



**Desarrollo de un modelo predictivo de morosidad de cartera de la línea de fianzas  
de una Empresa de Seguros**

Carvajal Castro, Natalia Fernanda

Vicerrectorado De Investigación, Innovación Y Transferencia Tecnológica

Centro de Posgrados

Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios

Trabajo de titulación, previo a la obtención del título de Magíster en Sistemas de  
Información e Inteligencia de Negocios

MSc. Jaramillo Vinuesa, Byron Orlando

15 de octubre del 2020



## Document Information

---

<b>Analyzed document</b>	Tesis_Natalia_Carvajal-urkund.docx (D84986816)
<b>Submitted</b>	11/12/2020 5:34:00 PM
<b>Submitted by</b>	Gualotuña Alvarez Tatiana Marisol
<b>Submitter email</b>	tmgualotunia@espe.edu.ec
<b>Similarity</b>	2%
<b>Analysis address</b>	tmgualotunia.espe@analysis.orkund.com

## Sources included in the report

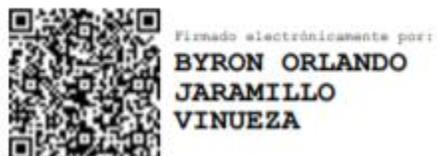
---

<b>SA</b>	<p><b>Universidad de las Fuerzas Armadas ESPE / TESIS_versionfinal01082019.doc</b>  Document TESIS_versionfinal01082019.doc (D54660967)  Submitted by: lgduque@espe.edu.ec  Receiver: lgduque.espe@analysis.orkund.com</p>
<b>SA</b>	<p><b>Universidad de las Fuerzas Armadas ESPE / TESIS KARINA MOLINA - ANALISIS URKUND.docx</b>  Document TESIS KARINA MOLINA - ANALISIS URKUND.docx (D79429885)  Submitted by: tmgualotunia@espe.edu.ec  Receiver: tmgualotunia.espe@analysis.orkund.com</p>
<b>SA</b>	<p><b>GARCES CAMACHO HANS SEBASTIAN - FINAL.docx</b>  Document GARCES CAMACHO HANS SEBASTIAN - FINAL.docx (D64066538)</p>
<b>SA</b>	<p><b>Universidad de las Fuerzas Armadas ESPE / Tesis Hugo Fiallios.docx</b>  Document Tesis Hugo Fiallios.docx (D60032137)  Submitted by: ERGOMEZ@ESPE.EDU.EC  Receiver: ergomez.espe@analysis.orkund.com</p>
<b>W</b>	<p>URL: <a href="https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/650407/Alarc%C3%B3n...">https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/650407/Alarc%C3%B3n...</a>  Fetched: 7/7/2020 12:04:34 AM</p>
<b>W</b>	<p>URL: <a href="https://docplayer.es/789942-Modelo-para-la-automatizacion-del-proceso-de-determina...">https://docplayer.es/789942-Modelo-para-la-automatizacion-del-proceso-de-determina...</a>  Fetched: 1/31/2020 6:17:42 PM</p>
<b>W</b>	<p>URL: <a href="https://www.analyticsvidhya.com/blog/2020/02/gartners-2020-magic-quadrant-for-data...">https://www.analyticsvidhya.com/blog/2020/02/gartners-2020-magic-quadrant-for-data...</a>  Fetched: 11/12/2020 11:18:00 PM</p>
<b>SA</b>	<p><b>Universidad de las Fuerzas Armadas ESPE / TesisFinal_FabianOrellana.docx</b>  Document TesisFinal_FabianOrellana.docx (D79194781)  Submitted by: frorellana3@espe.edu.ec  Receiver: vcparraga.espe@analysis.orkund.com</p>
<b>W</b>	<p>URL: <a href="https://info.microsoft.com/ww-landing-2020-gartner-magic-quadrant-for-analytics-an...">https://info.microsoft.com/ww-landing-2020-gartner-magic-quadrant-for-analytics-an...</a>  Fetched: 11/12/2020 11:18:00 PM</p>
<b>W</b>	<p>URL: <a href="https://www.academia.edu/25033602/INDICADORES_DE_GESTI%C3%93N_DE_COBRANZAS_COBRANZ...">https://www.academia.edu/25033602/INDICADORES_DE_GESTI%C3%93N_DE_COBRANZAS_COBRANZ...</a>  Fetched: 11/12/2020 11:18:00 PM</p>

Instances where selected sources appear:

23

Firma:



Msc. Jaramillo Vinueza, Byron Orlando  
**C.C.: 171455255**



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y  
TRANSFERENCIA DE TECNOLOGÍA  
CENTRO DE POSGRADOS**

**CERTIFICACIÓN**

Certifico que el trabajo de titulación, “**Desarrollo de un modelo predictivo de morosidad de cartera de la línea de fianzas de una Empresa de Seguros**” fue realizado por la señora **Carvajal Castro, Natalia Fernanda** el mismo que ha sido revisado y analizado en su totalidad, por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

**Sangolquí, 15 de octubre 2020**

Firma:



Firmado electrónicamente por:

**BYRON ORLANDO  
JARAMILLO  
VINUEZA**

Msc. Jaramillo Vinueza, Byron Orlando  
**C.C.: 1714555255**



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA  
DE TECNOLOGÍA  
CENTRO DE POSGRADOS**

**RESPONSABILIDAD DE AUTORÍA**

Yo **Carvajal Castro, Natalia Fernanda**, con cédula de ciudadanía n°1718346370, declaro que el contenido, ideas y criterios del trabajo de titulación: **Desarrollo de un modelo predictivo de morosidad de cartera de la línea de fianzas de una Empresa de Seguros** es de mí autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

**Sangolquí, 15 de octubre 2020**

Firma

**Carvajal Castro, Natalia Fernanda  
C.C.: 1718346370**



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA  
DE TECNOLOGÍA  
CENTRO DE POSGRADOS**

**AUTORIZACIÓN DE PUBLICACIÓN**

Yo **Carvajal Castro, Natalia Fernanda** autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **Desarrollo de un modelo predictivo de morosidad de cartera de la línea de fianzas de una Empresa de Seguros** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

**Sangolquí, 15 de octubre 2020**

Firma

Carvajal Castro, Natalia Fernanda  
C.C.:1718346370

## **Agradecimiento**

A Dios por su bondad, por la vida y todas las bendiciones que me ha brindado a lo largo de mi vida.

A mi esposo, por su apoyo, acompañamiento y paciencia incondicional.

A mis padres por guiar mi camino y estar presentes en cada momento de mi vida.

A mi director de tesis por su profesionalismo y guía en el desarrollo de este trabajo.

## **Dedicatoria**

A Dios, mi esposo, mis padres y mi familia, quienes son un pilar fundamental en mi vida.

A mi familia, quienes siempre están pendientes de cada paso que doy en mi vida.

## Tabla de Contenidos

Urkund.....	2
Certificación.....	4
Responsabilidad de autori.....	5
Autorización de publicación .....	6
Agradecimiento .....	7
Dedicatoria .....	8
Tabla de Contenidos .....	9
Índice de Tablas .....	11
Índice de Figuras.....	12
Resumen.....	13
Abstract.....	14
Capítulo 1 Generalidades .....	15
Antecedentes .....	15
Planteamiento del Problema.....	17
Objetivos de la Investigación .....	19
Objetivo General.....	19
Objetivos Específicos.....	19
Justificación, Importancia y Alcance del Proyecto.....	20
Preguntas de investigación.....	21
Hipótesis.....	22
Capítulo 2 Marco Teórico .....	23
Diseño de Investigación Científica (Desing Science Research,DSR.).....	23
Descubrimiento de Conocimiento en Base de Datos .....	23
Marco Teórico Referencial.....	24
Categorización de la Variable Independiente.....	24
Fundamentación de la Variable Independiente: .....	24
Técnicas de análisis de datos .....	24
Minería de Datos.....	25
Clasificación.....	26
Modelos Predictivos.....	26
Árbol de clasificación (variable por variable).....	27
Redes Neuronales.....	27
Deep Learning .....	27
Redes de Bayes .....	28
Categorización de la Variable Dependiente.....	28
Empresas de Seguros.....	29
Gerencia de Fianzas.....	29
Indicadores Comerciales.....	30
Indicadores de Cartera.....	30
Estado del Arte .....	31
Definición de la Problemática.....	31
Objetivo de la Investigación .....	32
Preguntas de Investigación.....	32
Importancia de Realizar el Estudio.....	32
Criterios de Inclusión y Exclusión.....	33
Criterios de Inclusión.....	33
Criterios de Exclusión.....	33
Grupos de Control.....	34
Cadena de Búsqueda.....	35

Proceso de Selección de Artículos.....	36
Síntesis de Resultados. ....	38
Conclusiones .....	40
Capítulo 3 Marco Metodológico .....	42
Selección de Metodología .....	42
Enfoque de la investigación .....	42
Categoría de la Investigación.....	43
Metodología.....	43
Descripción de cada fase.....	46
Definición del problema.....	46
Sugerencia.....	46
Desarrollo. ....	47
Interpretación y Evaluación. ....	49
Evaluación. ....	49
Conclusión. ....	49
Capítulo 4 Planteamiento de la Solución .....	50
Planteamiento de la Solución .....	50
Análisis del problema desde la perspectiva tecnológica.....	50
Infraestructura.....	50
Análisis de Plataformas de Data Science and Machine Learning.....	53
Análisis de Plataformas de Data Analytics and Business Intelligence .....	63
Alcance de la Solución.....	67
Capítulo 5 Desarrollo de la Investigación.....	69
Planteamiento de la Solución .....	69
Metodología KDD .....	69
Abstracción del escenario. ....	69
Selección de los datos. ....	70
Preprocesamiento(Limpieza de Datos). ....	72
Transformación de los Datos. ....	76
Minería de datos .....	80
División de la Base de Datos en entrenamiento y prueba .....	82
Aplicación de Algoritmos.....	83
Árbol de decisión .....	89
Redes neuronales (Deep learning).....	92
Redes de Bayes.....	95
Evaluación. ....	96
Capítulo 6 Pruebas y Evaluación con el Negocio .....	103
Análisis de Resultados.....	103
Evaluación de Resultados .....	104
Estrategias Generales .....	110
Estrategias Específicas.....	112
Capítulo 7 Conclusiones y Recomendaciones .....	115
Conclusiones y Recomendaciones .....	115
Conclusiones .....	115
Recomendaciones .....	117
Próximos pasos. ....	117
Lista de referencias .....	119
Bibliografía .....	119
Anexos .....	119

## Índice de tablas

Tabla 1 Grupos de Control .....	34
Tabla 2 Cadena de Búsqueda .....	35
Tabla 3 Cadena Seleccionada.....	36
Tabla 4 Selección de Artículos .....	37
Tabla 5 Listado de Artículos Primarios .....	38
Tabla 6 Ventajas y Desventajas de Herramientas de Data Science and Machine Lea ..	55
Tabla 7 Características de RapidMiner y Knime .....	58
Tabla 8 Productos de las Herramientas de Data Science and Machine Learning .....	59
Tabla 9 <i>Capacidades de Productos Seleccionados de Data Science</i> .....	60
Tabla 10 Comparativo de Experiencia al cliente Knime vs RapidMiner.....	61
Tabla 12 Comparativo de Plataformas de Análisis e Inteligencia Empresarial .....	64
Tabla 13 Variables del Reporte de Cartera.....	71
Tabla 14 Score de Clientes .....	74
Tabla 15 Variables Eliminadas .....	75
Tabla 16 Variables Creadas .....	76
Tabla 17 <i>Ventajas y Desventajas de los Modelos Seleccionados</i> .....	81
Tabla 18 Rangos de la Correlación .....	85
Tabla 19 Matriz de Confusión .....	98
Tabla 20 Guía de Interpretación Medidas de Precisión y Exactitud .....	100
Tabla 21 Calificación de Algoritmos.....	101
Tabla 22 Métricas Evaluadas por el Negocio.....	103
Tabla 23 Calificación de la empresa.....	104

## Índice de Figuras

Figura 1 Diagrama de Ishikawa .....	19
Figura 2. Variables Independientes .....	24
Figura 3. Variables Dependientes.....	28
Figura. 4 Diseño de Investigación.....	45
Figura 5. Relación del Diseño con los Objetivos.....	46
Figura 6. Metodología KDD .....	48
Figura 7. Infraestructura Actual .....	50
Figura 8. Cuadrante Mágico de Gartner de Data Science and Machine Learning.....	54
Figura 9. Comparativo Knime vs RapidMiner.....	62
Figura 10. Plataformas de Análisis e Inteligencia Empresarial.....	64
Figura 11. Arquitectura Tecnológica Propuesta .....	66
Figura. 12 Función usada en R .....	73
Figura. 13 Valores Perdidos .....	74
Figura. 14 Días Vencidos .....	79
Figura. 15 Estructura de la Data .....	80
Figura. 16 División de la Base .....	83
Figura. 17 División de la Base en Entrenamiento y Prueba .....	83
Figura. 18 Conexión MySql .....	84
Figura. 19 Matriz de Correlación.....	86
Figura. 20 Clúster Jerárquico .....	87
Figura. 21 Función Clúster Jerárquico .....	88
Figura. 22 Proceso de Modelamiento .....	89
Figura. 23 Parámetros del Árbol de Decisión.....	89
Figura. 24 Árbol de Decisión .....	90
Figura. 25 Parámetros Redes Neuronales.....	92
Figura. 26 Extracto de Modelo Deep Learning.....	94
Figura. 27 Importancia de Variables .....	95
Figura. 28 Parámetro de Redes de Bayes.....	96
Figura 29 Cross Validation .....	97
Figura. 30 Conexión Tableau .....	104
Figura. 31 Dashboard Pantalla 1 .....	105
Figura. 32 Gráfico Prima vs Ramo vs Endoso .....	106
Figura.33 Clientes Morosos por Sucursal .....	107
Figura. 34 Fecha vs Prima.....	108
Figura. 35 Dashboard.....	109
Figura. 36 Resultado de Nuevo Riesgo y Renovaciones .....	110

## Resumen

Un modelo predictivo permite extraer patrones de comportamiento para identificar oportunidades y riesgos. Las nuevas tecnologías de la información brindan a las empresas la oportunidad de optimizar sus recursos y tomar decisiones acertadas. El uso de técnicas de minería de datos apalancadas en las diferentes herramientas existentes en el mercado, facilitan el desarrollo de modelos predictivos que den respuestas a los interrogantes y necesidades de las compañías. El presente estudio combina la metodología de Diseño de investigación científica (Desing Science Research, DSR) con el Descubrimiento de Conocimiento en Bases de Datos o KDD (Knowledge Discovery in Databases), para aplicarla en el área de cartera de una línea de negocios de una Empresa de Seguros. A través de este trabajo se seleccionó un modelo predictivo que se ajuste a las necesidades de la compañía con el objetivo de optimizar los recursos disponibles para mejorar la gestión de cobranzas. Se demostró que el uso de herramientas de análisis y minería de datos son aliados de las compañías ya que a través de su uso se puede mejorar los resultados obtenidos por los ejecutivos y los mandos jerárquicos pueden tener un mayor control del estado de la cartera de la línea de negocio; además la creación de un dashboard proporciona a la línea de negocio una visión global de su cartera para que pueda anticiparse estableciendo estrategias y tomando decisiones de forma acertada.

Palabras Clave:

- **EMPRESA DE SEGUROS**
- **MODELO PREDICTIVO**
- **CARTERA**

## Abstract

A predictive model allows the extraction of behavior patterns to identify opportunities and risks. The new information technologies allow companies to optimize their resources and make the right decisions. The use of data mining techniques leveraged in the different tools on the market facilitates the development of predictive models that provide answers to companies' questions and needs. The present study combines the Design Science Research methodology with Knowledge Discovery in Databases or KDD (Knowledge Discovery in Databases) to be applied in the portfolio area of a company insurance business line. A predictive model was selected through this work that meets the company's needs to optimize the available resources to improve collections management. It was demonstrated that the use of analysis and data mining tools are allies of companies since, through their use, executives' results can be improved, and hierarchical managers can have greater control of the company's portfolio. In addition to creating a dashboard, a line of business provides a global vision of its portfolio to anticipate establishing strategies and making the right decisions.

Keywords:

- **INSURANCE COMPANY**
- **PREDICTIVE MODEL**
- **PURSE**

## Capítulo 1 Generalidades

### Antecedentes

Durante los años 2015 y 2016, las empresas de seguros sufrieron una contracción que se evidenció en su producción y en el número de compañías existentes en el mercado. A partir del año 2017 el sector asegurador empezó a mejorar, teniendo en el 2017 un crecimiento del 0,7% y en el 2018 del 3,6% (Zabala, Ranking de Seguros, 2019).

A partir de la aprobación del nuevo código monetario y las nuevas resoluciones emitidas por la Superintendencia de Compañías, se realizaron ajustes en las reservas de Riesgos en Curso – Primas no Devengadas, en la cual las empresas de seguros deben tomar el 80% de la prima neta retenida para las pólizas de vigencia anual y el 50% para las pólizas con cobertura menor a un año en los ramos de vehículos y accidentes personales. Se debe tomar el 75% y el 45% respectivamente para el cálculo de esta reserva (Calderón, 2019). Es decir: si una empresa asegura un vehículo cuya prima es de \$800 debe retener el 80% de este valor, es decir \$640 y contaría apenas con \$160 para cubrir gastos administrativos, comisiones, etc., los valores retenidos se van liberando mensualmente. Esta reforma obligó a las empresas de seguros a potenciar sus otros ramos para comercializar en el mercado ecuatoriano. La empresa decidió apostar a la línea de fianzas ya que se caracteriza por ser una de las más rentables y estructuradas de la compañía.

La línea de fianzas es una de las cinco líneas de negocio de la compañía. Cuenta con ejecutivos a nivel nacional y se encuentra estructurada de la siguiente manera: trece ejecutivos comerciales, uno de renovación, uno de recuperación de ejecuciones, uno de cartera, un jefe y la directora de la línea. Los ramos que comercializa fianzas son: Buen Uso del Anticipo, Fiel Cumplimiento de Contrato,

Seriedad de Oferta, Garantía Aduanera y Calidad de Materiales; siendo Buen uso del Anticipo y Fiel Cumplimiento del Contrato los que abarcan más del 80% de la producción. El presupuesto de la línea corresponde al 17% (Empresa de Seguros, 2019) del total del presupuesto de la empresa, lo cual le da una participación importante en la compañía.

Los clientes de la línea de fianzas son contratistas del estado, que, al adjudicarse un contrato, tienen por obligación presentar pólizas que garanticen que cumplirán con su contrato. Ellos se dedican principalmente a las siguientes actividades económicas: construcción, actividades de transporte, servicios profesionales, comercio, servicios de apoyo a las empresas, telecomunicaciones, actividades relacionadas con energía eléctrica y manufactura.

Los mejores meses en los cuales se incrementan las ventas de la línea es el último trimestre, sin embargo, el 2019 fue un año atípico, ya que las ventas no tuvieron el despunte que suelen tener; esto debido a que la contratación pública se contrajo, por la reducción del presupuesto general del estado para la inversión pública. Esto impactó directamente a la línea de negocio ya que el 90% de clientes de fianzas, corresponden a contratistas del estado. El resultado de estos factores fue el incumplimiento del presupuesto anual.

El proceso de una garantía o fianzas es diferente al de los seguros generales o vida. Este es un contrato en el cual se garantiza el cumplimiento de algo. Es un acuerdo suscrito entre tres partes. Las personas que intervienen en una fianza son: Beneficiario quien es el propietario del contrato y puede ser una persona natural, jurídica, pública o privada; quien busca garantizar que su contrato termine con éxito. En el caso de que el contratista incumpla busca tener un respaldo que le garantice que su contrato culmine con éxito. La Aseguradora, que es la empresa que se convierte en el garante del contratista, es decir el que se compromete con la otra parte y el Contratista quien tiene

como obligación cumplir con los parámetros estipulados en un contrato a favor de un tercero (beneficiario). Para que el contratista pueda acceder a la póliza de fianzas debe pasar por un proceso de calificación, en el cual la empresa de seguros evalúa su perfil, analiza su experiencia, respaldos financieros y solicita al contratista la entrega de toda la documentación que respalde su actividad.

El proceso de fianza consta de cinco pasos: Atención e información, calificación, emisión, renovación y liberación. Dentro de la renovación y la liberación se genera la cartera vencida de los clientes, ya que la renovación es automática una vez que lo solicite el cliente sin necesidad de solicitar el cobro de la misma.

El éxito de las empresas que ofertan los servicios de las fianzas es la rapidez, agilidad y buena suscripción para la emisión de las pólizas. Cabe mencionar que es necesario que cuenten con un proceso estandarizado y personal calificado que les permita responder de forma inmediata ante los requerimientos de sus clientes; además de implementar nuevas tecnologías de la información que les permita realizar su trabajo de una forma más eficiente y eficaz.

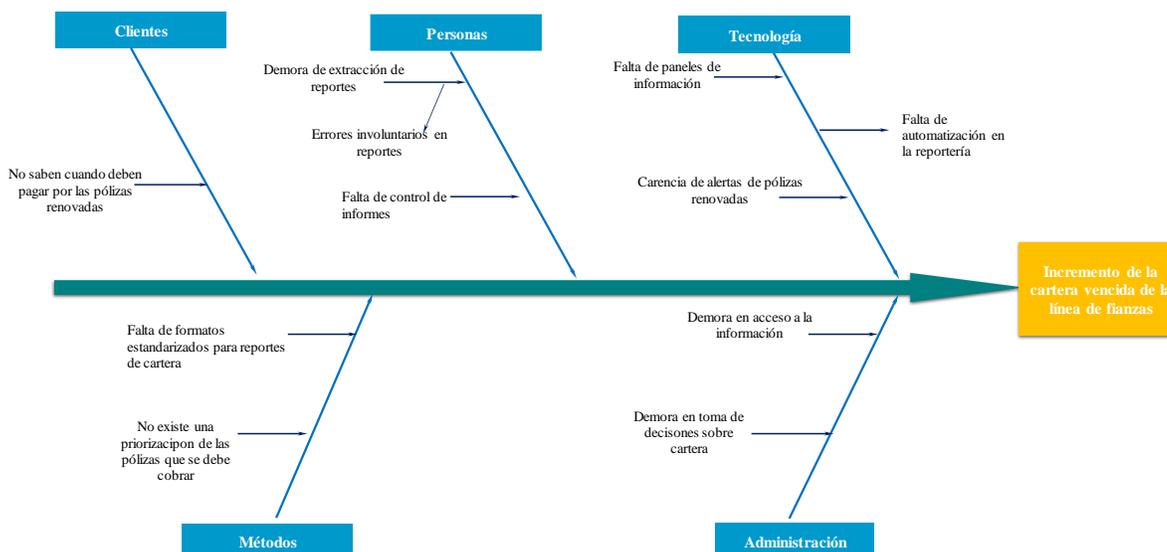
### **Planteamiento del Problema**

Todos los esfuerzos en mejorar la línea de fianzas se enfocaron en el proceso de comercialización. Mejorar los tiempos de respuesta y automatizar las actividades de calificación, permite a la línea tener mejores tiempos de respuesta, que le facultan ser más competitiva y atractiva en el mercado. Sin embargo, se dejó para una segunda y tercera fase los procesos de recuperación de cartera y contragarantías.

Actualmente la línea tiene una cartera vencida de 4 millones de dólares (Empresa de Seguros, 2019), lo que genera en impuestos un valor mayor a \$500.000 (Empresa de Seguros, 2019). Esto genera un gasto adicional para la compañía ya que al no originar ingresos de este rubro, debe desembolsar los valores correspondientes a sus obligaciones con el estado, por lo tanto reduce sus ingresos provocando una

disminución en su rentabilidad. Otro factor importante para tomar en cuenta en la cartera vencida de la línea de fianzas es la existencia de pólizas muy antiguas que se han ido renovando por años sin que nadie haga un proceso de recuperación adecuado. Estas pólizas representan un alto riesgo para la línea y la compañía ya que al haber sido renovadas indican que no se presentaron los documentos de liberación de la póliza o que el contrato no se concluyó, de manera que la convierte en una póliza con alto riesgo de ejecución. En caso de que una póliza de fianzas se ejecute, debe pagar los valores correspondientes al beneficiario de forma inmediata, de tal forma que puede reducir la liquidez de la empresa ya que pueden existir pólizas con altos montos de cobertura, lo que representa que las indemnizaciones sean por valores elevados.

La implementación de un modelo predictivo permitirá a la línea gestionar de una mejor manera su cartera vencida. Además, ahorrará tiempo a los ejecutivos de cartera, les permitirá centrarse en el objetivo de su puesto que es gestionar y recuperar los valores correspondientes de las pólizas que no se encuentran pagadas, en lugar de realizar reportes operativos para decidir a quién deben llamar para realizar el proceso de recuperación, de igual manera, a través del modelo podrán identificar aquellas pólizas que se encuentran más tiempo en cartera vencida para alertar a los ejecutivos comerciales de la línea de fianzas y tomar acciones que puedan evitar posibles ejecuciones. En la Figura 1 se muestra el diagrama de Ishikawa que permite visualizar un resumen de las principales causas del problema planteado:

**Figura 1***Diagrama de Ishikawa*

*Nota.* La figura es una representación visual del diagrama de causa - efecto basado en 5 aristas importantes para la línea.

## Objetivos de la Investigación

### **Objetivo General.**

Crear un modelo predictivo basado en técnicas de minería de datos que permita identificar el comportamiento de la morosidad de los clientes de la línea de fianzas para definir estrategias que permitan mejorar la gestión de cobro de las pólizas.

### **Objetivos Específicos.**

**OE1:** Identificar el estado actual de la cartera de la línea de fianzas para levantar los principales puntos de dolor y definir el principal problema que atraviesa la línea de fianzas.

**OE2:** Identificar las técnicas de modelamiento que son comúnmente empleadas en empresas de servicios para la recuperación de cartera a través de una revisión inicial de literatura.

**OE3:** Desarrollar un modelo predictivo de datos para identificar los patrones de comportamiento que generan un incremento de cartera de la línea de fianzas.

**OE4:** Seleccionar la mejor técnica de minería de datos en base a los mejores resultados obtenidos que permita definir las alertas para la recuperación de cartera vencida de la línea de fianza para el desarrollo de un dashboard que permita tener la información en un solo lugar.

**OE5:** Validar que el modelo seleccionado se adapta a las necesidades de la línea de fianzas para aportar en la reducción de la cartera de la línea de fianzas.

### **Justificación, Importancia y Alcance del Proyecto**

Los cambios en las estrategias de las empresas de seguros les han obligado a buscar nuevos nichos de mercado y no concentrarse únicamente en vehículos. Durante muchos años el principal seguro comercializado por las aseguradoras era el vehicular. Sin embargo, por las reformas establecidas, se vieron obligados a buscar nuevas alternativas que les permitan ser más competitivos y sobre todo rentables.

La línea de fianzas ha sido una de las mejores líneas de la empresa que gracias al profesionalismo de sus ejecutivos y líderes ha logrado conseguir importantes logros a lo largo de los años. Se ha mantenido dentro del top tres del ranking de las empresas de seguros en los ramos de fianzas y la rentabilidad generada por la línea ha aportado positivamente al cumplimiento de los objetivos de la compañía.

Los esfuerzos volcados hacia esta línea y la inversión realizada han dado frutos. No obstante, pese al repotenciamiento y los cambios generados en el proceso comercial, han quedado algunos temas pendientes por realizar. Uno de estos son las

mejoras en el área de cartera, un factor sumamente importante para completar el ciclo comercial de la línea y tener mejores indicadores tanto en rentabilidad como en siniestralidad.

Actualmente la cartera de fianzas asciende a los 4 millones. El 12,5% de este valor ha tenido que ser pagado al estado por concepto de impuestos de ley. Esto representa un problema a la empresa ya que debe desembolsar al estado un valor que aún no ha cobrado, lo cual provoca un impacto en sus indicadores de rentabilidad. La operatividad y la extracción de reportes de forma manual, ha hecho que la recuperación de cartera en la línea de fianzas sea un factor crítico. Es importante que la empresa busque alternativas que les permitan mejorar este proceso y lo automaticen, para poder optimizar los recursos que se disponen actualmente y hacer uso de las nuevas tecnologías para dar una solución a este problema.

Es por este motivo que se busca implementar un modelo predictivo que permita gestionar de una forma más eficiente el cobro de cartera de las pólizas no pagadas de la línea y tener un mejor control de aquellas que llevan largos períodos sin pagarse. A través de esto, se podrá mejorar la rentabilidad de la línea y la empresa podrá cerrar correctamente el ciclo del proceso de fianzas. El desarrollo de este proyecto tendrá un alcance a nivel nacional, y buscará mejorar los procesos de recuperación de cartera vencida que tiene actualmente la línea de fianzas.

### **Preguntas de investigación**

De acuerdo con los objetivos específicos planteados, se detallan a continuación las preguntas de investigación:

**OE1-RQ1:** ¿La aplicación de técnicas cualitativas de levantamiento de información son suficientes para poder determinar el estado actual de la cartera de la línea de fianzas?

**OE1-RQ2:** ¿El análisis de la información inicial permite identificar el principal problema de la recuperación de cartera de la línea de fianzas?

**OE2-RQ3:** ¿Los artículos de investigación encontrados permiten identificar técnicas para desarrollar modelos predictivos de recuperación de cartera?

**OE2-RQ4:** ¿En la revisión del estado del arte se podrá identificar estudios relacionados con modelos predictivos de recuperación de cartera en empresas de seguros?

**OE3-RQ5:** ¿El modelo predictivo permite identificar las variables que están relacionadas?

**OE3-RQ6:** ¿El desarrollo de un modelo predictivo permitirá determinar patrones de comportamiento de la cartera de fianzas?

**OE4-RQ7:** ¿La selección de la mejor técnica de minería de datos permitirá definir un modelo que se adapte a las necesidades de recuperación de cartera de la línea de fianzas?

**OE4-RQ8:** ¿El dashboard permitirá plasmar los resultados obtenidos del modelo seleccionado?

**OE5-RQ9:** ¿El modelo seleccionado se adapta a las necesidades de cobro de la línea de fianzas?

**OE5-RQ10:** ¿El modelo seleccionado permite mejorar los tiempos de cobro de cartera vencida en los ejecutivos de la línea de fianzas?

### **Hipótesis**

El desarrollo de un modelo predictivo para la cartera vencida de la línea de fianzas permitirá identificar el comportamiento de la morosidad de los clientes a través de la selección de un modelo que se ajuste a su realidad; el mismo que permitirá definir estrategias que mejoren la gestión de cobro de las pólizas a nivel nacional y la optimización de recursos asignados a la línea.

## **Capítulo 2 Marco Teórico**

Para el desarrollo de este estudio se han definido variables que facilitan el desarrollarlo de la solución para el presente problema, además se seleccionaron metodologías de investigación y modelado de datos, las cuales favorecen el desarrollo del presente proyecto de una forma más estructurada.

### **Diseño de Investigación Científica (Desing Science Research,DSR.)**

La investigación en ciencias del diseño es un conjunto de técnicas y perspectivas sintéticas y analíticas para realizar investigación en Sistemas de Información (SI). El DSR generalmente implica la creación de un artefacto y / o teoría del diseño que permita mejorar el estado actual de la práctica, así como el conocimiento de investigación existente (Vaishnavy, Kuechler, & Peter, desrist.org, 2019). Aken (2004) afirma que DSR busca desarrollar un conocimiento válido y confiable para diseñar soluciones, omite la utilización real y las capacidades de resolución de problemas. El DSR permite utilizar el conocimiento adquirido para resolver problemas, crear cambios o mejorar las soluciones existentes; generar nuevos conocimientos, ideas y explicaciones teóricas.

### **Descubrimiento de Conocimiento en Base de Datos (Knowledge Discovery in Databases (KDD)):**

Permite la extracción de conocimiento a partir de bases de datos. Permite la exploración informatizada de grandes cantidades de datos en el descubrimiento de relevantes patrones interesantes dentro de ellos. KDD Es utilizado como un mecanismo que resume y analiza los contenidos de los conjuntos de conceptos que aportan las bases de datos. (Díaz & Domínguez, Interpretación: Un reto en la investigación educativa , 2018)

## Marco Teórico Referencial

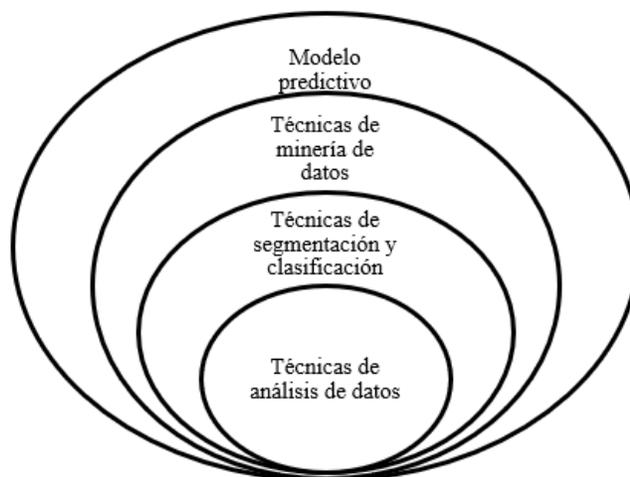
### ***Categorización de la Variable Independiente.***

Las Variables Independientes determinadas para el desarrollo del presente estudio

se ilustran en la Figura 2. En esta sección se detalla algunas de las técnicas de minería y segmentación, para finalmente llegar al modelo. Es importante conocer las definiciones y la teoría en la cual se basará el desarrollo de este proyecto.

### **Figura 2.**

*Variables Independientes*



*Nota.* La figura representa las variables independientes planteadas para el presente proyecto.

### ***Fundamentación de la Variable Independiente:***

#### ***Técnicas de análisis de datos***

El análisis de datos es fundamental para poder adquirir conocimiento y apalancarlo en la toma de decisiones de las empresas. El análisis de datos es una

técnica y por medio de ésta se inspeccionan, purifican y transforman datos, con la finalidad de destacar toda la información que sea de gran utilidad, a fin de poder elaborar conclusiones que sirvan de apoyo en la toma de decisiones (Concepto definición.de, 2019). Así mismo, se puede mencionar el siguiente concepto: El Análisis de Datos (Data Analysis, o DA) es la ciencia que examina datos en bruto con el propósito de obtener conclusiones sobre la información. El análisis de datos es usado en varias industrias para permitir que las compañías y las organizaciones tomen mejores decisiones empresariales. Además, es usado en las ciencias para verificar o reprobando modelos o teorías existentes. El análisis de datos se distingue de la extracción de datos por su alcance, su propósito y su enfoque sobre el análisis. (Rouse, 2012)

**Minería de Datos.** Proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos. Persiguen el conocimiento automático del conocimiento contenido en la información almacenada de modo ordenado en grandes bases de datos (Pérez & Santín, Minería de Datos, Técnicas y Herramientas , 2008). También se puede mencionar que es la extracción no trivial de información que reside de manera implícita en los datos. Prepara sondea y explora los datos para sustraer la información oculta en ellos. Sus bases radican en la Inteligencia Artificial y análisis estadístico. Los modelos resultantes de la minería permiten dar solución a problemas de predicción, clasificación y segmentación. **(Ruiz, 2017)**. Según Berson y Smith las herramientas de datos se han diseñado para ser empleadas directamente por los decisores o personal de negocio y no necesariamente técnicos. La minería de datos automatiza los procesos estadísticos y reduce la

operatividad para quien trabaja en ello. (Barreriro, Diez de Castro, Barreiro, Ruzo, & Losada, 2001).

**Clasificación.** La Clasificación mapea o asocia datos a grupos predefinidos, se encarga de encontrar modelos que describen clases o conceptos para futuras predicciones.

Las técnicas de clasificación pertenecen al grupo de técnicas predictivas (discriminante, árboles de decisión y redes neuronales) como a descriptivas (clustering y segmentación). Las técnicas de clasificación predictivas se denominan técnicas predictivas ad hoc ya que clasifican individuos u observaciones dentro de grupos previamente definidos. Las técnicas de clasificación descriptivas se denominan técnicas de clasificación post hoc porque realizan clasificación sin especificación previa de los grupos (Pérez & Santín, Minería de Datos Técnicas y herramientas, 2008).

**Modelos Predictivos.** Son modelos de datos, basados en estadísticas inferenciales, que se utilizan para predecir la respuesta a algo. En muchos casos el modelo se elige sobre la base de la teoría de la detección para tratar de predecir la probabilidad de un resultado dada una cantidad establecida de datos de entrada. Los modelos pueden utilizar uno o más clasificadores al intentar determinar la probabilidad de un conjunto de datos pertenecientes a otro conjunto (Aritmetics, s.f.).

El análisis predictivo es el conjunto de técnicas de inteligencia empresarial que elabora pronósticos a partir de lo aprendido de la experiencia de la organización. Analizan los datos del pasado para determinar la probabilidad de ocurrencia de un evento determinado. Es especialmente útil en situaciones en las que las empresas necesitan tomar decisiones rápidas basadas en grandes volúmenes de datos (Logicalis, 2015).

**Árbol de clasificación (variable por variable).** Es un dispositivo gráfico para atacar un modelo en el cual se deben tomar decisiones secuenciales. Estas decisiones están entremezcladas con eventos que tienen varios resultados posibles. El árbol completo representa todas las posibles secuencias (*Eppen & Gould, 2000*). Este modelo está conformado por varios nodos cuadrados, mismos que representan puntos de decisión y de los cuales surgen ramas que representan las diferentes alternativas. Las ramas que salen de nodos circulares, o casuales, representan los eventos. La probabilidad de cada evento causal, se indica sobre cada rama. La probabilidad de todas las ramas debe sumar 1

**Redes Neuronales.** Son modelos computacionales que procesan información que emulan el funcionamiento de redes neuronales del cerebro. La red aprende al examinar los registros individuales, al generar una predicción para cada registro y al realizar ajustes a las ponderaciones cuando efectúa una predicción incorrecta. Este proceso se repite muchas veces y la red mejora sus predicciones hasta alcanzar uno o varios criterios de parada (*IBM, s.f.*).

Son redes interconectadas masivamente en paralelo de elementos simples y con organización jerárquica que interactúan con los objetos del mundo real del mismo modo que lo hace el sistema nervioso biológico (*Flórez & Fernández, 2008*). Las capas que componen las Redes Neuronales son tres: capa de entrada, por donde ingresan los datos, oculta, por donde pasan y de salida por la que salen.

**Deep Learning.** Es una clase de algoritmos de Machine Learning basados en Redes Neuronales que se caracteriza por un procesamiento de datos en cascada. (*Telefonica Data Unit, 2020*). También se la conoce como aprendizaje profundo y hace referencia a que una red neuronal está formada por capas intermedias a las que se

llama capas ocultas o invisibles. (Monge, 2020). Es un algoritmo automático que imita la percepción humana inspirada en nuestro cerebro y la conexión entre neuronas. Es la técnica que más se acerca a la forma en la que aprenden los seres humanos. (Bismart, 2020)

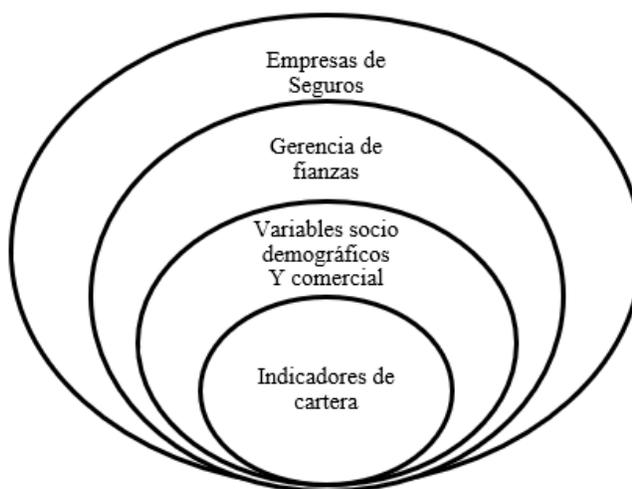
**Redes de Bayes.** Se elaboran en base a un conocimiento desarrollando diversas técnicas para aprender a partir de datos que estructuran parámetros asociados al modelo. Es un modelo probabilístico que relaciona un conjunto de variables aleatorias mediante un grafo dirigido. Son redes graficas sin ciclos en el que se representan variables aleatorias y las relaciones de probabilidad que existan entre ellas que permiten conseguir soluciones a problemas de decisión en casos de incertidumbre.

### ***Categorización de la Variable Dependiente.***

Las Variables dependientes determinadas para el desarrollo del presente proyecto se ilustran en la Figura 3:

### **Figura 3.**

*Variables Dependientes*



*Nota.* La figura representa las variables Dependientes planteadas.

**Empresas de Seguros.** De acuerdo a la Superintendencia de Compañías y Seguros en el Ecuador existen treinta y una compañías (Superintendencia de Compañías, Valores y Seguros, 2020) activas que ofertan sus servicios a los ecuatorianos. El aporte de este sector es del 3% sobre el Pib (Zabala, Especial de Seguros, 2018). Durante el 2018 el sector asegurador mostró una ligera recuperación luego de dos años de contracción. El sector financiero creció al 10,9%, dentro de este se encuentra el sector de seguros. El crecimiento del sector de seguros es apenas del 1%, sin embargo, es una buena señal para quienes forman parte de este. En el mundo de los seguros se clasifican sus servicios para: personas y daños. Dentro de los seguros para personas se encuentran los seguros de vida y en daños se encuentran los ramos generales que protegen el patrimonio de las personas. La mayor concentración se encuentra en el ramo de vehículos con un 23%, vida colectiva con un 21% y en tercer lugar incendio y líneas aliadas con un 8%. Los seguros de Buen Uso del Anticipo y fiel cumplimiento de contrato no tienen una participación importante en el mercado, no obstante, es uno de los ramos más rentables para la compañía. En el ranking de estos ramos la empresa de seguros se encuentra en el puesto tres y cuatro del mercado. Su principal competidor es confianza, una Aseguradora dedicada exclusivamente para ofertar ramos de fianzas.

**Gerencia de Fianzas.** La gerencia de fianzas se encarga de diseñar, coordinar, ejecutar, autorizar y supervisar la operación a nivel nacional de la línea de fianzas. Es la responsable de evaluar los grandes riesgos que se encuentran fuera de las facultades de los ejecutivos y jefes para autorizar las emisiones de las pólizas. Además, es el

responsable de dar seguimiento para el cumplimiento de presupuesto y buscar alianzas estratégicas con canales para poder colocar las pólizas.

**Indicadores Comerciales.** Según los define Bernard Marr, autor de The Intelligence Company, los indicadores son “la información más importante sobre el rendimiento que permite a los responsables y a los stakeholders conocer si la empresa está en el camino correcto o no”.

William Thomson Kelvin en el siglo XIX mencionó que, “Lo que no se define no se puede medir. Lo que no se mide, no se puede mejorar. Lo que no se mejora, se degrada siempre”. Por este motivo surgió la necesidad de diseñar y aplicar indicadores comerciales en la empresa que permitan descubrir las fortalezas y debilidades para poder tomar acciones de mejora. La Asociación Española para la Calidad define los indicadores como “el dato o un conjunto de datos que nos ayudan a medir objetivamente la evolución del sistema de gestión. Son medios, instrumentos o mecanismos para evaluar hasta qué punto o en qué medida se están logrando los objetivos estratégicos. Los indicadores comerciales son aquellos que están orientados a evaluar el área de Ventas, permitiendo a la empresa conocer la eficacia y eficiencia de sus procesos de comercialización y adoptando medidas que conduzcan a la compañía hacia una mejora continua que beneficie a todos los stakeholders, como parte de la filosofía del Capitalismo Consciente. (Capitalismo Consciente, 2017)

**Indicadores de Cartera.** Es una expresión de lo que se quiere medir, con base en factores y variables claves y tienen un objetivo y cliente predefinido. Los indicadores

más frecuentes para cobranzas son: De productividad, de efectividad, acciones de cobranza y calidad de servicio. (Escudero, 2008).

Lo más importante para medir la eficiencia en las cobranzas es contar con sistemas de información que permitan generar reportes e indicadores en tiempo real, los principales indicadores de eficiencia de cobranzas son: Plazo Promedio de Cobranzas, Antigüedad Promedio de la Deuda, Porcentaje de la Deuda Vencida, Porcentaje de Clientes Gestionados.

El uso de indicadores de cartera permite tener: retroalimentación inmediata, información para medidas correctivas y preventivas, facilitar mejoras a los procesos, identificación de tendencias y alertas y mejorar la comunicación entre niveles de la empresa (Credit Management solutions, 2007) Por ejemplo los días de cuentas por cobrar, representan el período medio de cobro de una empresa y los días deudores que representan el tiempo promedio de pago. (ServiceNow, 2020)

## **Estado del Arte**

### ***Definición de la Problemática***

La línea de fianzas de la empresa de Seguros se maneja de una forma muy tradicional y no busca explorar nuevas formas de comercializar sus ramos y realizar cambios innovadores dentro de la línea. Es una de las líneas más rentables para la compañía, sin embargo, actualmente no cuenta con un proceso automatizado de fidelización y recuperación de cartera, lo cual impide que la línea de negocios despunte completamente y sea más productiva al momento de realizar su gestión comercial.

La recuperación de cartera actualmente se la realiza de forma manual y no existen reportes automáticos que permitan hacer una correcta gestión de cobro de las pólizas pendientes. Acorde al sistema de información de la empresa (BI) existe alrededor de 4 millones de dólares por cobrar (Empresa de seguros, 2019). Estos

valores se encuentran en pólizas facturadas y generan gastos mensuales en comisiones e impuestos al estado. Motivo por el que resta rentabilidad a la línea y por lo tanto a la empresa. Debido a la gestión actual de cobro, el valor por recuperar nunca disminuye y conforme las renovaciones se van realizando de forma automática la cartera por cobrar crece cada día más (Bohórquez, 2019).

### ***Objetivo de la Investigación***

Identificar los modelos y esquemas de automatización del proceso de recuperación de cartera a través de la revisión inicial de la literatura disponible para determinar el estado del arte.

**Preguntas de Investigación.** Para desarrollar la presente investigación se plantearon las siguientes preguntas:

**RQ1:** ¿Qué técnicas son usadas para la recuperación de cartera en empresas de servicios?

**RQ2:** ¿Cómo se utiliza el análisis de datos en el proceso de recuperación de cartera?

**RQ3:** ¿Cómo influye la aplicación de modelos en la recuperación de cartera en empresas de servicios?

**RQ4:** ¿La aplicación de modelos ha permitido mejorar los índices de recuperación de cartera en empresas de servicios?

**Importancia de Realizar el Estudio.** Muchas empresas tienen problemas para realizar la recuperación de cartera y terminan contratando a una empresa externa para que realice la gestión de cobranza, generando costos adicionales y entregando información importante sobre los clientes a un tercero.

En el caso de las empresas de seguros que comercializan pólizas de garantías o fianzas, la recuperación de cartera se convierte en un problema, ya que se generan

renovaciones automáticas en el caso de que el cliente (contratista) no haya presentado la documentación correspondiente de la finalización del contrato y la solicite mientras la póliza se encuentre vigente. Esto implica que la empresa de seguros debe pagar todos los impuestos correspondientes a la facturación de las pólizas y el riesgo conforme pasa el tiempo se incrementa y puede sufrir una ejecución en caso de que no se tome las medidas correctivas a tiempo.

Por este motivo es importante analizar las diferentes técnicas, herramientas y soporte tecnológico que se puede aplicar en la recuperación de cartera para evitar problemas y gastos innecesarios para la compañía.

**Criterios de Inclusión y Exclusión.** Se plantearon criterios de inclusión y exclusión para poder seleccionar los estudios relacionados al tema de investigación.

***Criterios de Inclusión.***

- ✓ Artículos que contengan métodos de recuperación de cartera;
- ✓ Artículos que hablen de la aplicación de nuevas tecnologías en la recuperación de cartera;
- ✓ Artículos que hablen de procesos de recuperación de cartera en empresas de servicios;
- ✓ Artículos que contengan modelos predictivos para recuperación de cartera;
- ✓ Artículos que mencionen el uso de herramientas para toma de decisiones.

***Criterios de Exclusión.***

- ✓ Artículos que contengan información sobre creación de empresas de recuperación de cartera;
- ✓ Artículos que hablen sobre estrategias de recuperación de cartera de empresas que se dedican a ese giro de negocio;

- ✓ Artículos que contengan información sobre recuperación de cartera de tarjetas de crédito;
- ✓ Artículos que hablen sobre modelos de recuperación de cartera para créditos bancarios;
- ✓ Artículos que hablen sobre procesos manuales de recuperación de cartera.

**Grupos de Control.** Para el desarrollo del grupo de control se recolectaron estudios relacionados al tema de investigación para identificar las palabras claves que permitan armar la cadena de búsqueda de artículos relacionados, mismos que se describen en la Tabla 1.

**Tabla 1**

*Grupos de Control*

#	Estudios	Términos
EC1	Enhancing delinquent debt collection using statistical models of debt historical information and account events	Predictive model, debt holder, predict the amount to collect, historical data, delinquent debt accounts, collection methods
EC2	Method and computerized system for managing insurance receivable accounts	Method, insurance company, delinquent account information, recovery service,
EC3	Method and system for offering debt recovery products to a customer	Method, system, debt recovery, delinquent accounts, information,
EC4	Designing a Search Mechanism for Debt Collection	Search Mechanism, debt collection, process
EC5	Data-Driven Debt Collection Using Machine Learning and Predictive Analytics	data-driven, debt collection, machine learning, predictive analytics, big data, artificial intelligence, business processes, method, business, debt collection

#	Estudios	Términos
EC6	A decision support system for scoring distressed debts and planning their collection	System, scoring, debt, Decision Support System, artificial neural network, Analytic Hierarchy Process, scoring, regression tree, method Data Mining, Knowledge Discovery, Predictive Modeling, Pattern Recognition, Artificial Neural Networks, Link Analysis, Clustering, Nearest Neighbor, Telecommunications, Customers behavior, bad debt events
EC7	Customer's Relationship Segmentation Driving the Predictive Modeling for Bad Debt Events	

*Nota.* Esta tabla representa el resumen de los estudios hallados para el grupo de control.

**Cadena de Búsqueda.** Para la selección de la cadena de búsqueda se tomaron en cuenta los términos seleccionados en los grupos de control y el buscador académico IEEE, sin embargo, al ser un tema muy específico los primeros resultados arrojaron un número extenso de estudios, los mismos que se pueden visualizar en la tabla 2. Finalmente se seleccionó la cadena que contiene doce artículos relacionados al tema de investigación (ver Tabla 3).

**Tabla 2**

*Cadena de Búsqueda*

Cadena	# Estudios	Estudios del grupo de control	Cadena
C1	82064	EC3, EC1, EC5,	((((Debt recovery) AND predictive model) AND business intelligence) OR machine learning)
C2	697	EC1, EC5, EC6,	(((((debt) AND predictive model) OR machine learning) OR decision support system) AND scoring)
C3	234	EC1, EC5, EC2, EC5	(((((debt) AND predictive model) OR machine learning) AND company) OR insurance company) AND big data)

Cadena	# Estudios	Estudios del grupo de control	Cadena
C4	12	EC1, EC5, EC7, EC2, EC3	(((((debt) AND predictive model) OR machine learning) OR Pattern Recognition) AND insurance company) OR company) AND debt collection) OR debt recovery) AND method)
C5	10	EC1, EC3, EC2	((debt) AND recovery) AND method)
C6	6	E6, E5, E4	((Decision support system) AND debt) AND business processes)
C7	6	EC1, EC5, EC2, EC3	(((((debt) AND predictive model) OR machine learning) AND company) OR insurance company) AND debt collection) OR delinquent account)

*Nota.* Esta tabla representa el resumen de los estudios hallados para las cadenas de búsqueda establecidas.

**Tabla 3**

*Cadena Seleccionada*

Cadena	# Estudios	Cadena
C4	12	(((((debt) AND predictive model) OR machine learning) OR Pattern Recognition) AND insurance company) OR company) AND debt collection) OR debt recovery) AND method)

*Nota.* Esta tabla presenta la cadena de búsqueda seleccionada.

**Proceso de Selección de Artículos.** Para el proceso de selección de artículos se verificó si los doce artículos resultantes de la cadena de búsqueda seleccionada están relacionados al tema de investigación y podrán resolver las preguntas de investigación planteadas. De dichos artículos se eligieron seis que tienen relación con el

tema de investigación (ver tabla 4). A continuación, se detalla los estudios seleccionados y descartados.

**Tabla 4**

*Selección de Artículos*

#	Estudio	Seleccionado	Observación
1	A Neuro-fuzzy Approach to Bad Debt Recovery in Healthcare	Si	Responde a la pregunta de investigación de nuevas técnicas para la recuperación de cartera
2	Less Effort, More Outcomes: Optimising Debt Recovery with Decision Trees	Si	Responde a la pregunta de investigación de nuevas técnicas para la recuperación de cartera
3	A Bayesian Network Approach to Classifying Bad Debt in Hospitals	Si	Responde a la pregunta de investigación de nuevas técnicas para la recuperación de cartera
4	A Neuro-Fuzzy System with Semi-supervised Learning for Bad Debt Recovery in the Healthcare Industry	Si	Cómo influye la aplicación de modelos para poder recuperar cartera en empresas de servicios
5	Cost-Sensitive Learning for Imbalanced Bad Debt Datasets in Healthcare Industry	Si	Responde a la pregunta de investigación de ¿Cómo influye la aplicación de modelos para poder recuperar cartera en empresas de servicios?
6	Collateralized Debt Obligation Pricing with an Alpha-stable Copula	No	No responde a las preguntas de investigación
7	Reflections on Bootstrapping a Microwave Business [MicroBusiness]	No	Está dentro de los criterios de exclusión de tarjetas de crédito
8	Analysis of the inclusion of quantitative methods for the improvement of the effectiveness of collection actions in a power utility	Si	Responde a la pregunta de investigación de métodos para la recuperación de cartera
9	Reengineering of the patient flow process at the Western Sydney Area Health Service	No	No responde a las preguntas de investigación

#	Estudio	Seleccionado	Observación
10	Research on Asset Proposal Method Based on E-commerce	No	Está dentro de los criterios de exclusión de tarjetas de crédito
11	Research on multi-attribute auction based on rough set in non-performing asset disposal	No	Artículos que hablen sobre estrategias de recuperación de cartera de empresas que se dedican a ese giro de negocio
12	Social Engineering for Cyber Security	No	No responde a las preguntas de investigación

*Nota.* Esta tabla presenta el resumen de los artículos seleccionados

Como resultado se obtiene los estudios primarios para responder las preguntas de investigación como se describe en la tabla 5.

## Tabla 5

### Listado de Artículos Primarios

#	Estudio	Seleccionado
1	A Neuro-fuzzy Approach to Bad Debt Recovery in Healthcare	Si
2	Less Effort, More Outcomes: Optimising Debt Recovery with Decision Trees	Si
3	A Bayesian Network Approach to Classifying Bad Debt in Hospitals	Si
4	A Neuro-Fuzzy System with Semi-supervised Learning for Bad Debt Recovery in the Healthcare Industry	Si
5	Cost-Sensitive Learning for Imbalanced Bad Debt Datasets in Healthcare Industry	Si
6	Analysis of the inclusion of quantitative methods for the improvement of the effectiveness of collection actions in a power utility	Si

*Nota.* Esta tabla presenta el resumen de los artículos primarios.

**Síntesis de Resultados.** En base a la revisión preliminar de información se pudo seleccionar seis artículos relacionados al tema de investigación, los cuales fueron considerados ya que responden a ciertas preguntas de investigación.

A continuación, se detalla las respuestas encontradas en el estudio.

**RQ1: ¿Cuáles son las nuevas técnicas usadas para la recuperación de cartera en empresas de servicios?**

En base a la información recolectada, desarrollan modelos matemáticos como por ejemplo bayesianos o Redes Neuronales que permiten predecir y mejorar la recuperación de cartera. Se obtuvo como resultado algunos artículos referentes a servicios médicos, los cuáles se pueden replicar en una empresa de seguros ya que ambos brindan servicios. Otro ejemplo es el artículo “A Bayesian Network Approach to Classifying Bad Debt in Hospitals” cuyo objetivo es explorar la efectividad del uso de una red bayesiana para clasificar la deuda incobrable a través de la comparación con métodos alternativos en diferentes escenarios.

**RQ2: ¿Cómo se utiliza el análisis de datos para recuperar cartera?**

Se desarrolla modelos específicos en base a la información disponible, se encontraron algunos ejemplos en el sector de la salud, por ejemplo, en el artículo “A Neuro-fuzzy Approach to Bad Debt Recovery in Healthcare”, menciona sobre el uso de un modelo ANFIS, el cual permite en base a una muestra de datos desequilibrada simular y tener una interpretación profunda de los resultados.

**RQ3: ¿Cómo influye la aplicación de modelos para poder recuperar cartera en empresas de servicios?**

En los artículos seleccionados se demuestra que con la aplicación de modelos predictivos se ha obtenido mejores indicadores y se ha podido recuperar la cartera de una forma más eficiente. Por ejemplo, en el artículo “Less Effort, More Outcomes: Optimising Debt Recovery with Decision Trees”, se construyeron árboles de decisión para modelar la recuperación de deudas y predecir la respuesta de los clientes si se contactan por teléfono, el modelo es efectivo para clasificar a los clientes en su probabilidad de entrar en un plan de pago de recuperación de deuda exitoso. Los

principales beneficios dentro de la aplicación de los modelos es la segmentación y clasificación de las bases de datos para poder identificar quienes serán los potenciales clientes que pagarán sus deudas.

**RQ4: ¿La aplicación de modelos ha permitido mejorar los índices de recuperación de cartera en empresas de servicios?**

Sí, varios de los estudios encontrados se centran en servicios de la salud y mencionan que es complicado realizar cobros a ciertos pacientes, sin embargo, la aplicación de modelos, especialmente los bayesianos, ANFIS y Redes Neuronales han permitido segmentar de una mejor manera la cartera de sus clientes para dividir las cuentas incobrables y aquellas en las que sí se puede realizar la gestión de cobro. Esto ha permitido incrementar sus niveles de recuperación haciéndolos más efectivos.

**Conclusiones**

- ✓ Al no existir suficiente información relacionada con empresas de seguros se realizó investigación en base a empresas que brinden servicios.
- ✓ No existe mucha información disponible sobre la recuperación de cartera en empresas de seguros y no se hace referencia a la venta de pólizas de fianzas, sin embargo, existen artículos que hablan sobre la aplicación de modelos predictivos en el ámbito de la salud que ha reflejado cambios positivos y mejoras en los indicadores de recuperación de cartera.
- ✓ En los artículos revisados se ha tenido éxito con la aplicación de modelos basados en árboles de decisiones, Redes Neuronales y Bayes que han permitido identificar de una forma más eficaz los patrones de comportamiento de los clientes que se encuentran en morosidad.
- ✓ Los artículos seleccionados demuestran que la aplicación de modelos predictivos permite recuperar la cartera vencida de una forma más eficiente.

- ✓ A través de los artículos obtenidos se puede observar que a partir de los modelos predictivos seleccionados las empresas han podido tomar decisiones y aplicar estrategias para optimizar el cobro de la cartera vencida de sus servicios.

### **Capítulo 3 Marco Metodológico**

Es necesario seleccionar una metodología que permita alcanzar los objetivos planteados. El uso de una o varias metodologías permiten realizar la investigación de una forma estructurada y ordenada, además permite definir el uso de métodos, técnicas, instrumentos, estrategias y procedimientos que se aplicarán en el presente proyecto. La correcta selección de una metodología que se ajuste a las necesidades del proyecto de investigación es un factor fundamental, ya que facilita tener un mejor desarrollo de las actividades y responder adecuadamente a los objetivos planteados, en el presente capítulo se describe el marco metodológico usado para el presente estudio.

#### **Selección de Metodología**

##### ***Enfoque de la investigación***

El enfoque de investigación para el desarrollo de este proyecto es mixto, por la parte cualitativa permitirá contextualizar el fenómeno y obtener profundidad de significados, mientras que por la parte cuantitativa se aplicará predicciones y se analizará una realidad objetiva.

La parte cuantitativa ayudará a consolidar las creencias y establecer con exactitud patrones de comportamiento de una población; y el cualitativo, para que el investigador se forme creencias propias sobre el fenómeno estudiado, como lo sería un grupo de personas únicas o un proceso particular. (Hernández, Fernández, & Baptista, 2014)

Dentro del enfoque cuantitativo se aplicará un diseño de investigación histórica ya que permite reconstruir el pasado de la manera más objetiva y exacta posible, para lo cual de manera sistemática recolectar, evaluar, verificar y sintetizar evidencias que permitan obtener conclusiones válidas a menudo derivadas de hipótesis. (Monje, 2011).

Para la construcción del modelo se recopilará datos históricos de cartera de la línea de fianzas para poder definir un modelo predictivo que permita mejorar la gestión.

En el enfoque cualitativo se aplicará el estudio de casos evaluativo, ya que describe y explica pero además se orienta a la formulación de juicios de valor que constituyan la base para tomar decisiones. (Monje, 2011), el conocimiento y expertise de las personas que conforman la línea de negocios es fundamental para poder identificar cual sería el mejor modelo a aplicarse.

### ***Categoría de la Investigación***

La categoría de este estudio es exploratoria y descriptiva. El análisis exploratorio se realiza cuando el objetivo es examinar un tema o problema de investigación poco estudiado, del cual se tienen muchas dudas o no se ha abordado antes, y el descriptivo busca especificar propiedades y características importantes de cualquier fenómeno que se analice. Describe tendencias de un grupo o población (Hernández, Fernández, & Baptista, 2014). La categoría exploratoria permitirá tener una idea global del problema para posteriormente abarcarlo a través de la investigación descriptiva que permitirá analizar la situación actual de la cartera de fianzas para poder realizar predicciones que permitan tener un mejor comportamiento de los indicadores de cartera de la línea.

### ***Metodología***

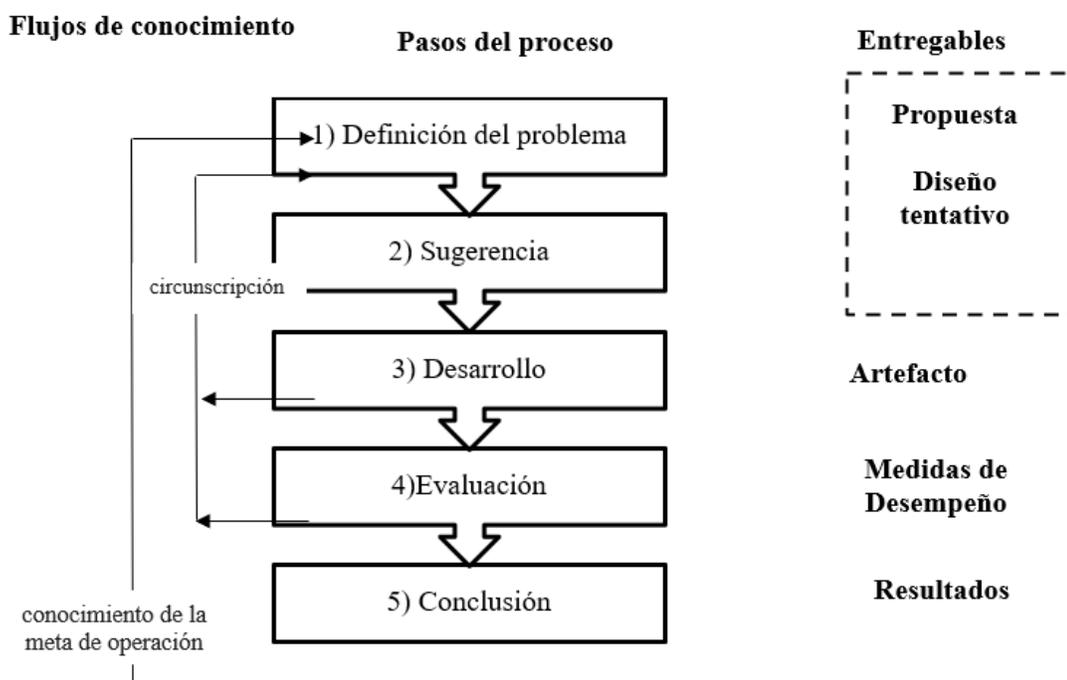
El proyecto está orientado a la metodología de información Diseño de Investigación Científica (Design Science Research, DSR). Andrés González, menciona que a través de esta metodología: "Somos científicos, porque lo que estamos haciendo es producir conocimiento relevante y riguroso de manera simultánea, solo que no compartimos la misma filosofía (con el método científico), y en lugar de forzarnos a pensar como físicos, químicos o matemáticos, pensamos como ingenieros y legitimamos nuestra actividad académica" (Rivera, 2016). Para entender de una mejor

manera esta metodología se puede evaluar por separado el “Design Science” y la parte metodológica que es el “Research”.

Diseño de investigación científica (Design Science Research, DSR) tiene como objetivo el conocimiento, la realización de la configuración, la estructura, la composición, el propósito, el valor y el significado en las cosas y sistemas creados por el hombre. La lógica básica de descubrimiento de la Metodología de Investigación en Ciencia del Diseño es deductiva, porque un problema no resuelto está tomando y trata de encontrar un conocimiento justificativo o una teoría del núcleo que ayude a resolver el problema (Piirainen y González, 2014). Dentro de este tipo de investigación uno de los factores más importantes es el diseño como la investigación del comportamiento; factores que se convierten en inseparables donde la verdad informa al diseño y la utilidad informa a la teoría.

Esta metodología busca determinar cómo deben ser las cosas, que propósito buscarán y cómo se alcanzarán. El proceso de esta metodología constituye una secuencia de actividades de expertos que produce un resultado innovador el cual puede ser probado para obtener mejores resultados. Los artefactos DSR pueden incluir: modelos, métodos, construcciones, instancias y teorías de diseño (March & Smith, 1995; Gregor 2002; March & Storey, 2008, Gregor y Hevner 2013).

Se aplicará esta metodología ya que consta de pasos estructurados los cuales permiten tener un mayor orden al momento de ejecutar el proyecto; además su principio fundamental es que el conocimiento, comprensión de un problema y su solución se adquiere en la aplicación y construcción de un artefacto; en el caso de este estudio se realizará a través de un modelo predictivo que permitirá identificar el comportamiento de pago de los clientes de la línea de fianzas y generar mejores indicadores para el proceso de recuperación de cartera. La metodología está conformada por cinco fases que se describen en la figura 4.

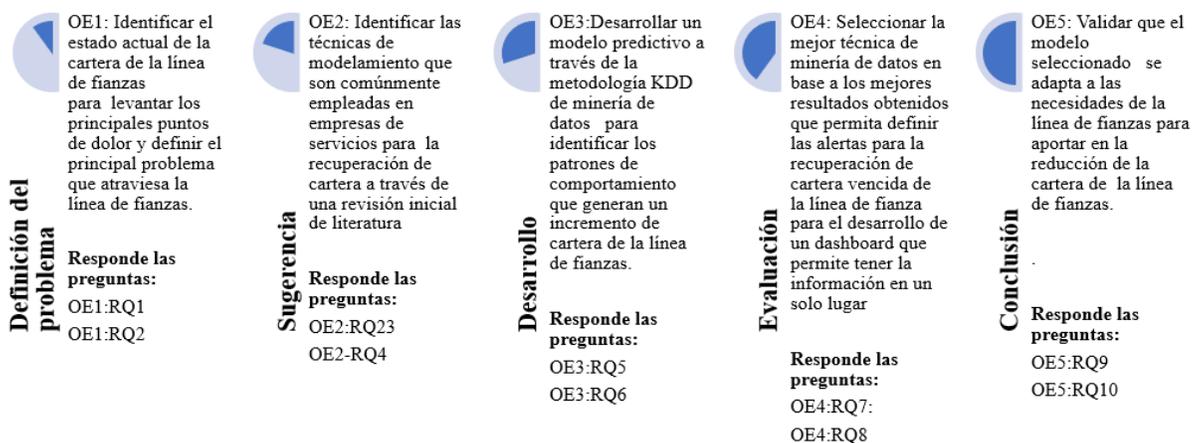
**Figura. 4***Diseño de Investigación*

**Nota.** La figura representa el diseño de investigación. Tomado de Design Science Research pasos generales (Vijay K. Vaishnavi, 2015).

La investigación debe estar alineada a los objetivos de investigación, es por este motivo que se debe demostrar la relación que existe de los objetivos con cada una de las fases planteadas de la metodología como lo muestra la figura 5.

Figura 5.

## Relación del Diseño con los Objetivos



*Nota.* La figura representa la relación entre los objetivos de investigación y las respuestas planteadas para el proyecto.

**Descripción de cada fase:**

**Definición del problema.** En esta fase se define cuál es el problema que atraviesa el área de cartera de la línea de negocio a través del método descriptivo. Se levantó la información usando fuentes secundarias por medio de la técnica de la entrevista al jefe de la línea de fianzas. Además, se analizó fuentes primarias de información como los reportes de cartera de la línea. Esta fase responde al objetivo uno ya que a través del conocimiento del estado actual de la cartera de la línea se identifica el problema.

**Sugerencia.** En esta fase se plantea la posible solución para el problema. Se analizó fuentes primarias a través de una revisión de literatura inicial aplicando un método descriptivo y se analizaron fuentes secundarias de información usando la técnica de la entrevista para determinar cuáles son las fuentes de extracción de data y

las variables disponibles para la construcción de un modelo. Esta fase permite dar respuesta al objetivo O2 del presente proyecto.

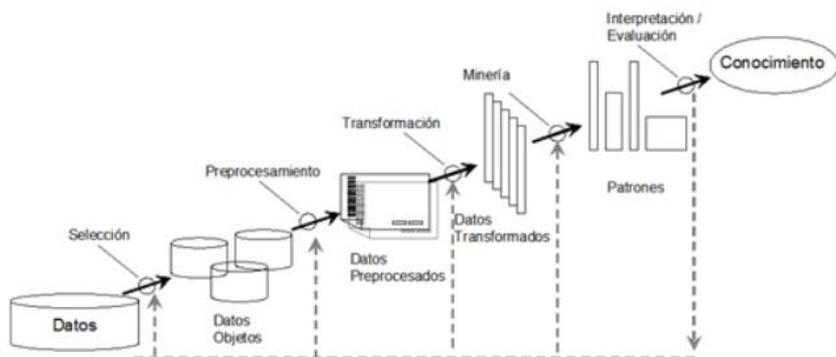
**Desarrollo.** En esta fase se desarrolla el modelo predictivo a través de la metodología de minería de datos KDD. Se aplicó el método exploratorio y el enfoque cuantitativo. A través de la construcción de los modelos se da respuesta a las preguntas del objetivo 3 O3.

La metodología KDD, Knowledge Discovery in Databases por sus siglas en inglés se centra en la exploración informatizada de grandes cantidades de datos y en el descubrimiento de patrones relevantes. Es básicamente un proceso automático en el que se combinan descubrimiento y análisis. Es utilizado como un mecanismo que resume y analiza los contenidos de los conjuntos de conceptos que aportan la base de datos. (Díaz & Domínguez, La interpretación: Un reto en la investigación educativa., 2018). Otra definición que podemos mencionar con respecto a la metodología KDD es “La tecnología KDD está basada en un bien definido proceso KDD de múltiples pasos, para el descubrimiento de conocimiento en grandes colecciones de datos. El proceso KDD es iterativo por naturaleza, y depende de la interacción para la toma de decisiones, de manera dinámica” [GUP]. Una ventaja del KDD es ser un proceso enfocado en el usuario, el cual tiene la capacidad de ser interactivo, y ser direccionado por las decisiones que toma el usuario por un agente inteligente, además de guiarlo para la correcta selección de herramientas y técnicas adecuadas que se ajusten para el cumplimiento de los objetivos. Puede ser usado en diferentes situaciones, como por ejemplo en la determinación de perfiles de clientes fraudulentos, relaciones entre síntomas y enfermedades, técnicas y diagnóstico de maquinaria, etc., Esto demuestra que esta metodología puede ser usada en distintos ámbitos como los negocios, finanzas, medicina, biología, etc.

El proceso KDD es interactivo e iterativo, está conformado por cinco pasos en los cuales interviene el usuario y debe tomar decisiones para llevar a cabo su proyecto de investigación. En la figura 6 se observa todas las fases de la metodología KDD.

**Figura 6.**

*Metodología KDD*



Nota. Tomado de Blog de Minería de Datos, <http://mineriadatos1.blogspot.com/2013/06/descubrimiento-del-conocimiento-kdd-el.html>).

**Selección de datos.** En esta etapa se determinan las fuentes de datos y el tipo de información a utilizar, se realiza la extracción de información de las diferentes fuentes de la empresa.

**Preprocesamiento.** En esta etapa se realiza la limpieza de los datos, eliminación de duplicados, campos vacíos, etc., el objetivo es dejar la base lista para poder hacer el modelado.

**Transformación.** En esta fase se realiza el tratamiento preliminar de los datos, transformación y generación de nuevas variables a partir de las ya existentes con una estructura de datos apropiada. Aquí se realizan operaciones de agregación o normalización, consolidando los datos de una forma necesaria para la fase siguiente.

**Data Mining.** En este paso se realiza el modelamiento a través de métodos inteligentes con el objetivo de extraer patrones previamente desconocidos. La selección del mejor modelo se lo realizará en la sección cuatro de la evaluación de la metodología DSR mediante los resultados generados dentro de esta fase.

**Interpretación y Evaluación.** Se identifican los patrones obtenidos y que son relevantes, basado en algunas medidas y la evaluación de los resultados obtenidos. Esta etapa se realiza en la metodología Diseño de investigación científica (Design Science Research, DSR) en el último punto.

**Evaluación.** En esta fase se realiza el análisis de escenario para poder evaluar cuál es la mejor opción que se debe escoger. Esta fase también se ata al punto 4 de la metodología DSR. Se aplica un método descriptivo y se analizan los modelos en base a los criterios de precisión. Se presentan los datos obtenidos al jefe de la línea para poder conocer su opinión respecto a los modelos obtenidos.

**Conclusión.** En esta fase se selecciona el modelo que se ajuste mejor a la compañía. Para la selección del modelo se analizan varias métricas y se escoge al que obtenga los mejores resultados. Una vez definido el modelo ganador, se establecen las estrategias para mejorar la gestión de cobro de cartera y se realiza la elaboración de un dashboard, el cual permite conocer la situación actual de los clientes y sus probabilidades de pago. En esta fase se da respuesta a las preguntas correspondientes del objetivo cinco (O5).

## Capítulo 4 Planteamiento de la Solución

En este capítulo se explica el planteamiento y alcance de la solución, además se realiza una evaluación de las diferentes herramientas de análisis y visualización de datos para el desarrollo de este estudio.

### Planteamiento de la Solución

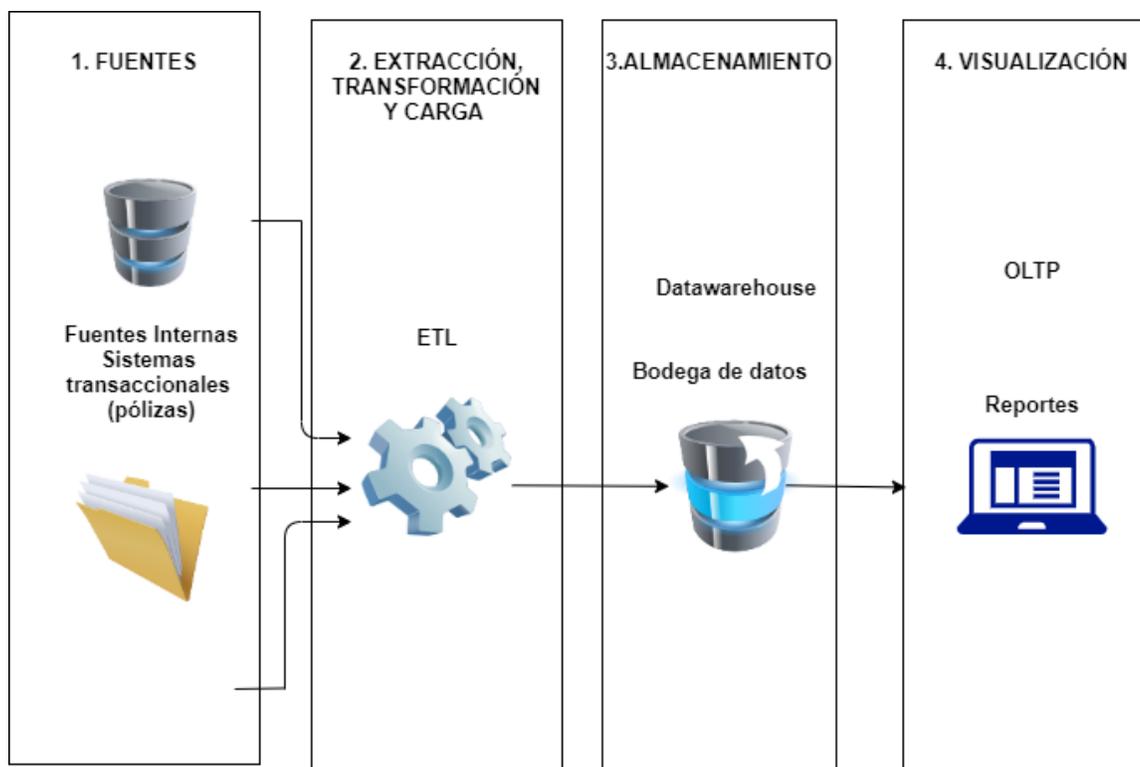
#### ***Análisis del problema desde la perspectiva tecnológica***

Las renovaciones automáticas de la línea de fianzas han provocado un aumento en su cartera por cobrar. La gestión operativa que debe hacer el ejecutivo para obtener los reportes en los cuales puede identificar los clientes que se encuentran en mora, hace que sus tiempos de llamadas y gestión de cobro se reduzcan significativamente y la eficacia de recuperar el dinero de las pólizas emitidas disminuye considerablemente, generando pérdidas a la línea de negocio y restando utilidad a la empresa. El proceso de extracción de la información se lo realiza a través de un reporte manual que arroja los siguientes datos: Datos del cliente, información de contactabilidad; información sobre la póliza: beneficiario, código, fecha de generación, fecha de vigencia, fecha de vencimiento, ramo al cual corresponde, estado, # de días vencidos, # de días de provisión, valor, saldo, prima y los impuestos de ley. Esta información es administrada por el ejecutivo de cartera de la línea y supervisada por la jefatura.

**Infraestructura.** Actualmente la información se la maneja de manera manual, y no se ha realizado ninguna mejora en la reportería correspondiente a la parte de cartera de la línea. La figura 7 muestra la infraestructura que se maneja actualmente en la compañía de seguros.

#### **Figura 7.**

*Infraestructura Actual*



*Nota.* La figura representa la infraestructura actual de la compañía.

En la figura 7 se puede apreciar que la infraestructura actual de la compañía no le permite realizar un proceso predictivo, sino que se basa en reportes manuales y descriptivos, ya que se obtiene la información de los dos sistemas transaccionales para enviarlos al Data Warehouse. De esta manera el ejecutivo puede obtener los reportes de cartera de la línea. Al contar con una reportería basada en sistemas OLTP, se puede observar que los datos son solamente almacenados y que no sufren ningún tipo de transformación o modelamiento, es decir se obtiene un reporte descriptivo, por lo que el ejecutivo de cartera al descargar el reporte obtiene un archivo en Excel que le da el listado de clientes que tienen morosidad, las pólizas y montos adeudados y los datos de contactabilidad. Sin embargo, no hay ningún modelo predictivo que le permita optimizar su trabajo guiándolo de una mejor manera y clasificando a los clientes que tienen una mayor probabilidad para el pago de sus deudas.

Automatización. Actualmente en la línea de fianzas no existe una automatización de reportes para la gestión de cobranzas, por lo que el ejecutivo debe descargar al menos una vez a la semana un reporte el cual arroja información en bruto, ya que ciertos campos son innecesarios para realizar su gestión de cobranzas; debe organizar la información de forma manual y decidir empíricamente a que cliente va a llamar primero para solicitar el cobro de las pólizas pendientes. Este proceso hace que ocupe su tiempo realizando una tarea manual en lugar de llamar a los clientes y obtener los pagos pendientes a través de su gestión. Tener la información de una forma automatizada, permitirá optimizar tiempos y poder mejorar los indicadores del área de cartera de la línea.

Proceso del ejecutivo de cartera. Actualmente el ejecutivo de cartera de la línea de fianzas debe generar y descargar un reporte de clientes, lo exporta a una hoja de cálculo en Excel, elimina los campos que no son necesarios y de esa manera obtiene un listado de clientes con el respectivo código de póliza, fechas de generación de renovaciones, montos a cobrar y días vencidos. En base a esto decide a qué clientes va a llamar; en algunas ocasiones se basa en el número de días vencidos o en los montos adeudados por el cliente. En algunos casos, pese a las llamadas realizadas no consigue recuperar cartera, por lo cual su gestión se ve penalizada al no generar los resultados requeridos por la compañía.

Planteamiento de la Solución. Las nuevas tecnologías, el desarrollo de nuevas herramientas y la facilidad para acceder a la información, han hecho que las empresas evolucionen y busquen nuevas estrategias para desarrollar sus procesos de una forma más eficiente y eficaz. En el capítulo dos se pudo observar varios casos de éxito de empresas de servicios tanto de medicina como de banca, los cuales han empleado nuevas herramientas y técnicas para mejorar sus indicadores de recuperación de

cartera; segmentando su cartera y clasificando a sus clientes para poder identificar quienes serán los mejores pagadores y poder iniciar con ellos los procesos de cobro.

Las empresas deben empezar a hacer uso de las nuevas tecnologías y aprovechar las ventajas que ofrecen. En la actualidad existe un sinnúmero de herramientas que permiten construir modelos matemáticos para identificar patrones de comportamiento a través de los datos de las empresas. Sin embargo, se debe analizar cuáles son las herramientas que ofrecen mejores beneficios y se adaptan a las necesidades de las empresas. Para poder identificar cuáles son las mejores herramientas de data science se puede hacer uso del cuadrante mágico de Gartner, que es el resultado de investigación realizado por la empresa que lleva su nombre y es una de las consultoras e investigadoras de tecnología de información más importante a nivel mundial. Gartner los clasifica en 4 cuadrantes, los cuales corresponden a:

**Líderes.** Poseen una gran cuota de mercado. Cuentan con una oferta madura y centralizada sin soluciones personalizadas.

**Retadores.** No poseen grandes cuotas de mercado, sin embargo, tienen productos sólidos y los recursos suficientes como para mantener un crecimiento continuo.

**Jugadores de Nicho.** Se especializan en segmentos específicos. Tienen mayor flexibilidad para ajustarse a las necesidades de sus clientes, poseen una alta orientación al usuario.

**Visionarios.** Son capaces de ofrecer productos innovadores, no poseen una cuota significativa de mercado y se puede encontrar a startups.

### ***Análisis de Plataformas de Data Science and Machine Learning***

La figura 8 muestra las plataformas que se encuentran dentro del cuadrante Mágico de Gartner de Analítica e Inteligencia de Negocio:

**Figura 8.**

*Cuadrante Mágico de Gartner de Data Science and Machine Learning*



Nota. Tomado de Gartner, 2020,

(<https://www.analyticsvidhya.com/blog/2020/02/gartners-2020-magic-quadrant-for-data-science-and-machine-learning-tools-check-out-the-new-leaders/>).

Para el desarrollo del presente proyecto se analizaron a aquellas empresas que se encuentran dentro del cuadrante de los visionarios; uno de los principales motivos para analizar a estas herramientas son los costos. La tabla 6 permite visualizar las ventajas y desventajas de las herramientas de Data Science y Machine Learning.

Tabla 6

*Ventajas y Desventajas de Herramientas de Data Science and Machine Learning*

<b>Empresa</b>	<b>Ventajas</b>	<b>Desventajas</b>
<b>Knime</b>	<ul style="list-style-type: none"> <li>• Código abierto con soporte comercial para capacidades de nivel empresarial</li> <li>• Es una solución DSML robusta, integral y de extremo a extremo</li> <li>• Facilita la orquestación en todo el proceso analítico de extremo a extremo, incluida una operacionalización simplificada de modelos.</li> <li>• Capacidad de crear y compartir, a través del KNIME Hub, componentes reutilizables con múltiples tipos de usuarios</li> </ul> <p>Visión analítica aumentada madura</p>	<ul style="list-style-type: none"> <li>• Visibilidad moderada y crecimiento relativo lento ya que no actualizan a la plataforma comercial donde puedan aprovechar las capacidades de extremo a extremo.</li> <li>• Deficiencias en la capacitación Problemas de rendimiento: cuando se trabaja con grandes conjuntos de datos, flujos de trabajo largos o el uso complejo de múltiples extensiones, que a veces requieren movimientos o conversiones de datos importante</li> </ul>
<b>DataRobot</b>	<ul style="list-style-type: none"> <li>• Lidera la carga de incorporar analítica aumentada dentro de DSML.</li> <li>• Proporciona capacidades beneficiosas para una amplia variedad de roles, incluidos desarrolladores, científicos de datos, estadísticos y analistas de negocios.</li> <li>• Excelente soporte continuo a través de CFDS y asistencia en el producto.</li> <li>• Agrega rápidamente mejoras de productos y responde a las solicitudes de los clientes.</li> </ul>	<ul style="list-style-type: none"> <li>• Percibido como una herramienta introductoria y elemental para los no expertos</li> <li>• El enfoque del Plan de Éxito AI requiere muchos recursos</li> <li>• Percepción de altos costos</li> </ul>

Empresa	Ventajas	Desventajas
<b>Microsoft</b>	<ul style="list-style-type: none"> <li>• Funcionalidad atractiva y madura para múltiples niveles de habilidades y personas analíticas</li> <li>• Capacidad para actualizar y mejorar el producto en la nube con una cadencia rápida.</li> <li>• Alto puntaje en escalabilidad y rendimiento</li> </ul>	<ul style="list-style-type: none"> <li>• Muchas capacidades dentro de la cartera de Azure ML se vuelven más débiles o complicadas en entornos híbridos, multinube o locales.</li> <li>• enfrentan problemas de licencia y costo total de propiedad.</li> <li>• DSML aumentado sólido, pero no líder</li> </ul>
<b>Google</b>	<ul style="list-style-type: none"> <li>• Énfasis en los problemas de ciencia de datos de vanguardia</li> <li>• Ofrece escalabilidad de alta gama y capacidades integradas de etiquetado de datos.</li> <li>• Ofrece un fuerte soporte para la transmisión de datos con Cloud Dataproc y Cloud Dataflow.</li> <li>• Oferta una infraestructura de nube administrada para escalar automáticamente diferentes configuraciones de hardware, desde medianas hasta ultra grandes</li> </ul>	<ul style="list-style-type: none"> <li>• Gran huella y curva de aprendizaje</li> <li>• la cadena de herramientas aún no está adaptada homogéneamente a las necesidades de los proyectos de ciencia de datos</li> <li>• Beta persistente y capacidades locales limitadas</li> </ul>
<b>RapidMiner</b>	<ul style="list-style-type: none"> <li>• El proveedor ofrece específicamente un programa de certificación a través de su Academia RapidMiner para que los analistas que no son expertos en el manejo datos comprendan el producto, el desarrollo del modelo y la operacionalización.</li> <li>• Flexibilidad, agilidad y escalabilidad moderna: RapidMiner es muy compatible con una amplia gama de tecnologías permite el cumplimiento y la auditoría de modelos tanto en las instalaciones como en la nube.</li> </ul>	<ul style="list-style-type: none"> <li>• Tiene un crecimiento relativo lento en comparación con otros proveedores con propuestas de valor comparables</li> <li>• Su equipo de ventas y el tamaño promedio del negocio son pequeños en comparación con otros proveedores de tamaño similar</li> <li>• El costo total de propiedad puede ser difícil de calcular</li> </ul>

Empresa	Ventajas	Desventajas
<b>Microsoft</b>	<ul style="list-style-type: none"> <li>• Funcionalidad atractiva y madura para múltiples niveles de habilidades y personas analíticas</li> <li>• Capacidad para actualizar y mejorar el producto en la nube con una cadencia rápida.</li> <li>• Alto puntaje en escalabilidad y rendimiento</li> </ul>	<ul style="list-style-type: none"> <li>• Muchas capacidades dentro de la cartera de Azure ML se vuelven más débiles o complicadas en entornos híbridos, multinube o locales. Enfrentan problemas de licencia y costo total de propiedad.</li> <li>• DSML aumentado sólido, pero no líder</li> </ul>
<b>Google</b>	<ul style="list-style-type: none"> <li>• Énfasis en los problemas de ciencia de datos de vanguardia</li> <li>• Ofrece escalabilidad de alta gama y capacidades integradas de etiquetado de datos.</li> <li>• Ofrece un fuerte soporte para la transmisión de datos con Cloud Dataproc y Cloud Dataflow.</li> <li>• Oferta una infraestructura de nube administrada para escalar automáticamente diferentes configuraciones de hardware, desde medianas hasta grandes</li> </ul>	<ul style="list-style-type: none"> <li>• Gran huella y curva de aprendizaje la cadena de herramientas aún no está adaptada homogéneamente a las necesidades de los proyectos de ciencia de datos</li> <li>• Beta persistente y capacidades locales limitadas</li> </ul>

*Nota.* La tabla presenta la recopilación de las principales ventajas y desventajas de las herramientas de Data Science and Machine Learning, tomado de Peer Insight, 2020, ([https://www.gartner.com/reviews/home?\\_ga=2.38304739.191729287.1597644004-643297307.1597644004](https://www.gartner.com/reviews/home?_ga=2.38304739.191729287.1597644004-643297307.1597644004)).

En base al análisis de ventajas y desventajas, se preseleccionaron dos herramientas: RapidMiner y Knime, debido a que ambas herramientas tienen características que se adaptan a las necesidades de la empresa y además existen versiones gratuitas, lo cual representa un ahorro. En la tabla 7 se detallan las principales funcionalidades y características de las herramientas.

**Tabla 7***Características de RapidMiner y Knime*

<b>KNIME</b>	<b>RAPIDMINER</b>
✓ Visualización de datos	✓ Análisis Estadístico
✓ Transformación de datos	✓ Detección de Fraude
✓ Análisis de datos	✓ Gestión de datos vinculados
✓ Modelamiento de datos	✓ Aprendizaje Automático
✓ Analítica predictiva	✓ Exploración de textos
✓ Big Data	✓ Modelado Predictivo
✓ Api	✓ Extracción de datos
	✓ Visualización

*Nota.* La tabla presenta la recopilación de las principales características de las herramientas preseleccionadas, tomado de Peer Insight, 2020, ([https://www.gartner.com/reviews/home?\\_ga=2.38304739.191729287.1597644004-643297307.1597644004](https://www.gartner.com/reviews/home?_ga=2.38304739.191729287.1597644004-643297307.1597644004)).

Para la selección de la herramienta se utilizaron comparadores de empresas especialistas en la venta y análisis de Software, una de ellas es Capterra que es una empresa fundada en 1999 que se dedica a la comparación de softwares, actualmente cuenta con más de 1.170.000 de reseñas verificadas en su página y miles más añadidas mensualmente. En 2015 fueron adquiridos por Gartner se unieron a la familia de Digital Markets. Esta división, formada por Capterra, Software Advice y GetApp. Y también la página de Gartner peer Insights, Las distinciones Gartner Peer Insights Customer 'Choice reconocen a los proveedores y productos que son altamente calificados por sus clientes. Los datos que recopilan representan una síntesis de alto nivel de los productos de software de proveedores más valorados por los profesionales

de TI Enterprise. (Gartner Peer Insights, 2020) En la tabla 8 se visualiza el comparativo de ambas herramientas tomando en cuenta los productos que poseen, las capacidades de la herramienta y la experiencia del usuario.

**Tabla 8**

*Productos de las Herramientas de Data Science and Machine Learning*

<b>KNIME</b>	<b>RAPIDMINER</b>
✓ Analytics Platform	✓ RapidMiner Studio
✓ Knime Server	✓ RapidMiner Auto Model
	✓ RapidMiner Server
	✓ RapidMiner Radoop
	✓ RapidMiner Turbo
	✓ RapidMiner Real Time-Scoring

*Nota.* La tabla presenta el portafolio de productos de las herramientas preseleccionadas, tomado de Peer Insight, 2020, ([https://www.gartner.com/reviews/home?\\_ga=2.38304739.191729287.1597644004-643297307.1597644004](https://www.gartner.com/reviews/home?_ga=2.38304739.191729287.1597644004-643297307.1597644004)).

RapidMiner posee un portafolio más amplio de productos comparado con Knime. En la tabla 9 se observa las capacidades de ambos productos.

**Tabla 9**

*Capacidades de los Productos Seleccionados de Data Science and Machine Learning*

<b>VARIABLE</b>	<b>KNIME</b>	<b>RAPIDMINER</b>
Acceso a datos y manipulación	4.4	4.5
Exploración y visualización de datos	4.2	4.2
Analítica Predictiva	4.4	4.7
Previsión	4.7	4.2
Mejoramiento	4.3	4.2
Simulación	4.5	4.4
Entrega, integración e implementación	4.5	4.7
Plataforma y gestión de proyectos	4.2	4.1
Rendimiento y escalabilidad	4.3	4.4
Experiencia del Usuario	4.6	4
Acceso a los datos	4.7	4.5
Preparación de datos	4.3	4.4
Aumento (Automatización)	4.3	4.6
Interfaz de usuario	4.7	4.5
Aprendizaje Automático	3.3	3.8
Otras analíticas Avanzadas	4.7	4.3
Flexibilidad y Apertura	4.3	4.5
Entrega	4.7	4.4
Gestión de Modelo	4.7	4.3
Soluciones pre-enlatadas	3.7	4.4
Colaboración	4	4.4
Coherencia	4.3	4.5

*Nota.* La tabla presenta los puntajes obtenidos por las herramientas preseleccionados.

Tomado de Peer Insight, 2020,

([https://www.gartner.com/reviews/home?\\_ga=2.38304739.191729287.1597644004-643297307.1597644004](https://www.gartner.com/reviews/home?_ga=2.38304739.191729287.1597644004-643297307.1597644004)).

De las veinte y dos variables analizadas, RapidMiner obtiene un puntaje más alto en once, mientras que Knime en diez y una variable tienen un puntaje similar. El promedio obtenido por Rapidmer es de 4.36, mientras que Knime 4.35, lo cual las hace

ser herramientas robustas con ventajas en puntos específicos sobre la otra. Mientras Knime es más fuerte en la experiencia al usuario, RapidMiner obtiene mejores puntajes en aprendizaje automático y analítica predictiva. En la tabla 10 se detalla el comparativo de las dos herramientas.

**Tabla 10**

*Comparativo de Experiencia al Cliente Knime vs RapidMiner*

<b>VARIABLE</b>	<b>KNIME</b>	<b>RAPIDMINER</b>
Flexibilidad de precios	4.5	4.3
Facilidad de implementación	4.6	4.3
Puntualidad de la respuesta del vendedor	4.5	4.4
Calidad de soporte técnico	4.4	4.6

*Nota.* La tabla presenta el comparativo de experiencia al cliente por las herramientas preseleccionados. Tomado de Peer Insight, 2020, ([https://www.gartner.com/reviews/home?\\_ga=2.38304739.191729287.1597644004-643297307.1597644004](https://www.gartner.com/reviews/home?_ga=2.38304739.191729287.1597644004-643297307.1597644004)).

Knime tiene una ventaja en las variables de experiencia al usuario, ya que obtiene mejores puntajes en las variables analizadas, sin embargo RapidMiner es superior en la calidad de soporte técnico, lo cual representa una ventaja para las personas que lo usan. La tabla 11 muestra los puntajes obtenidos por las herramientas en su evaluación de factores.

**Tabla 11**

*Evaluación de Principales Factores de RapidMiner y Knime*

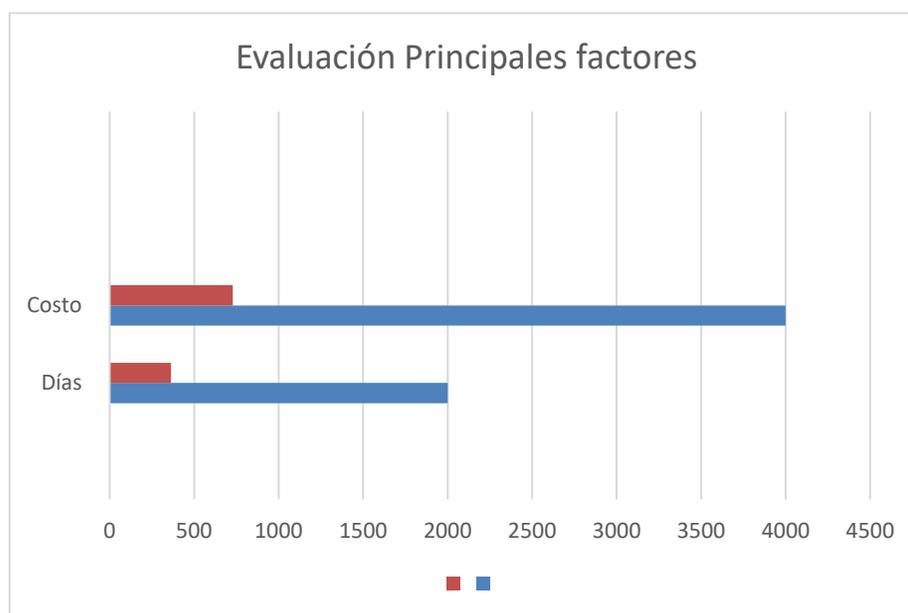
VARIABLE	KNIME	RAPIDMINER
Capacidades del producto	4.5	4.6
Evaluación y contratación	4.7	4.6
Integración e Implementación	4.3	4.3
Servicio de ayuda	4.4	4.4
Calificación general de pares	4.5	4.6

*Nota.* La tabla presenta la evaluación de los principales factores de las herramientas preseleccionados. Tomado de Peer Insight, 2020, ([https://www.gartner.com/reviews/home?\\_ga=2.38304739.191729287.1597644004-643297307.1597644004](https://www.gartner.com/reviews/home?_ga=2.38304739.191729287.1597644004-643297307.1597644004)).

Knime y RapidMiner obtienen calificaciones muy similares en los principales factores analizados, el promedio obtenido por RapidMiner es de 4,5 mientras q de Knime de 4,48. La figura 9 muestra la evaluación de los principales factores.

**Figura 9.**

*Comparativo Knime vs RapidMiner.*



*Nota.* La figura representa un comparativo entre las herramientas de RapidMiner y Knime basado en costo y días de implementación. Basado en Peer Insight Gartner, 2020, ([https://www.gartner.com/reviews/home?\\_ga=2.201962129.1493004894.1597636726-792984529.1597636726](https://www.gartner.com/reviews/home?_ga=2.201962129.1493004894.1597636726-792984529.1597636726)).

Estar en el Cuadrante Mágico garantiza la alta calidad de los servicios ofrecidos en materia tecnológica y un amplio conocimiento de las necesidades reales del mercado. Luego de analizar las ventajas y desventajas de cada una de las herramientas del cuadrante de visionarios se seleccionará la herramienta RapidMiner, por las diferentes ventajas que presenta comparada con el resto, pese a que obtuvieron puntajes muy similares, RapidMiner se proyecta con ciertas ventajas sobre Knime en las variables de analítica predictiva y aprendizaje automático, lo cual agrega valor para los requerimientos de la empresa.

### ***Análisis de Plataformas de Data Analytics and Business Intelligence***

Luego de analizar las herramientas para modelar los datos del presente estudio; es importante seleccionar una herramienta que permita visualizar la información de una forma amigable y que obtenga gráficos y reportes para poder optimizar los tiempos de respuesta y tomar mejores

Para la selección de esta herramienta se vuelve a analizar el cuadrante de Gartner (ver Figura 10), de plataformas de análisis e inteligencia empresarial. Las plataformas modernas de análisis e inteligencia empresarial (ABI) se caracterizan por una funcionalidad fácil de usar que admite un flujo de trabajo analítico completo, desde la preparación de datos hasta la exploración visual y la generación de información, con énfasis en el autoservicio. En la tabla 12 se puede observar el comparativo realizado de dichas herramientas.

Figura 10.

## Plataformas De Análisis E Inteligencia Empresarial



Nota: Tomado de Gartner, 2020, (<https://info.microsoft.com/ww-landing-2020-gartner-magic-quadrant-for-analytics-and-business-intelligence.html?LCID=EN-US>).

Tabla 11

## Comparativo de Plataformas de Análisis e Inteligencia Empresarial

Empresa	Ventajas	Desventajas
Microsoft	<ul style="list-style-type: none"> <li>✓ Bajos costos para los usuarios y se encuentra incluido en Office 365 E5</li> <li>✓ Servicio en la nube Power BI Pro superó a la mayoría de sus competidores en términos de funcionalidad.</li> <li>✓ Incluye capacidades innovadoras para el análisis aumentado y ML automatizado</li> <li>✓ Servicios basados en inteligencia artificial, como análisis de texto, sentimientos basados en capacidades de Azure.</li> </ul>	<ul style="list-style-type: none"> <li>✓ No es flexible de elegir una oferta de IaaS en la nube. Se ejecuta en Azure.</li> <li>✓ El rendimiento de la consulta de las puertas de enlace de datos locales es variable y requiere un esfuerzo para optimizarlo.</li> <li>✓ La conectividad a las consultas directas de SAP BW y HANA es problemática</li> </ul>

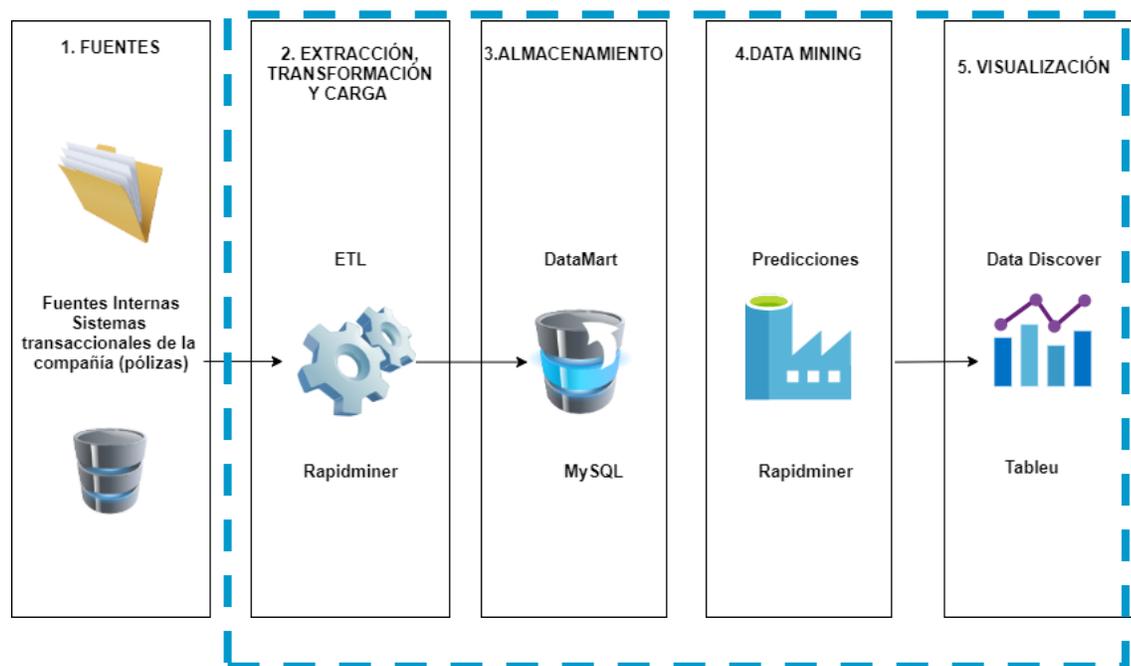
Empresa	Ventajas	Desventajas
<b>Qlik</b>	<ul style="list-style-type: none"> <li>✓ Flexibilidad para implementar en las instalaciones, o con cualquier proveedor importante de la nube</li> <li>✓ Ofrece capacidades de descubrimiento de datos visuales de autoservicio para analistas o usuarios comerciales, al tiempo que admite análisis integrados de desarrolladores desde la misma plataforma</li> <li>✓ Proyecto de alfabetización de datos de Qlik ayuda a los usuarios a comprender y utilizar mejor los datos.</li> </ul>	<ul style="list-style-type: none"> <li>✓ La experiencia de migración de QlikView a Qlik Sense, es una preocupación para sus clientes</li> <li>✓ La mayoría de sus usuarios utiliza paneles parametrizados, reduciendo el uso del autoservicio</li> </ul>
<b>Tableau</b>	<ul style="list-style-type: none"> <li>✓ Alto % de satisfacción de sus clientes</li> <li>✓ Facilidad de exploración visual y manipulación de datos</li> <li>✓ Sus clientes se actualizan a la nueva versión y mencionan experiencias positivas de su experiencia de migración</li> </ul>	<ul style="list-style-type: none"> <li>✓ Puede surgir competidores más disruptivos</li> <li>✓ Percepciones de gobierno y administración débiles</li> <li>✓ Los usuarios deben pagar más por nuevas funciones</li> </ul>
<b>ThoughtSpot</b>	<ul style="list-style-type: none"> <li>✓ Especialidad en búsqueda e inteligencia artificial a escala</li> <li>✓ Capacidad analítica aumentada de ThoughtSpot, descubre anomalías, correlaciones y análisis comparativos entre puntos de datos sin necesidad de codificación</li> <li>✓ Visión del producto orientada al consumidor y aumentada.</li> <li>✓ El conocimiento del mercado de su propuesta de valor basada en la búsqueda es alto</li> </ul>	<ul style="list-style-type: none"> <li>✓ Los datos deben prepararse y limpiarse utilizando herramientas de terceros</li> <li>✓ El software no permite a los usuarios manipular fácilmente los datos en grupos o contenedores sin usar fórmulas; los paneles son básicos y carecen de funciones de mapeo enriquecidas.</li> <li>✓ Requiere preparación de datos y mapeo de datos por adelantado. Genera mayor esfuerzo del área de TI</li> </ul>

Nota: Tomado de Gartner, 2020, (<https://info.microsoft.com/ww-landing-2020-gartner-magic-quadrant-for-analytics-and-business-intelligence.html?LCID=EN-US>)

Para este estudio se seleccionó Tableau, gracias a sus ventajas, su fácil manejo e integración con diversas fuentes de datos, incluidas Excel. Además, es el líder en su categoría, lugar que lo ha mantenido durante varios años y lo ha hecho convertirse en una de las primeras opciones de las compañías. Adicionalmente, la empresa tuvo un acercamiento en el 2019 para la contratación de la plataforma. Se utilizará un motor de base de datos MySQL, debido a que ésta es la usada por la compañía. En la Figura 11, se ilustra la arquitectura tecnológica para el presente proyecto:

**Figura 11.**

*Arquitectura Tecnológica Propuesta*



*Nota.* La figura muestra la propuesta de arquitectura para el área de cartera de la línea de fianzas.

La figura 11 permite visualizar el cambio que se propone con el presente proyecto. Como se puede observar se busca dar un valor agregado a los reportes que maneja actualmente la línea, incluyendo un proceso de data mining, que permita predecir aquellos clientes que tienen probabilidades más altas de pago y colocar esta

información en un dashboard, para que el ejecutivo de cartera y las jefaturas de la línea, puedan tener un mejor control de la gestión de cobranzas que se realiza. Para poder llevar a cabo esta automatización se propuso el uso de la herramienta RapidMiner para la creación de un datamart a través de una base MySQL para la obtención de un dashboard de control.

Entregar información adecuada y oportuna al ejecutivo de cartera de la línea de fianzas permitirá que sus tiempos de gestión de cobranzas sean óptimos y que sus esfuerzos sean enfocados en el cobro y no en obtener reportes manuales. El análisis predictivo es un área de la minería de datos que consiste en la obtención de datos y su utilización para la predicción de tendencias o patrones de comportamiento. Este análisis se basa en la identificación entre variables en eventos pasados para poder predecir situaciones futuras. Actualmente existen herramientas que facilitan el desarrollo de estos modelos, los cuales pueden ser implementados en las empresas para obtener grandes resultados. La línea de fianzas cuenta con una base de datos importante, que permitirá aplicar técnicas de minería para obtener un modelo predictivo que permita segmentar a los clientes en base a su comportamiento. De esa manera se identifica los que tienen probabilidades de pago más altas con el fin de buscar la optimización de la generación de estrategias y mejorar los indicadores de gestión de cobranzas de la línea y de rentabilidad de la empresa.

### ***Alcance de la Solución***

El desarrollo del presente proyecto propone mejorar la gestión en los procesos de recuperación de cartera de la línea de fianzas a nivel nacional a través del desarrollo de un modelo predictivo seleccionado matemáticamente y por el negocio. Además, la creación de un dashboard en el cual se consolide toda la información correspondiente a los clientes que concentran cartera vencida para poder identificar a aquellos que tienen una probabilidad más alta de pago y empezar la gestión por ellos. El análisis de datos

corresponde al período de enero a diciembre de 2017 a 2019. El presente estudio culminó con la selección del modelo más preciso, que se acopló a las necesidades de la empresa y la definición de estrategias que se establecieron en base a la información obtenida.

La línea de fianzas se caracteriza por tener una política de pago inmediato y con contadas excepciones se brinda plazos de pago dependiendo del tipo de cliente, el tipo de contrato, el monto asegurado y el valor de la póliza. Para las renovaciones, en el caso de que el cliente lo solicite, es un proceso automático, por lo cual se dificulta recaudar el pago de forma inmediata fomentando el incremento de la cartera vencida.

## Capítulo 5 Desarrollo de la Investigación

En este capítulo se muestra el desarrollo de la solución aplicada al problema de cartera de la línea de negocio. Como se describió en el capítulo 3, la metodología aplicada fue la de Ciencia del Diseño (DSR) y en su cuarta fase de desarrollo se combina con la técnica KDD de minería de datos. A continuación, se muestra la aplicación de la técnica para la obtención de los modelos predictivos.

### Planteamiento de la Solución

#### *Metodología KDD*

Permite extraer conocimiento a partir de grandes cantidades de datos. Para llevar a cabo esta metodología se deben seguir cinco pasos. A continuación, se detalla los pasos y su aplicación en el desarrollo del presente proyecto.

**Abstracción del escenario.** Actualmente el área de cartera de la línea de fianzas de la empresa se maneja de una forma manual. Pese a las modificaciones y mejoras por las cuales ha atravesado, no se ha tenido una automatización para los cobros de las pólizas. En la figura 7 del capítulo 4, se pudo observar el flujo que se maneja en la compañía en la que se observa que los reportes son descriptivos y no predictivos. La empresa cuenta con dos sistemas transaccionales, los cuales permiten que los datos se almacenen en un data warehouse, para posteriormente poder obtener reportes operativos sobre las pólizas que se encuentran con falta de pago. Sin embargo, no cuentan con una inteligencia por detrás, ya que el reporte que se obtiene es un listado con datos como: cliente, APS (asesor productor de seguros), beneficiario, fechas de inicio y fin de vigencia, valor de la póliza, código de la póliza, saldo adeudado, número de días de mora, etc. Como consecuencia, no permite al ejecutivo de cartera

contar con una base consolidada, que le facilite identificar cuáles son los clientes que tienen una probabilidad de pago más alta para iniciar la gestión con ellos.

El ejecutivo de la línea una vez por semana obtiene el reporte y emplea varias horas de un día en organizar los datos, para posteriormente realizar la gestión de cobro realizando llamadas a los clientes que bajo su criterio son los que considere más aptos para obtener el pago de sus deudas.

Utilizar las nuevas tecnologías y herramientas existentes en el mercado brindan la oportunidad de mejorar los resultados en la gestión de cartera de la línea, ya que se automatizará el reporte y se aplicará un modelo que proporciona información sobre quiénes son los clientes que tienen una mayor probabilidad de pago, esto optimizará los tiempos del ejecutivo y favorecerá el desarrollo de su trabajo al hacerlo más eficiente.

**Selección de los datos.** Para la selección de datos de este estudio se analizó los diferentes reportes de la línea en conjunto con el personal responsable de fianzas. En base a la información que contenía cada uno de los reportes, se decidió usar el reporte transaccional de cartera que fue obtenido y entregado por la línea de negocio. Esta información proviene de los sistemas transaccionales de la empresa y contiene de forma consolidada el reporte histórico de los clientes morosos. Al ser un reporte consolidado, se encuentran datos del año 2008 hasta el 2020.

La información tomada en cuenta corresponde a los años completos del 2017, 2018 y 2019. El motivo por el que se seleccionó este período se debe a que en el año 2016 el sector asegurador fue afectado por varias normativas, las cuales fueron explicadas en el Capítulo 1. Estas normativas hicieron que la producción sea irregular y exista inestabilidad. Otro de los motivos fue el cambio de gobierno que se realizó en el año 2017. Para la línea de fianzas cuando existe un cambio de gobierno provoca que se

incrementen ejecuciones y emisiones de pólizas, lo cual hace que su producción sea irregular comparada con otros años.

A partir del 2017, se puede evidenciar una mayor estabilidad y recuperación en el mercado asegurador. Para tomar un período de tres años completos se realizó un corte al 31 de diciembre de 2019. La información del 2020 fue omitida debido a que el primer trimestre del año sus ventas fueron bajas y los efectos provocados por la pandemia Covid\_19 se vieron reflejados en la reducción de ventas de la línea de negocio. El reporte de la línea consta de las variables detalladas en la tabla 13:

**Tabla 12**

*Variables del Reporte de Cartera*

<b>Variable</b>	<b>Descripción</b>
<b>Sucursal</b>	Agencia en la cual se emitió la póliza
<b>Código de APS</b>	código del bróker
<b>Cliente</b>	Cliente que adquiere la póliza de fianzas (contratista)
<b>Cod beneficiario</b>	Código del beneficiario
<b>Score del Cliente</b>	Calificación otorgada al cliente
<b>Cobrador</b>	Estado en el que se encuentra la póliza
<b>Rp</b>	Código único de la póliza
<b>Tipo documento</b>	Estado de la póliza
<b>Fecha generación</b>	Fecha en la que se generó la orden de emisión de la póliza
<b>Fecha vigencia</b>	Vigencia de la póliza, indica la fecha de inicio y fin en la que tiene cobertura
<b>Fuente</b>	Sistema del cual se extraen los datos
<b>Agente</b>	Bróker de Seguros
<b>Cliente</b>	Cliente que adquiere la póliza de fianzas (contratista)
<b>Beneficiario</b>	Beneficiario
<b>Dirección</b>	Lugar de domicilio o donde se encuentra localizada la empresa del cliente
<b>Teléfonos</b>	Corresponde a los números de teléfono convencionales y celulares del cliente
<b>Correo electrónico</b>	Dirección de correo electrónico del cliente
<b>Orden de documento</b>	Corresponde al tipo de documento de la póliza
<b>Garante</b>	Persona que avala que el contratista cumplirá con su contrato. Prendará sus bienes a nombre del contratista
<b>Grupo económico</b>	Tipo de emisión de la póliza
<b>Línea de Negocio</b>	Línea bajo la cual se emite la póliza
<b>Cob</b>	Código del cobrador

<b>Variable</b>	<b>Descripción</b>
<b>Factura real</b>	Número de la factura
<b>Fecha vencimiento</b>	Fecha en la que finaliza la cobertura de la póliza
<b>Ramo</b>	Ramo de la póliza emitida
<b>Póliza</b>	Número de la póliza emitida
<b>Endoso</b>	Tipo de póliza
<b>Estado póliza</b>	Estado de póliza
<b>Días vencidos</b>	# de días en los que se encuentra vencida la póliza
<b>Provisión</b>	Estimación de las pérdidas potenciales de una empresa por el riesgo asumido por el crédito dado a clientes en ventas de productos. Las empresas con frecuencia tienen un porcentaje de clientes que no pagarán sus deudas, entonces usan este dato para anticiparlas y protegerse del impacto de esos fondos perdidos.
<b>Vencimiento</b>	Estado de vencimiento de las pólizas
<b>Valor</b>	Valor adeudado
<b>Saldo</b>	Valor que se debe recuperar
<b>Prima</b>	Valor correspondiente a la póliza emitida
<b>Iva</b>	Impuesto del valor agregado

*Nota.* La tabla presenta el resumen de las variables establecidas para el desarrollo del modelo.

**Preprocesamiento (Limpieza de Datos).** La limpieza de datos permite asegurar la calidad de los datos que posteriormente se usarán para el modelamiento. Este paso es fundamental para reducir errores y obtener modelos más confiables.

RapidMiner es una herramienta que permite integrar otros lenguajes de programación, en este caso y en base a los conocimientos obtenidos en clases se utilizó R para llevar a cabo este paso. Se utilizó un método estadístico en el cual se analizó la data a través de métricas estadísticas que permitieron identificar errores. El paquete usado fue Summary tools. Este paquete muestra una visión global sobre las variables: frecuencia, datos perdidos y distribución, lo cual facilitó decidir las variables que se eliminarían del modelo (ver Figura 12).

## Figura. 12

### Función usada en R

```
# SE RETIRAN VARIABLES CONSTANTES O TODAS DIFERENTES
indice <- as.numeric()
for(i in seq_len(NCOL(X))){
  if(NROW(unique(X[,i, with = FALSE]))==NROW(X) | (NROW(unique(X[,i, with = FALSE]))==1)) {
    indice=c(indice,i)
  }
}
X <- X[,-c(indice), with = FALSE]

# ELIMINACION DE LAS VARIABLES CON ALTA CORRELACION
matrix_cor <- cor(X)
diag(matrix_cor) <- 999
ind_cor <- unique(row.names(which(abs(matrix_cor) > 0.70 & matrix_cor != 999, arr.ind=TRUE)))
```

*Nota.* La figura, muestra el código utilizado para realizar el proceso de eliminación de variables.

En este caso se eliminó la variable de Score, debido a que existía una ausencia de data superior al 20%, ya que el 26.8% de los clientes no contaban con una calificación; esto en base a los principios de Pareto, el cual señala que en un proceso o una situación el 20% de las causas generan el 80% de los resultados. La ausencia de esta información se debe a que en algunos casos los clientes son consorcios, los cuales no cuentan con una calificación debido a que son una agrupación de varias empresas y cada una de estas cuentan con una calificación individual. Otro factor que afecta a esta variable son las migraciones de sistemas por las que ha atravesado la línea de negocio. A partir del año 2017 se empezó a migrar la data de los clientes de un sistema transaccional a otro. En el sistema antiguo la creación de clientes se los hacía en base a su ruc o cédula y existían muchos duplicados. En el nuevo sistema registra a los clientes únicamente con cédula de identidad de identidad en caso de ser persona natural con o sin ruc y con ruc a las personas jurídicas, motivo por el que en muchos casos al cruzar la data existen faltantes. La tabla 14 muestra los porcentajes de calificaciones de clientes y aquellos que no cuentan con una calificación.

Tabla 13

Score de Clientes

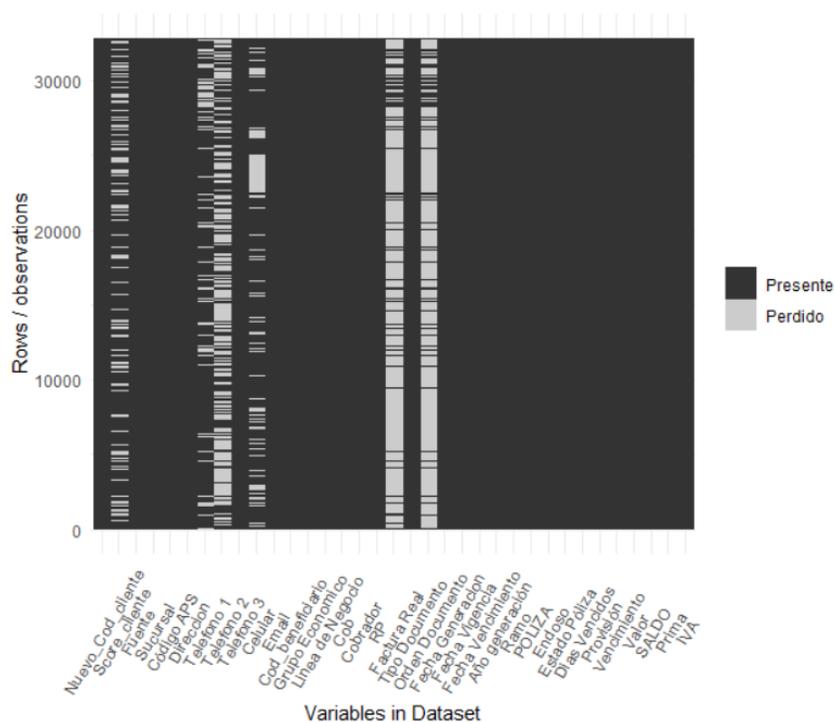
Score	% de participación
A	1.5%
AA	1.8%
AAA	1%
B	21.8%
BB	13.2%
BBB	5.1%
C	24.8%
D	30.8%
Sin Score	26.8%
<b>Total general</b>	<b>100%</b>

Nota. La tabla muestra el porcentaje de participación de los clientes en base a su score.

Para complementar el análisis se realizó un gráfico de valores perdidos para identificar los valores faltantes de la base de datos (ver Figura 13).

Figura. 13

Valores perdidos



Nota. La figura muestra los valores perdidos en el set de datos.

Los sectores grises representan la ausencia de valores en las variables analizadas.

En base al análisis realizado se pudo descartar las variables que no aportaban valor para el desarrollo del modelo, tal como se describe en la tabla 15. La información correspondiente al análisis estadístico realizado puede revisarse en el anexo 1.

**Tabla 14**

*Variables Eliminadas*

<b>Variable</b>	<b>Descripción</b>	<b>Motivo de eliminación</b>
<b>Fuente</b>	Sistema del cual se extraen los datos	Toda la información se consolida en el Data Warehouse, no tiene incidencia el sistema del que proviene la información, el 85% e proviene de un Sistema.
<b>Agente</b>	Bróker de Seguros	Se codificó la variable y se dejó únicamente el campo de código Agente. Se requiere anonimizar la data
<b>Cliente</b>	Cliente que adquiere la póliza de fianzas (contratista)	Se codificó la variable y se dejó únicamente el campo de código cliente. Se requiere anonimizar la data
<b>Beneficiario</b>	Beneficiario	Se codificó la variable y se dejó únicamente el campo de código cliente
<b>Dirección</b>	Lugar de domicilio o donde se encuentra localizada la empresa del cliente	Dato de contactabilidad del cliente, la información de la empresa se encuentra depurada y validada en un 75%
<b>Teléfonos (1,2 y 3)</b>	Corresponde a los números de teléfono convencionales y celulares del cliente	Dato de contactabilidad del cliente, la información de la empresa se encuentra depurada y validada en un 75%
<b>Correo electrónico</b>	Dirección de correo electrónico del cliente	Dato de contactabilidad del cliente, no incide en el modelo. La tasa de apertura de mails es menos del 19%
<b>Orden de documento</b>	Corresponde al tipo de documento de la póliza	97.3% de la base está vacía

Variable	Descripción	Motivo de eliminación
<b>Grupo económico</b>	Tipo de emisión de la póliza	El 97,2% corresponde a individual
<b>Línea de Negocio</b>	Línea bajo la cual se emite la póliza	Todas las pólizas corresponden a la línea de fianzas, sin embargo, existe un error en la extracción del reporte, ya que sale otro tipo de información
<b>Cob</b>	Código del cobrador	Es el código del cobrador
<b>Factura real</b>	Número de la factura	Indica el número de factura de la póliza. El 83.3% no tiene información
<b>Tipo de documento</b>	Indica el estado en la que se encuentra la póliza	Se revisó con fianzas e indicaron que es un campo que no se encuentra bien llenado y que no se tome en cuenta a esta información porqué no es real

*Nota.* La tabla muestra las variables eliminadas con su respectivo motivo de eliminación.

**Transformación de los Datos.** Dentro de esta fase se procedió a crear variables para cada uno de los clientes en base a los datos obtenidos de la base depurada. El número de variables obtenidos en este proceso fue de setenta y cinco.

Para estos análisis se tomaron en cuenta tres grandes categorías: Sucursal, las cuales se dividieron en Quito (concentra el 60% de la producción) y se agrupó al resto de las sucursales en la variable otras. El endoso, el cual indica si es: renovación, nuevo riesgo u otra (dentro de la variable otra se agrupó: cancelaciones, anulaciones, aumentos, extensiones) y el ramo: Buen Uso del Anticipo y Cumplimiento de Contrato. En la tabla 16 se muestra el detalle de las variables.

**Tabla 15**

*Variables Creadas*

Variable	Definición
Cliente	Código del cliente
Nuevo riesgoBACuenta de RP	Número de pólizas de nuevo riesgo de Buen uso del anticipo

<b>Variable</b>	<b>Definición</b>
Nuevo riesgoBAPromedio de Prima2	Promedio de prima de nuevo riesgo de buen uso del anticipo
Nuevo riesgoBAPromedio de Provisión	Promedio de provisión de nuevo riesgo de buen uso del anticipo
Nuevo riesgoBAPromedio de Saldo2	Promedio de saldo de nuevo riesgo de buen uso del anticipo
Nuevo riesgoBASuma de Prima	Suma de prima de nuevo riesgo de buen uso del anticipo
Nuevo riesgoBASuma de Saldo	Suma de saldo de nuevo riesgo de buen uso del anticipo
Nuevo riesgoCUCuenta de RP	Número de pólizas de nuevo riesgo de cumplimiento de contrato
Nuevo riesgoCUPromedio de Prima2	Promedio de prima de nuevo riesgo de cumplimiento de contrato
Nuevo riesgoCUPromedio de Provisión	Promedio de provisión de nuevo riesgo de cumplimiento de contrato
Nuevo riesgoCUPromedio de saldo2	Promedio de saldo de nuevo riesgo de cumplimiento de contrato
Nuevo riesgoCUSuma de Prima	Suma de prima de nuevo riesgo de cumplimiento de contrato
Nuevo riesgoCUSuma de saldo	Suma de saldo de nuevo riesgo de cumplimiento de contrato
OtroBACuenta de RP	Número de primas de otro tipo de endoso de buen uso del anticipo
OtroBAPromedio de Prima2	Promedio de prima de otro tipo de endoso de buen uso del anticipo
OtroBAPromedio de Provisión	Promedio de provisión de otro tipo de endoso de buen uso del anticipo
OtroBAPromedio de Saldo2	Promedio de saldo de otro tipo de endoso de buen uso del anticipo
OtroBASuma de Prima	Suma de prima de otro tipo de endoso de buen uso del anticipo
OtroBASuma de saldo	Suma de saldo de otro tipo de endoso de buen uso del anticipo
OtroCUCuenta de RP	Número de pólizas de otro tipo de endoso de cumplimiento de contrato
OtroCUPromedio de Prima2	Promedio de prima de otro tipo de endoso de cumplimiento de contrato
OtroCUPromedio de Provisión	Promedio de provisión de otro tipo de endoso de cumplimiento de contrato
OtroCUPromedio de saldo2	Promedio de saldo de otro tipo de endoso de cumplimiento de contrato
OtroCUSuma de Prima	Suma de prima de otro tipo de endoso de cumplimiento de contrato
OtroCUSuma de saldo	Suma de saldo de otro tipo de endoso de cumplimiento de contrato
RenovaciónBACuenta de RP	Número de pólizas de renovación de Buen uso del anticipo

<b>Variable</b>	<b>Definición</b>
RenovaciónBAPromedio de Prima <sup>2</sup>	Promedio de prima de renovación de buen uso del anticipo
RenovaciónBAPromedio de Provisión	Promedio de provisión de renovación de buen uso del anticipo
RenovaciónBAPromedio de Saldo <sup>2</sup>	Promedio de saldo de renovación de buen uso del anticipo
RenovaciónBASuma de Prima	Suma de prima de renovación de buen uso del anticipo
RenovaciónBASuma de Saldo	Suma de saldo de renovación de buen uso del anticipo
RenovaciónCUCuenta de RP	Número de pólizas de renovación de cumplimiento de contrato
RenovaciónCUPromedio de Prima <sup>2</sup>	Promedio de prima de renovación de cumplimiento de contrato
RenovaciónCUPromedio de Provisión	Promedio de provisión de renovación de cumplimiento de contrato
RenovaciónCUPromedio de saldo <sup>2</sup>	Promedio de saldo de renovación de cumplimiento de contrato
RenovaciónCUSuma de Prima	Suma de prima de renovación de cumplimiento de contrato
RenovaciónCUSuma de Saldo	Suma de saldo de renovación de cumplimiento de contrato
Promedio días vencidos	Promedio de días vencidos de las pólizas

*Nota.* La tabla muestra el resumen de las variables que se crearon para el desarrollo del modelo.

En base a esto, se realizaron los cálculos de: promedios, días máximos y sumas de las variables que se mantenían en un principio, para poder identificar los comportamientos de los clientes morosos de la línea de fianzas.

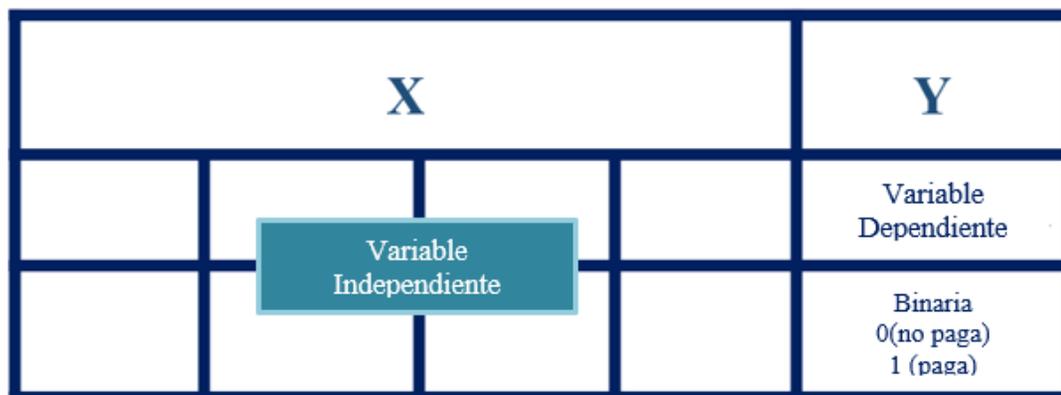
Se procedió a definir la variable dependiente en conjunto con el área de negocio responsable, para lo cual se realizó un análisis basado en el número de días que mantenía deuda. Para el análisis de esta variable se aplicó “change detection” o detección de cambio, dicho algoritmo permite identificar los tiempos en que la distribución de probabilidad de un proceso estocástico o una serie temporal cambia. Es decir, permite visualizar el punto de inflexión de la data. Para la identificación de esto se creó un histograma tomando en cuenta el número de días promedio que mantienen la deuda los clientes; cuyos resultados se muestran en la Figura 14:

**Figura. 14***Días Vencidos*

*Nota.* La figura permite ver el punto de inflexión de la data analizada.

En base a las políticas del negocio y al análisis realizado, se definió que el rango de días a partir de los que se considerará como un cliente que no paga son 90 (uno paga, cero no paga). Sobre la base de los principios del punto de inflexión, en donde los puntos de la gráfica de una función cambian de concavidad. Este valor fue validado con el negocio y las políticas internas de pago.

Una vez realizados los procesos de limpieza y la definición de la variable dependiente se obtuvo el conjunto final para usarlo en el modelamiento de datos en la herramienta RapidMiner. En la figura 15 se puede visualizar la estructura de los datos:

**Figura. 15***Estructura de la Data*

*Nota.* La figura muestra la estructura que se usó para realizar el modelamiento de la base de datos.

**Minería de datos.** Conforme a la metodología seleccionada para el desarrollo de este estudio, se siguieron los pasos establecidos por KDD los cuales se detallan a continuación:

**Selección de algoritmos.** Se aplicaron tres diferentes algoritmos: Árbol de Clasificación, Redes de Bayes y Redes Neuronales. Se definió el uso de estos algoritmos en base a la investigación realizada en el Capítulo 2, en el que se evidenció que empresas de seguros y del sector financiero obtuvieron resultados positivos aplicando estos modelos.

En la tabla 17 se presentan las ventajas y desventajas de los modelos anteriormente mencionados para poder identificar las fortalezas y debilidades de cada uno de ellos.

Tabla 16

*Ventajas y Desventajas de los Modelos Seleccionados*

<b>Modelo</b>	<b>Ventajas</b>	<b>Desventajas</b>
Redes Neuronales	<ul style="list-style-type: none"> <li>✓ Aprendizaje Adaptativo. Capacidad de aprender a realizar tareas basadas en una experiencia inicial o en un entrenamiento</li> <li>✓ Tolerancia a fallos. Almacena la información de forma redundante, lo que le permite seguir respondiendo de forma aceptable aun si se daña parcialmente.</li> <li>✓ Operación en tiempo real. La estructura es paralela, lo que le permite al conectarse con dispositivos electrónicos obtener resultados en tiempo real.</li> <li>✓ Auto organización: Crea su propia representación de la información en su interior, reduciendo carga en el usuario.</li> <li>✓ Flexibilidad: Permite realizar cambios no importantes en la información de entrada, como señales con ruido u otros cambios en la entrada</li> </ul>	<ul style="list-style-type: none"> <li>✓ Complejidad de aprendizaje, mientras más cosas se necesiten que aprenda se vuelve más complejo enseñarle.</li> <li>✓ Requiere de tiempos más extensos de entrenamiento</li> <li>✓ Tiene el efecto "Black box", lo cual hace más difícil la comprensión de las relaciones entre variables dependientes e independientes</li> <li>✓ Requiere de muchos datos para el entrenamiento</li> <li>✓ No cuenta con reglas definitorias que ayuden a realizar una red para un problema dado.</li> </ul>
Árbol de decisión	<ul style="list-style-type: none"> <li>✓ Son fáciles de entender</li> <li>✓ Analiza las consecuencias de llevar a cabo una alternativa.</li> <li>✓ Facilita la interpretación de la decisión adoptada.</li> <li>✓ Muestra un esquema de reglas que permiten realizar estrategias</li> <li>✓ Plantea el problema para que todas las opciones sean analizadas</li> <li>✓ Se pueden agregar nuevas opciones a los árboles existente</li> </ul>	<ul style="list-style-type: none"> <li>✓ Dificultad para elegir un árbol óptimo.</li> <li>✓ Ausencia de una función global de las variables y por lo tanto pérdida de la representación.</li> <li>✓ Los árboles de decisión requieren un gran número de datos de los que muchas veces no se dispone.</li> </ul>
Red de Bayes	<ul style="list-style-type: none"> <li>✓ Es una poderosa herramienta de comunicación</li> <li>✓ Relaciones causa-efecto se visualizan fácilmente sin</li> </ul>	<ul style="list-style-type: none"> <li>✓ Depende mucho del conocimiento de la persona que modela</li> </ul>

Modelo	Ventajas	Desventajas
	<ul style="list-style-type: none"> <li>✓ necesidad de cálculo de probabilidades</li> <li>✓ Puede modelar sistemas complejos</li> <li>✓ Se puede actualizar la red fácilmente por cambio en la información</li> <li>✓ Están basadas en algoritmos de propagación de evidencias que actualizan dinámicamente el modelo con datos actuales</li> <li>✓ Maneja algoritmos que permite manejar de manera eficiente la información faltante</li> <li>✓ Permiten reducir el sobre ajuste de los datos.</li> <li>✓ Posibilitan el descubrimiento de la estructura causal subyacente en un conjunto de datos.</li> </ul>	<ul style="list-style-type: none"> <li>✓ Es difícil describir una estructura compleja</li> <li>✓ Posee variables ocultas</li> <li>✓ Posee probabilidades inconsistentes</li> <li>✓ Es útil sólo si se puede confiar en los parámetros. Una mala estimación de los parámetros distorsiona toda la red e invalida los resultados.</li> <li>✓ existen numerosas fuentes de error como por ejemplo en la discretización de variables, ruido, valores perdidos en la base de datos</li> </ul>

*Nota.* La tabla muestra las principales ventajas y desventajas de los modelos preseleccionados.

**División de la Base de Datos en entrenamiento y prueba.** La división de la base de datos permite separar la base en dos conjuntos. El primero que sirve para entrenar el modelo y la segunda parte que habilita evaluar el modelo. Para llevar a cabo esto, se utilizó la herramienta RapidMiner para dividir un porcentaje de la base en entrenamiento y validación. De esta manera, se colocó el 80% para entrenamiento y el 20% para validación. Se utilizan estos porcentajes debido a los parámetros sugeridos por los autores de la herramienta y porque en la práctica son los valores más comunes que se utilizan. La partición de la base permite tener dos conjuntos de datos (ver figura 16):

- **Conjunto de entrenamiento:** Es un subconjunto para entrenar el modelo.
- **Conjunto de prueba:** Es un subconjunto para probar el modelo entrenado.

**Figura. 16**

*División de la Base*

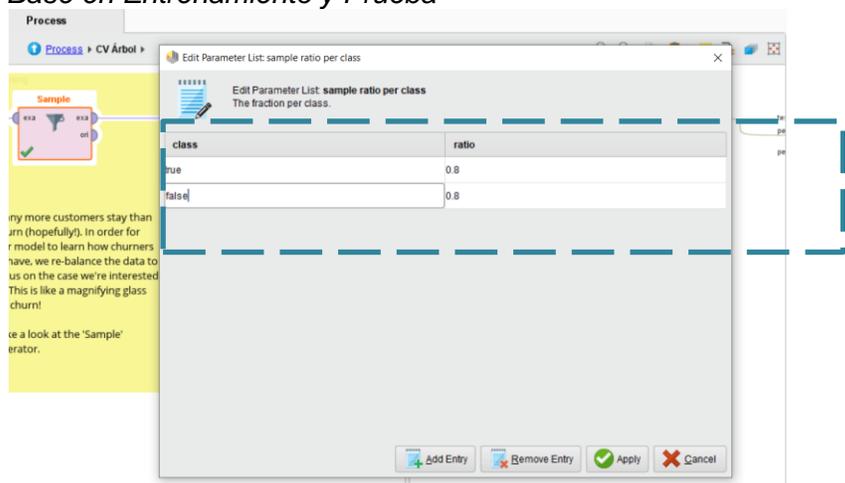
<b>X</b> Entrenamiento	80%	<b>Y</b> Entrenamiento
<b>X</b> Prueba	20%	<b>Y</b> Prueba

*Nota.* La figura muestra la división de la base de datos de una forma gráfica.

La partición se la realizó utilizando el componente sample de la herramienta, el cual permite colocar los porcentajes que corresponderán a entrenamiento y prueba. En la figura 17 se muestra la pantalla de la herramienta.

**Figura. 17**

*División de la Base en Entrenamiento y Prueba*



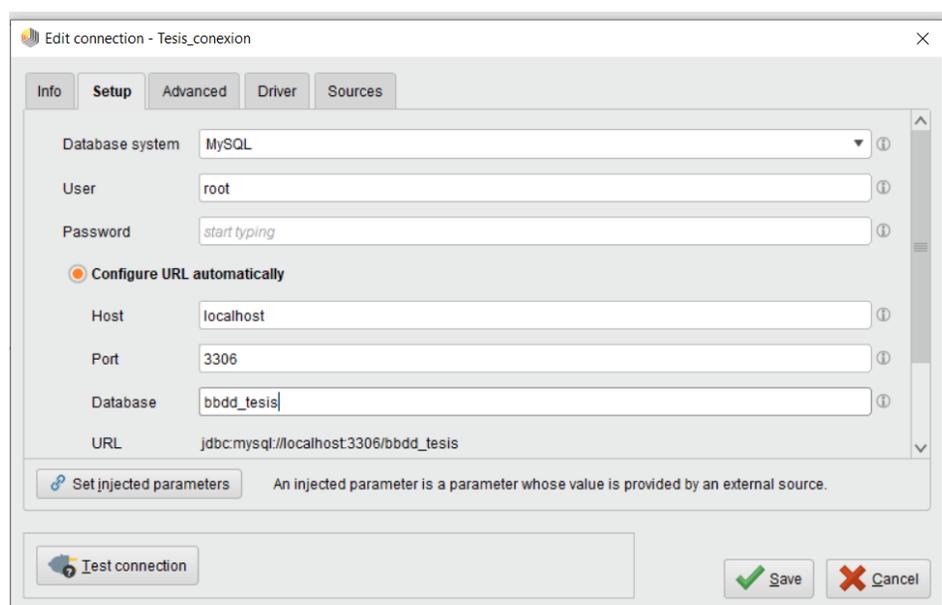
*Nota.* La figura permite visualizar la partición realizada en la herramienta de la base en entrenamiento y prueba.

**Aplicación de Algoritmos.** Como se pudo observar en el capítulo 4, RapidMiner cuenta con plantillas preestablecidas para poder modelar y se aprovechó este beneficio para el desarrollo de los modelos. En la pantalla de inicio se seleccionó

la opción Churn Modeling, que se especializa en realizar modelos a través de árboles de decisión. Se realizó una conexión a través de Workbench, que es una herramienta visual para diseño de base de datos. A partir de esta herramienta se creó la base en MySQL con el objetivo de que la herramienta RapidMiner pueda trabajar a partir de la base. En la herramienta RapidMiner se creó el repositorio Tesis\_repositorio\_2 y en la figura 18 se puede apreciar la conexión de RapidMiner con MySQL.

**Figura. 18**

### *Conexión MySQL*



*Nota.* La figura muestra la conexión realizada a través de la herramienta RapidMiner con MySQL.

Se elaboró un flujo en la mesa de trabajo para poder correr los tres modelos preseleccionados. Para obtener un mejor resultado se aplicó Cross Validation en cada uno de los modelos.

Para evitar la correlación de variables se utilizó el coeficiente de Pearson, el cual es una prueba que mide la relación estadística entre dos variables continuas (Question

Pro, 2020). El coeficiente de Pearson puede ubicarse entre 1 y -1; si el valor es 0 quiere decir que no existe asociación entre las variables. En la tabla 18 se detalla los rangos de correlación de las variables.

**Tabla 17**

*Rangos de Correlación*

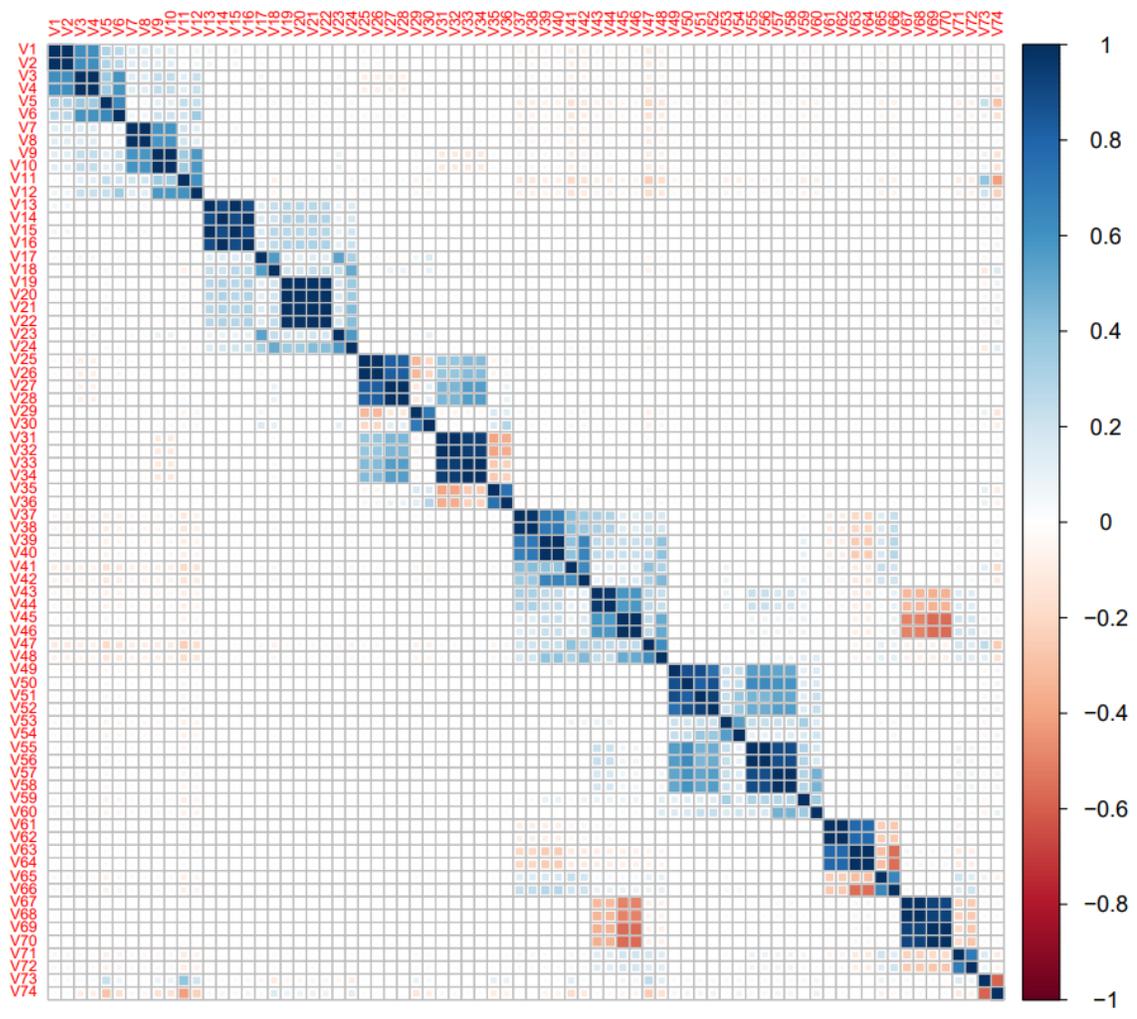
<b>Rango</b>	<b>Correlación</b>
0.6 a 0.7	Correlación positiva moderada
0.7 a 0.9	Correlación positiva fuerte
0.9 a 1	Correlación positiva muy fuerte

*Nota.* La tabla muestra los rangos establecidos para definir los tipos de correlaciones, tomado de Métricas de la web social para bibliotecas.

El proceso realizado para la eliminación de variables constó de tres pasos. En primer lugar, se creó una matriz de correlación, que permite evaluar la fuerza y dirección de la relación entre dos variables, como se muestra en la figura 19. El detalle de las variables se encuentra en el anexo 2.

Figura. 19

Matriz de Correlación

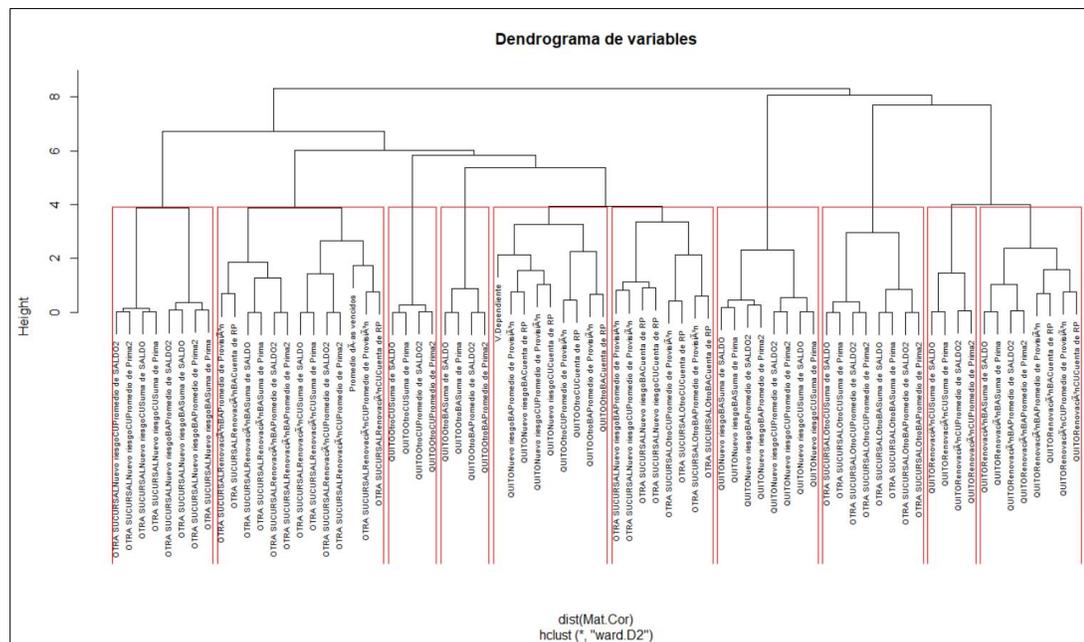


Nota. La figura muestra las correlaciones existentes entre las variables de la base de datos.

Esta matriz permite visualizar que las variables que se encuentran en los extremos tienen una fuerte correlación. En segundo lugar, se aplicó el algoritmo de clúster jerárquico, el cual permite agrupar a las variables de forma homogénea y asegura que entre grupos exista heterogeneidad. La figura 20 muestra el dendograma obtenido:

Figura. 20

## Clúster Jerárquico



Nota. Se puede apreciar el dendrograma realizado a partir de la base de datos establecida.

Se puede observar que, dentro de los cuadrados marcados en rojo, la herramienta ha agrupado a aquellas variables que son homogéneas. En base al análisis de esta data se obtuvo diez clústers. Los clústers se agrupan basados en la sucursal y los ramos analizados. Las variables que toma en cuenta principalmente para hacer los grupos homogéneos son: saldos, prima, renovaciones y nuevo riesgo.

Finalmente se utilizó la técnica de Análisis de Componentes Principales (ACP), mismo que describe un conjunto de datos en términos de nuevas variables («componentes») no correlacionadas. Los componentes se ordenan por la cantidad de varianza original que describen. La figura 21 muestra la función usada para obtener los nuevos componentes.

## Figura. 21

### Función Clúster Jerárquico

```
#cluster jerárquico (agrupar segentos que tienen más valor)

library(factoextra)
library(cluster)
library(ggpubr)

alg.agnes <- agnes(x=Mat.Cor, stand=TRUE, metric="euclidean", method="ward")
fviz_dend(alg.agnes, k = 5, cex = 0.5,
          # k_colors = c("#2E9FDF", "#b000a1", "#E7B800", "#FC4E07", "#09d102"),
          color_labels_by_k = TRUE, rect = TRUE)

clusters <- hclust(Mat.Cor, method = "ward.D2" )
hclust_cor <- hclust(dist(Mat.Cor), method = "ward.D2")
plot(hclust_cor, cex = 0.6, main = "Dendrograma de variables")
rect.hclust(hclust_cor, k = 10)
```

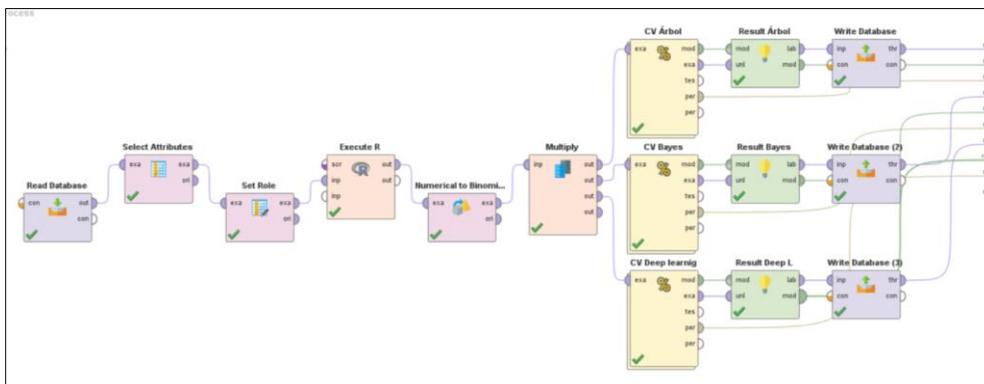
Nota. La figura muestra el código usado en R estudio para la aplicación para clúster Jerárquico.

Una vez que se realizaron todos los pasos previos, se obtuvo la base depurada, y se identificaron treinta y dos variables con las cuales se realizó el modelamiento. El flujo presentado a continuación permitió obtener el modelo ganador.

La primera parte del modelo permite leer la base de datos desde Workbench, y seleccionar todos los atributos que intervendrán en el modelamiento (32 variables), y la eliminación de las variables que tienen mayor correlación. La segunda parte del modelo corresponde a la aplicación de logaritmos, donde se puede configurar a cada uno de ellos para obtener los resultados. Finalmente, la última parte del modelo arroja los resultados. En la figura 22 se muestra el proceso de modelamiento utilizado.

Figura. 22

Proceso de Modelamiento

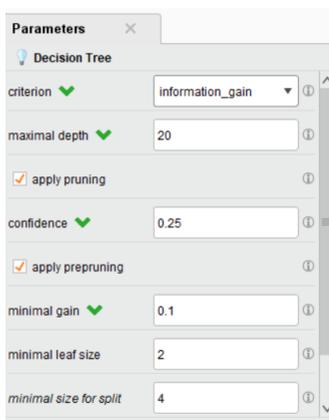


Nota. La figura muestra el proceso realizado en RapidMiner para la obtención de los modelos.

**Árbol de decisión.** Los parámetros definidos para este modelo son los detallados en la Figura 23.

Figura. 23

Parámetros del Árbol de Decisión



Nota. La figura muestra los parámetros que fueron usados para la elaboración del modelo de árbol de decisión.

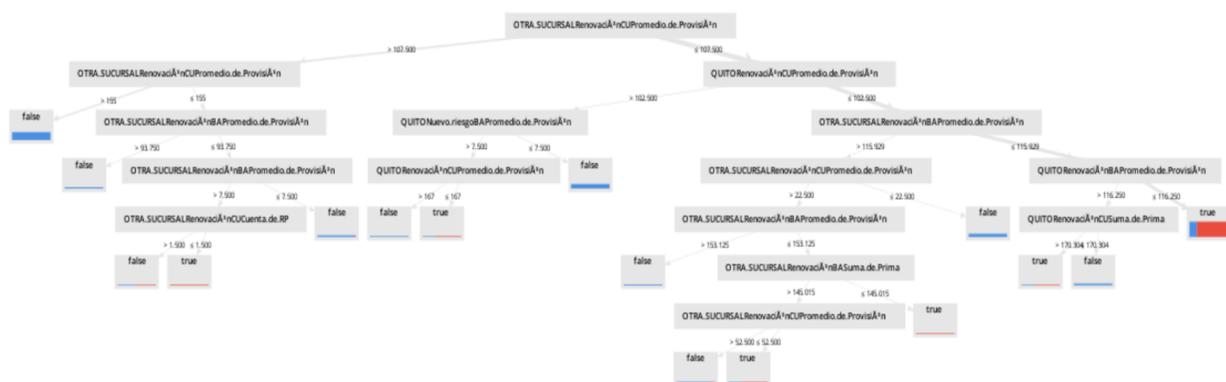
Para usar este modelo se debe configurar siete parámetros. La herramienta coloca sugerencias de valores, los cuales pueden ser modificados, dependiendo del

criterio de la persona que está desarrollando el modelo. En este caso la mayor parte de los criterios se mantuvieron en base a lo que se encontraba predefinido. Se escogió el criterio basado en “Information Gain”. Este operador calcula la relevancia de los atributos basándose en la ganancia de información y les asigna ponderaciones en consecuencia cuanto mayor sea el peso del atributo, más relevante se considerará. (RapidMiner Studio Core, 2020)

RapidMiner cuenta con un módulo de parámetros el cual permite configurar los operadores para efectuar un algoritmo. Tomando en cuenta esto se decidió mantener las sugerencias de la herramienta: veinte niveles de profundidad, función de poda del árbol activada y un intervalo de confianza (tolerancia) para aberturas de ramas de 0,25. Para que una rama se extienda se mantiene la sugerencia de la herramienta de 0,1 al igual que para que el nodo final exista se mantiene dos elementos y el número mínimo de elementos que deben permanecer para abrir un nodo cuatro. En base a estos criterios se obtuvo los resultados que se detallan en la Figura 24:

**Figura. 24**

### Árbol de Decisión



*Nota.* La figura muestra el árbol de decisión obtenido en la herramienta.

Una de las ventajas del árbol de decisión es que entrega las reglas en las cuales se basó el modelo. Estas reglas nos permiten identificar condiciones que pueden ser aplicadas para la elaboración de las estrategias. A continuación, se detallan las principales reglas obtenidas por el modelo.

```

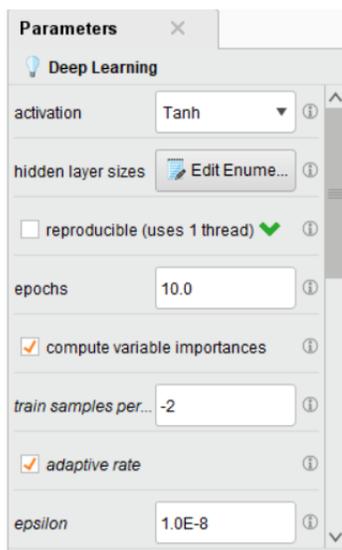
OTRA.SUCURSALRenovaciOnCUPromedio.de.ProvisiOn > 107.500
| OTRA.SUCURSALRenovaciOnCUPromedio.de.ProvisiOn > 155: false {false=563,
true=0}
| OTRA.SUCURSALRenovaciOnCUPromedio.de.ProvisiOn ≤ 155
| | OTRA.SUCURSALRenovaciOnBAPromedio.de.ProvisiOn > 93.750: false
{false=33, true=0}
| | OTRA.SUCURSALRenovaciOnBAPromedio.de.ProvisiOn ≤ 93.750
| | | OTRA.SUCURSALRenovaciOnBAPromedio.de.ProvisiOn > 7.500
| | | | OTRA.SUCURSALRenovaciOnCUCuenta.de.RP > 1.500: false {false=1,
true=1}
| | | | OTRA.SUCURSALRenovaciOnCUCuenta.de.RP ≤ 1.500: true {false=0,
true=5}
| | | OTRA.SUCURSALRenovaciOnBAPromedio.de.ProvisiOn ≤ 7.500: false
{false=65, true=2}
OTRA.SUCURSALRenovaciOnCUPromedio.de.ProvisiOn ≤ 107.500
| QUITORenovaciOnCUPromedio.de.ProvisiOn > 102.500
| | QUITONuevo.riesgoBAPromedio.de.ProvisiOn > 7.500
| | | QUITORenovaciOnCUPromedio.de.ProvisiOn > 167: false {false=7, true=0}
| | | QUITORenovaciOnCUPromedio.de.ProvisiOn ≤ 167: true {false=1, true=2}
| | | QUITONuevo.riesgoBAPromedio.de.ProvisiOn ≤ 7.500: false {false=290, true=1}
| | QUITORenovaciOnCUPromedio.de.ProvisiOn ≤ 102.500
| | | OTRA.SUCURSALRenovaciOnBAPromedio.de.ProvisiOn > 115.929
| | | | OTRA.SUCURSALRenovaciOnCUPromedio.de.ProvisiOn > 22.500
| | | | | OTRA.SUCURSALRenovaciOnBAPromedio.de.ProvisiOn > 153.125: false
{false=18, true=0}
| | | | | OTRA.SUCURSALRenovaciOnBAPromedio.de.ProvisiOn ≤ 153.125
| | | | | | OTRA.SUCURSALRenovaciOnBASuma.de.Prima > 145.015
| | | | | | | OTRA.SUCURSALRenovaciOnCUPromedio.de.ProvisiOn > 52.500: false
{false=9, true=1}
| | | | | | | OTRA.SUCURSALRenovaciOnCUPromedio.de.ProvisiOn ≤ 52.500: true
{false=1, true=2}
| | | | | | | OTRA.SUCURSALRenovaciOnBASuma.de.Prima ≤ 145.015: true {false=0,
true=2}
| | | | OTRA.SUCURSALRenovaciOnCUPromedio.de.ProvisiOn ≤ 22.500: false
{false=189, true=0}
| | | OTRA.SUCURSALRenovaciOnBAPromedio.de.ProvisiOn ≤ 115.929
| | | | QUITORenovaciOnBAPromedio.de.ProvisiOn > 116.250
| | | | | QUITORenovaciOnCUSuma.de.Prima > 170.304: true {false=1, true=2}
| | | | | QUITORenovaciOnCUSuma.de.Prima ≤ 170.304: false {false=79, true=1}
| | | | QUITORenovaciOnBAPromedio.de.ProvisiOn ≤ 116.250: true {false=221,
true=929}

```

**Redes neuronales (Deep learning).** De igual manera que en el modelamiento del árbol de decisiones se puede modificar la configuración de los parámetros de este modelo. Para este desarrollo se debe tomar en cuenta doce parámetros. Se tomó la sugerencia realizada por la herramienta como se observa en la figura 25.

**Figura. 25**

*Parámetros Redes Neuronales*



*Nota.* La figura muestra los parámetros que fueron usados para la elaboración del modelo de Redes Neuronales.

La función de activación se encarga de devolver una salida a partir de un valor de entrada. Para el desarrollo de este modelo es Tanh, misma que se utiliza para decidir entre una opción y la contraria y tiene un buen desempeño en redes recurrentes. Esta función en conjunto con la sigmoide, ReLU y softmax, son las más usadas en Redes Neuronales. (Calvo, 2018). Esta función es muy útil en la capa final de salida al final de la red neuronal, no solo para clasificar con valores categóricos, sino también para intentar predecir las probabilidades de pertenencia a cada categoría, donde se conoce que la probabilidad de un suceso imposible es 0 y la de un suceso seguro es 1

(Villanueva, 2019). En el resto de los valores se mantienen las sugerencias de la herramienta: Epoch, lo cual corresponde al número de veces que se usarán los datos en el proceso de aprendizaje: 10. Se mantiene activa la función de cálculo de importancia de variables el cual considera el peso que conecta las características de entrada a las dos primeras capas ocultas. La muestra de tren de iteración, la cual controla la frecuencia con la que puede ocurrir la puntuación y la cancelación del modelo, se mantiene en -2 debido a que los valores especiales son 0 para una época por iteración, -1 para procesar la cantidad máxima de datos por iteración y el valor especial de -2 vueltas en modo automático. La tasa de adaptación se mantiene activa ya que evita una convergencia lenta.

La herramienta trabaja las Redes Neuronales bajo el aprendizaje profundo de H2O, el cual incluye muchos algoritmos comunes de aprendizaje automático. H2O implementa los mejores algoritmos de su clase a escala, como bosque aleatorio distribuido, aumento de gradiente, y aprendizaje profundo. Además, incluye un método de conjuntos apilados, que encuentra la combinación óptima de una colección de algoritmos de predicción utilizando un proceso conocido como "apilamiento". Con este método se puede construir miles de modelos y comparar los resultados para obtener las mejores predicciones. (Candel & LeDell, 2018)

A continuación, se detalla un extracto del modelo en la Figura 26. El modelo completo se lo puede ver en el Anexo 3

**Figura. 26***Extracto de Modelo Deep Learning*

```

Deep Learning Model
Model Metrics Type: Binomial
Description: Metrics reported on full training frame
model id: rm-h2o-model-deep_learning-567793
frame id: rm-h2o-frame-deep_learning-59544
MSE: 0.061255682
RMSE: 0.24749885
AUC: 0.9762637
pr_auc: 0.9550174
logloss: 0.20021443
mean_per_class_error: 0.06888728
default threshold: 0.5788865685462952
CM: Confusion Matrix (Row labels: Actual class; Column labels: Predicted class):
      false true Error Rate
false 1357 121 0.0819 121 / 1,478
true   53  895 0.0559  53 / 948
Totals 1410 1016 0.0717 174 / 2,426
Gains/Lift Table (Avg response rate: 39.08 %, avg score: 41.57 %):

```

Nota. La figura muestra un extracto del resumen elaborado por la herramienta del modelo de Redes Neuronales.

A través de este resumen se puede visualizar los resultados obtenidos de las métricas del performance del modelo y la matriz de confusión. Los valores arrojados en este resumen servirán para analizar los resultados posteriormente y seleccionar el mejor modelo para la empresa.

A pesar de que Redes Neuronales se lo conoce como “una caja negra” RapidMiner permite obtener a través de un componente las principales variables que inciden en el modelo, y se visualizan en la Figura 27:

**Figura. 27***Importancia de Variables*

Variable	Relative Importance	Scaled Importance	Percentage
OTRA.SUCURSALRenovaciÃ³nBAPromedio.de.ProvisiÃ³n	1.000000	1.000000	0.057734
OTRA.SUCURSALRenovaciÃ³nCUPromedio.de.ProvisiÃ³n	0.926407	0.926407	0.053486
QUITOREnovaciÃ³nCUPromedio.de.ProvisiÃ³n	0.786393	0.786393	0.045402
OTRA.SUCURSALRenovaciÃ³nCUCuenta.de.RP	0.778446	0.778446	0.044943
QUITOREnovaciÃ³nBAPromedio.de.ProvisiÃ³n	0.722507	0.722507	0.041713
OTRA.SUCURSALNuevo.riesgoBAPromedio.de.ProvisiÃ³n	0.698819	0.698819	0.040346
QUITONuevo.riesgoBAPromedio.de.ProvisiÃ³n	0.677417	0.677417	0.039110
OTRA.SUCURSALRenovaciÃ³nBACuenta.de.RP	0.665622	0.665622	0.038429
QUITOOtroCUSuma.de.SALDO	0.643142	0.643142	0.037131
QUITOOtroBAPromedio.de.ProvisiÃ³n	0.625320	0.625320	0.036102
---			
QUITOREnovaciÃ³nBASuma.de.SALDO	0.557345	0.557345	0.032178
OTRA.SUCURSALOtroCUCuenta.de.RP	0.556187	0.556187	0.032111
OTRA.SUCURSALNuevo.riesgoBACuenta.de.RP	0.546788	0.546788	0.031568
QUITOOtroBACuenta.de.RP	0.539140	0.539140	0.031127
QUITOOtroBAPromedio.de.SALDO2	0.504219	0.504219	0.029111
QUITOREnovaciÃ³nBACuenta.de.RP	0.503712	0.503712	0.029082
QUITONuevo.riesgoCUSuma.de.SALDO	0.502459	0.502459	0.029009
OTRA.SUCURSALNuevo.riesgoCUPromedio.de.SALDO2	0.481499	0.481499	0.027799
QUITOREnovaciÃ³nCUSuma.de.Prima	0.466613	0.466613	0.026940
OTRA.SUCURSALRenovaciÃ³nBASuma.de.Prima	0.466049	0.466049	0.026907

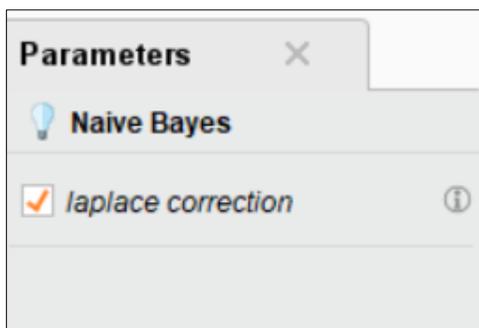
*Nota.* LA figura muestra los resultados obtenidos de las variables más importantes para el desarrollo del modelo.

En base a los resultados obtenidos, se puede visualizar que las variables que tienen una mayor incidencia en el modelo corresponden a aquellas que están dentro de la categoría Renovaciones, otras sucursales, cumplimiento de contrato y buen uso del anticipo en base a la provisión. También tiene una incidencia considerable las variables que se encuentran dentro de nuevo riesgo y de la provisión. Con una incidencia media se puede observar que se toma en cuenta las variables que se refieren al número de pólizas y saldo.

**Redes de Bayes.** Para el desarrollo de este modelo se establecieron los parámetros detallados en la Figura 28:

**Figura. 28**

*Parámetros de Redes de Bayes*



*Nota.* La figura muestra el parámetro seleccionado para el desarrollo del modelo de Redes de Bayes.

Este parámetro permite obtener resultados precisos ya que en el caso de que en el entrenamiento un valor de atributo la probabilidad condicional se puede establecer en cero. Cuando este valor cero se multiplican junto con otras probabilidades, esos valores también se establecen en cero y los resultados no serán los más adecuados. La corrección de Laplace permite evitar este problema, agregando uno a cada conteo para evitar la ocurrencia de valores cero. Para la mayoría de los conjuntos de entrenamiento, agregar uno a cada conteo solo tiene un efecto insignificante en las probabilidades estimadas (RapidMiner Studio Core, 2020).

**Evaluación.** Una vez obtenidos los modelos, es importante realizar su evaluación para poder definir cuál se adapta más a las necesidades del negocio. En este paso se da respuesta al objetivo cuatro de este proyecto. Para la selección del modelo se usaron las siguientes técnicas y métricas de validación: validación cruzada, la técnica de la matriz de confusión; a partir de esta técnica se obtuvieron las siguientes

métricas: exactitud, especificidad y sensibilidad. También se obtendrá las siguientes métricas: el lift, El espacio ROC (AUC), recall.

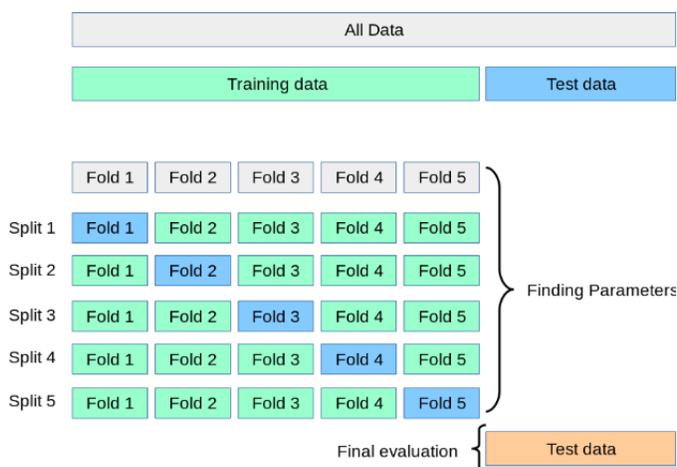
A continuación, se describe cada una de las herramientas y métricas a ser utilizadas, para poder definir las variables con las cuáles se evaluará los resultados se presentó una propuesta y se corroboró con los responsables que las métricas a evaluar son las adecuadas:

**Validación Cruzada (cross validation).** La técnica de Validación

*Cruzada* permite medir el comportamiento de los modelos creados y nos ayudará a encontrar un mejor modelo rápidamente. Esta técnica se encarga de repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones. Es decir, quita un dato muestral y lo predice mediante el modelo determinado y compara el valor obtenido de la predicción con el valor real; el procedimiento se repite varias veces. La evaluación de las diferentes validaciones cruzadas se da por el error obtenido, el que posee el valor más bajo es el modelo más acertado, como se puede observar en la figura 29.

**Figura 29 .**

*Cross Validation*



*Nota:* Tomado de scikit-learn.org, (<https://scikit-learn.org/stable/index.html>).

**Matriz de Confusión.** Es una herramienta útil para evaluar el desempeño de un algoritmo de clasificación a partir del conteo de aciertos y errores de cada una de las clases en la clasificación. Esto comprueba si el algoritmo está clasificando mal, las clases no son adecuadas y en qué medida.

**Tabla 18**

*Matriz de Confusión*

	Predijo Si	Predijo No
Positivos (Real Si)	Verdadero Positivo (VP)	Falso Positivo (FN)
Negativos (Real No)	Falso Negativo (FP)	Verdadero Negativo (VN)

Nota. La tabla muestra el resumen de la matriz de confusión, tomado de Ecured, 2020, [https://www.ecured.cu/Matrices\\_de\\_confusi%C3%B3n](https://www.ecured.cu/Matrices_de_confusi%C3%B3n).

**VP.** Cantidad de positivos que fueron clasificados correctamente como positivos por el modelo.

**VN.** Cantidad de negativos que fueron clasificados correctamente como negativos por el modelo.

**FN.** Cantidad de positivos que fueron clasificados incorrectamente como negativos. Error tipo 2 (Falsos Negativos).

**FP.** Cantidad de negativos que fueron clasificados incorrectamente como positivos. Error tipo 1 (Falsos positivos).

La matriz de confusión nos permite obtener indicadores adicionales que se detallan a continuación:

**Exactitud:** Proporción del número total de predicciones que fueron correctas:

**Precisión (P).** Proporción de casos predichos positivos que fueron correctos

**Error de Clasificación.** Corresponde a la diferencia entre el valor medido y el "valor verdadero".

**Sensibilidad.** Se conoce como Tasa de Verdaderos Positivos (True Positive Rate) ó TP. Es la proporción de casos positivos que fueron correctamente identificadas por el algoritmo.

**Especificidad.** Conocida como Tasa de Verdaderos Negativos, ("true negative rate") o TN. Son los casos negativos que el algoritmo ha clasificado correctamente.

**Lift** Mide el cambio en términos de una puntuación de elevación. Al comparar los puntajes de elevación para diferentes modelos, puede determinar qué modelo es el mejor.

**Recall o Exhaustividad,** Informa sobre la cantidad que el modelo de Machine Learning es capaz de identificar. Es decir, responde a la pregunta: ¿Qué proporción de positivos reales se identificó correctamente?

**Espacio ROC.** Un espacio ROC se define por FPR y VPR como ejes x e y respectivamente, y representa los intercambios entre verdaderos positivos (en principio, beneficios) y falsos positivos (en principio, costes). Dado que VPR es equivalente a sensibilidad y FPR es igual a 1-especificidad, el gráfico ROC también es conocido como la representación de sensibilidad frente a (1-especificidad). Cada resultado de predicción o instancia de la matriz de confusión representa un punto en el espacio ROC.

En base a las métricas revisadas y a la validación con los expertos en la empresa se definió los rangos de calificación de las medidas de precisión y exactitud, el resto de las medidas que se tomaron en cuenta para evaluación, se consideraron los valores óptimos. En la tabla 20, se detalla los parámetros que se analizaron para la evaluación del modelo.

**Tabla 19***Guía de Interpretación Medidas de Precisión y Exactitud*

<b>Rangos</b>	<b>Detalles</b>
<b>0.5 a 0.6</b>	Test deficiente
<b>0.61 a 0.75</b>	Test Regular
<b>0.76 a 0.9</b>	Test Bueno
<b>0.9 a 0.97</b>	Test muy bueno
<b>0.97 a 1</b>	Test Excelente

Nota. La tabla muestra los niveles de aceptación de los modelos basados en las métricas de precisión y exactitud, levantado con los actuarios de la empresa.

El modelo fue seleccionado en base a lo revisado anteriormente. Además, se realizó un análisis comparativo seleccionando el modelo que obtuvo los mejores resultados.

El análisis predictivo proporciona múltiples oportunidades para la evolución de la empresa y la aplicación de esta técnica contribuye a que la línea de fianzas obtenga mejores resultados en su gestión de cobro.

Para la elección de algoritmos se realizó un análisis comparativo entre los indicadores arrojados por la herramienta. Para evaluar el algoritmo se analizaron nueve indicadores que se muestran en la tabla 21 y para su selección se escogió el algoritmo que obtuvo mejores indicadores. El algoritmo que mejor se ajusta para las necesidades del negocio fue el de “Redes Neuronales”.

Tabla 20

## Calificación de Algoritmos

Criterio	Árbol de Decisión	Redes de Bayes	Redes Neuronales	Ganador
Accuracy	88.47% +/- 1.72% (micro average: 88.46%)	86.12% +/- 2.36% (micro average: 86.12%)	92.35% +/- 1.51% (micro average: 92.35%)	Redes Neuronales
Classification error	11.53% +/- 1.72% (micro average: 11.54%)	13.88% +/- 2.36% (micro average: 13.88%)	7.65% +/- 1.51% (micro average: 7.65%)	Redes Neuronales
AUC	0.935 +/- 0.017 (micro average: 0.935) (positive class: true)	0.913 +/- 0.019 (micro average: 0.913) (positive class: true)	0.969 +/- 0.014 (micro average: 0.969) (positive class: true)	Redes Neuronales
Precision	78.31% +/- 2.46% (micro average: 78.24%) (positive class: true)	77.36% +/- 3.57% (micro average: 77.22%) (positive class: true)	90.17% +/- 4.92% (micro average: 89.75%) (positive class: true)	Redes Neuronales
Recall	97.65% +/- 1.99% (micro average: 97.64%) (positive class: true)	91.49% +/- 2.49% (micro average: 91.48%) (positive class: true)	90.82% +/- 4.75% (micro average: 90.81%) (positive class: true)	Árbol de Decisión
Lift	200.35% +/- 6.84% (micro average: 200.16%) (positive class: true)	197.93% +/- 9.44% (micro average: 197.55%) (positive class: true)	230.68% +/- 12.62% (micro average: 229.60%) (positive class: true)	Redes Neuronales
Sensitivity	97.65% +/- 1.99% (micro average: 97.64%) (positive class: true)	91.49% +/- 2.49% (micro average: 91.48%) (positive class: true)	90.82% +/- 4.75% (micro average: 90.81%) (positive class: true)	Árbol de Decisión
Specificity	82.57% +/- 2.50% (micro average: 82.58%) (positive class: true)	82.68% +/- 3.50% (micro average: 82.68%) (positive class: true)	93.34% +/- 3.94% (micro average: 93.34%) (positive class: true)	Redes Neuronales
<b>Ganador General</b>				<b>Redes Neuronales</b>

*Nota.* La tabla muestra el resumen de los resultados de las métricas obtenidos en el desarrollo de los modelos.

En base a las métricas obtenidas y el comparativo que se realizó de los tres modelos obtenidos en la herramienta, se definió como ganador a Redes Neuronales, puesto que en seis de los ocho indicadores analizados obtuvieron mejores resultados. Se puede apreciar que es un modelo sólido debido a la alta precisión que posee (92,35%). El ranking de los modelos obtenidos queda de la siguiente manera:

- 1) Redes Neuronales
- 2) Árbol de Decisión
- 3) Redes de Bayes

Sobre la base de la investigación realizada y expuesta en el Capítulo 2, Estado del Arte se pudo evidenciar que las redes neuronales habían tenido muy buenos resultados al aplicarlos en empresas de seguros médicos y del sector financiero; motivo por el cual en este proyecto se puede corroborar que es un modelo que se ajusta a las necesidades de la empresa y que podría traer muy buenos resultados en la recuperación de cartera.

## Capítulo 6 Pruebas y Evaluación con el Negocio

En este Capítulo se revisaron los resultados en conjunto con la empresa para definir el modelo ganador. Acorde a lo revisado en el Capítulo 5 y tomando en cuenta la metodología de evaluación el modelo ganador desde el punto de vista técnico, se aplicaron las Redes Neuronales.

En el presente capítulo se presenta los resultados obtenidos y se define las estrategias que se seguirán en base a los hallazgos identificados en los modelos.

### **Análisis de Resultados**

En base a la información arrojada a través de las métricas que se usaron para comparar los modelos, se presentó el modelo ganador, el cual fue Redes Neuronales. Basado en la experiencia de los encargados de modelos matemáticos de la empresa se corroboró que Redes Neuronales es un modelo robusto y que brinda buenos resultados. Las métricas analizadas por los actuarios fueron la precisión y la exactitud como se puede ver en la tabla 22.

**Tabla 21**

*Métricas Evaluadas por el Negocio*

<b>Criterio</b>	<b>Árbol de Decisión</b>	<b>Redes de Bayes</b>	<b>Redes Neuronales</b>	<b>Ganador</b>
Accuracy	88.47% +/- 1.72% (micro average: 88.46%)	86.12% +/- 2.36% (micro average: 86.12%)	92.35% +/- 1.51% (micro average: 92.35%)	Redes Neuronales
Classification error	11.53% +/- 1.72% (micro average: 11.54%)	13.88% +/- 2.36% (micro average: 13.88%)	7.65% +/- 1.51% (micro average: 7.65%)	Redes Neuronales
Precision	78.31% +/- 2.46% (micro average: 78.24%) (positive class: true)	77.36% +/- 3.57% (micro average: 77.22%) (positive class: true)	90.17% +/- 4.92% (micro average: 89.75%) (positive class: true)	Redes Neuronales

*Nota.* La tabla muestra el resumen de los resultados de las métricas obtenidos en el desarrollo de los modelos.

Los rangos establecidos por la empresa se pueden ver en el Capítulo 5. La tabla 23 muestra la calificación de la empresa de los criterios de los modelos obtenidos.

**Tabla 22**

*Calificación de la Empresa*

Criterio	Árbol de Decisión	Redes de Bayes	Redes Neuronales
Accuracy	Test Bueno	Test bueno	Test muy bueno
Classification error	Valor elevado	Valor elevado	Aceptable
Precision	Test Bueno	Test bueno	Test muy bueno

*Nota.* La tabla muestra el cruce de los resultados obtenidos con la tabla realizada en base a los criterios de los actuarios.

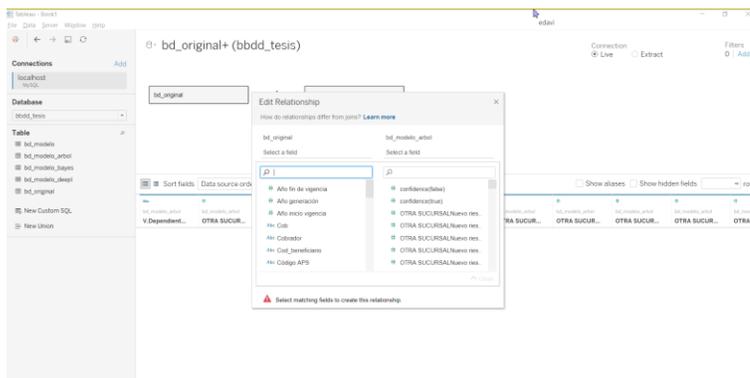
A través del análisis realizado, se corrobora los resultados obtenidos en el capítulo 5 y se valida que el modelo ganador es el de Redes Neuronales.

**Evaluación de Resultados**

Una vez definido el modelo ganador se procedió a elaborar un dashboard utilizando la herramienta seleccionada en el capítulo 4: Tableau. Para poder realizar el dashboard se realizó una conexión a través de Workbench con la base MySQL para que pueda obtener los resultados del modelo ganador en la herramienta RapidMiner. La Figura 30 muestra la conexión de la base con la herramienta.

**Figura. 30**

*Conexión Tableau*



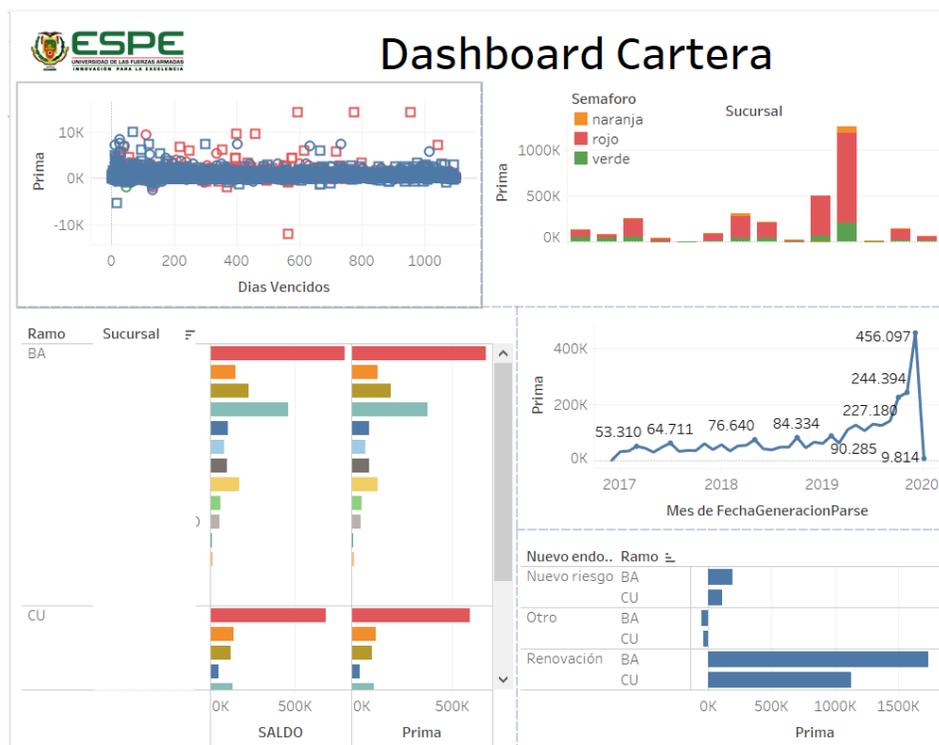
*Nota.* La figura permite ver la página principal de Tableau, donde se realiza la conexión a la base de datos.

El dashboard de clientes morosos permite visualizar la información del comportamiento. Esto es un gran avance para la línea ya que les permite tener consolidada la información en un solo lugar y dejar de realizar los reportes manuales, lo que antes se tardaba alrededor de tres horas ahora lo puede tener a tan solo un clic.

Se desarrollaron dos paneles, el primero que consta de información explicativa sobre el comportamiento de los clientes morosos, en el cual se puede ver las sucursales, prima, endoso y ramo y la evolución de la cartera en los tres años (ver Figura 31).

**Figura. 31**

*Dashboard Pantalla 1*



*Nota.* La figura permite visualizar el dashboard con la información del estado actual de la cartera de fianzas.

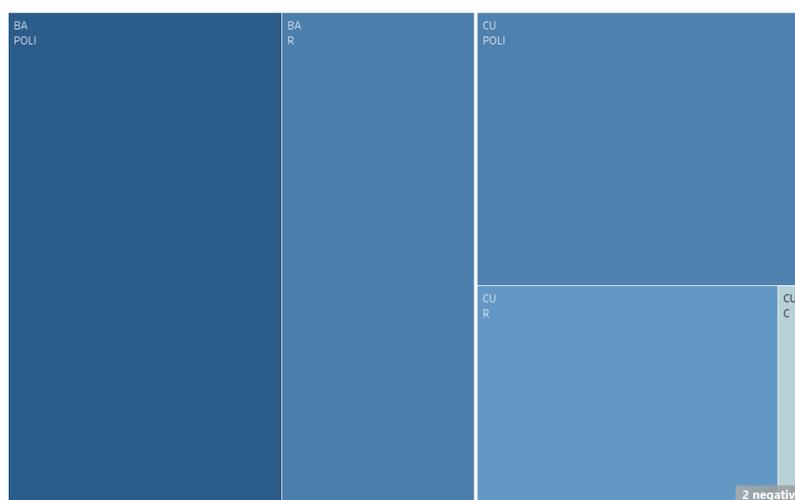
A través de este dashboard, el personal de la línea de fianzas tiene una visión global de cómo se encuentra concentrada su cartera. De hecho, reemplaza al reporte manual que debía obtener cada semana el ejecutivo de cartera y ahorra el tiempo que se tardaba en analizar el archivo en Excel para obtener los resultados de los clientes morosos de la línea de fianzas.

Este dashboard facilita el trabajo del ejecutivo y brinda a los líderes de la línea una visión global del estado de cartera de sus clientes. A través de esta pantalla pueden observar el comportamiento por ramo, tipo de endoso (si es una renovación o nuevo riesgo), analizar las sucursales que concentran mayor cartera y ver los montos acumulados por días vencidos.

Este panel permite a los líderes tomar decisiones más rápidas para establecer estrategias que permitan mejorar la recuperación de la línea y de igual manera tomar acciones preventivas que les permita evitar que los clientes caigan en mora. La Figura 32 muestra la concentración de la cartera por ramo y endoso.

### Figura. 32

*Gráfico Prima vs Ramo vs Endoso*



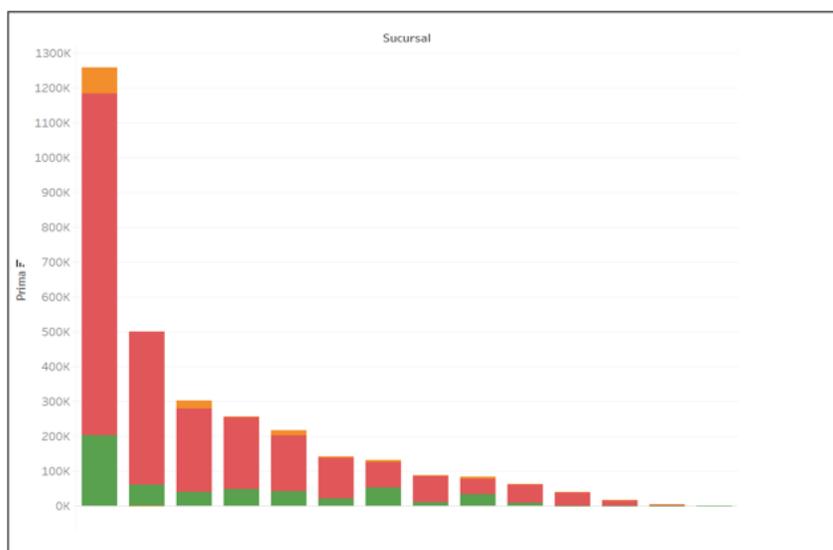
*Nota.* La Figura permite ver el comportamiento de la cartera de la línea de fianzas en base al ramo, la prima y el endoso.

El comportamiento de la cartera de fianzas se concentra en el ramo de Buen Uso del Anticipo (BA) Se cuenta con una prima de aproximadamente \$400.000. La línea tenía la hipótesis de que su principal concentración estaba en las renovaciones; sin embargo, la concentración se encuentra en el endoso de nuevo riesgo.

Desde el punto de vista de sucursales; Quito refleja la mayor cantidad de clientes moroso, lo cual está acorde a las hipótesis que mantenía la línea. La siguiente sucursal con concentración de cartera es Portoviejo, seguido por Cuenca y Loja. El gráfico se detalla a continuación.

**Figura.33**

Clientes Morosos por Sucursal



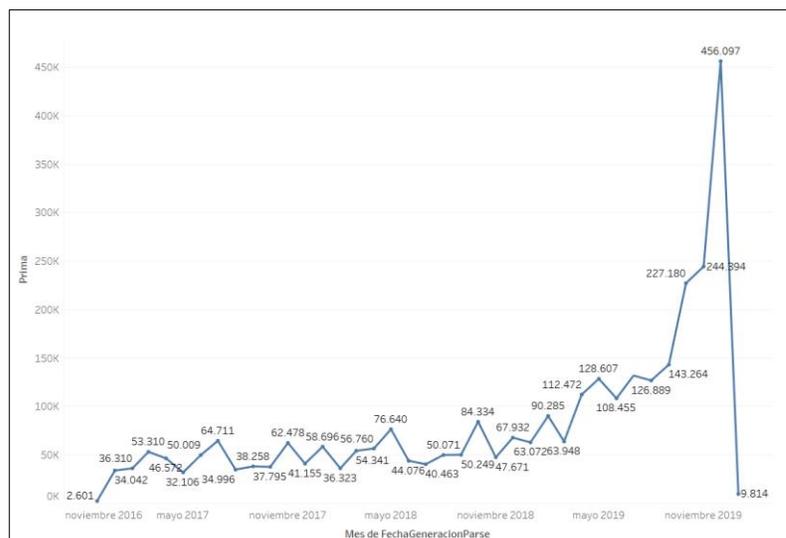
Nota. La ilustración permite visualizar por sucursales la concentración de clientes moroso.

De igual manera en la Figura 33 se puede apreciar que los clientes que tienen probabilidades de pago bajas son más altos, comparados con los que se encuentran con verde y naranja. La línea pensaba que iba a existir un porcentaje más alto en naranja. En base a esta información se puede decir que las sucursales que deben

enfocarse, redefinir y fortalecer sus estrategias de cobro de cartera son las que presentan mayores montos de morosidad. Se puede establecer una estrategia unificada o hacer un análisis más profundo para definir estrategias por sucursal. En la Figura 34 se observa la estacionalidad de los clientes morosos de la línea de fianzas:

**Figura. 34**

*Fechas vs Prima*



*Nota.* La ilustración permite visualizar el comportamiento de la cartera a través de los últimos tres años.

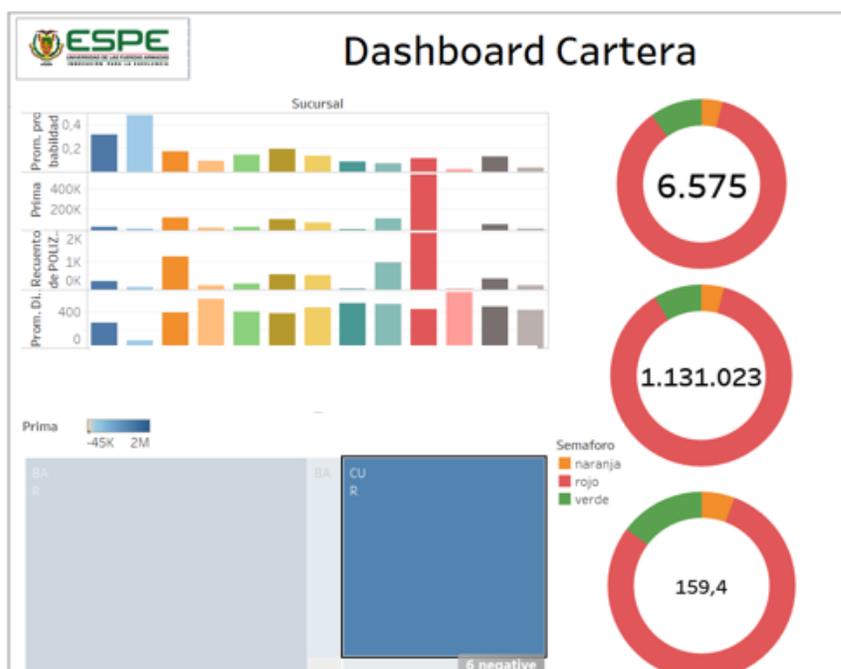
A través de esta gráfica se puede apreciar el comportamiento de la cartera en función de las fechas. Se puede ver que en los meses de mayo hay picos al igual que en noviembre, y se aprecia un incremento considerable a partir de mayo 2019, con una tendencia creciente y despuntando en el mes de noviembre 2019.

El segundo dashboard recopila la información obtenida en el modelo ganador “Redes Neuronales” y permite visualizar el estado de los clientes morosos, al generar alertas que indican aquellos que concentran las probabilidades más altas de no pagar los valores adeudados. Para realizar este semáforo se categorizaron tres niveles: Rojo (clientes con una probabilidad de pago menor al 50%), naranja (entre 51% a 80%) y

verde (>80%). Los resultados obtenidos, corroboran la hipótesis de la línea, la cual apuntaba a que uno de sus principales problemas corresponde a las renovaciones. Adicional los porcentajes obtenidos en los modelos como incobrables o probabilidades de pago muy bajos, sorprendieron a la línea.

**Figura. 35**

*Dashboard Pantalla 2*



*Nota.* La figura permite ver las probabilidades de pago de los clientes a través de un sistema de alertas (semáforo).

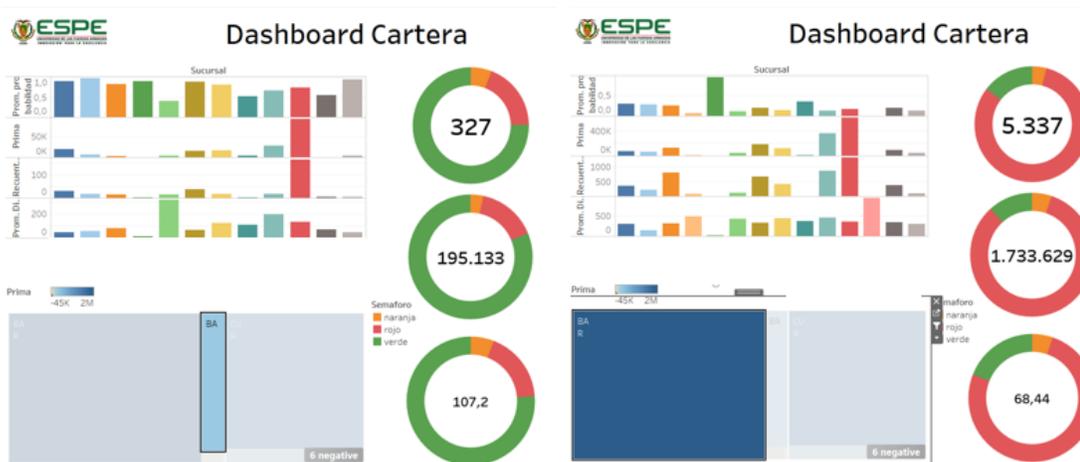
Este dashboard consta de tres categorías. La primera ubicada en la parte izquierda que nos permite visualizar el comportamiento de la cartera por sucursal. La Figura 36 permite visualizar el comportamiento de la cartera por ramo y endoso. Los tacómetros de la parte derecha generan las alertas en base a la información disponible. El primer tacómetro permite visualizar el número de pólizas que se encuentran dentro de un segmento. El segundo tacómetro muestra el valor de la prima que acumula ese

segmento y el tercero, el promedio de días vencidos. Toda esta información se la puede obtener por sucursal, ramo y endoso. Esto facilita definir las estrategias que se tomaran para la recuperación.

Para el desarrollo de estrategias se analizaron los filtros realizados en el tablero. Es decir, estrategias por ramo y tipo de endoso. Las pólizas correspondientes a nuevo riesgo tienen una probabilidad más alta de pago, mientras que la probabilidad de no pago de renovaciones, es bastante alta.

**Figura. 36**

*Resultado de Nuevo Riesgo y Renovaciones*



*Nota.* La figura permite ver la semaforización en base a nuevo riesgo y renovaciones

### **Estrategias Generales**

En base a las reglas y criterios que se definieron se procedió a crear las siguientes estrategias, que fueron validadas con el negocio con el objetivo de tener herramientas que permitan disminuir el porcentaje de clientes que probablemente no van a pagar. El estudio se basó en los resultados del sistema de alertas creado con Tableau. Los tableros de control realizados muestran una visión global del comportamiento actual y de las probabilidades de pago de los clientes.

Con esta información se mejorará la recuperación a través del desarrollo de diferentes estrategias, las cuales se clasifican en generales y específicas.

**Certificación del dashboard.** Validar con los usuarios que la información presentada es la adecuada. Verificar que la información presentada sea la correcta y que esté alineada al negocio.

**Establecer estrategias en el proceso de comercialización.** Determinar mejores controles en la comercialización de las pólizas, sobre todo en el proceso de renovaciones de las pólizas para notificar de forma inmediata al cliente sobre los pagos pendientes y evitar la acumulación de la cartera.

**Definición de indicadores de recuperación y gestión.** Establecer metas mensuales de recuperación e indicadores de gestión.

**Definición de nuevas políticas de cobro.** Se deberán levantar nuevas políticas de cobro, para evitar el incremento valores adeudados.

**Proceso de Recuperación.** Para realizar el proceso de recuperación se sugiere las siguientes actividades:

**Revisión del dashboard de cartera.** Obtener los clientes que se encuentran dentro del semáforo verde para realizar las llamadas la primera semana del mes. Para los clientes que se encuentran en naranja y rojo se realizará las estrategias definidas en la sección de estrategias específicas.

**Realizar un cronograma de llamadas.** Definir un cronograma para realizar las llamadas. Las llamadas serán registradas en el CRM de la empresa con los principales comentarios del cliente.

**Realizar un seguimiento del cliente.** El ejecutivo de cartera deberá llamar al cliente en caso de que no realice el pago de su cartera la primera vez luego de siete días. La segunda vez luego de siete días adicionales para dar seguimiento y solicitar el

pago. En caso de que el cliente no realice el pago se deberá realizar una carta escrita la cual será enviada a través de correo electrónico solicitando el pago al cliente.

**Levantamiento de protocolos de recuperación.** Definir los scripts y protocolos que se usarán cuando los ejecutivos de la línea de fianzas realicen la gestión de cobro de cartera.

**Capacitación.** Realizar una capacitación al ejecutivo de cartera sobre el uso de la herramienta (dashboard). Debe ser capaz de aplicar adecuadamente los filtros establecidos, conocer el significado de los gráficos y utilizar el protocolo de recuperación de cartera establecido.

**Motivación al equipo.** Definir incentivos en base al cumplimiento y sobrecumplimiento de la recuperación de cartera de la línea de fianzas.

### ***Estrategias Específicas***

**Clientes en semáforo Verde.** A los clientes que se encuentren en semáforo verde se les realizará seguimiento telefónico y una campaña través de correo electrónico y SMS, con el detalle de cuáles son los montos adeudados y los medios de pago que pueden usar para realizar sus cobros. El objetivo con estos clientes es recuperar en su totalidad los valores adeudados.

Se realizará campañas estacionales con beneficios para promover el pago de la cartera.

**Clientes en semáforo Naranja.** Se realizará un seguimiento telefónico y a aquellos que presenten saldos mayores a \$5.000 los realizará el jefe de la línea o de la sucursal. De igual manera se enviará correos y SMS con los valores adeudados y el detalle de medios de pago. Las acciones determinadas para este segmento son:

- a) **Reestructuración de deuda y plan de pagos.** Se dará la opción al cliente de que obtenga plazos para poder pagar los valores adeudados.

- b) **Gestión de llamadas a través del call center de la empresa:** Se entregará el listado de clientes y el protocolo a seguir para este segmento de clientes
- c) **Bloqueo de cliente para obtener pólizas:** Se bloqueará temporalmente al cliente para que pueda obtener una póliza de fianzas hasta que haya pagado la totalidad o un 80% de sus valores adeudados.

**Clientes en semáforo rojo.** Para estos clientes se debe realizar un análisis a fondo y definir cuáles son los incobrables para no desperdiciar recursos en recuperación de cartera. Para el análisis de estos clientes se verificará:

- Tiempo de mora
- Estado del cliente (vivo o fallecido)
- Estado de la póliza (vigente o vencida)

En base a estos parámetros una vez definidos los clientes que pueden pagar sus deudas se establecerán las siguientes acciones.

- a) **Reestructuración de deuda y plan de pagos.** Se dará la opción al cliente de que obtenga plazos para poder pagar los valores adeudados.
- b) **Contratación de una empresa especialista en cobro de cartera.** Se tercerizará la gestión de cobro.

**Bloqueo definitivo del cliente.** Los clientes que se encuentren en rojo y presenten montos superiores a \$5.000 no podrán obtener pólizas de la línea de fianzas.

En base a las estrategias mencionadas, y la revisión realizada con el área, se establece como principales las siguientes:

**a) Establecer estrategias sobre la comercialización:**

Es necesario establecer mayores controles para evitar que la cartera se acumule en las renovaciones. Este es un punto de dolor que aqueja a la línea debido a las renovaciones automáticas por solicitud del cliente. Para mitigar esto se debe enviar un

SMS, correo y realizar una llamada al cliente cuando se procede a realizar una renovación para solicitar el pago pendiente.

**b) Automatizar los reportes**

La implementación del dashboard permitirá optimizar los tiempos de gestión de cobros y mejorar la gestión de recuperación.

**c) Establecer metas e indicadores en la gestión de cobros**

Definir las metas que se deben alcanzar entregando todas las herramientas necesarias para la gestión de cobros es indispensable para la línea.

**d) Repotenciar el proceso de gestión de cobro**

Modificar y repotenciar el proceso de cartera de la línea a través del uso de nuevas tecnologías.

## Capítulo 7 Conclusiones y Recomendaciones

### Conclusiones y Recomendaciones

En el presente capítulo se presentará las principales conclusiones y recomendaciones del desarrollo del presente proyecto.

#### **Conclusiones**

- El análisis inicial y levantamiento de información cualitativa a la línea, permitió definir los principales problemas y tener un conocimiento global del manejo actual de cobranzas. También se pudo evidenciar las hipótesis que mantenía la línea respecto a su cartera, los procesos manejados actualmente, y la forma en la que se extraen y elaboran los reportes correspondientes a clientes morosos. Esta información fue muy valiosa ya que gracias a todo lo identificado se pudo desarrollar de una mejor manera el proceso de modelamiento y de selección del algoritmo adecuado.
- En respuesta a los objetivos dos y tres y basado en los análisis realizados en el estudio del arte y en la investigación desarrollada para el presente proyecto se corrieron tres tipos de modelos: Bayes, Redes Neuronales y Árbol de decisión, de los cuales las Redes Neuronales obtuvo los mejores resultados con una precisión de 92.35%. La herramienta seleccionada para el desarrollo de la tesis fue RapidMiner; la misma que demostró ser sólida e intuitiva, que permite realizar múltiples modelos y que ofrece a sus usuarios un sinnúmero de funciones para facilitar su trabajo. Además, se usó Tableau como una plataforma de análisis visual, la cual demostró sus fortalezas para representar y entender la información que fue obtenida en este proyecto.

- Para la selección del mejor modelo se analizaron ocho métricas, las cuales se obtuvieron por cada uno de los modelos que fueron elaborados. Al analizar estos indicadores se seleccionó el modelo de Redes Neuronales, ya que, de ocho indicadores evaluados, seis obtuvieron los mejores resultados. El modelo ganador obtuvo una exactitud del 92,35%, una precisión del 90,17% y un error del 7,65%. Por tanto, se demostró su solidez y buen desempeño. Posterior a esto se validó con el negocio los resultados obtenidos y se corroboró que es una muy buena alternativa para ponerla en práctica en la empresa.
- El modelo seleccionado responde a las necesidades de la empresa y le permite anticiparse y tomar decisiones más ágiles, además de poder definir estrategias en base a los segmentos establecidos. Contar con alertas permite que la empresa sea más proactiva y no se maneje de la forma reactiva que lo ha venido haciendo.
- Las nuevas tecnologías de la información permiten a las empresas optimizar sus recursos y tomar decisiones más acertadas. El uso de técnicas de minería apalancadas en las diferentes herramientas existentes en el mercado, permiten desarrollar modelos robustos que den respuestas a los interrogantes de las compañías. En el caso del presente proyecto, se pudo solventar la gestión de cobranzas para brindar herramientas que faciliten su gestión y mejoren los indicadores correspondientes a la morosidad.

**Recomendaciones**

- Revisar los procesos actuales que involucran el proceso de recuperación de cartera.
- Automatizar los reportes de la línea y crear un dashboard que permita al ejecutivo de cartera optimizar su tiempo para realizar la tarea de recuperación.
- Establecer mayores controles en el proceso de comercialización del ramo de Buen Uso del Anticipo, para evitar que la cartera se acumule en las pólizas de nuevo riesgo.
- Establecer alertas hacia el cliente como (SMS, correos y llamadas), cuando se renuevan las pólizas, para poder realizar una gestión de cobro más ágil y evitar que se acumule la cartera vencida.
- Definir estrategias de acuerdo a la segmentación o clasificación de clientes, y dar prioridad de cobro a aquellos que se encuentran en verde para obtener mejores resultados. Aquellos que se encuentran en rojo y naranja, prestar atención y dar seguimiento, para recuperar los valores adeudados.
- Establecer metas de cobro y desarrollar programas estacionales de incentivos que permitan motivar al personal de recuperación de cartera para que pueda sobre cumplir sus metas.
- Contar con un aliado estratégico de recuperación de cartera.
- Apoyarse en las nuevas tecnologías para facilitar la gestión y optimizar los recursos de la empresa.

**Próximos pasos.**

La empresa cuenta con un área de Inteligencia de Negocio, que es la responsable de priorizar los diferentes proyectos de la compañía para poder asignar los

recursos y establecer los tiempos de entrega de los proyectos. Una vez los proyectos se encuentren priorizados, se debe realizar algunas actividades previas para la constitución del proyecto.

***Análisis de factibilidad de implementación.*** Permite identificar el impacto que tendrá en el negocio la implementación del proyecto y en base a ese impacto se le asignará una prioridad para poder definir la fecha en la cual se lo pondrá en marcha.

***Levantamiento del proyecto.*** Una vez el proyecto haya sido priorizado se levantan los requerimientos para la elaboración del proyecto. Para este punto se realiza un acta de constitución en la cual se coloca la información más relevante del proyecto.

***Asignación de recursos.*** Una vez levantado el proyecto se asignan las actividades a un equipo multidisciplinario, para lo cual se realiza una reunión en la que se explica el alcance del proyecto.

***Levantamiento de cronograma de implementación.*** Una vez que se encuentre socializado el proyecto con los involucrados.

***Kick off del proyecto.*** Se realiza una reunión con todos los involucrados para la socialización del proyecto y la presentación de los cronogramas de trabajo

***Seguimiento al proyecto.*** Se realiza un seguimiento para controlar los avances de la implementación del proyecto.

## Lista de referencias

### Bibliografía

- (s.f.).  
 American Psychological Association. (2010). *Manual de Publicaciones de la American Psychological Association* (6 ed.). (M. G. Frías, Trad.) México, México: El Manual Moderno.
- Aritmetics. (s.f.). *Aritmetics*. Obtenido de <https://www.aritmetics.com/glosario-digital/modelo-predictivo>
- Barreriro, J. M., Diez de Castro, J. A., Barreiro, B., Ruzo, E., & Losada, F. (2001). *Gestión Científica Empresarial*. La Coruña: Netbiblo.
- Bismart. (2020). *Bismart*. Obtenido de <https://bismart.com/en/home-2/>
- Bohórquez, C. (26 de 02 de 2019). Levantamiento de problemas en cartera. (N. Carvajal, Entrevistador)
- Calderón, P. (25 de 1 de 2019). Indagación de información de Seguros. (N. Carvajal, Entrevistador)
- Calvo, D. (7 de Diciembre de 2018). *Análisis de Datos*. Obtenido de <https://www.diegocalvo.es/funcion-de-activacion-redes-neuronales/>
- Candel, A., & LeDell, E. (2018). *Deep Learning with H2O*. Estados Unidos: H2O.ai, Inc.
- Capitalismo Consciente. (Noviembre de 2017). *Capitalismo Consciente*. Obtenido de <https://capitalismoconsciente.es/blog/indicadores-comerciales-puedo-medir-la-eficiencia-ventas/>
- Concepto definición.de*. (2019). Obtenido de <https://conceptodefinicion.de/analisis-de-datos/>
- Credit Management solutions. (2007). *Indicadores Clave del Proceso de Cobranzas*. Lima.
- Díaz, Á., & Domínguez, C. (2018). *Interpretación: Un reto en la investigación educativa*. Tlaxcala, México: Newton.
- Díaz, Á., & Dominguez, C. (2018). *La interpretación: Un reto en la investigación educativa*. Tlaxclala: Newton.
- Ecured*. (s.f.). Obtenido de <https://www.ecured.cu/Clustering>
- Empresa de Seguros. (2019). *Estado de cuenta fianzas*. Quito.
- Empresa de seguros. (15 de 03 de 2019). *Oracle BI*. Obtenido de Reporte de Fianzas: <http://172.16.1.114:9502/analytics/saw.dll?Dashboard>
- Empresa de Seguros. (19 de 03 de 2019). *Oracle BI*. Obtenido de <http://172.16.1.114:9502/analytics/saw.dll?Dashboard>
- Eppen, G., & Gould, F. (2000). *Investigación de operaciones en la ciencia administrativa: construcción de modelos para la toma de decisiones con hojas de cálculo electrónicas*. México: Pearson.
- Escudero, R. (2008). *Academia Edu*. Obtenido de [https://www.academia.edu/25033602/INDICADORES\\_DE\\_GESTI%C3%93N\\_D E\\_COBRANZAS\\_COBRANZAS](https://www.academia.edu/25033602/INDICADORES_DE_GESTI%C3%93N_D E_COBRANZAS_COBRANZAS)
- Fiuza, M. D., & JC, R. (diciembre de 2000). *revistanefrologia*. Obtenido de <https://www.revistanefrologia.com/es-relacionados-la-regresion-logistica-una-herramienta-versatil-articulo-X0211699500035664>
- Flórez, R., & Fernández, J. M. (2008). *Las Redes Neuronales Artificiales*. La Coruña: Netbiblo.
- Gartner Peer Insights. (2020). *Gartner Peer*. Obtenido de <https://www.gartner.com/reviews/customers-choice-landing-page>

- Hernández, R., Fernández, C., & Baptista, M. d. (2014). *Metodología de la Investigación*. México D.F: Mc Graw Hill.
- Hurtado, J. (2000). *Retos y alternativas en la formación de investigadores*. Caracas : Sypal.
- IBM. (s.f.). *IBM Knowledge Center*. Obtenido de [https://www.ibm.com/support/knowledgecenter/es/SS3RA7\\_sub/modeler\\_mainhlp\\_client\\_ddita/components/neuralnet/neuralnet\\_model.html](https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhlp_client_ddita/components/neuralnet/neuralnet_model.html)
- Krajewski, L. J., & P., R. L. (2000). *Administración de operaciones: estrategia y análisis*. México: Pearson education.
- Lévy, J. P., & Varela, J. (2006). *Modelización con Estructuras de Covarianzas en Ciencias Sociales*. Madrid : Netbiblo.
- Logicalis. (2015). *Predictive Analytics El impacto de la predicción en una organización*. LOGICALIS.
- Mochón, F., & González, J. C. (2016). *Big Data: Una gestión inteligente de los datps*. García Maroto Editores.
- Monge, L. (8 de octubre de 2020). *LM*. Obtenido de <https://luca-d3.com/es/data-speaks/diccionario-tecnologico/aprendizaje-profundo>
- Monje, C. (2011). *Metodología de la investigación Cuantitativa y Cualitativa*. Neiva: Universidad Surcolombiana.
- Nettleton, D. (2003). *Análisis de datos comerciales* . Madrid: Díaz de Satos.
- Pérez, C., & Santín, D. (2008). *Minería de Datos Técnicas y herramientas*. Madrid: Thomson.
- Pérez, C., & Santín, D. (2008). *Minería de Datos, Técnicas y Herramientas* . Madrid: Thomson.
- Porter, M. (1999). Los clusters y la competencia . *Harvard Business Review* , 130-145.
- Question Pro. (2020). *Question Pro*. Obtenido de <https://www.questionpro.com/blog/es/coeficiente-de-correlacion-de-pearson/>
- RapidMiner Studio Core. (2020). *Rapidminer Documentation*. Obtenido de [https://docs.rapidminer.com/latest/studio/operators/modeling/feature\\_weights/weight\\_by\\_information\\_gain.html](https://docs.rapidminer.com/latest/studio/operators/modeling/feature_weights/weight_by_information_gain.html)
- Raquel, F., & Fernández, J. M. (2008). *Las redes Neuronales Artificiales Fundamentos Teóricos y aplicaciones prácticas*. Madrid: Netbiblo.
- Rivera, R. A. (29 de marzo de 2016). DSR: A un paso más de la investigación científica. (USAT, Entrevistador)
- Rouse, M. (Noviembre de 2012). *Search Data center*. Obtenido de Search Data Center .
- Ruiz, E. (2017). *Nuevas tendencias en los sistemas informáticos*. Madrid: Editorial Universitaria Ramón Areces.
- ServiceNow. (2020). *KPI Library*. Obtenido de <http://kpilibrary.com/categories/accounts-payable-receivable>
- Superintendencia de Compañías, Valores y Seguros. (2020). *Supercias*. Obtenido de <https://reporteria.supercias.gob.ec/portal/cgi-bin/cognos.cgi>
- Telefonica Data Unit. (2020). *Luca AI Powered Decision*. Obtenido de <https://luca-d3.com/es/data-speaks/diccionario-tecnologico/aprendizaje-profundo>
- Vaishnavy, V., Kuechler, B., & Peter, S. (30 de Junio de 2019). *desrist.org*. Obtenido de [desrist.org](http://desrist.org): <http://desrist.org/desrist/content/design-science-research-in-information-systems.pdf>
- Vaishnavy, V., Kuechler, B., & Petter, S. (30 de 06 de 2019). *Desrist.org*. Obtenido de [Desrist.org](http://desrist.org): <http://desrist.org/desrist/content/design-science-research-in-information-systems.pdf>

- Vijay K. Vaishnavi, V. K. (2015). *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*. Boca Ratón: CRC Press.
- Villanueva, J. (2019). *IArtificial.net*. Obtenido de <https://www.iartificial.net/redes-neuronales-desde-cero-i-introduccion/>
- Wunsch, D. (2009). *Clustering*. Hoboken: Wiley.
- Zabala, V. (2018). Especial de Seguros. *Ekos*, 68-74.
- Zabala, V. (2019). Ranking de Seguros. *Ekos*, 46-73.