentro de la Categoría Normal(0) y CPP(1) tienden hacer buenos clientes.

➤ **Categorización de la variable Hábito de pago promedio**

Figura 25. Porcentaje de clientes malos vs categoría Hábito de pago promedio



Fuente: Propia

Tabla 18. Porcentaje de clientes malos vs categoría hábito de pago promedio

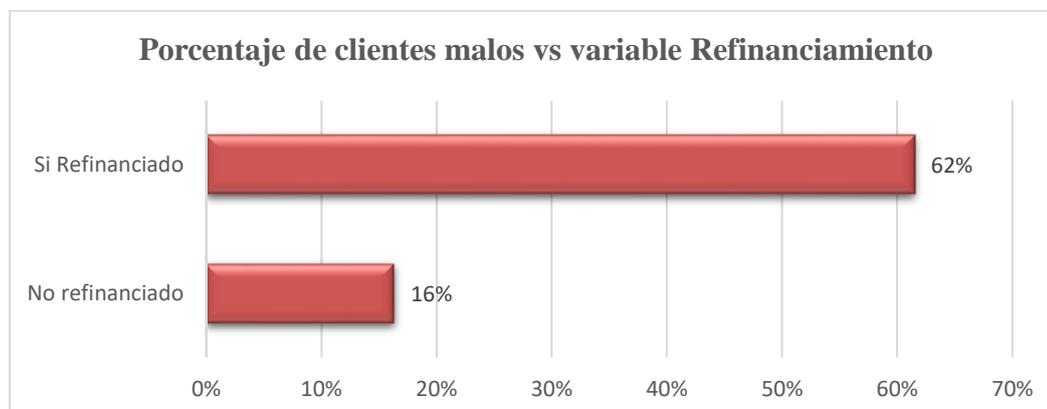
			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	.554	.032	8.745	.000
		HabitoPagoPromedio Independiente	.810	.036	8.745	.000
		TipoCliente dependiente	.421	.039	8.745	.000
a. Asumiendo la hipótesis alternativa.						
b. Empleando el error típico asintótico basado en la hipótesis nula.						

Fuente: Propia

Interpretación: “Considerando a la variable Hábito de Pago Promedio como independiente y la variable Tipo de Cliente como dependiente se obtiene  $d = 42,1\%$ . Lo cual indica que hay una asociación positiva fuerte entre las mencionadas variables”. Los clientes que están dentro del hábito promedio en la Categoría Normal (0) tienden hacer buenos clientes.

### ➤ Categorización de la variable Refinanciamiento

Figura 26. Porcentaje de clientes malos vs variable Refinanciamiento



Fuente: Propia

Tabla 19. Medidas direccionales de la variable refinanciamiento

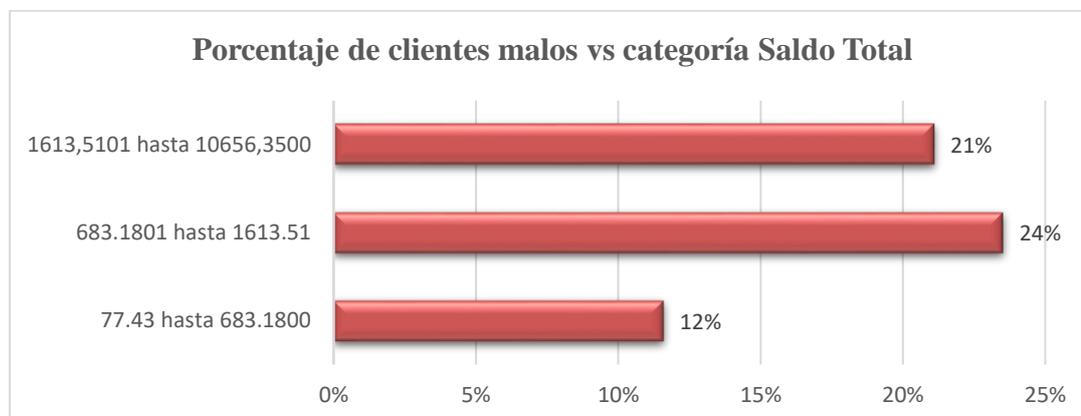
			Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	.235	.076	2.530	.011
		Refinanciado Independiente	.159	.061	2.530	.011
		TipoCliente dependiente	<b>.452</b>	.137	2.530	.011
a. Asumiendo la hipótesis alternativa.						
b. Empleando el error típico asintótico basado en la hipótesis nula.						

Fuente: Propia

Interpretación: “Considerando a la variable Refinanciado como independiente y la variable Tipo de Cliente como dependiente se obtiene  $d=42,2$ . Lo cual indica que hay una asociación positiva fuerte entre las mencionadas variables”. Los clientes que no refinancian de cuenta tienden hacer buenos clientes.

➤ **Categorización de la variable Saldo Total**

Figura 27. Porcentaje de clientes malos vs categoría Saldo Total



Fuente: Propia

Tabla 20. Medidas direccionales de la variable SaldoTotal\_Cat

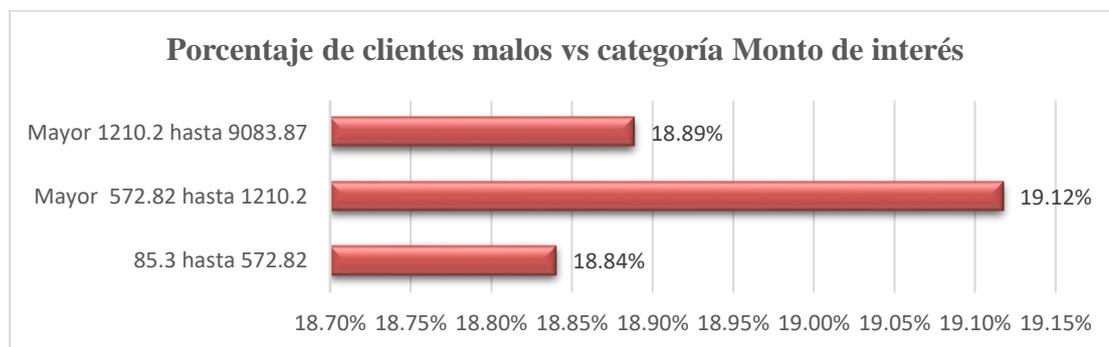
			Valor	Error típ. asint.a	T aproximada <sup>b</sup>	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	.104	.055	1.858	.063
		SaldoTotal_Cat Independiente	.164	.087	1.858	.063
		TipoCliente dependiente	.076	.041	1.858	.063
a. Asumiendo la hipótesis alternativa.						
b. Empleando el error típico asintótico basado en la hipótesis nula.						

Fuente: Propia

Interpretación: “Considerando a la variable Saldo Total\_Cat como independiente y a Tipo de Cliente como dependiente se obtiene  $d=7,6\%$ , indica una asociación positiva (aunque débil) entre ambas variables”. En otras palabras, a hay una tendencia en los extremos del cliente a ser bueno, menor igual a S/ 683,18 y mayor igual a 1613.5101.

➤ **Categorización de la variable Monto de Interés en los 6 meses**

Figura 28. Porcentaje de clientes malos vs categoría Monto de interés



Fuente: Propia

Tabla 21. Medidas direccionales de la variable MontoInteres\_Cat

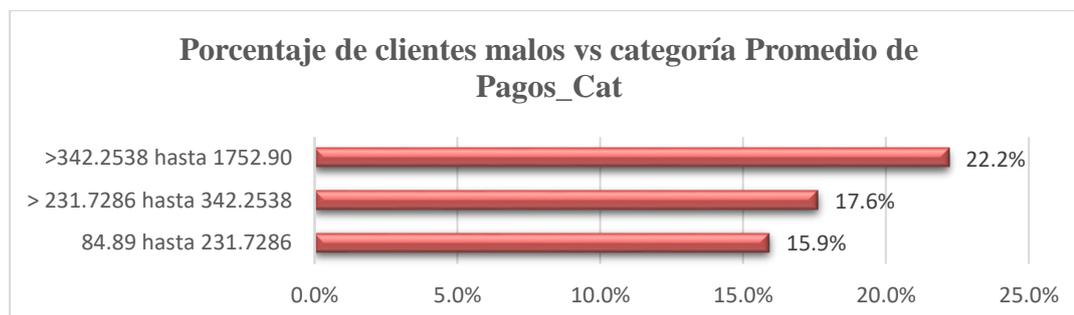
			Valor	Error típ. asint.a	T aproximada	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	.002	.058	.041	.967
		MontoInteres_Cat Independiente	.004	.092	.041	.967
		TipoCliente dependiente	.002	.043	.041	.967

Fuente: Propia

Interpretación: “Considerando a la variable Monto Interés Categoría como independiente y a Tipo de Cliente como dependiente se obtiene  $d=0,2\%$ , indica una asociación positiva (aunque muy débil) entre ambas variables”. En otras palabras, hay una ligera tendencia en los extremos que los clientes son bueno, menor igual a S/ 572.82 y mayor igual a 1210.2.

➤ **Categorización de la variable Promedio de Pagos**

Figura 29. Porcentaje de clientes malos vs categoría Promedio de Pagos\_Cat



Fuente: Propia

Tabla 22. Medidas direccionales de la variable promedio de Pagos\_Cat

			Valor	Error típ. asint.a	T aproximada	Sig. aproximada
Ordinal por ordinal	d de Somers	Simétrica	.070	.058	1.198	.231
		PromedioPagos_Cat Independiente	.111	.092	1.198	.231
		TipoCliente dependiente	.052	.043	1.198	.231
a. Asumiendo la hipótesis alternativa.						
b. Empleando el error típico asintótico basado en la hipótesis nula.						

Fuente: Propia

#### 5.1.4.2. Desarrollo del modelo

La puntuación de crédito se construyó en base a una Regresión Logística binaria, en donde la variable de respuesta puede tomar dos posibles valores: Ser Bueno o ser Malo.

A diferencia de la Regresión Múltiple, en este modelo las variables explicativas pueden ser categóricas y no necesitan la condición de tener una distribución conjunta normal multivariada. Se planteó dos escenarios para encontrar el factor de ponderación y corrección como se aprecia en el cuadro 15.

Cuadro 15. Ecuación del score en base a dos escenarios

	Score	=	$\text{Ln}(p/1-p)$	x	Factor de ponderación	+	Factor de corrección
<b>Escenario 1</b>	45	=	3.912023005	x	Factor de ponderación	+	Factor de corrección (I)
<b>Escenario 2</b>	50	=	4.605170186	x	Factor de ponderación	+	Factor de corrección (II)
	(II) - (I)						
	5	=	0.693147181	x	Factor de ponderación		
<b>Factor ponderación =</b>			<b>7.21347520</b>				
<b>Factor corrección =</b>			<b>16.78071905</b>				

Fuente: Propia

Luego se procedió a realizar varias corridas para encontrar la estimación ideal (ver apéndice C) de los parámetros para tener el mejor modelo Logístico Binario, como se puede evidenciar en el cuadro 16.

Cuadro 16. Variables en la ecuación del modelo

Variables en la ecuación: Paso 3 <sup>a</sup>	Categoría/Nivel	B	E.T.	Wald	gl	Sig.	Exp(B)	Elegido
Grado de Instrucción	Técnica o superior			8.653	3	.034		SI
	Secundaria	1.015685	.650	2.445	1	.118	2.761	
	Primaria	-0.177008	.895	.039	1	.843	.838	
	Sin instrucción	2.071743	.771	7.214	1	.007	7.939	
Estado de Crédito	Credito pagado con precancelacion			8.225	3	.042		SI
	Credito pagado regularmente	.813857	1.074	.575	1	.448	2.257	
	Credito creado (activo)	1.350997	.836	2.611	1	.106	3.861	
	Credito reprogramado (refinanciamiento) finalizado	3.172322	1.132	7.847	1	.005	23.863	
Monto de Interés	Desde 85.3 hasta 572.82			1.584	2	.453		NO
	Mayor a 572.82 hasta 1210.2	.181	.684	.070	1	.791	1.198	
	Mayor a 1210.2 hasta 9083.87	-.602	.692	.758	1	.384	.548	
VecesMoraMayora0		.420616	.220	3.671	1	.055	1.523	SI
HabitoPago3_Cat		2.469862	.549	20.240	1	.000	11.821	SI
Constante		-7.342525	1.493	24.181	1	.000	.001	SI

Fuente: Propia

B: Valor estimado para cada variable del modelo (Por método de MV).

ET: Representa los errores estándar de los coeficientes.

Wald: Corresponde al estadístico de Wald, la cual contrasta la hipótesis de si los coeficientes son iguales a cero (No significativos), la cual además sigue una distribución chi-cuadrado.

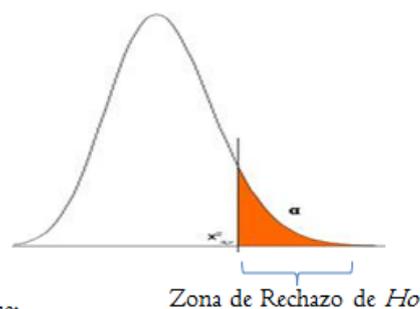
gl: Indica los grados de libertad del estadístico de Wald.

Sig.: Nivel de significancia real del estadístico de Wald.

$$Wald = \frac{(\text{Coeficiente})^2}{(\text{ErrorEs tan dar})^2} \approx \chi^2$$

$H_0 : \beta = 0$  Coeficiente no significativo

A un nivel de significancia:  $\alpha = 0.05$  Tenemos que:



Si el nivel de Significancia Real  $< 0.05$ , entonces se Rechaza  $H_0$ , es decir el coeficiente si es significativo para el modelo, en cambio si el nivel de la significancia real  $> 0.05$  entonces el coeficiente no es significativo para el modelo. (Chambi Canahuire & Arias Huamani, 2014)

Se puede determinar que (observar la cuadro 15), las variables categóricas Grado de instrucción y Estado de crédito como las variables escalares Veces de mora mayor a cero y Hábito de pago 3 son las que se encuentran en Zona de Rechazo de  $H_0$ .

Remplazando los coeficientes calculados, las ecuaciones para llegar al Score final:

$$\text{Probabilidad\_Calculada} = \frac{1}{1 + \exp^{((-1)*(g_i)}}$$

$$\begin{aligned} g_i = & 1.015685 * \text{GradoInstruccion\_Cat\_1} - 0.177 * \text{GradoInstruccion\_Cat\_2} \\ & + 2.07174 * \text{GradoInstruccion\_Cat\_3} + 0.813856 \\ & * \text{EstadoCredito\_Cat\_1} + 1.350995 * \text{EstadoCredito\_Cat\_2} \\ & + 3.172322 * \text{EstadoCredito\_Cat\_3} + 0.420616 \\ & * \text{VecesMoraMayora0} + 2.469862 * \text{HabitoPago3\_Cat} \\ & - 7.34252525 \end{aligned}$$

$$LN\_ODDS = LN\left(\frac{Probabilidad\_calculada}{1 - Probabilidad\_calculada}\right)$$

$$SCORE\_FINAL = LN\_ODDS * 7.2134752 + 16.78072$$

Para poder leer el puntaje en forma ascendente de aquellos clientes que tiene mayor puntaje tiende hacer buenos clientes y no haya puntaje negativo se modificó el Score\_final.

$$SCORE\_FINAL = 60 - SCORE\_FINAL$$

La construcción del modelo se desarrolló y simuló en el programa SPSS y para mejorar la presentación de las tablas y gráficos se utilizó el Excel para interpretación de los resultados. El programa SPSS tiene la ventana de sintaxis que sirve para programar el algoritmo. A continuación, se muestra lo más importante del algoritmo de la construcción del modelo y el resto del algoritmo se detalla en apéndice D.

```
compute Probabilidad_Calculada=1/(1+exp((-1)*
1.015685*GradoInstruccion_Cat_1-0.177*GradoInstruccion_Cat_2
+2.07174*GradoInstruccion_Cat_3+0.813856*EstadoCredito_Cat_1+1.350995*
EstadoCredito_Cat_2      +      3.172322*EstadoCredito_Cat_3      +
0.420616*VecesMoraMayora0 + 2.469862*HabitoPago3_Cat - 7.34252525))).

execute.

compute LN_ODDS=ln(Probabilidad_Calculada/(1-
Probabilidad_Calculada)).

execute.

***SCORE_FINAL:

compute SCORE_FINAL=LN_ODDS*7.2134752+16.78072.

execute.
```

```
compute SCORE_FINAL=60-SCORE_FINAL.  
execute.  
  
if SCORE_FINAL>100 SCORE_FINAL=100.  
  
if SCORE_FINAL<0 SCORE_FINAL=0.  
  
execute.
```

### 5.1.5. Evaluación

Una vez estimado los parámetros del modelo y realizado el cálculo del Score\_Final, se evalúa el poder predictivo del mismo, aplicando dicho modelo en la muestra de validación. Con el resultado obtenido, se verifica si el modelo que se ha realizado es confiable. Al aplicar la prueba KS (Kolmogorov – Smirnov) se observa en la tabla 23 la diferencia máxima (Máxima separación) entre la distribución acumulada de clientes buenos y la distribución acumulada de clientes malos.

Tabla 23. Prueba K-S Aplicado a la muestra de validación

Score Final Agrupado		Tipo de Cliente		Total	Malos Acumulados	Buenos Acumulados	%Malos Acumulados	%Buenos Acumulados	Diferencia Acumulada
		Malos	Buenos						
C1	<= 33,88	1	0	2	1	0	20.00%	0.00%	20.00%
C2	33,89 - 41,09	1	0	1	2	0	40.00%	0.00%	40.00%
C3	41,10 - 42,38	0	1	1	2	1	40.00%	4.76%	35.24%
C4	42,39 - 46,66	1	1	2	3	2	60.00%	9.52%	50.48%
C5	46,67 - 46,97	0	1	1	3	3	60.00%	14.29%	45.71%
C6	46,98 - 50,42	0	1	1	3	4	60.00%	19.05%	40.95%
C7	50,43 - 56,07	1	1	2	4	5	80.00%	23.81%	56.19%
C8	56,08 - 58,91	1	1	1	5	6	100.00%	28.57%	71.43%
C9	58,92 - 59,36	0	1	1	5	7	100.00%	33.33%	66.67%
C10	59,37 - 63,94	0	2	2	5	9	100.00%	42.86%	57.14%
C11	63,95 - 66,98	0	1	1	5	10	100.00%	47.62%	52.38%
C12	66,99 - 68,46	0	1	1	5	11	100.00%	52.38%	47.62%
C13	68,47 - 71,27	0	1	1	5	12	100.00%	57.14%	42.86%
C14	71,28 - 71,49	0	2	2	5	14	100.00%	66.67%	33.33%
C15	71,50 - 72,55	0	1	1	5	15	100.00%	71.43%	28.57%
C16	72,56 - 73,04	0	1	1	5	16	100.00%	76.19%	23.81%
C17	73,05 - 81,21	0	2	2	5	18	100.00%	85.71%	14.29%
C18	81,22 - 84,25	0	1	1	5	19	100.00%	90.48%	9.52%
C19	84,26 - 88,86	0	1	1	5	20	100.00%	95.24%	4.76%
C20	88,87+	0	1	1	5	21	100.00%	100.00%	0.00%
Total		5	21	26					

Fuente: Propia

En la tabla 23 se observa que la máxima separación es: 71.43%

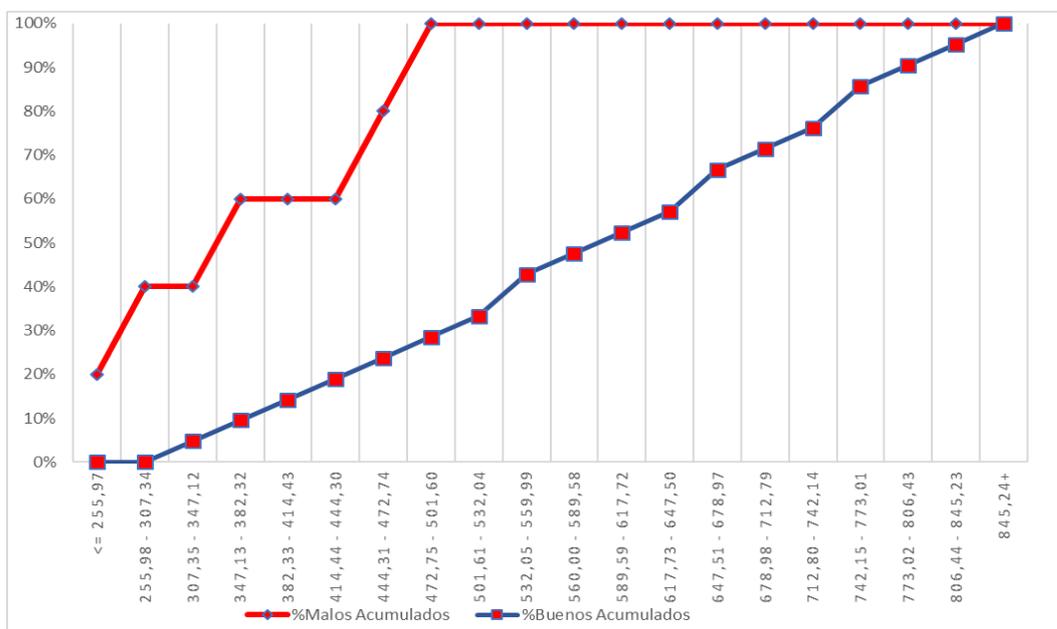
Cuadro 17. Niveles de score para KS.

Si KS	Es menor que 20	Malo
	Entre 20 y 40	Aceptable
	Entre 40 y 60	Bueno
	60 y 75	Excelente
	más de 75	Perfecto o sospechoso

Fuente: Chambi Canahuire & Arias Huamani (2014)

Por lo tanto, el valor de K-S está en el nivel de 60 y 75 según el cuadro 16 que pertenece a la categoría excelente, para ello se elaboró la figura 30, donde se aprecia que es idéntico a la teoría de la validación de prueba K-S, por consecuencia se acepta que el modelo encontrado es válido - excelente.

Figura 30. Prueba K-S



Fuente: Propia

## 5.2. Análisis, interpretación y discusión de resultados

En este punto, se describe la manera cómo se han obtenido, tratado y analizado los resultados con respecto a los indicadores planteados en la operacionalización de variables de esta investigación, para ello se aplicó técnicas estadísticas y cuestionario antes y después de aplicar el modelo, instrumento que ha permitido medir el impacto del modelo de minería de datos para generar el puntaje de cumplimiento de pago y demostrar la hipótesis de la presente investigación.

### 5.2.1. Indicadores de la variable Independiente

#### 5.2.1.1. Nivel de respuesta al usuario

Dicho indicador mide el tiempo requerido para obtener los datos solicitados en la etapa de requerimientos (cuadros 6 hasta 10) antes y después de aplicar minería de datos como se observa en la tabla 24.

Tabla 24. Tiempo de requerido por requerimiento

Requerimientos	Tiempo Minutos		Diferencia
	Pre-Test	Post-Test	
R01	0.033	0.01667	-0.01633
R02	0.033	0.01667	-0.01633
R03	30	0.05	-29.95
R04	60	1	-59
R05	1440	30	-1410
Total	1530.066	31.08334	-1498.98266

Fuente: Propia

La tabla 24 muestra que hubo una reducción de tiempo de 1498.98266 minutos en la obtención de los datos al aplicar minería de datos, Por lo tanto, este indicador es favorable para dicha investigación.

### 5.2.1.2. Nivel de conformidad

Este indicador se realizó mediante la aplicación del instrumento de un cuestionario de acuerdo al apéndice E, aplicadas a las 2 personas implicadas en el desarrollo y mantenimiento del Sistema de Crédito de la Micro Financiera Progreso. El cuestionario está basado en los criterios de determinar la satisfacción del usuario, con respecto a la información procesada por minería de datos, en marcados en la usabilidad y funcionalidad del modelo.

Se ha considerado a cada pregunta un peso a las respuestas como indica el cuadro 18, dicho peso está en base a la escala de Liker.

Cuadro 18. Criterios de evaluación del cuestionario del Pre y Post -Test

Criterio	Abreviatura	Puntaje
Siempre	S	5
Casi Siempre	CS	4
A veces	AV	3
Casi Nunca	CN	2
Nunca	N	1

Fuente: Propia

Se presentan los resultados del Pre-test y Post - Test en las tablas 25 y 26, donde “F.” representa la frecuencia para un determinado criterio, “ $\bar{X}_1$ ” el promedio ponderado de la pregunta en el pre - test y “ $\bar{X}_2$ ” el promedio ponderado de la pregunta en el post - test respectivamente.

Tabla 25. Resultado del Pre - Test

Item	Descripción de la Pregunta	S	CS	AV	CN	N	X1
		F	F	F	F	F	
1	Las sentencias SQL utilizadas para obtener el Data Mart del Sistema de Créditos y los códigos de programación utilizados en SPSS se ejecutan sin ningún inconveniente.	0	0	0	1	1	1.50
2	¿El modelo de minería de datos procesa todos los datos almacenados por el Data Mart del Sistema de Crédito?	0	0	0	1	1	1.50
3	¿El modelo de minería de datos del Sistema de Créditos describe la asociación de la variable dependiente tipo de cliente con respecto a las variables cualitativas que intervienen en el cálculo del puntaje de cumplimiento de pagos de los clientes?	0	0	0	0	2	1.00
4	¿El modelo de minería de datos del Sistema de Créditos describe la asociación de la variable dependiente tipo de cliente con respecto a las variables cuantitativas que intervienen en el cálculo del puntaje de cumplimiento de pagos de los clientes?	0	0	1	0	1	2.00
5	¿El modelo de minería de datos del Sistema de Créditos utiliza las variables cualitativas y cuantitativas en la regresión logística binaria para el cálculo del puntaje de cumplimiento de pagos de los clientes?	0	0	0	0	2	1.00
6	¿El modelo de minería de datos del Sistema de Créditos está validada correctamente con otra muestra?	0	0	0	0	2	1.00
7	¿El resultado que calcula el puntaje de cumplimiento de pagos de los clientes a través del modelo de minería de datos del Sistema de Créditos es fácil de interpretar para tomar decisiones en la predicción del otorgamiento de préstamos de los buenos clientes?	0	0	0	0	2	1.00

Fuente: Propia

Tabla 26. Resultado del Post – Test

Item	Descripción de la Pregunta	S	CS	AV	CN	N	$\bar{X}_2$
		F	F	F	F	F	
1	Las sentencias SQL utilizadas para obtener el Data Mart del Sistema de Créditos y los códigos de programación utilizados en SPSS se ejecutan sin ningún inconveniente.	1	1	0	0	0	4.50
2	¿El modelo de minería de datos procesa todos los datos almacenados por el Data Mart del Sistema de Crédito?	0	2	1	0	0	3.67
3	¿El modelo de minería de datos del Sistema de Créditos describe la asociación de la variable dependiente tipo de cliente con respecto a las variables cualitativas que intervienen en el cálculo del puntaje de cumplimiento de pagos de los clientes?	2	0	0	0	0	5.00
4	¿El modelo de minería de datos del Sistema de Créditos describe la asociación de la variable dependiente tipo de cliente con respecto a las variables cuantitativas que intervienen en el cálculo del puntaje de cumplimiento de pagos de los clientes?	1	1	0	0	0	4.50
5	¿El modelo de minería de datos del Sistema de Créditos utiliza las variables cualitativas y cuantitativas en la regresión logística binaria para el cálculo del puntaje de cumplimiento de pagos de los clientes?	2	0	0	0	0	5.00
6	¿El modelo de minería de datos del Sistema de Créditos está validada correctamente con otra muestra?	1	1	0	0	0	4.50
7	¿El resultado que calcula el puntaje de cumplimiento de pagos de los clientes a través del modelo de minería de datos del Sistema de Créditos es fácil de interpretar para tomar decisiones en la predicción del otorgamiento de préstamos de los buenos clientes?	1	1	0	0	0	4.50

Fuente: Propia

Luego se procede a realizar la tabla 27 donde se calcula la diferencia de medias

Pre – Test y Post – Test.

Tabla 27. Diferencia de medias del Pre y Post - Test

N° Pregunta	$\bar{X}_1$ : Pre Test	$\bar{X}_2$ : Post Test	$\bar{X}_2 - \bar{X}_1$
1	1.50	4.50	3.00
2	1.50	3.67	2.17
3	1.00	5.00	4.00
4	2.00	4.50	2.50
5	1.00	5.00	4.00
6	1.00	4.50	3.50
7	1.00	4.50	3.50
Total			22.67

Fuente: Propia

El resultado de la diferencia de medias es un valor positivo de 22.67 puntos de 28, que indica que los usuarios de la implementación de minería de datos tienen una buena aceptación del nivel de conformidad que da como resultado final la puntuación crediticia de los clientes de la Micro Financiera Progreso e involucran las variables explicativas (patrones) que intervienen en el cálculo de dicha puntuación.

## **5.2.2. Indicadores de la variable dependiente**

### **5.2.2.1. Riesgo crediticio**

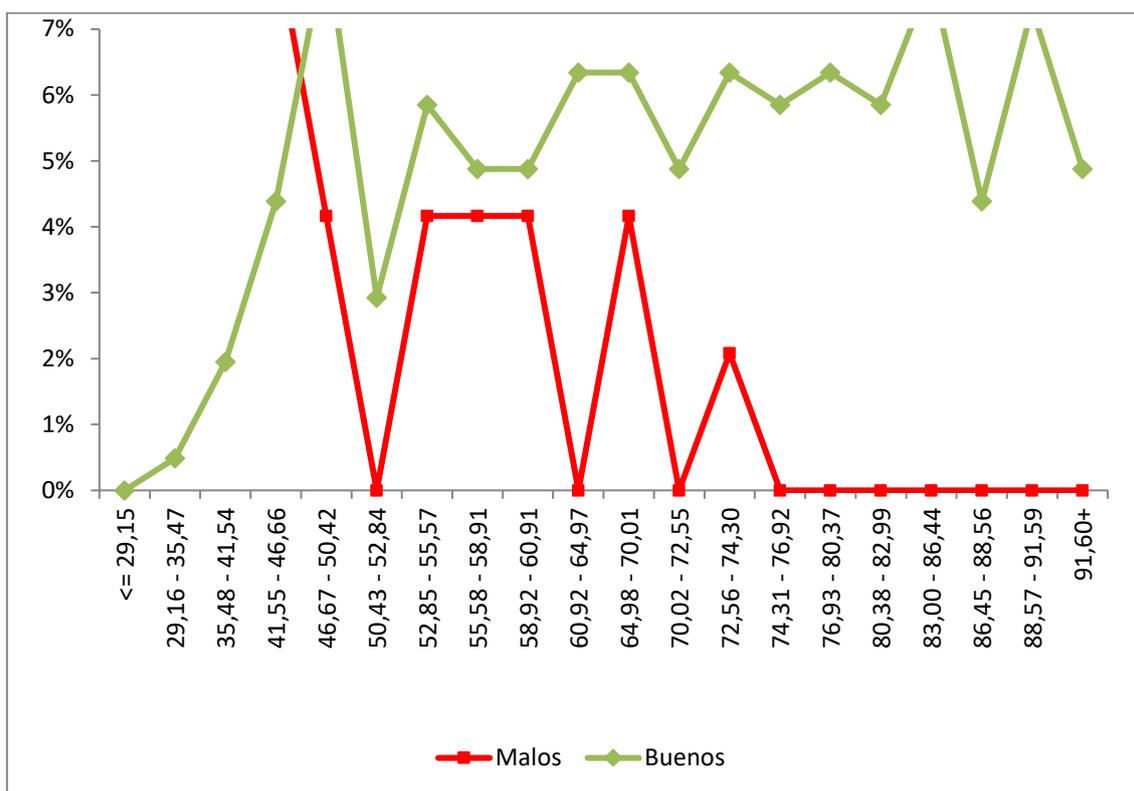
Para calcular el indicador de Riesgo Crediticio primeramente se elaboró un análisis de Poder Discriminatorio Univariado, ver tabla 28, donde se observa el comportamiento de los malos con respecto a los buenos clientes, además se aprecia que hasta un puntaje de 46,66 puntos son clientes con alto riesgo de pago y los clientes que tiene un puntaje mayor igual a 46.67 puntos son clientes que tienen una mejor probabilidad de que paguen en los 6 primeros meses.

Tabla 28. Poder Discriminatorio Univariado

		TipoCliente		Número Clientes	Malos	Buenos	N° Clientes Acumulados
		Buenos	Malos				
Score_Final (agrupado)	<= 29,15	0	13	13	27%	0%	13
	29,16 - 35,47	1	15	16	31%	0%	29
	35,48 - 41,54	4	5	9	10%	2%	38
	41,55 - 46,66	9	4	13	8%	4%	51
	46,67 - 50,42	17	2	19	4%	8%	19
	50,43 - 52,84	6	0	6	0%	3%	25
	52,85 - 55,57	12	2	14	4%	6%	39
	55,58 - 58,91	10	2	12	4%	5%	51
	58,92 - 60,91	10	2	12	4%	5%	12
	60,92 - 64,97	13	0	13	0%	6%	25
	64,98 - 70,01	13	2	15	4%	6%	40
	70,02 - 72,55	10	0	10	0%	5%	50
	72,56 - 74,30	13	1	14	2%	6%	14
	74,31 - 76,92	12	0	12	0%	6%	26
	76,93 - 80,37	13	0	13	0%	6%	39
	80,38 - 82,99	12	0	12	0%	6%	51
	83,00 - 86,44	16	0	16	0%	8%	16
	86,45 - 88,56	9	0	9	0%	4%	25
88,57 - 91,59	15	0	15	0%	7%	40	
91,60+	10	0	10	0%	5%	50	
Total		205	48	253	100%	100%	253

Fuente: Propia

Figura 31. Comportamiento de los bueno y malos clientes



Fuente Propia

Una vez determinado la intersección del comportamiento de los bueno y malos clientes como se observa en la figura 31, se procedió a calcular los cuantos clientes tiene un puntaje menor igual a 46.66 que haciende a 51, por lo tanto, se reemplaza los valores en la fórmula del indicador.

$$\text{Riesgo Crediticio} = \frac{\text{N}^\circ \text{ de Clientes alto riesgo de Pago} * 100}{\text{N}^\circ \text{ total de Cliente}}$$

$$\text{Riesgo Crediticio} = \frac{51}{253} * 100 = 20.16\%$$

En un periodo de tiempo se reducirá 20.16% de clientes de alto riesgo, al no otorgarles ningún tipo de crédito incluyendo refinanciamiento.

### 5.2.2.2. Recuperación de capital

Al haber determinado el puntaje máximo que se encuentran los clientes de alto riesgo de crédito se realizó la tabla de contingencia del puntaje agrupado por Saldo Total hasta la fecha de corte de dicha investigación.

Tabla 29. Contingencia del puntaje agrupado por Saldo Total

Categoría		Saldo Total	Hi	Hi
Score_final (agrupado)	<= 29,15	20824.8600	4.76%	4.76%
	29,16 - 35,47	30441.0000	6.95%	11.71%
	35,48 - 41,54	17336.0700	3.96%	15.67%
	41,55 - 46,66	18260.5700	4.17%	19.85%
	46,67 - 50,42	29110.7000	6.65%	26.50%
	50,43 - 52,84	15838.7200	3.62%	30.12%
	52,85 - 55,57	20008.6100	4.57%	34.69%
	55,58 - 58,91	24369.2600	5.57%	40.25%
	58,92 - 60,91	27970.3000	6.39%	46.65%
	60,92 - 64,97	36732.7900	8.39%	55.04%
	64,98 - 70,01	20517.7800	4.69%	59.73%
	70,02 - 72,55	11341.8800	2.59%	62.32%
	72,56 - 74,30	25534.1800	5.83%	68.15%
	74,31 - 76,92	18536.5700	4.24%	72.39%
	76,93 - 80,37	24270.5600	5.55%	77.93%
	80,38 - 82,99	15724.2400	3.59%	81.52%
	83,00 - 86,44	31867.9300	7.28%	88.80%
	86,45 - 88,56	11164.4800	2.55%	91.35%
	88,57 - 91,59	18851.8400	4.31%	95.66%
91,60+	18986.7000	4.34%	100.00%	
Total		437689.0400		

Fuente: Propia

El porcentaje acumulado del saldo total de los clientes riesgosos representa el 19.85%; por complemento  $100\% - 19.85\% = 80.15\%$  representa el pronóstico de recuperación del capital que se va a realizar sin problemas en el futuro.

### 5.3. Contrastación de hipótesis

En la presente investigación se aplicó la Prueba de Hipótesis para dos muestras dependiente, dichas muestras contienen los mismos individuos en dos condiciones que se trata de diferenciar. Por lo cual se usa la siguiente formula:

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}}$$

Dónde:

$\bar{d}$ : Es la media de las diferencias entre los valores de las muestras del pre – test y post – test.

$S_d$ : Es la desviación estándar de las diferencias de medias.

$n$ : Número de elementos o preguntas de las encuestas,  $n = 7$

El valor  $t$  para comparar, se obtiene la tabla  $t$  – Student. En este caso se tiene que los grados de libertad es de  $gl = n - 1 = 6$ . Para este estudio se consideró un nivel de significancia de  $\alpha = 0.05$  o 5%. Por tanto  $\frac{\alpha}{2} = 0.025$ .

#### 5.3.1. Formulación de la Hipótesis

Hipótesis nula ( $H_0$ ): La implementación de minería de datos a través del modelo planteado no mejorará la calificación de crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca.

$$H_0: \mu_1 = \mu_2 \therefore \mu_1 - \mu_2 = 0$$

Hipótesis alternativa ( $H_1$ ): La implementación de minería de datos a través del modelo planteado mejorará la calificación de crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca.

$$H_1: \mu_1 \neq \mu_2 \therefore \mu_1 > \mu_2 \text{ ó } \mu_1 < \mu_2$$

### 5.3.2. Ubicación de la región crítica

Para la región de la hipótesis nula o de rechazo se tiene:

$$t: < -t_{\frac{\alpha}{2}}; t_{\frac{\alpha}{2}} \geq < -0.025; 0.025 >$$

$$\therefore t: < -2.44; 2.44 >$$

Para la región de aceptación de la hipótesis se tiene:

$$t: < -\infty; -t_{\frac{\alpha}{2}} ], [ t_{\frac{\alpha}{2}}; \infty+ > = < -\infty; -0.025 ], [ 0.025; \infty+ >$$

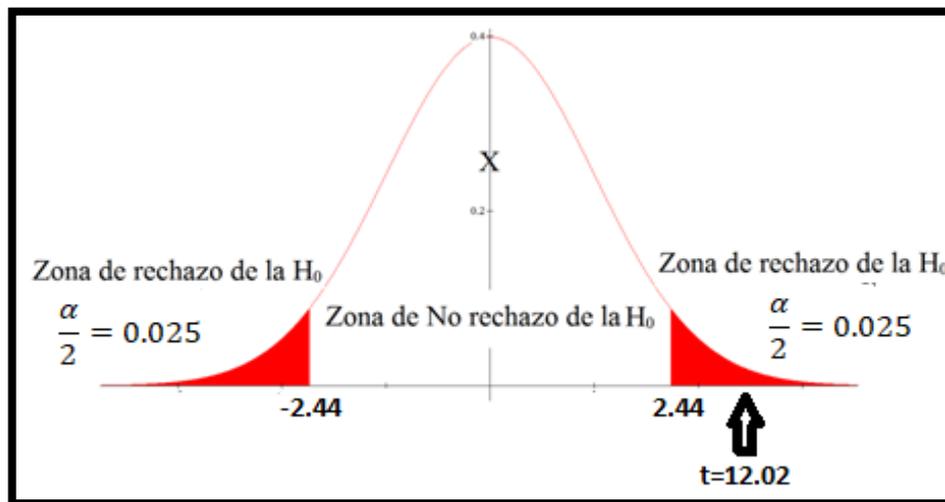
$$\therefore t: < -\infty; -2.44 ], [ 2.44; \infty+ >$$

### 5.3.3. Determinación y ubicación de valor esperado en la región crítica

Luego de aplicar la fórmula para el cálculo de  $t$  sé que:

$$t = \frac{\frac{3.24}{0.7126}}{\sqrt{7}} = 12.02$$

Figura 32. Ubicación del valor esperado en la región crítica



Fuente: Propia

Según la figura 32 se cumple  $H_1: \mu_1 \neq \mu_2 \therefore \mu_1 > \mu_2$  ó  $\mu_1 < \mu_2$

#### 5.3.4. Aceptación o rechazo de la hipótesis

Luego de ubicar el valor de  $t = 12.02$  en la curva se tiene que el valor es mayor a 2.44, por lo tanto se encuentra en la zona de rechazo de  $H_0$  entonces se rechaza la hipótesis nula  $H_0$ : "La implementación de minería de datos a través del modelo planteado no mejorará la calificación de crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca" y se acepta la hipótesis  $H_1$ : "La implementación de minería de datos a través del modelo planteado mejorará la calificación de crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca".

#### 5.3.5. Discusión de resultados

Realizado el análisis, interpretación y prueba de hipótesis podemos concluir:

Que el efecto de la implementación de minería de datos en la calificación de crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca se basa en la aceptación positiva por el área usuario demostrado por la prueba de hipótesis de  $t=12.02$  mayor al  $t_{0.025}=2.44$ , las razones de dicha aceptación es porque la implementación de minería de datos está programado para que se pueda ejecutar en un tiempo aproximado de 31.083 minutos frente a 1498.98266 minutos si lo desearían realizar con reportes del sistema de información de créditos que actualmente usan; Además los usuarios pudieron evidenciar que a través de la implementación de minería de datos se logró la puntuación de calificación a futuro de los clientes de la Micro Financiera Progreso en base a atributos cualitativos y cuantitativos del historial de sus créditos y sus respectivos pagos. Con dicha puntuación de calificación de los clientes se determinó los clientes con alto nivel de riesgo (no puedan pagar su deuda) que

representan el 20.16% por lo que se puede asegurar el retorno del capital en un 80.15% que representa el pronóstico de recuperación del capital sin problemas.

Igualmente se realizó un análisis y contraste de los estudios realizados en las tesis presentadas como antecedentes, determinando lo siguiente:

- En el estudio realizado por Figueroa (2006), se observa que ha desarrollado un modelo de credit scoring que sirva como soporte del proceso de decisión si otorgar o rechazar una solicitud de crédito de consumo utilizando características propias del solicitante. Este desarrollo es similar con la variante que la construcción de un modelo de credit scoring es parte de la metodología de minería de datos, resaltando que un Data Mart es programado y se puede ir acumulando los datos preparados de forma automática y esto pueda servir en el futuro volver a generar el modelo y encontrar nuevos patrones. Figueroa (2006) estudia a los individuos ecuatorianos que han recibido un crédito en la institución involucrada que hasta el corte de la investigación no tienen cuotas pendientes. En este punto la presente tesis involucra a clientes que radican en la provincia de Cajamarca – Perú y que tiene pagos pendientes hasta la fecha de corte con el fin de ver averiguar el comportamiento de los malos y buenos clientes. Además, la presente tesis se validó el modelo, se calculó el nivel de riesgo y pronosticó la recuperación del capital.
- Peralta (2009), en su informe de tesis al igual que la presente investigación, implementó minería de datos para predecir los clientes de alto riesgo (insolventes). En cambio, la presente investigación validó el modelo con un porcentaje de prueba K-S de 71.43% para encontrar clientes riesgosos por

otorgamiento de crédito, frente al porcentaje de precisión 72 que calculó Peralta en su primer entrenamiento que le permite predecir clientes insolventes referente al servicio de telefonía fija.

- Lara y Camino (2010) en su trabajo de investigación, desarrollaron un modelo de credit scoring aplicando técnica paramétrica de regresión logística para una institución sometida a supervisión y especialización en microcréditos similar a la presente investigación, aunque ambas investigaciones utilizan diferentes metodologías en el desarrollo. Lara y camino realizan un modelo de calificación estadística capaz de predecir correctamente en 78.3% de los créditos de la cartera de la Edyme Proempesa diferente a la que se obtuvo en la actual investigación con porcentaje de prueba K-S de 71.43% para encontrar clientes riesgosos por otorgamiento de crédito. Otra diferencia es que no utiliza algoritmos para automatizar el repositorio de datos que servirá para poder actualizar el modelo y encontrar nuevos patrones.

## CONCLUSIONES

1. El análisis y validez de la información de la base de datos transaccional se realizó primeramente comprendiendo los datos de los atributos que contiene las tablas de la base de datos transaccional que genera el sistema de información de crédito de la Micro Financiera Progreso, acorde con los 5 requerimientos planteados para el logro del objetivo del negocio, obteniendo como resultado un diseño de esquema de las tablas de la Base de datos temporal(Staging Area) que contiene trece tablas y la migración de los datos del origen a la base de datos temporal a través de sentencias SQL de tal manera que no se pierda o altere su valor de los datos.
2. Se diseñó la Ventana de Tiempo donde se conceptualiza la separa los buenos y malos clientes teniendo como criterio a los días de morosidad y que el cliente tenga compromiso de pago hasta la fecha de corte de la investigación, en seguida se realizó el modelo dimensional conformado por OSD\_DIM\_CREDITO, OSD\_DIM\_CLIENTE y OSD\_DIM\_SOLICITUD con las medidas: número de días de mora de una cuota a pagar en un mes, número de veces que el cliente demoró en pagar en determinados meses, antigüedad de días y meses que es cliente, saldo total actual y monto de interés. Posteriormente se elaboró el Diseño físico del Data Mart y se implementó el desarrollo del proceso ETL de las dimensiones y fact table con sus respectivas programaciones de los flujos de datos de cada una de las dimensiones definidas.
3. El diseño del modelo de calificación se elaboró a través del desarrollo de un modelo de regresión logística binario donde se identificó las variables que inciden en el cumplimiento o morosidad de los préstamos a los clientes, resultado las variables

de tipo categorización Grado Instrucción con coeficientes de 1.015685, - 0.177 y 2.07174 por cada tipo y Estado de Créditos con sus respectivos coeficiente 0.813856, 1.350995 y 3.172322 por tipo; también se descubrieron las variables escalares Veces de Mora Mayor a Cero con coeficiente de 0.420616 y el hábito de Pago 3 con coeficiente de 2.469862, cabe indicar que hay una constante de valor - 7.34252525.

4. Por cada cliente se determinó un puntaje de riesgo de morosidad, dicho puntaje se obtuvo mediante el modelo de regresión binaria, donde aquellos clientes que obtuvieron un puntaje menor o igual a 46.66 son de alto riesgo de poder cumplir sus compromisos de pago; los clientes que están en el rango de 46.66 y 50.42 puntos tiene la probabilidad del 50% de poder pagar, y los clientes que tiene más de 50.42 puntos tiene una alta probabilidad de que cumplan sus pagarés, en otros términos son clientes que no representan riesgos a la Micro Financiera Progreso. También se pudo pronosticar que el nivel de riesgo crediticio igual a 20.16 % de clientes de alto riesgo de incumplimiento de pago y por ende se tiene asegurado el 80.15% del capital prestado a sus clientes.

## RECOMENDACIONES Y/O SUGERENCIAS

1. Se sugiere al área de desarrollo y soporte del sistema de información de crédito de la Micro Financiera Progreso, a utilizar estándares de categorización de los datos de las entidades profesión, distrito y zona de residencia, similar o igual al que tienen la RENIEC o el INEI, debido a que no se encontró ninguna asociación de variable. Esto servirá para realizar futuras actualizaciones, dentro de un periodo máximo 6 meses, del modelo de calificación de riesgo.
2. Se recomienda a la Micro Financiera Progreso que los clientes que han sido identificado de alto riesgo ya no se les otorgue por ningún motivo nuevos créditos porque comprometen al capital a no ser recuperado. Así mismo por más bueno que sea el modelo de calificación nunca va a ser mejorado para futuras actualizaciones.
3. Se aconseja a la Micro Financiera Progreso, para que el presente estudio sirva como base de antecedente, para realizar estudios similares donde se diseñe e implemente algoritmos de asociación de variables categóricas y escalares en el lenguaje de programación que esté desarrollado el sistema de información de crédito, sin necesidad de tener dependencia con ningún software estadístico para poder minimizar el periodo de actualización del modelo.

## REFERENCIAS BIBLIOGRÁFICAS

- Asociación de Supervisores Bancarios de las Américas. (s.f.). *http://www.asba-supervision.org*. Obtenido de *http://www.asba-supervision.org/PEF/el-cliente-bancario/quien-es-el-cliente-bancario.shtml*
- Chambi Canahuire, J., & Arias Huamani, W. (15 de 02 de 2014). Curso Taller Minería de Datos. Lima, Lima, Perú.
- Chapman, P. J., Clinton, R. K., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. CRISP-DM consortium.
- Córdova, A. (2009). *Análisis de Riesgo de la Cartera de Microcrédito de la Cooperativa San Francisco Ltda. Sucursal Salcedo para la Toma de Decisiones*. Ambato.
- Del Valle Córdova, E. E. (1998). *Crédito y Cobranzas*. México: Universidad Nacional Autónoma de México.
- Duran Lopez., M. (11 de 07 de 2012). *buenastareas.com*. Obtenido de *http://www.buenastareas.com/ensayos/Factores-Biosocioculturales/4756375.html*
- Empresas. (29 de Julio de 2015). *Portafolio*. Obtenido de *http://www.portafolio.co/negocios/empresas/mineria-datos-mundo-b2b-35754*
- Figuerola, M. (2006). *Minería de Datos Aplicada a Credit Scoring*. QUITO: UNIVERSIDAD SAN FRANCISCO DE QUITO.
- Forex. (s.f.). *Diccionario forex*. Recuperado el 30 de 08 de 2017, de *https://efxto.com/diccionario/calificacion-de-credito*
- García Herrero, J., & Molina López, J. M. (2012). *Técnicas de análisis de datos*. Madrid: Universidad Carlos III de Madrid.
- Group, K. (31 de Diciembre de 2015). *Acerca de nosotros: Kimball Group*. Obtenido de *http://www.kimballgroup.com/*
- Gutiérrez Montes, A. (20 de Febrero de 2012). *buenastareas.com*. Obtenido de *http://www.buenastareas.com/ensayos/Factores-Economicos/4445572.html*
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. D. (2010). *Metodología de la investigación*. Mexico: MCGRAW-HILL INTERAMERICANA EDITORES S.A. DE C.V.
- Inmon, B. (1998). Data Marts Data Warehouse. *DM Review*.

- Juste, C. A. (10 de Noviembre de 2015). *economipedia.com*. Obtenido de <http://economipedia.com/definiciones/calidad-crediticia.html>
- Laudon, K. C., & Laudon, J. P. (2012). *Sistemas de información gerencial* (Décimosegunda ed.). México: Pearson.
- Micro Financiera Progreso. (2015). *Resumen Económico*. Cajamarca: Progreso.
- Micro Financiera Progreso. (2016). *Manual de políticas para el otorgamiento y recuperación de créditos*. Cajamarca: Progreso.
- Morales Castro, J. A., & Morales Castro, A. (2014). *Crédito y cobranza*. México: Grupo Editorial PATRIA.
- Moreno García, M., Miguel Quintales, L., García Peñalvo, F., & Polo Martín, J. (2009). *Aplicación de técnicas de minería de datos en la construcción y validación de modelos predictivos y asociativos a partir de especificaciones de requisitos de software*. Salamanca - España: Universidad de Salamanca Departamento de Informática y Automática.
- Peralta Cochancela, D. E. (2009). *Proyecto de minería de datos para el análisis del comportamiento de los clientes de telecomunicaciones*. Cuenca: Universidad Politecnica Salesiana.
- Rayo Cantón, S., Lara rubio, J., & Camino Blasco, D. (2010). *Un Modelo de Credit scoring para instituciones de Microfinanzas en el marco de Basilea II*. Lima Perú: Journal of Economics, Finance and Administrative Science.
- Robbins, S. P., & Coulter, M. (2005). *Administración* (Vol. 8 Edición). México: Pearson Educación.
- Sinergia e Inteligencia de Negocio. (Enero de 2007-2016). *Sinnexus*. Obtenido de [http://www.sinnexus.com/business\\_intelligence/datamart.aspx](http://www.sinnexus.com/business_intelligence/datamart.aspx)
- Stair, R. M., & Reynolds, G. W. (2010). *Principios de Sistemas de Información un Enfoque Administrativo* (novena ed.). México: CENGAGE Learning.
- Valenzuela Ramírez, I. E. (2009). *La Profundidad de Alcance en un contexto de comercialización de las microfinanzas en el Perú*. Lima: Pontificia Universidad Católica del Perú.
- Velasquez, J. D., & Pelade, V. (2008). *Adaptive Web site: A KnowledgeExtraction from Web Data Approach*. IOS Press.

## **APÉNDICES**

**APÉNDICE A.** Implementación de la base de datos temporal

Solo se muestra las sentencias SQL más importantes como son la creación de la base de datos y de la tabla de crédito.

```
USE [master]
```

```
GO
```

```
/****** Object: Database [STG] Script Date: 19/12/2017 15:47:16 *****/
```

```
CREATE DATABASE [STG]
```

```
CONTAINMENT = NONE
```

```
ON PRIMARY
```

```
( NAME = N'STG', FILENAME = N'C:\Program Files\Microsoft SQL
```

```
Server\MSSQL11.MSSQLSERVER\MSSQL\DATA\STG.mdf' , SIZE = 55296KB ,
```

```
MAXSIZE = UNLIMITED, FILEGROWTH = 1024KB )
```

```
LOG ON
```

```
( NAME = N'STG_log', FILENAME = N'C:\Program Files\Microsoft SQL
```

```
Server\MSSQL11.MSSQLSERVER\MSSQL\DATA\STG_log.ldf' , SIZE = 20096KB ,
```

```
MAXSIZE = 2048GB , FILEGROWTH = 10%)
```

```
GO
```

```
ALTER DATABASE [STG] SET COMPATIBILITY_LEVEL = 110
```

```
GO
```

```
IF (1 = FULLTEXTSERVICEPROPERTY('IsFullTextInstalled'))
```

```
begin
```

```
EXEC [STG].[dbo].[sp_fulltext_database] @action = 'enable'
```

end

/\*\*\*\*\* Object: Table [dbo].[q] Script Date: 19/12/2017 15:47:17 \*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE TABLE [dbo].[q](

[cod\_credito] [int] NULL,

[Mora 2017-1] [int] NULL,

[Mora 2017-2] [int] NULL,

[Mora 2017-3] [int] NULL,

[Mora 2017-4] [int] NULL,

[Mora 2017-5] [int] NULL,

[Mora 2017-6] [int] NULL

) ON [PRIMARY]

/\*\*\*\*\* Object: Table [dbo].[STG\_Credito] Script Date: 19/12/2017 15:47:17  
\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

SET ANSI\_PADDING ON

GO

CREATE TABLE [dbo].[STG\_Credito](

[codigoCredito] [int] NULL,  
[codigoCliente] [int] NULL,  
[codigoEstadoCredito] [int] NULL,  
[codigoDetalleProductoCredito] [int] NULL,  
[codigoSubTipoCredito] [int] NULL,  
[descripcionDestinoCredito] [varchar](500) NULL,  
[montoOriginalCredito] [decimal](18, 2) NULL,  
[montoCredito] [decimal](18, 2) NULL,  
[numeroCuotas] [int] NULL,  
[montoCuota] [decimal](30, 20) NULL,  
[tasaInteresNominal] [decimal](18, 8) NULL,  
[comisionFlat] [decimal](18, 15) NULL,  
[fechaDesembolso] [date] NULL,  
[fechaPrimerVencimiento] [date] NULL,  
[observacionCredito] [varchar](1000) NULL,  
[codigoCreacionEmpleado] [int] NULL,  
[fechaHoraCreacion] [datetime] NULL,  
[codigoModificacionEmpleado] [int] NULL,  
[fechaHoraModificacion] [datetime] NULL,  
[codigoBajaEmpleado] [int] NULL,  
[fechaHoraBaja] [datetime] NULL,  
[motivoBaja] [varchar](500) NULL,  
[igvCredito] [decimal](18, 8) NULL,  
[factorActualizacion] [float] NULL,  
[numeroCuotasOriginal] [int] NULL,  
[direccionCliente] [varchar](500) NULL,  
[codigoZonaCliente] [int] NULL,  
[codigoLugarCliente] [int] NULL,  
[codigoConyugue] [int] NULL,  
[tasaInteresMoraDiaria] [decimal](18, 10) NULL,  
[codigoCreditoAntiguo] [varchar](8) NULL,  
[estadoMigrado] [bit] NULL,

[codigoTipoCreditoAntiguo] [varchar](2) NULL,  
[codigoFuenteAntiguo] [varchar](2) NULL,  
[montoPreCancelacion] [decimal](18, 2) NULL,  
[fechaHoraMigracion] [datetime] NULL,  
[usuarioMigracion] [varchar](150) NULL,  
[fechaPagoTotal] [date] NULL,  
[fechaOperacionPagoTotal] [datetime] NULL,  
[contabilidadEstado] [char](1) NULL,  
[estadoReprogramado] [bit] NULL,  
[fechaReprogramacion] [date] NULL,  
[fechaOperacionReprogramacion] [datetime] NULL,  
[montoAmortizacionReprogramacion] [decimal](18, 2) NULL,  
[tipoReprogramacion] [varchar](2) NULL,  
[diasAtrasadosReprogramacion] [int] NULL,  
[diasAdicionalesReprogramacion] [int] NULL,  
[codigoSubCreditoReprogramado] [int] NULL,  
[interesReprogramacion] [decimal](18, 8) NULL,  
[igvReprogramacion] [decimal](18, 8) NULL,  
[moraReprogramacion] [decimal](18, 8) NULL,  
[cobranzaReprogramacion] [decimal](18, 8) NULL,  
[codigoCreditoReprogramado] [int] NULL,  
[estadoPermisoReproRefinLimitacion] [bit] NULL,  
[comentarioPermisoReproRefinLimitacion] [varchar](250) NULL,  
[comentarioReprogramacion] [varchar](250) NULL,  
[fechaCastigo] [date] NULL,  
[fechaOperacionCastigo] [datetime] NULL,  
[usuarioCastigo] [varchar](150) NULL,  
[comentarioCastigo] [varchar](5000) NULL,  
[observacionJefeCredito] [varchar](5000) NULL,  
[estadoNuevaAsignacion] [bit] NULL,  
[estable] [bit] NULL,  
[montoCreditoOriginalRepro] [decimal](18, 2) NULL,

```
[codigoCreditoOriginalRepro] [int] NULL,  
[estadoMigradoRepro] [bit] NULL,  
[acumuladoCapitalDetalle] [decimal](18, 5) NULL,  
[tasaEfectivaAnual] [decimal](10, 2) NULL,  
[bitDesencadenador] [bit] NULL,  
[bitDesencadenadorPago] [bit] NULL,  
[observacionReprogramacionEspecial] [varchar](500) NULL,  
[estadoPermisoReproRefinMoraAnterior] [bit] NULL,  
[comentarioPermisoReproMoraAnterior] [varchar](250) NULL,  
[estadoMontoDiferido] [bit] NULL,  
[LOAD_DATE] [date] NULL  
) ON [PRIMARY]
```

**APÉNDICE B.** Tablas complementarias de la categorización de variables

**Tabla de contingencia Sexo por Tipo de Cliente**

		Tipo de Cliente		Total	Tasa de Malos
		Buenos	Malos		
Sexo	MASCULINO	119	31	150	21%
	FEMENINO	65	12	77	16%
Total		184	43	227	19%

	Sexo_Cat	Porcentaje de Clientes Malos	Peso de Categoría
*+ Riesgo	Masculino	21%	1
*-Riesgo	Femenino	16%	0

**Tabla de contingencia Estado Civil por Tipo Cliente**

		Tipo de Cliente		Total	Tasa de Malos
		Buenos	Malos		
EstadoCivil	Conviviente	34	12	46	26%
	Soltero	105	26	131	20%
	Casado	39	5	44	11%
	Divorciado	2	0	2	0%
	Viudo	4	0	4	0%
Total		184	43	227	

	EstadoCivil_Cat	Porcentaje de Clientes Malos	Peso de Categoría
*+ Riesgo	Conviviente	26%	4
	Soltero	20%	3
	Casado	11%	2
	Divorciado	0%	1
*-Riesgo	Viudo	0%	0

**Tabla de contingencia Distrito\_Cat \* TipoCliente**

		Tipo de Cliente		Total	Tasa de Malos
		Buenos	Malos		
Distrito_Cat	Namora	2	0	2	0%
	Jesús	2	0	2	0%
	Llacanora	1	0	1	0%
	Encañada	6	0	6	0%
	Cajamarca	170	40	210	19%
	Los baños del inca	2	2	4	50%
	Asuncion	1	1	2	50%
Total		184	43	227	

	Distrito_Cat	Porcentaje de Clientes Malos	Peso de Categoría
*- Riesgo	Namora	0%	0
	Jesús	0%	1
	Llacanora	0%	2
	Encañada	0%	3
	Cajamarca	19%	4
	Los baños del inca	50%	5
*+Riesgo	Asunción	50%	6

**Tabla de contingencia Zona de Residencia por Tipo de Cliente**

		Tipo de Cliente		Total	Tasa de Malos
		Buenos	Malos		
Zona	El molino	0	1	1	100%
	Horacio zevallos	0	1	1	100%
	La merced	0	1	1	100%
	Pj. Maria parado de belloido	0	1	1	100%
	Shudal	0	1	1	100%
	Urba el molino	0	1	1	100%
	Urubamba	0	1	1	100%
	Villa universitaria	0	1	1	100%
	San Martin de porras	1	2	3	67%
	Baños del inca	3	2	5	40%
	La florida	3	2	5	40%
	Mollepampa	2	1	3	33%
	Cajamarca	115	28	143	20%

Agopampa baja	1	0	1	0%
Ajoscancha	1	0	1	0%
Aranjuez	1	0	1	0%
Tartar grande	2	0	2	0%
Tres molinos	1	0	1	0%
Urb. Los jazmines	1	0	1	0%
Calispuquio	3	0	3	0%
Cas. Huacariz	1	0	1	0%
Chamis	1	0	1	0%
Chinchimarca	1	0	1	0%
Chontapaccha	1	0	1	0%
Cruz blanca	1	0	1	0%
Cumbe mayo	1	0	1	0%
Dos de mayo	1	0	1	0%
El estanco	2	0	2	0%
Encañada	3	0	3	0%
Huambocancha	1	0	1	0%
Jesus	4	0	4	0%
La encañada	1	0	1	0%
La molina	2	0	2	0%
La paccha	5	0	5	0%
La tulpuna	1	0	1	0%
Lot santa rosa	1	0	1	0%
Lotizacion columbo	1	0	1	0%
Lucmacucho	1	0	1	0%
Lucmacucho alto	1	0	1	0%
Marcopampa	3	0	3	0%
Miraflores	1	0	1	0%
Mollepampa alta	3	0	3	0%
Mollepampa baja	2	0	2	0%
Moyococha	1	0	1	0%
Namora	1	0	1	0%
P.j. magna vallejo	1	0	1	0%
P.j. v.r. haya de la torre	1	0	1	0%
Polloc	1	0	1	0%
San Martin	2	0	2	0%
San sebastian	2	0	2	0%
Santa barbara	1	0	1	0%
Santa elena	1	0	1	0%
<b>Total</b>	<b>184</b>	<b>43</b>	<b>227</b>	<b>19%</b>

**Tabla de contingencia GradoInstrucción\_Cat \* TipoCliente**

		Tipo Cliente		Total	Tasa de Malos
		Buenos	Malos		
GradoInstrucción_Cat	Técnica o superior	71	14	85	16%
	Secundaria	75	15	90	17%
	Primaria	22	3	25	12%
	Sin instrucción	16	11	27	41%
Total		184	43	227	

	GradoInstruccion_Cat	Porcentaje de Clientes Malos	Peso de Categoría
*- Riesgo	Técnica o superior	16%	0
	Secundaria	17%	1
	Primaria	12%	2
*+Riesgo	Sin instrucción	41%	3

**Tabla de contingencia Profesión por Tipo de Cliente**

		Tipo de Cliente		Total	Tasa de Malos
		Buenos	Malos		
Profesión	(Otros)	33	13	46	28%
	Administrador	2	0	2	0%
	Administrativo	0	2	2	100%
	Agricultor	1	0	1	0%
	Agrónomo	1	1	2	50%
	Alquiler de vivienda	2	1	3	33%
	Ayudante general	1	0	1	0%
	Carpintería	3	0	3	0%
	Chofer de carga pesada	3	0	3	0%
	Chofer de combi	1	0	1	0%
	Chofer de moto taxi	1	1	2	50%
	Chofer de vehiculo liviano	5	0	5	0%
	Comerciante	41	5	46	11%
	Conductor de taxi	2	0	2	0%

Conductor moto taxi	4	1	5	20%
Consultora de productos de belleza	1	0	1	0%
Contadora	1	0	1	0%
Empleado	13	1	14	7%
Empresario	1	0	1	0%
Enfermero(a)	0	1	1	100%
Estibador	2	0	2	0%
Ganadero	5	1	6	17%
Maestro de obra	25	2	27	7%
Mecánico automotriz	1	0	1	0%
Mecánico de soldadura	2	0	2	0%
Moto taxista	1	0	1	0%
Null	15	10	25	40%
Obrero	1	0	1	0%
Obrero de construcción	3	0	3	0%
Operador	0	1	1	100%
Operario de construcción	1	0	1	0%
Panadería	0	1	1	100%
Policía nacional del Perú	2	0	2	0%
Profesor	3	1	4	25%
Proveedor de leche	2	0	2	0%
Servicios generales	1	0	1	0%
Transporte publico	1	1	2	50%
Venta de carne	1	0	1	0%
Vigilante	2	0	2	0%
<b>Total</b>	<b>184</b>	<b>43</b>	<b>227</b>	<b>19%</b>

**Tabla de contingencia EstadoCredito\_Cat \* TipoCliente**

		Tipo de Cliente		Total	Tasa de Malos
		Buenos	Malos		
EstadoCredito_Cat	Crédito pagado con precancelación	69	2	71	3%
	Crédito pagado regularmente	29	6	35	17%
	Crédito creado (activo)	81	27	108	25%
	Crédito reprogramado (refinanciamiento) finalizado	5	8	13	62%
Total		184	43	227	

	EstadoCredito_Cat	Porcentaje de Clientes Malos	Peso de Categoría
*- Riesgo	Crédito pagado con precancelacion	3%	0
	Crédito pagado regularmente	17%	1
	Crédito creado (activo)	25%	2
*+Riesgo	Crédito reprogramado (refinanciamiento) finalizado	62%	3

**Tabla de contingencia AntigüedadMeses (agrupado) \*  
Tipo de Cliente**

		Tipo de Cliente		Total	Tasa de Malos	Agrupación
		Buenos	Malos			
Antigüedad Meses (agrupado)	<= 26	21	9	30	30%	18%
	27 - 28	17	3	20	15%	
	29 - 30	21	1	22	5%	
	31 - 32	19	4	23	17%	21%
	33 - 33	21	2	23	9%	
	34 - 34	15	7	22	32%	
	35 - 36	18	6	24	25%	
	37 - 37	13	5	18	28%	17%
	38 - 42	19	6	25	24%	
	43+	20	0	20	0%	
Total		184	43	227	19%	

	AntigüedadMeses_Cat	Porcentaje de Clientes Malos	Peso de Categoría
*-Riesgo	<=dos años y medio	18%	1
*+Riesgo	dos año y medio - 3 años	21%	2
*-Riesgo	> 3 años	17%	0

**Tabla de contingencia MaximoDíasMora6Meses (agrupado) \* TipoCliente**

		Tipo de Cliente		Total	Tasa de Malos
		Buenos	Malos		
MáximoDíasMora6Meses (agrupado)	<= 0	25	0	25	0%
	1 - 1	26	0	26	0%
	2 - 3	18	0	18	0%
	4 - 6	23	0	23	0%
	7 - 12	23	0	23	0%
	13 - 16	23	0	23	0%
	17 - 21	21	0	21	0%
	22 - 28	23	0	23	0%
	29 - 45	2	21	23	91%
	46+	0	22	22	100%
Total		184	43	227	19%

	MáximoDíasMora6Mes	Porcentaje de Clientes Malos	Peso de Categoría
*-Riesgo	0 hasta 30	0%	0
*+Riesgo	31 hasta 86	100%	1

**Tabla de contingencia PeorHabitopago\_6meses \* Tipo de Cliente**

		Tipo de Cliente		Total	Tasa de Malos
		Buenos	Malos		
PeorHabitopago_6meses	Normal	97	0	97	0%
	CPP	87	0	87	0%
	Deficiente	0	34	34	100%
	Dudoso	0	9	9	100%
Total		184	43	227	

	Peor hábito de pago	Porcentaje de Clientes Malos	Peso de Categoría
*-Riesgo	Normal	0%	0
	CPP	0%	1
	Deficiente	100%	2
*+Riesgo	Dudoso	100%	3

**Tabla de contingencia HabitopagoPromedio \* TipoCliente**

		Tipo de Cliente		Total	Tasa de Malos
		Buenos	Malos		
HábitopagoPromedio	Normal	105	0	105	0%
	CPP	79	19	98	19%
	Deficiente	0	21	21	100%
	Dudoso	0	3	3	100%
Total		184	43	227	

	Hábito de Pago Promedio	Porcentaje de Clientes Malos	Peso de Categoría
*-Riesgo	Normal	0%	0
	CPP	19%	1
	Deficiente	100%	2
*+Riesgo	Dudoso	100%	3

**Tabla de contingencia Flag Refinanciado \* TipoCliente**

		Tipo de Cliente		Total	Tasa de Malos
		Buenos	Malos		
FlagRefinanciado	No refinanciado	179	35	214	16%
	Si Refinanciado	5	8	13	62%
Total		184	43	227	

	Refinanciamiento	Porcentaje de Clientes Malos	Peso de Categoría
*-Riesgo	No refinanciado	16%	0
*+Riesgo	Si Refinanciado	62%	1

**Tabla de contingencia SaldoTotal (agrupado) \* TipoCliente**

		Tipo de Cliente		Total	Tasa de Malos	Agrupación
		Buenos	Malos			
Saldo Total (agrupado)	<= 215,0600	19	4	23	17%	12%
	215,0601 - 327,4800	22	1	23	4%	
	327,4801 - 683,1800	20	3	23	13%	
	683,1801 - 908,6700	17	5	22	23%	24%
	908,6701 - 1266,1300	15	8	23	35%	
	1266,1301 - 1613,5100	20	3	23	13%	
	1613,5101 - 2208,4400	17	5	22	23%	21%
	2208,4401 - 2917,9000	16	7	23	30%	
	2917,9001 - 4333,4900	18	5	23	22%	
	4333,4901+	20	2	22	9%	
Total		184	43	227		

	SaldoTotal_Cat	Porcentaje de Clientes Malos	Peso de Categoría
*-Riesgo	77.43 hasta 683.1800	12%	0
*+Riesgo	683.1801 hasta 1613.51	24%	2
*-Riesgo	1613,5101 hasta 10656,3500	21%	1

**Tabla de contingencia MontoInteres (agrupado) \* Tipo de Cliente**

		Tipo de Cliente		Total	Tasa de Malos	Agrupación
		Buenos	Malos			
Monto Interés (agrupado)	<= 326,8000	20	3	23	13%	18.8%
	326,8001 - 448,1000	17	6	23	26%	
	448,1001 - 572,8200	19	4	23	17%	
	572,8201 - 738,9000	18	4	22	18%	19.1%
	738,9001 - 970,3000	20	3	23	13%	
	970,3001 - 1210,2000	17	6	23	26%	
	1210,2001 - 1475,5000	16	6	22	27%	18.9%
	1475,5001 - 1866,7000	22	1	23	4%	
	1866,7001 - 2501,3800	17	6	23	26%	
	2501,3801+	18	4	22	18%	
<b>Total</b>		<b>184</b>	<b>43</b>	<b>227</b>		

	<b>Monto de Interes_Cat</b>	<b>Porcentaje de Clientes Malos</b>	<b>Peso de Categoría</b>
*-Riesgo	85.3 hasta 572.82	18.84%	0
*+Riesgo	Mayor 572.82 hasta 1210.2	19.12%	2
*-Riesgo	Mayor 1210.2 hasta 9083.87	18.89%	1

**Tabla de contingencia Promedio Pagos (agrupado) \* Tipo de Cliente**

		Tipo de Cliente		Total	Tasa de Malos	Agrupación
		Buenos	Malos			
Promedio Pagos (agrupado)	<= 152,2143	20	3	23	13%	15.9%
	152,2144 - 206,3429	17	6	23	26%	
	206,3430 - 231,7286	21	2	23	9%	
	231,7287 - 267,1000	18	4	22	18%	17.6%
	267,1001 - 294,4333	20	3	23	13%	
	294,4334 - 342,2538	18	5	23	22%	
	342,2539 - 419,1700	17	5	22	23%	22.2%
	419,1701 - 501,3071	18	5	23	22%	
	501,3072 - 656,2667	18	5	23	22%	
	656,2668+	17	5	22	23%	
Total		184	43	227	19%	

	Promedio de Pagos_Cat	Porcentaje de Clientes Malos	Peso de Categoría
*-Riesgo	84.89 hasta 231.7286	15.9%	0
	> 231.7286 hasta 342.2538	17.6%	1
*+Riesgo	>342.2538 hasta 1752.90	22.2%	2

**APÉNDICE C.** Corridas para encontrar las variables en la ecuación del modelo.

**Variables en la ecuación del modelo**

Variables en la ecuación	Categoría/Nivel	B	E.T.	Wald	gl	Sig.	Exp(B)	Elegido
Grado de Instrucción	TECNICA o SUPERIOR			11.042	3	.012		
	Secundaria	.984	.570	2.977	1	.084	2.674	
	Primaria	.357	.861	.172	1	.679	1.429	
	Sin instrucción	2.412	.744	10.512	1	.001	11.154	SI
Estado de Crédito	Crédito pagado con precancelacion			12.033	3	.007		
	Crédito pagado regularmente	2.609	1.058	6.085	1	.014	13.581	SI
	Crédito creado (activo)	1.628	.812	4.020	1	.045	5.095	
	Crédito reprogramado (refinanciamiento) finalizado	3.432	1.050	10.675	1	.001	30.925	
Saldo Total	Desde 77.43 hasta 683.1800			5.400	2	.067		
	Desde 683.1801 hasta 1613.51	1.427	.840	2.885	1	.089	4.167	
	Desde 1613,5101 hasta 10656,3500	2.292	.994	5.314	1	.550	9.899	NO
Monto de Interés	Desde 85.3 hasta 572.82			5.796	2	.055		
	Mayor a 572.82 hasta 1210.2	-.859	.722	1.413	1	.235	.424	
	Mayor a 1210.2 hasta 9083.87	-1.883	.814	5.346	1	.021	.152	SI
Hábito de Pago 1		1.256	.446	7.950	1	.005	3.513	SI
Num Veces Mayor a Cero		.677	.209	10.474	1	.001	1.968	SI
Constante		-8.189	1.537	28.378	1	.000	.000	SI

2.4233

### Variables en la ecuación del modelo

Variables en la ecuación: Paso 6a	Categoría/Nivel	B	E.T.	Wald	gl	Sig.	Exp(B)	Elegido
Grado de Instrucción	Técnica o Superior			11.042	3	.012		SI
	Secundaria	-2.412	.744	10.512	1	.001	.090	
	Primaria	-1.428	.700	4.168	1	.041	.240	
	Sin instrucción	-2.055	.949	4.694	1	.030	.128	
Estado de Crédito	Crédito pagado con precancelación			12.033	3	.007		SI
	Crédito pagado regularmente	-3.432	1.050	10.675	1	.001	.032	
	Crédito creado (activo)	-.823	1.027	.642	1	.423	.439	
	Crédito reprogramado (refinanciamiento) finalizado	-1.803	.787	5.245	1	.022	.165	
Saldo Total	Desde 77.43 hasta 683.1800			5.400	2	.067		NO
	Desde 683.1801 hasta 1613.51	.865	.582	2.207	1	.137	2.376	
	Desde 1613,5101 hasta 10656,3500	-1.427	.840	2.885	1	.089	.240	
Monto de Interés	Desde 85.3 hasta 572.82			5.796	2	.055		NO
	Mayor a 572.82 hasta 1210.2	-1.024	.585	3.069	1	.080	.359	
	Mayor a 1210.2 hasta 9083.87	.859	.722	1.413	1	.235	2.360	
VecesMoraMayora0		.677	.209	10.474	1	.001	1.968	SI
HabitoPago1_Cat		1.256	.446	7.950	1	.005	3.513	SI
Constante		-1.777	1.451	1.500	1	.221	.169	NO

-11.797

**APÉNDICE C.** Algoritmo en SPSS para la construcción del modelo

```
***-----
```

```
***Construcción de variables categóricas
```

```
***-----
```

```
*****EDAD
```

```
if (edad<=38) Edad_Cat=3.
```

```
if (edad>=39 and edad<=45) Edad_Cat=2.
```

```
if (edad>=46 and edad<=52) Edad_Cat=1.
```

```
if (edad>=53) Edad_Cat=0.
```

```
execute.
```

```
value labels Edad_Cat
```

```
0 'Desde 53 años a Mas'
```

```
1 'Desde 46 hasta 52 años'
```

```
2 'Desde 39 hasta 45 años'
```

```
3 'Desde 21 hasta 38 años'
```

```
***SEXO
```

```
if Sexo='FEMENINO' Sexo_Cat=0.
```

```
if Sexo='MASCULINO' Sexo_Cat=1.
```

```
execute.
```

```
value labels Sexo_Cat
```

```
1 'MASCULINO'
```

```
0 'FEMENINO'
```

```
***Estado Civil
```

```
if EstadoCivil='VIUDO' EstadoCivil_Cat=0.
```

```
if EstadoCivil='DIVORCIADO' EstadoCivil_Cat=1.
```

```
if EstadoCivil='CASADO' EstadoCivil_Cat=2.  
if EstadoCivil='SOLTERO' EstadoCivil_Cat=3.  
if EstadoCivil='CONVIVIENTE' EstadoCivil_Cat=4.
```

```
execute.
```

```
value labels EstadoCivil_Cat
```

```
0 'VIUDO'
```

```
1 'DIVORCIADO'
```

```
2 'CASADO'
```

```
3 'SOLTERO'
```

```
4 'CONVIVIENTE'
```

```
*** Distrito
```

```
if Distrito='ASUNCION' Distrito_Cat=6.
```

```
if Distrito='LOS BAÑOS DEL INCA' Distrito_Cat=5.
```

```
if Distrito='CAJAMARCA' Distrito_Cat=4.
```

```
if Distrito='ENCAÑADA' Distrito_Cat=3.
```

```
if Distrito='LLACANORA' Distrito_Cat=2.
```

```
if Distrito='JESUS' Distrito_Cat=1.
```

```
if Distrito='NAMORA' Distrito_Cat=0.
```

```
execute.
```

```
value labels Distrito_Cat
```

```
6 'ASUNCION'
```

```
5 'LOS BAÑOS DEL INCA'
```

```
4 'CAJAMARCA'
```

```
3 'ENCAÑADA'
```

```
2 'LLACANORA'
```

1 'JESUS'

0 'NAMORA'

\*\*\*Grado de Instrucción

if GradoInstruccion='SIN INSTRUCCION (ILETRADO)' GradoInstruccion\_Cat=3.

if GradoInstruccion='NULL' GradoInstruccion\_Cat=3.

if GradoInstruccion='PRIMARIA' GradoInstruccion\_Cat=2.

if GradoInstruccion='SECUNDARIA' GradoInstruccion\_Cat=1.

if GradoInstruccion='TECNICA' GradoInstruccion\_Cat=0.

if GradoInstruccion='SUPERIOR' GradoInstruccion\_Cat=0.

execute.

value labels GradoInstruccion\_Cat

3 'SIN INSTRUCCION'

2 'PRIMARIA'

1 'SECUNDARIA'

0 'TECNICA O SUPERIOR'

\*\*\*Estado de Credito

if EstadoCredito='CREDITO REPROGRAMADO (REFINANCIAMIENTO)  
FINALIZADO' EstadoCredito\_Cat=3.

if EstadoCredito='CREDITO CREADO (ACTIVO)' EstadoCredito\_Cat=2.

if EstadoCredito='CREDITO PAGADO REGULARMENTE' EstadoCredito\_Cat=1.

if EstadoCredito='CREDITO PAGADO CON PRECANCELACION'  
EstadoCredito\_Cat=0.

execute.

value labels EstadoCredito\_Cat

3 'CREDITO REPROGRAMADO (REFINANCIAMIENTO) FINALIZADO'

2 'CREDITO CREADO (ACTIVO)'

1 'CREDITO PAGADO REGULARMENTE'

0 'CREDITO PAGADO CON PRECANCELACION'

\*\*\*ANTIGUEDAD

if AntiguedadMeses<=30 AntiguedadMeses\_Cat=1.

if AntiguedadMeses>=31 and AntiguedadMeses<=36 AntiguedadMeses\_Cat=2.

if AntiguedadMeses>=37 AntiguedadMeses\_Cat=0.

execute.

value labels AntiguedadMeses\_Cat

1 '<=dos años y medio'

2 '> dos año y medio - 3 años'

0 '> 3 años '

\*\*\*RANGO MORA MAXIMA (ULTIMOS 6 MESES)

IF (MaximoDiasMora6meses<=30) RangoMoraMaxima=0.

IF (MaximoDiasMora6meses>=31) RangoMoraMaxima=1.

EXECUTE.

value labels RangoMoraMaxima

0 '0 hasta 30 días'

1 '31 hasta 86 días'

\*\*\*\*\*Habito de Pago

if HabitoPago1='Normal' HabitoPago1\_Cat=0.

if HabitoPago1='CPP' HabitoPago1\_Cat=1.

if HabitoPago1='Deficiente' HabitoPago1\_Cat=2.

if HabitoPago1='Dudoso' HabitoPago1\_Cat=3.

if HabitoPago1='Perdido' HabitoPago1\_Cat=4.

EXECUTE.

if HabitoPago2='Normal' HabitoPago2\_Cat=0.

if HabitoPago2='CPP' HabitoPago2\_Cat=1.

if HabitoPago2='Deficiente' HabitoPago2\_Cat=2.

if HabitoPago2='Dudoso' HabitoPago2\_Cat=3.

if HabitoPago2='Perdido' HabitoPago2\_Cat=4.

EXECUTE.

if HabitoPago3='Normal' HabitoPago3\_Cat=0.

if HabitoPago3='CPP' HabitoPago3\_Cat=1.

if HabitoPago3='Deficiente' HabitoPago3\_Cat=2.

if HabitoPago3='Dudoso' HabitoPago3\_Cat=3.

if HabitoPago3='Perdido' HabitoPago3\_Cat=4.

EXECUTE.

if HabitoPago4='Normal' HabitoPago4\_Cat=0.

if HabitoPago4='CPP' HabitoPago4\_Cat=1.

if HabitoPago4='Deficiente' HabitoPago4\_Cat=2.

if HabitoPago4='Dudoso' HabitoPago4\_Cat=3.

if HabitoPago4='Perdido' HabitoPago4\_Cat=4.

EXECUTE.

if HabitoPago5='Normal' HabitoPago5\_Cat=0.

if HabitoPago5='CPP' HabitoPago5\_Cat=1.

if HabitoPago5='Deficiente' HabitoPago5\_Cat=2.

if HabitoPago5='Dudoso' HabitoPago5\_Cat=3.

if HabitoPago5='Perdido' HabitoPago5\_Cat=4.

EXECUTE.

if HabitoPago6='Normal' HabitoPago6\_Cat=0.

if HabitoPago6='CPP' HabitoPago6\_Cat=1.

if HabitoPago6='Deficiente' HabitoPago6\_Cat=2.

if HabitoPago6='Dudoso' HabitoPago6\_Cat=3.

if HabitoPago6='Perdido' HabitoPago6\_Cat=4.

EXECUTE.

COMPUTE

PeorHabitoPago\_6meses=MAX(HabitoPago1\_Cat,HabitoPago2\_Cat,HabitoPago3\_Cat,  
HabitoPago4\_Cat,HabitoPago5\_Cat,HabitoPago6\_Cat).

EXECUTE.

value labels PeorHabitoPago\_6meses

0 'Normal'

1 'CPP'

2 'Deficiente'

3 'Dudoso'

\*\*\*\*Habito de Pago Promedio

compute

HabitoPagoPromedio=SUM(HabitoPago1\_Cat,HabitoPago2\_Cat,HabitoPago3\_Cat,Hab  
itoPago4\_Cat,HabitoPago5\_Cat,HabitoPago6\_Cat)/6.

execute.

compute

HabitoPagoPromedio=RND(HabitoPagoPromedio).

execute.

value labels HabitoPagoPromedio

0 'Normal'

1 'CPP'

2 'Deficiente'

3 'Dudoso'

\*\*\*\*Refinanciamiento

value labels FlagRefinanciado

0 'No refinanciado'

1 'Si Refinanciado'

\*\*\*Saldo Total

if SaldoTotal<=683.18 SaldoTotal\_Cat=0.

if SaldoTotal>683.18 and SaldoTotal<=1613.51 SaldoTotal\_Cat=2.

if SaldoTotal>1613.51 SaldoTotal\_Cat=1.

execute.

value labels SaldoTotal\_Cat

0 '77.43 hasta 683.1800'

2 '683.1801 hasta 1613.51'

1 '1613,5101 hasta 10656,3500'

\*\*Monto de Interes

if MontoInteres<= 572.82 MontoInteres\_Cat=0.

if MontoInteres>572.82 and MontoInteres<=1210.2 MontoInteres\_Cat=2.

if MontoInteres>1210.2 MontoInteres\_Cat=1.

execute.

value labels MontoInteres\_Cat

0 '85.3 hasta 572.82'

2 '> 572.82 hasta 1210.2'

1 '>1210.2 hasta 9083.87'

\*\*Promedio Pagos

if PromedioPagos<= 231.7286 PromedioPagos\_Cat=0.

if PromedioPagos>231.7286 and PromedioPagos<= 342.2538 PromedioPagos\_Cat=1.

if PromedioPagos> 342.2538 PromedioPagos\_Cat=2.

execute.

value labels PromedioPagos\_Cat

0 '84.89 hasta 231.7286'

1 '> 231.7286 hasta 342.2538'

2 '>342.2538 hasta 1752.90'

\*\*\*

\*Estimar las variables para el modelo

\*\*\*\_-----

\*\*\* Selección de las variables en el modelo de regresión logístico

\*\*\*\_-----

\*\*\*:

LOGISTIC REGRESSION VARIABLES Flag\_Cliente

/METHOD=BSSTEP(WALD) Edad\_Cat GradoInstruccion\_Cat EstadoCredito\_Cat

SaldoTotal\_Cat MontoInteres\_Cat VecesMoraMayora0 HabitoPago3\_Cat

/CONTRAST (Edad\_Cat)=Indicator

/CONTRAST (GradoInstruccion\_Cat)=Indicator(1)

/CONTRAST (EstadoCredito\_Cat)=Indicator(1)

/CONTRAST (SaldoTotal\_Cat)=Indicator(1)

/CONTRAST (MontoInteres\_Cat)=Indicator(1)

```

/SAVE=PRED PGROUP COOK LEVER DFBETA RESID LRESID SRESID
ZRESID DEV

/CLASSPLOT

/PRINT=GOODFIT CORR ITER(1)

/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.4).

if GradoInstruccion_Cat=1 GradoInstruccion_Cat_1 =1.

if GradoInstruccion_Cat=2 GradoInstruccion_Cat_2 =1.

if GradoInstruccion_Cat=3 GradoInstruccion_Cat_3 =1.

execute.

if EstadoCredito_Cat=1 EstadoCredito_Cat_1=1.

if EstadoCredito_Cat=2 EstadoCredito_Cat_2=1.

if EstadoCredito_Cat=3 EstadoCredito_Cat_3=1.

execute.

Recode GradoInstruccion_Cat_1 GradoInstruccion_Cat_2 GradoInstruccion_Cat_3
      EstadoCredito_Cat_1 EstadoCredito_Cat_2 EstadoCredito_Cat_3
      (MISSING=0).

execute.

***Calculo de la probabilidad calculada (con los parámetros estimados)

compute Probabilidad_Calculada=1/(1+exp((-1)*(
1.015685*GradoInstruccion_Cat_1          -          0.177*GradoInstruccion_Cat_2
+2.07174*GradoInstruccion_Cat_3
+ 0.813856*EstadoCredito_Cat_1          +          1.350995*EstadoCredito_Cat_2          +
3.172322*EstadoCredito_Cat_3
+0.420616*VecesMoraMayora0 + 2.469862*HabitoPago3_Cat - 7.34252525))).

execute.

```

```
compute LN_ODDS=ln(Probabilidad_Calculada/(1-Probabilidad_Calculada)).
```

```
execute.
```

```
***score final:
```

```
compute SCORE_FINAL=LN_ODDS*7.2134752+16.78072.
```

```
execute.
```

```
compute SCORE_FINAL=60-SCORE_FINAL.
```

```
execute.
```

```
if SCORE_FINAL>100 SCORE_FINAL=100.
```

```
if SCORE_FINAL<0 SCORE_FINAL=0.
```

```
execute.
```

## APÉNDICE E. Validación y llenado del cuestionario

### CUESTIONARIO I

El presente cuestionario tiene por objetivo, determinar la satisfacción del usuario con respecto a la información procesada por la minería de datos del Sistema de Información del Créditos de la Microfinanciera Progreso antes de aplicar minería de Datos.

### INSTRUCCIONES

Lea las preguntas y/o afirmaciones que a continuación se muestran e indicar una respuesta, según el nivel de acuerdo o desacuerdo que tenga.

1. ¿Las sentencias SQL utilizadas para obtener el Data Mart del Sistema de Créditos y los códigos de programación utilizados en SPSS se ejecutan sin ningún inconveniente?
 

a) Siempre	b) Casi siempre	c) A veces
d) Casi nunca	<input checked="" type="checkbox"/> Nunca	
  
2. ¿El modelo de minera de datos procesa todos los datos almacenados por el Data Mart del Sistema de Crédito?
 

a) Siempre	b) Casi siempre	c) A veces
d) Casi nunca	<input checked="" type="checkbox"/> Nunca	
  
3. ¿El modelo de minería de datos del Sistema de Créditos describe la asociación de la variable dependiente tipo de cliente con respecto a las variables cualitativas que intervienen en el cálculo del puntaje de cumplimiento de pagos de los clientes?
 

a) Siempre	b) Casi siempre	c) A veces
d) Casi nunca	<input checked="" type="checkbox"/> Nunca	
  
4. ¿El modelo de minería de datos del Sistema de Créditos describe la asociación de la variable dependiente tipo de cliente con respecto a las variables cuantitativas que intervienen en el cálculo del puntaje de cumplimiento de pagos de los clientes?
 

a) Siempre	b) Casi siempre	c) A veces
d) Casi nunca	<input checked="" type="checkbox"/> Nunca	

5. ¿El modelo de minería de datos del Sistema de Créditos utiliza las variables cualitativas y cuantitativas en la regresión logística binaria para el cálculo del puntaje de cumplimiento de pagos de los clientes?
- a) Siempre                      b) Casi siempre                      c) A veces
- d) Casi nunca                       Nunca
6. ¿El modelo de minería de datos del Sistema de Créditos está validada correctamente con otra muestra?
- a) Siempre                      b) Casi siempre                      c) A veces
- d) Casi nunca                       Nunca
7. ¿El resultado que calcula el puntaje de cumplimiento de pagos de los clientes a través del modelo de minería de datos del Sistema de Créditos es fácil de interpretar para tomar decisiones en la predicción del otorgamiento de préstamos de los buenos clientes?
- a) Siempre                      b) Casi siempre                      c) A veces
- d) Casi nunca                       Nunca

  
Ing. Walter Pérez Estrada  
DNI 45589949  
Área de Sistemas

## CUESTIONARIO I

El presente cuestionario tiene por objetivo, determinar la satisfacción del usuario con respecto a la información procesada por la minería de datos del Sistema de Información del Créditos de la Microfinanciera Progreso después de aplicar minería de Datos.

### INSTRUCCIONES

Lea las preguntas y/o afirmaciones que a continuación se muestran e indicar una respuesta, según el nivel de acuerdo o desacuerdo que tenga.

1. ¿Las sentencias SQL utilizadas para obtener el Data Mart del Sistema de Créditos y los códigos de programación utilizados en SPSS se ejecutan sin ningún inconveniente?  
a) Siempre                       b) Casi siempre                      c) A veces  
d) Casi nunca                      e) Nunca
2. ¿El modelo de minería de datos procesa todos los datos almacenados por el Data Mart del Sistema de Crédito?  
a) Siempre                       b) Casi siempre                      c) A veces  
d) Casi nunca                      e) Nunca
3. ¿El modelo de minería de datos del Sistema de Créditos describe la asociación de la variable dependiente tipo de cliente con respecto a las variables cualitativas que intervienen en el cálculo del puntaje de cumplimiento de pagos de los clientes?  
 a) Siempre                      b) Casi siempre                      c) A veces  
d) Casi nunca                      e) Nunca
4. ¿El modelo de minería de datos del Sistema de Créditos describe la asociación de la variable dependiente tipo de cliente con respecto a las variables cuantitativas que intervienen en el cálculo del puntaje de cumplimiento de pagos de los clientes?  
 a) Siempre                      b) Casi siempre                      c) A veces  
d) Casi nunca                      e) Nunca

5. ¿El modelo de minería de datos del Sistema de Créditos utiliza las variables cualitativas y cuantitativas en la regresión logística binaria para el cálculo del puntaje de cumplimiento de pagos de los clientes?
- a) Siempre                      b) Casi siempre                      c) A veces
- d) Casi nunca                      e) Nunca
6. ¿El modelo de minería de datos del Sistema de Créditos está validada correctamente con otra muestra?
- a) Siempre                      b) Casi siempre                      c) A veces
- d) Casi nunca                      e) Nunca
7. ¿El resultado que calcula el puntaje de cumplimiento de pagos de los clientes a través del modelo de minería de datos del Sistema de Créditos es fácil de interpretar para tomar decisiones en la predicción del otorgamiento de préstamos de los buenos clientes?
- a) Siempre                       b) Casi siempre                      c) A veces
- d) Casi nunca                      e) Nunca



Ing. Walter Pérez Estrada

45589949

## CUESTIONARIO I

El presente cuestionario tiene por objetivo, determinar la satisfacción del usuario con respecto a la información procesada por la minería de datos del Sistema de Información del Créditos de la Microfinanciera Progreso antes de aplicar minería de Datos.

### INSTRUCCIONES

Lea las preguntas y/o afirmaciones que a continuación se muestran e indicar una respuesta, según el nivel de acuerdo o desacuerdo que tenga.

1. ¿Las sentencias SQL utilizadas para obtener el Data Mart del Sistema de Créditos y los códigos de programación utilizados en SPSS se ejecutan sin ningún inconveniente?
 

a) Siempre	b) Casi siempre	c) A veces
<del>d) Casi nunca</del>	e) Nunca	
  
2. ¿El modelo de minería de datos procesa todos los datos almacenados por el Data Mart del Sistema de Crédito?
 

a) Siempre	b) Casi siempre	c) A veces
<del>d) Casi nunca</del>	e) Nunca	
  
3. ¿El modelo de minería de datos del Sistema de Créditos describe la asociación de la variable dependiente tipo de cliente con respecto a las variables cualitativas que intervienen en el cálculo del puntaje de cumplimiento de pagos de los clientes?
 

a) Siempre	b) Casi siempre	c) A veces
d) Casi nunca	<del>e) Nunca</del>	
  
4. ¿El modelo de minería de datos del Sistema de Créditos describe la asociación de la variable dependiente tipo de cliente con respecto a las variables cuantitativas que intervienen en el cálculo del puntaje de cumplimiento de pagos de los clientes?
 

a) Siempre	b) Casi siempre	<del>c) A veces</del>
d) Casi nunca	e) Nunca	

5. ¿El modelo de minería de datos del Sistema de Créditos utiliza las variables cualitativas y cuantitativas en la regresión logística binaria para el cálculo del puntaje de cumplimiento de pagos de los clientes?
- a) Siempre                      b) Casi siempre                      c) A veces
- d) Casi nunca                      ~~e) Nunca~~
6. ¿El modelo de minería de datos del Sistema de Créditos está validada correctamente con otra muestra?
- a) Siempre                      b) Casi siempre                      c) A veces
- d) Casi nunca                      ~~e) Nunca~~
7. ¿El resultado que calcula el puntaje de cumplimiento de pagos de los clientes a través del modelo de minería de datos del Sistema de Créditos es fácil de interpretar para tomar decisiones en la predicción del otorgamiento de préstamos de los buenos clientes?
- a) Siempre                      b) Casi siempre                      c) A veces
- d) Casi nunca                      ~~e) Nunca~~



Ing. Gilmer Cayotopa Jara

## CUESTIONARIO I

El presente cuestionario tiene por objetivo, determinar la satisfacción del usuario con respecto a la información procesada por la minería de datos del Sistema de Información del Créditos de la Microfinanciera Progreso después de aplicar minería de Datos.

### INSTRUCCIONES

Lea las preguntas y/o afirmaciones que a continuación se muestran e indicar una respuesta, según el nivel de acuerdo o desacuerdo que tenga.

1. ¿Las sentencias SQL utilizadas para obtener el Data Mart del Sistema de Créditos y los códigos de programación utilizados en SPSS se ejecutan sin ningún inconveniente?  
a) Siempre                      b) Casi siempre                      c) A veces  
d) Casi nunca                      e) Nunca
2. ¿El modelo de minería de datos procesa todos los datos almacenados por el Data Mart del Sistema de Crédito?  
a) Siempre                      b) Casi siempre                      c) A veces  
d) Casi nunca                      e) Nunca
3. ¿El modelo de minería de datos del Sistema de Créditos describe la asociación de la variable dependiente tipo de cliente con respecto a las variables cualitativas que intervienen en el cálculo del puntaje de cumplimiento de pagos de los clientes?  
a) Siempre                      b) Casi siempre                      c) A veces  
d) Casi nunca                      e) Nunca
4. ¿El modelo de minería de datos del Sistema de Créditos describe la asociación de la variable dependiente tipo de cliente con respecto a las variables cuantitativas que intervienen en el cálculo del puntaje de cumplimiento de pagos de los clientes?  
a) Siempre                      b) Casi siempre                      c) A veces  
d) Casi nunca                      e) Nunca

5. ¿El modelo de minería de datos del Sistema de Créditos utiliza las variables cualitativas y cuantitativas en la regresión logística binaria para el cálculo del puntaje de cumplimiento de pagos de los clientes?
- ~~a) Siempre~~                      b) Casi siempre                      c) A veces  
d) Casi nunca                      e) Nunca
6. ¿El modelo de minería de datos del Sistema de Créditos está validada correctamente con otra muestra?
- a) Siempre                      b) Casi siempre                      c) A veces  
d) Casi nunca                      e) Nunca
7. ¿El resultado que calcula el puntaje de cumplimiento de pagos de los clientes a través del modelo de minería de datos del Sistema de Créditos es fácil de interpretar para tomar decisiones en la predicción del otorgamiento de préstamos de los buenos clientes?
- ~~a) Siempre~~                      b) Casi siempre                      c) A veces  
d) Casi nunca                      e) Nunca



Ins. Gilmer Cayotola J.

**SOLICITUD DE VALIDACIÓN DIRIGIDA A EXPERTO EN TÉCNICAS  
ESTADÍSTICAS DE MINERÍA DE DATOS**

Cajamarca, 9 de enero del 2018

**Docente del Departamento Sistemas Estadística e Informática - FI - UNC**

**Mg. Lic. Enzo Bravo Burgos**

Presente

Reciba un cordial saludo

Motivado a su reconocida formación en el tema de técnicas estadísticas en Minería de Datos, me complace dirigirme a usted en solicitud de su valiosa colaboración para la validación de los cuestionarios que anexo, los mismos que servirán para recolectar información relativa a la investigación denominada: **Efecto de la implementación de minería de datos en la calificación de crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca**, que será presentado para optar el Grado de Maestro en Ciencias.

Asimismo, anexo el Instrumento para la validación de los cuestionarios y el cuadro: De Variables e Indicadoras para una rápida comprensión y validación de los instrumentos.

Agradeciendo su valiosa colaboración en el desarrollo e impulso de la investigación, me suscribo.

Muy Cordialmente,

**Ing. Jaime Amador Meza Huamán**  
DNI.26720960

## INSTRUMENTO PARA LA VALIDACIÓN

### Cuestionario I

Dirigido al docente del Departamento de Sistemas Estadística e Informática de la  
Universidad Nacional de Cajamarca

APRECIACIÓN CUALITATIVA				
CRITERIOS	Excelente	Bueno	Regular	Deficiente
Presentación del Instrumento	X			
Claridad en la redacción de los ítems	X			
Pertinencia de la variable con los indicadores	X			
Relevancia del contenido	X			
Factibilidad de la aplicación		X		

Observaciones:

Validado por: ENZO ALDO BRAVO BURGOS	DNI. N°: 17834976
Profesión: LIC. ESTADÍSTICA	
Lugar de Trabajo: UNIVERSIDAD NACIONAL DE CAJAMARCA.	
Cargo que desempeña: DOCENTE.	
Lugar y fecha de validación: 09-01-2018	
Firma:	 ..... <b>Lic. ENZO ALDO BRAVO BURGOS</b> COE SPE N° 931 <b>Colegio de Estadísticos del Perú</b>

**SOLICITUD DE VALIDACIÓN DIRIGIDA A EXPERTOS DE MINERÍA DE DATOS**

Cajamarca, 9 de enero del 2018

**Director de la Escuela Académico Profesional de Ingeniería de Sistemas**

**Mg. Ing. Carlos Enrique Aparicio Arteaga**

Presente

Reciba un cordial saludo

Motivado a su reconocida formación en el tema de Minería de Datos, me complace dirigirme a usted en solicitud de su valiosa colaboración para la validación de los cuestionarios que anexo, los mismos que servirán para recolectar información relativa a la investigación denominada: **Efecto de la implementación de minería de datos en la calificación de crédito a los clientes de la Micro Financiera Progreso de la ciudad de Cajamarca**, que será presentado para optar el Grado de Maestro en Ciencias.

Asimismo, anexo el Instrumento para la validación de los cuestionarios y el cuadro: De Variables e Indicadoras para una rápida comprensión y validación de los instrumentos.

Agradeciendo su valiosa colaboración en el desarrollo e impulso de la investigación, me suscribo.

Muy Cordialmente,

**Ing. Jaime Amador Meza Huamán**  
DNI.26720960

## INSTRUMENTO PARA LA VALIDACIÓN

### Cuestionario I

Dirigido al Director de la Escuela Académico Profesional de Ingeniería de Sistemas de la  
Universidad Nacional de Cajamarca

CRITERIOS	APRECIACIÓN CUALITATIVA			
	Excelente	Bueno	Regular	Deficiente
Presentación del Instrumento	X			
Claridad en la redacción de los ítems	X			
Pertinencia de la variable con los indicadores	X			
Relevancia del contenido	X			
Factibilidad de la aplicación	X			

Observaciones:

Validado por: CARLOS APARICIO ARICAGA	DNI. N°: 18071268
Profesión: ING. DE COMPUTACIÓN Y SISTEMAS / ING. SISTEMAS	
Lugar de Trabajo: UNIVERSIDAD NACIONAL DE CAJAMARCA	
Cargo que desempeña: DIRECTOR DE LA EAP. DE INGENIERÍA DE SISTEMAS	
Lugar y fecha de validación: Cajamarca, 9 de mayo	
Firma:	