



FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE COMPUTACIÓN Y SISTEMAS

**PREDICCIÓN DE LA TASA DE VENTA PARA RENTAS
VITALICIAS EN UNA EMPRESA ASEGURADORA**

PRESENTADA POR

**JUAN CARLOS CONDE HUMAREDA
SEBASTIÁN RODRIGO OSORIO SÁNCHEZ**

**TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE
INGENIERO DE COMPUTACIÓN Y SISTEMAS**

LIMA – PERÚ

2015



**Reconocimiento - No comercial - Compartir igual
CC BY-NC-SA**

El autor permite transformar (traducir, adaptar o compilar) a partir de esta obra con fines no comerciales, siempre y cuando se reconozca la autoría y las nuevas creaciones estén bajo una licencia con los mismos términos.

<http://creativecommons.org/licenses/by-nc-sa/4.0/>



USMP
UNIVERSIDAD DE
SAN MARTÍN DE PORRÉS

**FACULTAD DE
INGENIERÍA Y ARQUITECTURA**

**ESCUELA PROFESIONAL DE INGENIERÍA DE COMPUTACIÓN Y
SISTEMAS**

**PREDICCIÓN DE LA TASA DE VENTA PARA RENTAS
VITALICIAS EN UNA EMPRESA ASEGURADORA**

TESIS

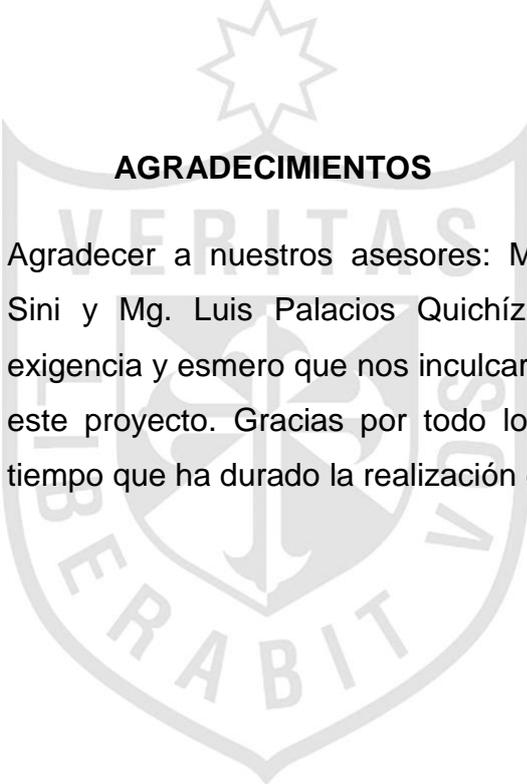
**PARA OPTAR EL TÍTULO DE
INGENIERO DE COMPUTACIÓN Y SISTEMAS**

PRESENTADO POR

**CONDE HUMAREDA, JUAN CARLOS
OSORIO SÁNCHEZ, SEBASTIÁN RODRIGO**

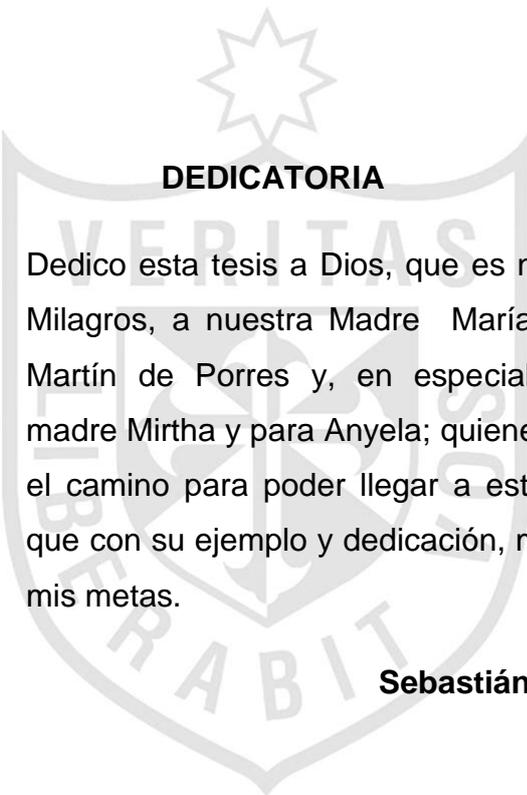
Lima – Perú

2015



AGRADECIMIENTOS

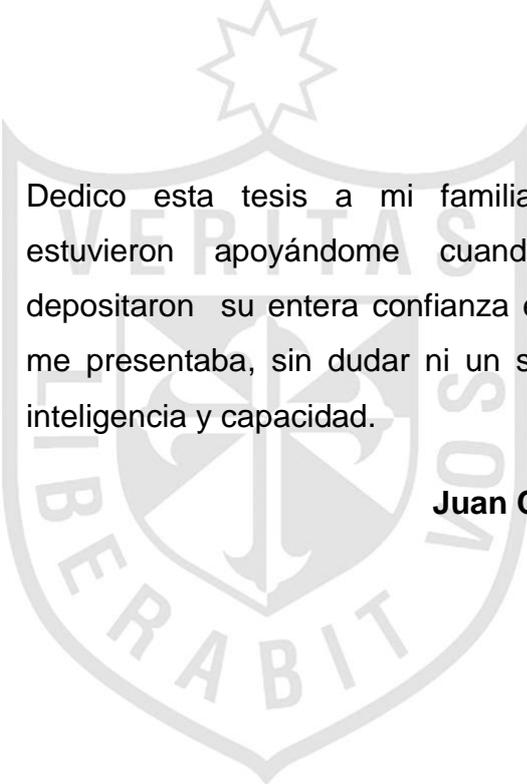
Agradecer a nuestros asesores: Mg. Gustavo Castillo Sini y Mg. Luis Palacios Quichíz por la dedicación, exigencia y esmero que nos inculcaron desde el inicio de este proyecto. Gracias por todo lo recibido durante el tiempo que ha durado la realización de la tesis.



DEDICATORIA

Dedico esta tesis a Dios, que es nuestro Señor de los Milagros, a nuestra Madre María Auxiliadora, a San Martín de Porres y, en especial mención, para mi madre Mirtha y para Anyela; quienes han sido mi guía y el camino para poder llegar a este punto de mi vida, que con su ejemplo y dedicación, me alentaron a lograr mis metas.

Sebastián R. Osorio Sánchez



Dedico esta tesis a mi familia, quienes siempre estuvieron apoyándome cuando lo necesité y depositaron su entera confianza en cada reto que se me presentaba, sin dudar ni un solo momento en mi inteligencia y capacidad.

Juan C. Conde Humareda



ÍNDICE

	Página
RESUMEN	x
ABSTRACT	xi
INTRODUCCIÓN	xii
CAPÍTULO I MARCO TEÓRICO	1
1.1 Antecedentes	1
1.2 Bases teóricas	7
1.3 Definición de términos básicos	17
CAPÍTULO II METODOLOGÍA	25
2.1 Materiales	25
2.2 Métodos	31
CAPÍTULO III DESARROLLO DEL PROYECTO	38
3.1 Desarrollo de la evaluación predictiva (PREDICTOR)	38
3.2 Desarrollo del Data Mart	52
3.3 Factores de riesgo	59

CAPÍTULO IV PRUEBAS Y RESULTADOS	61
4.1 Pruebas	61
4.2 Resultados	64
CAPÍTULO V DISCUSIÓN Y APLICACIÓN	68
5.1 Discusión	68
5.2 Aplicaciones	71
CONCLUSIONES	73
RECOMENDACIONES	74
FUENTES DE INFORMACIÓN	76
ANEXOS	81



ÍNDICE DE TABLAS

	Página
Tabla 1.1: Número de nuevos pensionistas de jubilación.	8
Tabla 1.2: Tipo de jubilación: Requisitos principales.	19
Tabla 1.3: Diferencias entre SPP y SNP.	23
Tabla 2.1: Recursos de hardware	26
Tabla 2.2: Recursos de software.	26
Tabla 2.3: Costo de los recursos humanos.	27
Tabla 2.4: Costos directos.	27
Tabla 2.5: Costos indirectos.	28
Tabla 2.6: Costos de operación.	28
Tabla 2.7: Estimación de las primas de seguros anuales.	29
Tabla 2.8: Estimación del margen de contribución del proyecto.	29
Tabla 2.9: Estimación del VAN.	30
Tabla 2.10: Comparación de metodologías para Data Mart.	33
Tabla 2.11: Ponderación por metodología Data Mart.	33
Tabla 2.12: Comparación de metodologías para Data Mining.	36
Tabla 2.13: Ponderación por metodología.	37
Tabla 3.1: Dimensiones del Data Mart.	54
Tabla 3.2: Tablas de Hechos (facts) del Data Mart.	55
Tabla 3.3: Bases de datos transaccionales del proceso de Rentas Vitalicias	57
Tabla 4.1: Cantidad de solicitudes enviadas a pruebas.	64
Tabla 4.2: Resultados de cierres – Rentas Vitalicias.	65
Tabla 4.3: Variación de TRA.	66
Tabla 4.4: Variación de Tasa de Venta.	67

ÍNDICE DE FIGURAS

	Página
Figura 1: Proceso de solicitud de jubilación.	xiv
Figura 2: Tendencia de la tasa de venta en el año 2014.	xv
Figura 1.1: Jubilados del SPP.	2
Figura 1.2: Número de nuevos pensionistas.	9
Figura 1.3: Egresos de fondo de pensiones para el pago de pensionistas.	9
Figura 1.4: Componentes de Inteligencia de Negocios.	11
Figura 1.5: Retiro Programado.	20
Figura 1.6: Renta Vitalicia Familiar.	20
Figura 1.7: Renta Temporal con Renta Vitalicia Diferida.	21
Figura 1.8: Renta Mixta.	21
Figura 1.9: Renta Bimoneda.	22
Figura 2.1: Etapas de la Metodología Kimball.	31
Figura 2.2: Fases del modelo de referencia CRISP-DM.	34
Figura 3.1: Flujo de venta de renta vitalicia.	39
Figura 3.2: Proceso de generación de base de muestra.	42
Figura 3.3: Base muestra – Base predictiva.	43
Figura 3.4: Proceso de generación de cuartiles en IBM SPSS.	45
Figura 3.5: Resultado de la generación de cuartiles en IBM SPSS.	45
Figura 3.6: Proceso de limpieza de muestra – Identificación de Outliers.	46
Figura 3.7: Resumen de método predictivo.	48
Figura 3.8: Representación de Árbol de Decisión.	49
Figura 3.9: Algoritmo predictivo y validación de perfiles.	50
Figura 3.10: Registros reales para prueba.	51
Figura 3.11: Aplicación para utilización de la predicción en los casos de prueba.	52

Figura 3.12: Modelo Lógico del Data Mart.	55
Figura 3.13: Arquitectura del Data Mart.	58
Figura 4.1: Proceso de pruebas Predictor.	63
Figura 5.1: Cotización actual.	68
Figura 5.2: Calculo de la tasa de venta ofrecida por compañía.	69
Figura 5.3: Selección de la compañía driver.	69
Figura 5.4: Aplicación de la Tasa de Rentabilidad Ajustada.	70
Figura 5.5: Tasa de venta final a cotizar.	70
Figura 5.6: Aplicaciones del Método Predictivo.	71



RESUMEN

La presente tesis consiste en realizar una propuesta de predicción de la tasa de venta, utilizada en la generación de cotizaciones de rentas vitalicias, dentro de una empresa aseguradora. El mercado de rentas vitalicias ha experimentado un crecimiento constante; ante esta situación, es necesaria la mejora de los procesos que intervienen, siendo la generación de cotizaciones, uno de los subprocesos más importantes.

Una de las formas de mejorar la generación de cotizaciones es integrando herramientas de Inteligencia de Negocios: como Data Mart y Data Mining. Como resultado, se obtuvo un algoritmo capaz de predecir la tasa de venta, logrando la optimización del proceso e incrementando la rentabilidad de la empresa. Se concluye que el uso de árboles de decisión optimiza el proceso de clasificación de las variables que intervienen directamente en el cálculo de la tasa de venta, mejorando el control de la inversión y la toma de decisiones.

Palabras clave: Renta Vitalicias, Arboles de decisión, Minería de datos, Data Mart.

ABSTRACT

This thesis is to make a proposal prediction *sale-rate* used in the generation of annuities quotes within an insurance company. The annuities market has grown steadily; in this situation, improving the processes involved is necessary, with the generation of quotations, one of the most important threads.

One way to improve quote generation is integrating business intelligence tools: as Data Mart and Data Mining. As a result, an algorithm capable of predicting the *sale-rate* was obtained, achieving process optimization and increasing profitability. It is concluded that the use of decision trees streamlines the process of classification of variables involved directly in the calculation of the *sale-rate*, improving investment control and the take of choose.

Keywords: Annuities, Decision trees, Data Mining, Data Mart

INTRODUCCIÓN

Según la publicación hecha por el Diario Gestión, en enero del año 2015, se sabe que el mercado de rentas vitalicias es muy competitivo y de crecimiento constante, en el cual se tienen hasta 6 o 7 empresas compitiendo por la captación de afiliados a las distintas AFP y que desean pasar a la jubilación, pasando a ser llamados “pensionados”.

De acuerdo con las cifras obtenidas por la Superintendencia de Banca, Seguros y AFP (SBS) al cierre del año 2014, las primas de rentas vitalicias habrían registrado una nueva cifra de crecimiento del 24.9%, aproximadamente, con respecto al año 2013, que representaron un total de S/1,974.9 millones de nuevos soles al cierre del año; que colocó a la Empresa Aseguradora donde se realiza esta tesis, en primer lugar, con un Market Share del 26% en el mercado peruano, seguido por Pacífico Seguros con un 22%.

En la venta de una Renta Vitalicia, el cliente realiza la elección de la modalidad de pensión que desea recibir en base a las cotizaciones remitidas por las diferentes empresas aseguradoras, en la cual, para el cálculo de la pensión del futuro jubilado, interviene la tasa de venta, pero ¿qué es la Tasa de Venta? – Según la Asociación de AFP: “Es la tasa utilizada por la empresa de seguros para calcular el capital requerido de las pensiones bajo la modalidad de renta vitalicia a otorgar al afiliado y/o sus beneficiarios, de ser el caso. Esta tasa es fijada libremente por las empresas de seguros”.

La presente tesis busca desarrollar una propuesta de método del cálculo para poder predecir la mejor tasa de venta al momento en que el afiliado busque jubilarse. Como se ha mencionado, la tasa de venta es fijada por las empresas de seguros e informada al momento de remitir la cotización a la AFP y que, con el paso de los meses, sufre variaciones dependiendo de diversos factores, tales como el incremento de los aportantes, el incremento de IPC (Índice Precio-Consumidor), entre otros.

La situación problemática reside en la evaluación realizada por Class & Asociados S.A. - Clasificadora de Riesgo (2015), donde se detalla que la empresa aseguradora en la cual se desarrolla este proyecto se ha convertido en una de las principales aseguradoras peruanas especializada en rentas vitalicias. La empresa aseguradora apuesta por la constante capacitación de sus agentes de ventas, proporcionándoles toda una gama de herramientas, know-how, entre otras herramientas, para alcanzar las metas y la mejor participación en el mercado. Dichas herramientas son aplicadas al flujo que siguen los afiliados a una AFP (próximos a jubilarse) al momento de pedir su jubilación, tal como se muestra en la Figura 1.

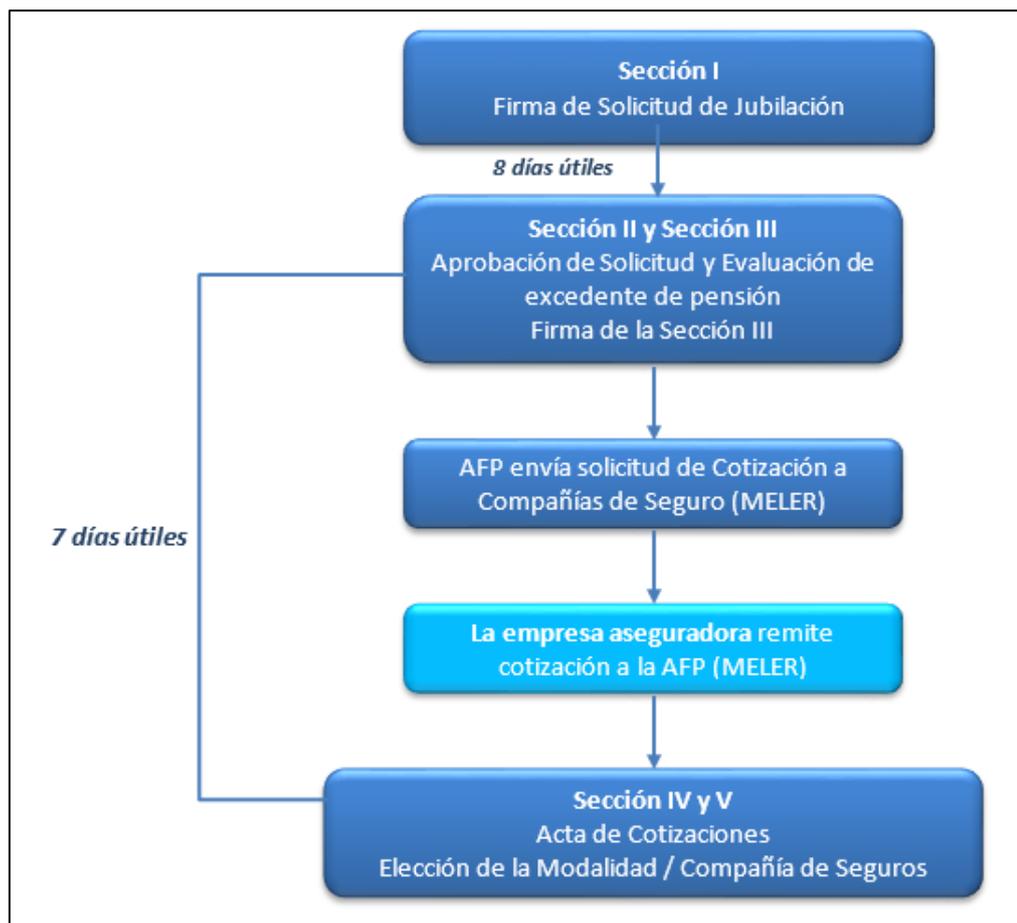


Figura 1: Proceso de solicitud de jubilación.
Fuente: Interseguro, 2014.

Dentro de la generación de cotizaciones de rentas vitalicias, existe el cálculo de la tasa de venta que, según el reporte (Superintendencia de Banca, Seguros y AFP, 2014), y tal como se muestra en la Figura 2, se puede apreciar la existencia de la variación con el paso del tiempo. Como ejemplo de esta variación, se muestra la tendencia promedio de la tasa de venta ofrecida a las personas que se jubilaron por “Edad legal” (o Jubilación Ordinaria), en el tipo de moneda “Soles Ajustados” en el año 2014.

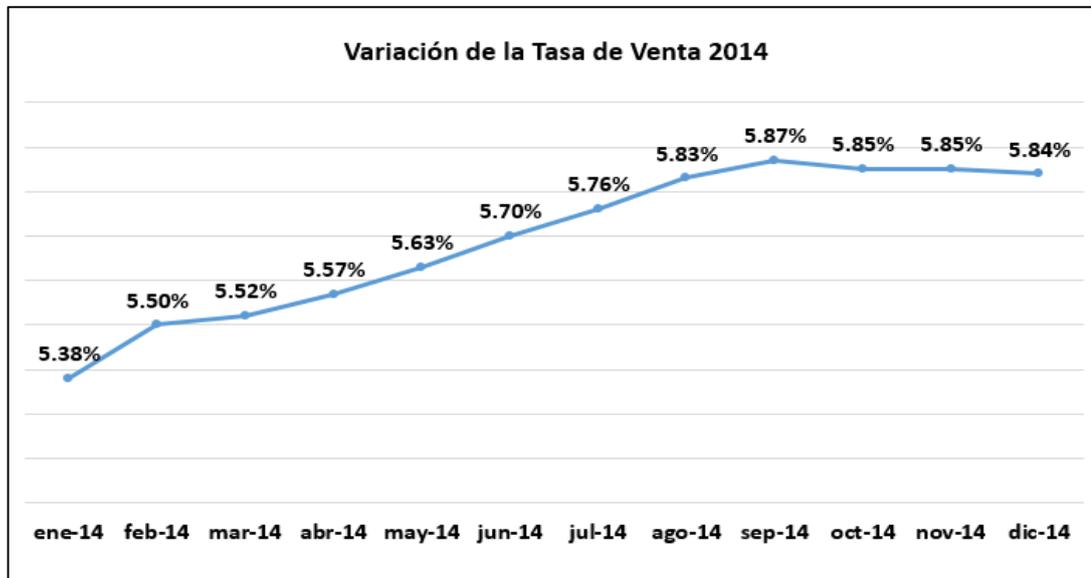


Figura 2: Tendencia de la tasa de venta en el año 2014.

Elaboración: los autores.

A la variación de la tasa de venta, dentro del proceso de generación de cotizaciones no existe estructuración de la información, los datos utilizados para el cálculo de la tasa de venta son ingresados y calculados de forma mecanizada por los sistemas transaccionales. (Ver Anexo 1: Entrevistas a expertos).

El problema general es la limitada funcionalidad de las herramientas tecnológicas usadas en la generación de cotizaciones para Rentas Vitalicias en las empresas aseguradoras del Perú.

El objetivo general es proponer la predicción de la tasa de venta para mejorar la generación de cotizaciones e incrementar la rentabilidad de la empresa.

Como objetivos específicos se precisan: Diseñar y construir un Data Mart que integre la información proveniente de los sistemas transaccionales que se utilizan en el área Comercial y alimenten la información para la predicción de la tasa de venta, elaborar un cálculo de predicción, utilizando técnicas de Data Mining, para obtener la tasa de venta más óptima en el

proceso de generación de cotizaciones para Rentas Vitalicias, ajustar el resultado de la predicción de la tasa de venta en el proceso de generación de cotizaciones para Rentas Vitalicias.

La justificación teórica se basa en la utilización de la técnica de árboles de decisión, para la predicción de tasas de venta en el proceso generación de cotizaciones al momento de vender una renta vitalicia a un afiliado que desee jubilarse; con la capacidad de ofrecer la mejor tasa de venta para competir con las demás empresas aseguradoras, dado que en los últimos años el mercado de rentas vitalicias está experimentando un crecimiento constante y con proyección de mayor crecimiento.

La justificación práctica implica que se logrará mejorar y automatizar la obtención de la tasa de venta en el proceso generación de cotizaciones de rentas vitalicias (siendo el producto “core” de la empresa) mediante técnicas de Data Mining. Esto conlleva a una reducción de tiempo en la ejecución del proceso mencionado y un ahorro, según sea el caso, en el desembolso del dinero adicional que la empresa entrega al futuro pensionista o jubilado por llevar su Cuenta Individual de Capitalización (CIC) con la empresa aseguradora.

Además, con la creación del Data Mart se logrará integrar la información de las áreas que participan en el negocio de rentas vitalicias, con lo que se obtiene una información confiable de los clientes y permitiendo de esta forma la generación de campañas del producto con una mayor prospección.

CAPÍTULO I MARCO TEÓRICO

1.1 Antecedentes

1.1.1 Sistema Privado de Pensiones (SPP)

De acuerdo con Kritzer (2011), en América Latina se desarrollaron las cuentas individuales de capitalización (CIC) destinadas a complementar o sustituir las prestaciones de los sistemas nacionales de pensión. Con la creación de las Administradoras de fondos de pensiones (AFP) 1981 en Chile, otros países de América Latina incorporaron la gestión privada de cuentas individuales como parte de sus sistemas de ingresos de jubilación a partir de la década de 1990.

En la última década, el mundo ha seguido observando a América, ya que estos sistemas de pensiones enfrentan retos de política en curso relacionadas con la cobertura, la tasa de contribución, los costos y la competencia.

Según los reportes mostrados por la SBS (2015), en los últimos años, el mercado asegurador peruano viene mostrando un crecimiento significativo. En el año 2014, la industria de seguros experimentó un crecimiento de 12% en primas netas con respecto al año anterior, representado S/. 10,154.1 millones de nuevos soles en el sector. Dicha evolución es explicada por el crecimiento del SPP, seguido de los Seguros

de Vida y los Seguros Generales que alcanzaron un incremento de 21%, 18% y 7%, respectivamente.

Cabe resaltar que los dos seguros más representativos, en términos de primas anuales, son las rentas vitalicias, que representaron el 19.1% de las primas vendidas, y los Seguros Vehiculares, que representaron el 12.3% de las ventas totales de seguros en 2014. Al cierre del año 2014, el patrimonio de las empresas que participaron en el mercado ascendió a S/.5,533.6 millones de nuevos soles, esto es 21.2% por encima de los S/.4,564.3 millones de nuevos soles registrados en 2013. El Retorno sobre Patrimonio (ROE) del mercado mostró un incremento, pasando de 15.19% registrado en 2013, a 18.26% en 2014. De manera similar, el Retorno sobre Activos (ROA) pasó de 2.6% a 3.0% en el mismo periodo. Esto se explica, en parte, por un aumento en el rendimiento del portafolio de inversiones, que mostró en promedio 7.59%; cifra ligeramente mayor a la que se obtuvo en el 2013 donde la rentabilidad promedio alcanzó el 7.55%.

Para el 2015 se pronostica un incremento en la cantidad de jubilados en el SPP, como se puede ver en la Figura 1.1, además de estabilidad para el sector dadas las adecuadas posiciones de capital y estabilidad de las empresas aseguradoras.



Figura 1.1: Jubilados del SPP.

Fuente: Asociación de AFP, 2015.

Asimismo, se prevé un mayor dinamismo por la recuperación del mercado interno y la creciente expansión de una cultura preventiva que permitiría una rápida evolución de las primas de seguros brutas en especial, dado el bajo nivel de penetración del mercado de seguros en el país comparado con el resto de los niveles promedio de América Latina.

En la actualidad, existen empresas, entre nacionales e internacionales, que se dedican a la creación de sistemas dedicados al rubro asegurador en el Perú.

Como se sabe, el Perú tiene un modelo de pensiones muy parecido al de Chile, es por ello que muchos de los sistemas que se utilizan en el país del sur, también se pueden aplicar para el nuestro. Ejemplo de ello, son las herramientas tecnológicas desarrolladas por empresas de Latinoamérica como se muestra a continuación:

- ACMASOFT

Empresa argentina especializada en el desarrollo a medida de aplicaciones para diversos rubros, en especial para el mercado de seguros. Se ha expandido a varios países de la región desarrollando diversos aplicativos como son:

Argentina:

Sistema de administración de Rentas Vitalicias Previsionales, HSBC-New York Life Seguros de Retiro (Argentina) S.A, SMG LIFE Compañía de Seguros de Retiro S.A.

Bolivia:

SIVAR. Sistema Integral de Valuación Actuarial de Reservas, Sistema de administración de Seguros Vitalicios de Provida Seguros.

Cabe mencionar que los sistemas mencionados, no contemplan la tendencia o la historia de las tasa de ventas al momento de cotizar, solo consideran los datos referente al cliente en cuestión.

1.1.2 Casos de éxito: Aplicación de Data Mart y Data Mining

La Inteligencia de Negocios y la Minería de Datos pueden ser empleadas por las empresas para incrementar su rendimiento, productividad y competitividad (INMON., 2005), esto la ha hecho trascender en todos los campos del mercado y ser adaptable a cualquier tipo o rubro de negocio: sector salud, sector financiero, sector educativo, etc.

Los avances de la tecnología de almacenamiento de datos han contribuido en el aumento del volumen y variedad de información que se encuentra en las bases de datos, esto ha hecho que el análisis de un gran volumen de información se haya convertido en una tarea rutinaria, compleja (Bissantz & Hagedorn, 2009; Kamath, 2006) y difícil de realizar manualmente (Goebel & Gruenwald, 1999).

La minería de datos trabaja de la mano con los almacenes de datos, sobre todo en los casos de volúmenes de datos muy grandes o de interrelaciones entre los datos complejas; es decir, que no puedan ser expresadas en una tabla plana. La minería de datos entrega información útil de una base de datos (CHANG, 2006).

Existen muchos casos de éxito en la aplicación de esta tecnología del tratamiento de la información, se muestran los siguientes:

- Predicción de fuga de clientes desde un enfoque de competencia (López Herrera, Leandro. Universidad de Chile, 2013): “Para abordar el problema de retención de clientes, las firmas utilizan enfoques basados en técnicas tradicionales de la estadística y el Data Mining. Para construir estos modelos, se utilizan datos históricos sobre el comportamiento y características de la cartera de clientes. Este problema también puede ser abordado desde un enfoque microeconómico, modelando la competencia de la industria y entendiendo la interacción y decisiones de los agentes participantes. Las firmas del mercado compiten en precios, descuentos y esfuerzos de marketing para atraer y retener a los consumidores.”

- Modelo de Activación de Tarjetas de Crédito en el mercado crediticio ecuatoriano (Pérez Tatamués, Alex. Escuela Politécnica Nacional - Ecuador, 2014): “Describe una metodología estadística basada en medidas de divergencia, medidas de asociación, árboles de decisión y regresión logística, que es muy empleada en la construcción de modelos de activación de tarjetas de crédito. Estos modelos toman en consideración el hábito de consumo actual e histórico que tiene un sujeto, con el fin de predecir la probabilidad de que dicho sujeto realice al menos un consumo con una nueva tarjeta, en una ventana de tiempo determinada posterior a la fecha de apertura de la misma.”
- Credit Scoring: Evaluación del riesgo crediticio de la cartera de microcréditos en una institución financiera en Uruguay (Arenas Díaz, María. Universidad de la República - Uruguay, 2012): “El uso de técnicas estadísticas de credit scoring permite estimar la probabilidad de incumplimiento de solicitantes de crédito así como contribuir a diseñar políticas crediticias acordes con el nivel de riesgo que puede asumir la institución financiera. Con el fin de contribuir a la discusión acerca de si estos modelos son útiles para el sector microfinanciero, se ha confeccionado un modelo de credit scoring para predecir el riesgo crediticio de la cartera de microcréditos de una institución financiera en Uruguay, utilizando de forma complementaria la técnica estadística de árboles de decisión y un modelo de regresión logística.”
- Decision Tree-Based technology credit scoring for startup firms: Korean Case: El gobierno de Corea del Sur asume una política de financiamiento a las nuevas empresas y/o *startups* con potencial de crecimiento y tecnología superior. La investigación que realizó (So Young Sohn, 2012) utilizó tecnologías basadas en árboles de decisión para hallar la calificación del score de créditos para las nuevas empresas. El estudio se realizó con 3,347 casos entre los años 1,999 y 2004 y se realizaron comparaciones actuariales utilizando 16 variables independientes, 9 indicadores económicos y 10 características de empresas; esto se hizo usando diferentes conjuntos predictores.

Consecuencia de este estudio, es la mejor capacidad de predicción de este score; esto se dio al comparar diferentes métodos de predicción, tales como son los árboles de decisión, la regresión logística, análisis de discriminantes, etc.

Los resultados más representativos lo obtuvo el árbol de decisión (77% de precisión) y la regresión logística (66% de precisión). Cabe mencionar que, como todo modelo de predicción basado en árboles de decisión, la muestra no representa toda la diversidad de tipos por empresas nuevas (para el estudio se utilizó empresas del tipo: servicios e industriales), a esto se le añade los indicadores económicos de impacto para las empresas en la realidad Coreana, por ello, la posibilidad de utilizar estos resultados, en otros países, es limitada.

- Classification trees in consumer studies for combining both product attributes and consumer preferences with additional consumer characteristics: En la investigación presentada por (Rosario Romano, 2014), se utiliza el enfoque de árboles de decisión para la combinación de los atributos del producto y las preferencias de los consumidores, centrándose en la segmentación de consumidores en términos de datos para determinar la aceptación de los productos en cada grupo de consumidores. Los resultados obtenidos del estudio revelaron que el uso de los árboles de decisión depende de la muestra que permite la eliminación de los valores atípicos y son fáciles de interpretar dado que las ramificaciones del árbol sugieren perfiles de consumidores tentativos a futuro.
- Desarrollo de un Modelo de Recomendación de Compra para Clientes de una Empresa de Seguros chilena (Méndez, Marcela del Pilar (2013): En la tesis, se propone el desarrollo de un modelo predictivo de compra de seguro, con la data e información disponible para los clientes de la línea vida, a fin de mejorar la efectividad de las campañas de venta cruzada de los seguros de automóvil sobre la cartera de clientes de los

productos de vida, a través de la identificación y focalización de recursos sobre los clientes con mayor probabilidad de contratación.

Se aplicó el enfoque KDD de minería de datos del cual, se consideró el desarrollo de diferentes alternativas de técnicas de modelación a fin de comparar los resultados obtenidos, el modelo final estuvo basado en árboles de decisión, elegido tanto por el ajuste de sus resultados como por decisión de la empresa.

El resultado del modelo construido, permite identificar y posicionar de manera periódica a los clientes que serán referidos al call center de la empresa, desde donde son contactados para ofrecerles la contratación del seguro.

1.2 Bases teóricas

1.2.1 Sistema Privado de Pensiones

El Sistema Privado de Pensiones (SPP) se creó mediante Ley N° 25897: “Ley de creación del Sistema Privado de Administración de Fondos de Pensiones” en 1992, que inicia sus operaciones en junio de 1993. Fue creado como alternativa al Sistema Nacional de Pensiones (SNP) que fue administrado por el Instituto Peruano de Seguridad Social (IPSS), y que se encontraba en declive por las malas administraciones y el uso de la liquidez del mismo para los déficits fiscales de los gobiernos de turno.

La principal característica del Sistema Privado de Pensiones es que administra un Fondo de Pensiones bajo la modalidad de Cuentas Individuales de Capitalización (CIC), por lo que la persona recibirá una pensión según los aportes efectuados durante su vida laboral y la rentabilidad generada por la Administradora de Fondo de Pensiones elegida.

La Administradora de Fondo de Pensiones (AFP) tiene como objetivo brindar a sus afiliados las prestaciones de pensiones de Jubilación, Invalidez, Sobrevivencia y reembolsos de Gastos de Sepelio. La primera prestación se brinda a partir de los 65 años cumplidos por el afiliado o en

caso el afiliado desee el régimen de Jubilación anticipada. Por otra parte, las prestaciones de Invalidez y Sobrevivencia se da a través de una Compañía de Seguros que consiste en la obtención de una pensión para los beneficiarios, en caso de que el afiliado sufra una invalidez temporal o permanente o fallezca.

Según el reporte de principales indicadores del sistema privado de pensiones (SBS, 2015), tanto la cantidad de jubilados y/o pensionistas, así como los montos desembolsados para el pago de pensiones han ido creciendo con el transcurrir de los últimos 5 años. Como se puede apreciar en la Tabla 1.1, entre los meses de enero – febrero 2015, ha tenido un incremento del 24% por encima de los mismos meses en el año 2014, mientras que en la Figura 1.2 se muestra la tendencia al alza con el que el mercado de rentas vitalicias se proyecta a años futuros.

Tabla 1.1: Número de nuevos pensionistas de jubilación.

	Acumulado			Mensual		
	Ene-Feb 2013	Ene-Feb 2014	Ene-Feb 2015	dic-2014	ene-2015	feb-2015
Jubilación Anticipada para Desempleados Ley 29426	318	291	281	121	133	148
Jubilación Legal	964	873	1 070	579	484	586
Jubilación Anticipada Ordinaria	40	44	39	18	16	23
Pensión Mínima Ley 27617	162	274	348	269	228	120
Jubilación Anticipada por Riesgo - Régimen Extraordinario	36	31	87	40	59	28
Jubilación Anticipada para Desempleados Ley 27617	1	9	0	-	-	-
Jubilación Anticipada por Riesgo - Régimen Genérico	5	10	41	14	20	21
Jubilación Anticipada para Desempleados Ley 28991	1	0	0	-	-	-
Jubilación Adelantada del Decreto Ley 19990	1	21	65	45	35	30
Total	1 528	1 553	1 931	1 086	975	956

Fuente: SBS, 2015.

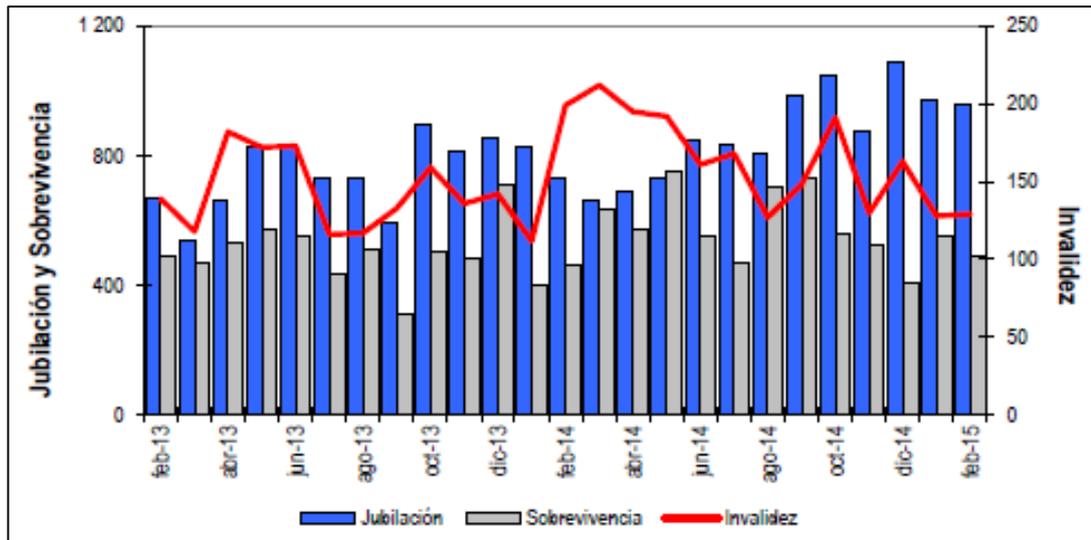


Figura 1.2: Número de nuevos pensionistas.

Fuente: SBS, 2015.

Como se ha mencionado anteriormente, el monto de lo desembolsado para el pago de los pensionistas ha ido incrementando en 21% con respecto a los primeros meses de los años 2014 y 2015, respectivamente debido al alza en el número de pensionistas de rentas vitalicias, tal como se muestra en la Figura 1.3.

	(En millones de nuevos soles)			
	2014	Acumulado		Var(%)
		Ene-Feb 2014	Ene-Feb 2015	
Renta Temporal con Renta Vitalicia Diferida	1712,3	258,5	292,4	13,1
REJA 29426 - Renta Vitalicia Diferida	376,9	54,4	68,5	25,9
Jubilación Anticipada - Régimen Extraordinario	109,7	11,3	40,1	254,5
Refiro Programado	118,4	19,2	20,8	8,4
REJA 29426 - Renta Temporal	109,8	18,5	18,8	1,6
Renta Vitalicia Familiar	89,7	8,8	16,7	89,3
REJA 29426 - Refiro Programado	50,3	8,2	8,8	6,7
Otros Egresos	75,9	12,0	9,0	(24,7)
Total	2 643,1	390,9	474,9	21,5

Figura 1.3: Egresos de fondo de pensiones para el pago de pensionistas.

Fuente: SBS, 2015.

1.2.2 Inteligencia de negocios

El primero que acuñó el término de Inteligencia de Negocios o “Business Intelligence” (BI) fue Howard Dresner (Gartner, 2012) para describir un conjunto de conceptos y métodos que mejorarán la toma de decisiones. Mediante el uso de las tecnologías y metodologías del BI, se pretende convertir datos en información y a partir de la información, ser capaces de descubrir conocimiento. Del glosario de términos de Gartner (2012) se extrae la siguiente definición: “BI es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un “datawarehouse”), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones. El proceso de BI incluye la comunicación de los descubrimientos y efectuar los cambios. Las áreas incluyen clientes, proveedores, productos, servicios y competidores.”

De acuerdo con Ranjan, Jayanthi (2009) el objetivo principal del BI es ayudar a las empresas a tomar decisiones que mejoren su rendimiento e impulsen su ventaja competitiva en el mercado. Pero es común una larga brecha entre la información que los responsables en la toma de decisiones requieren, y las grandes cantidades de datos que las organizaciones recopilan cada día. Para saltar de esta brecha, las organizaciones hacen significativas inversiones en desarrollar sistemas de BI para convertir los datos originales en información de utilidad.

El BI es útil para todas aquellas personas de una organización que tengan que tomar decisiones, depende de qué preguntas requieran responder, se establecerá el modelo de BI necesario y para hacerlo posible, se requiere de una serie de componentes. En la Figura 1.4, observamos los distintos componentes de la Inteligencia de Negocios:

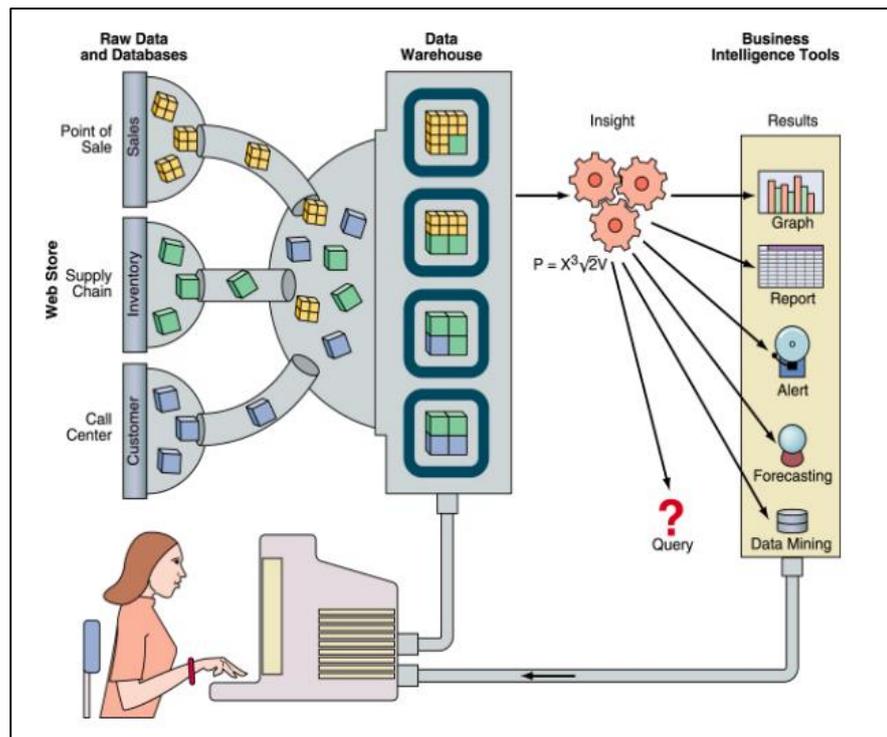


Figura 1.4: Componentes de Inteligencia de Negocios.

Fuente: Ranjan, Jayanthi, 2009

- OLAP (On-line analytical processing): Se refiere a la manera cómo los usuarios del negocio pueden utilizar herramientas sofisticadas que permitan la navegación entre las dimensiones. Provee vistas de negocio multidimensional.
- Análisis avanzado: Hace referencia a la minería de datos, la prospección y el análisis predictivo; este análisis toma ventaja de las técnicas de análisis estadístico para predecir o proveer medidas acertadas sobre los hechos.
- Administración del rendimiento corporativo: Es una categoría en general, utilizada para mostrar los *dashboards* que contienen las métricas combinadas del negocio.
- Inteligencia de Negocios en tiempo real: Permite a la organización la distribución de las métricas a través de emails, mensajería, sistemas y/o pantallas interactivas.
- DataWarehouse: Soporta la propagación física de los datos almacenando todos los datos de la organización para su integración, limpieza, agregación y consulta. También puede contener los datos

operacionales que pueden ser definidos como un conjunto actualizable de datos integrados para ser usado en la toma de decisiones tácticas a nivel organizacional.

- Data Marts: Es una colección de datos organizados para soportar la toma de decisiones dentro de un departamento o área en específico en la organización
- Data Source: Las fuentes de datos o “Data Source” pueden ser bases de datos relacionales u otras estructuras de datos que soporten las aplicaciones utilizadas en la organización como tablas, archivos planos o información no estructurada como archivos planos o información multimedia.

1.2.2.1 DataWarehouse y Data Mart

De acuerdo con (Reig, J et al, 2004) un Data Warehouse proporciona una visión global, común e integrada de los datos de la organización, independientemente de cómo se vayan a utilizar, posteriormente, por los consumidores o usuarios, con las propiedades siguientes: estable, coherente, fiable y con información histórica.

Al abarcar un ámbito global de la organización y con un amplio alcance histórico, el volumen de datos puede ser muy grande (centenas de terabytes), mientras que un Data Mart como un subconjunto de los datos del DataWarehouse, tiene el objetivo de responder a un determinado análisis, función o necesidad y con una población de usuarios específica. Al igual que en un DataWarehouse, los datos están estructurados en esquemas de estrella o copo de nieve y puede ser dependiente o independiente de un DataWarehouse.

Ante lo expuesto, se puede decir que la diferencia entre ambos conceptos es su alcance. El Data Mart está pensado para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización. Es el almacén natural para los datos

departamentales. En cambio, el ámbito del DataWarehouse es la organización en su conjunto. (De los Angeles, 2005).

Así se van solucionando los problemas de negocio más específicos, conociendo aspectos importantes como quién o quiénes son los clientes que más beneficios producen. Ofreciendo este concepto no es necesario remitirse a toda la información, se trata de obtener respuestas a preguntas. Para un mayor entendimiento de un Data Mart, se puede mencionar los elementos que lo conforman:

- Metadata: detalla las tablas que existen, así como las columnas y atributos de cada una de ellas y facilitan la labor de entendimiento y comprensión a los programadores del Data Mart.
- Middleware: asegura la correcta unión y conectividad entre los componentes de la estructura del Data Mart, es decir, que la información pueda integrarse.
- Mecanismos de extracción: realizan consultas a la información almacenada en el repositorio con el fin de conocer tendencias, proyecciones, comportamientos, estadísticas o indicadores de un tema o entidad en particular que sean necesarios en la toma de decisiones de los usuarios finales.
- Mecanismos de carga: para la carga de datos hacia un Data Mart, se pueden utilizar procesos batch, procesos ETL, procesos OLAP o simples queries de lenguaje SQL.

1.2.2.2 Data Mining

La minería de datos se refiere a la extracción de conocimiento a partir de grandes cantidades de datos, sobre los que se aplican métodos inteligentes con el fin de extraer patrones interesantes. Dichos datos pueden estar almacenados en bases de datos, almacenes de datos u otros repositorios de información.

El proceso de descubrimiento del conocimiento incluye la limpieza, integración, selección, transformación y minería de los datos para una posterior evaluación de patrones y presentación del conocimiento. (Valcárcel, 2004).

La minería de datos provee métodos automáticos para descubrir patrones en los datos. Las aplicaciones de minería de datos absorben y correlacionan los datos de una manera comprensible, contrario a lo que hacen los analistas humanos quienes son guiados por sus instintos de una manera secuencial. Las organizaciones que comúnmente se benefician de la minería de datos son las que:

- Cuentan con un gran volumen de datos.
- Tienen comunidades de colaboradores que necesitan entender los datos, pero no están preparados en temas de estadística.
- Tienen datos organizacionales complejos en su naturaleza, es decir, detallados y multifacéticos, con complejas relaciones entre ellos.

1.2.3 Metodologías

1.2.3.1 Metodologías para Data Mart

a) Metodología de Kimball

La metodología Kimball et al. (1998), se basa en el ciclo de Vida Dimensional del Negocio. Tiene cuatro principios básicos:

- Centrarse en el negocio: en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio.
- Construir una infraestructura de información adecuada: diseñar una base de información única, integrada, fácil de usar, de alto rendimiento.
- Realizar entregas en incrementos significativos: crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses.
- Ofrecer la solución completa: proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios.

b) Metodología de Inmon

En la metodología propuesta por (Inmon, 2002), el diseño de un Datawarehouse comienza con la introducción de datos en el mismo, debido a las grandes cargas de datos que deben hacerse antes de su introducción

en el Datawarehouse, dependiendo de ello la eficiencia de estos sistemas para acceder a los datos.

Esta metodología es utilizada con frecuencia para los Datawarehouse a nivel empresarial, que involucran desde un inicio todo el ámbito corporativo, sin centrarse en un incremento específico hasta después de haber terminado completamente el diseño del Datawarehouse.

En su filosofía, un Data Mart es solo una de las partes del Datawarehouse y son dependientes del depósito central de datos y, por lo tanto, se construyen después de él. El enfoque de Inmon de desarrollar una estrategia de Datawarehouse es identificar las áreas principales, desde el inicio del proyecto para asegurar una solución integral, ya que esto ayuda a evitar la aparición de situaciones inesperadas que puedan poner en peligro el proyecto, debido a que se conoce con antelación y exactitud la estructura que presentarán los principales núcleos del desarrollo. Ello permitirá enfocar los esfuerzos del desarrollo actual para ser compatibles con los subsiguientes.

1.2.3.2 Metodologías para Data Mining

a) Metodología CRISP-DM (Cross-Industry Standard Process for Data Mining)

La metodología propuesta por (Chapman et al, 2000), está estructurada en un proceso jerárquico, compuesto por tareas descritas en cuatro niveles diferentes de abstracción, que van desde lo general a lo específico.

CRISP-DM propone, en el nivel más alto, seis fases para el proceso de minería de datos: entendimiento del negocio, entendimiento de los datos, preparación de los datos, modelado, evaluación e implementación.

b) Metodología KDD (Knowledge Discovery in Databases)

Como se presenta en (Fayyad et al, 1996), el descubrimiento de conocimiento en bases de datos "KDD", se utilizan métodos de Data Mining para extraer lo que se considera conocimiento de acuerdo con la

especificación de medidas y umbrales, y se utiliza una base de datos en conjunto con el pre-procesamiento requerido, el muestreo, y la transformación de la base de datos.

Se consideran cinco etapas:

Selección: esta etapa consiste en la creación de un conjunto de datos de destino o concentrarse en un subconjunto de variables o muestras de datos, en el que se va a realizar el descubrimiento.

Pre-procesamiento: esta etapa consiste en la limpieza de datos de destino y el pre procesamiento con el fin de obtener datos consistentes.

Transformación: consiste en la transformación de los datos mediante métodos de reducción de dimensionales o de transformación.

Minería de datos: esta etapa consiste en la búsqueda de patrones de interés en una forma de representación particular, dependiendo del objetivo de Data Mining (por lo general, la predicción).

Interpretación / Evaluación: consiste en la interpretación y evaluación de los patrones extraídos.

c) Metodología SEMMA

(Sample, Explore, Modify, Model and Assess)

El acrónimo SEMMA significa Muestra, Explorar, Modificar, Modelar, Evaluar, y se refiere al proceso de realización de un proyecto Data Mining. El Instituto SAS propone un ciclo con cinco etapas para el proceso:

Muestra: consiste en el muestreo de los datos mediante la extracción de una porción de un gran conjunto de datos suficientemente grande que contenga información significativa, pero lo suficientemente pequeño para manipularla rápidamente.

Explorar: consiste en la exploración de los datos mediante la búsqueda de tendencias y anomalías imprevistas con el fin de obtener una mayor comprensión.

Modificar: consiste en la modificación de los datos mediante la creación, selección y transformación de las variables para enfocar el proceso de selección de modelo.

Modelar: consiste en el modelado de los datos al permitir que el software busque automáticamente una combinación de datos que predice de forma fiable el resultado deseado.

Evaluar: consiste en la evaluación de los datos mediante la evaluación de la utilidad y la fiabilidad de los resultados del proceso de Data Mining y estimar su funcionalidad.

1.3 Definición de términos básicos

1.3.1 Sistema Privado de Pensiones

- **Administradora de Fondos de Pensiones:**

Empresa privada constituida con el objeto exclusivo de administrar los Fondos de Pensiones de los trabajadores que se encuentran afiliados a esta y de otorgar las prestaciones de jubilación, invalidez, sobrevivencia y gastos de sepelio en el Sistema Privado de Pensiones (SPP)

- **Bono de reconocimiento:**

Documento (Título Valor) mediante el cual el Estado reconoce al trabajador que decidió pasarse al SPP, las aportaciones que realizó al SNP.

- **Cuenta individual de capitalización:**

El afiliado, a lo largo de su vida laboral aporta el dinero que servirán directamente para el pago de su pensión (CIC), este siempre permanece en una cuenta personal por cada trabajador.

- **Jubilación:**

Beneficio al cual acceden los afiliados al SPP que cumplen con alcanzar los sesenta y cinco (65) años de edad o, alternativamente, cumplen los requisitos de acceso a una jubilación anticipada, con garantía estatal o sin ella. En virtud de la jubilación, el afiliado elige libremente contratar una pensión o bien con su AFP o con alguna empresa de seguros que le pueda ofrecer un producto que se adecue a sus necesidades particulares.

- **Mercado electrónico de rentas y retiros (MELER):** Plataforma electrónica de contratación, de propiedad de la Superintendencia, a la

cual los afiliados o beneficiarios, a través de las AFP, y las empresas de seguros, concurren a fin de contratar algún producto previsional materializado en una pensión.

- **Tasa de cotización de rentas vitalicias (o Tasa de venta):** Es la tasa utilizada por la empresa de seguros para calcular el capital requerido de las pensiones bajo la modalidad de renta vitalicia a otorgar al afiliado y/o sus beneficiarios, de ser el caso. Esta tasa es fijada libremente por las empresas de seguros.

- **Tipos de jubilación:**

Como se puede apreciar en la Tabla 1.2, los tipos de jubilación se diferencian, principalmente, en el cumplimiento en los requisitos que estos solicitan. Los tipos de jubilación que tiene el Sistema Privado de Pensiones son los siguientes:

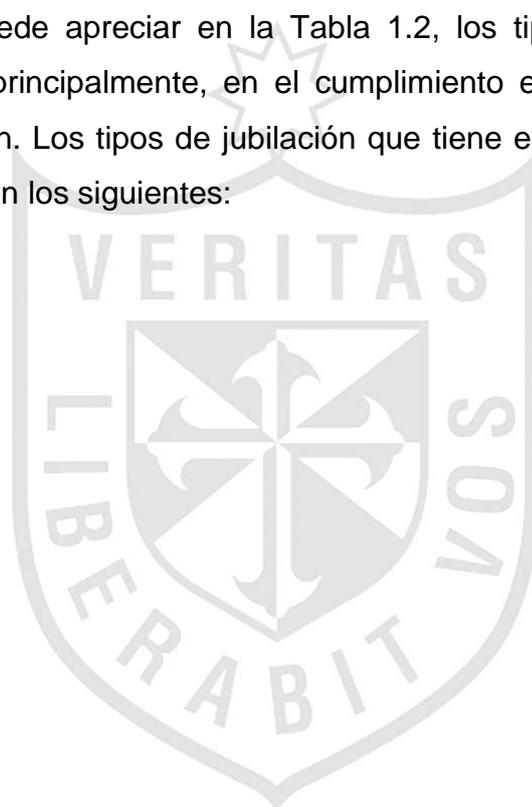


Tabla 1.2: Tipo de jubilación: Requisitos principales.

Tipo de Jubilación	Requisitos principales	Datos adicionales
Jubilación por edad legal	Estar afiliado a una AFP y Tener 65 años de edad.	Los años cumplidos en meses y días.
Jubilación Anticipada Ordinaria	Cálculo de pensión \geq al 50% del promedio de las remuneraciones de últimos 120 meses. Densidad de Cotización : 60%	Evaluación de acceso en base al Retiro Programado.
Jubilación Anticipada por Desempleo (REJA)	Edad: 55 años (varones) y 50 años (mujeres) Desempleado: 12 meses consecutivos o más. Pensión calculada resulte \geq a RMV.	Si pensión es $<$ a RMV: Devolución 50% de la CIC.
Jubilación Anticipada Riesgo – Régimen Genérico	Realizar aportes complementarios a la CIC a razón de: 4% en caso de Minería (reduce 2 años respecto de la edad legal de jubilación) 2% en caso de Construcción Civil (reduce 1 año respecto de la edad legal de jubilación)	Mismo universo de trabajadores del régimen anterior. Reduce la edad de jubilación por cada 36 meses de aporte.

Elaboración: los autores.

- **Modalidades de Jubilación:**

En el Sistema Privado de Pensiones existen 5 modalidades de pensión vigentes, estas influyen en el modo en el que el jubilado recibirá su pensión mensual y la rentabilidad que esta última tendrá a lo largo del tiempo.

➤ **Retiro Programado:** El afiliado mantiene propiedad de la CIC. Efectúa retiros mensuales contra el saldo de dicha cuenta. Los fondos que no se lleguen a utilizar constituyen herencia siempre y cuando no queden beneficiarios. Es otorgado por la AFP y es la única opción revocable.

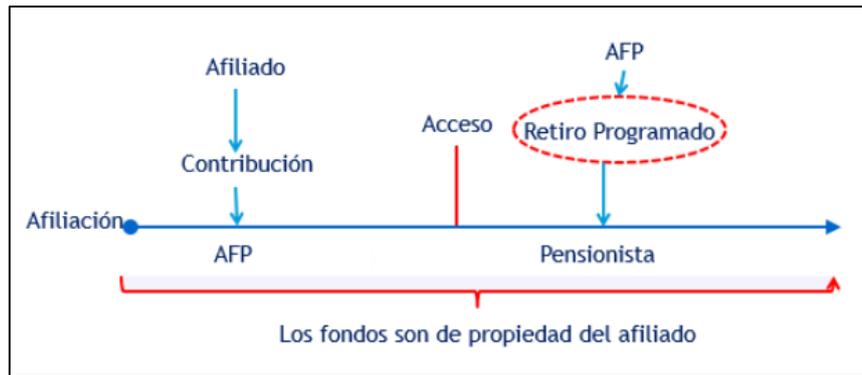


Figura 1.5: Retiro Programado.

Elaboración: los autores.

- **Renta Vitalicia Familiar:** Se contrata con una empresa de seguros el pago de una renta mensual hasta el fallecimiento del afiliado y el pago de pensiones de sobrevivencia al fallecimiento del titular de la cuenta. La CIC pasa a ser propiedad de la Empresa de Seguros.



Figura 1.6: Renta Vitalicia Familiar.

Elaboración: los autores.

- **Renta Temporal con Renta Vitalicia Diferida:** El fondo se divide en dos: Una parte queda en la AFP para que pague una Renta Temporal por máximo 2 años; y vencido este plazo empieza a pagar (Renta Vitalicia Familiar) una compañía de seguros. Es irrevocable.

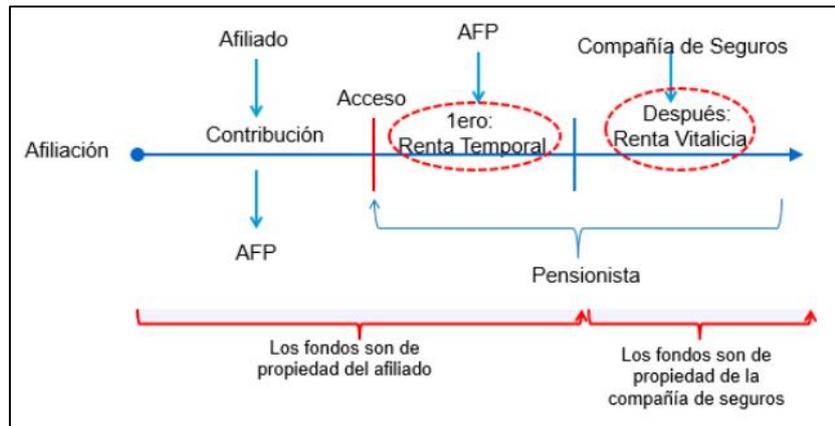


Figura 1.7: Renta Temporal con Renta Vitalicia Diferida.

Elaboración: los autores.

- **Renta Mixta:** El fondo de la CIC se divide en partes iguales. El 50% queda en la AFP (Retiro Programado) y al mismo tiempo, con el otro 50% se paga una Renta Vitalicia a cargo de una empresa de seguros. La pensión total corresponderá a la suma de los montos de pensión percibidos por cada una de las modalidades. Es irrevocable.



Figura 1.8: Renta Mixta.

Elaboración: los autores.

- **Renta bimoneda:** El afiliado contrata dos (2) Rentas Vitalicias de manera simultánea; Una en moneda nacional (S/) y la otra en dólares americanos (US\$), ambas otorgadas por la misma empresa de seguros. La pensión total corresponderá a la suma de los montos percibidos por cada una de las monedas. Es irrevocable.

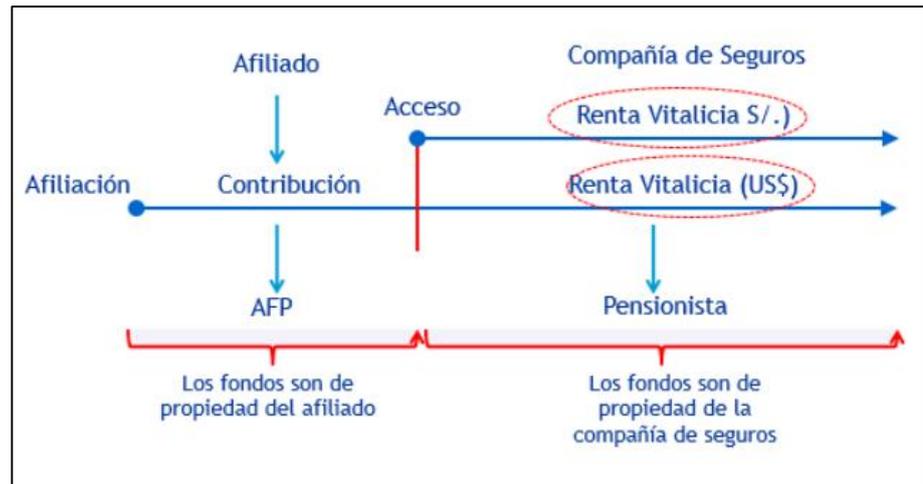


Figura 1.9: Renta Bimoneda.

Elaboración: los autores.

- **¿Sistema Privado de Pensiones u Sistema Nacional de Pensiones?**

Las diferencias más notorias entre ambos sistemas de pensiones se pueden observar en la Tabla 1.3.

Tabla 1.3: Diferencias entre SPP y SNP.

Características	Sistema Nacional de Pensiones	Sistema Privado de Pensiones
Modelo	Reparto Simple (Beneficio Definido)	Capitalización Individual
Administración	Oficina de Normalización Previsional (ONP)	AFP's
Alcance	En ambos casos: Trabajadores, cualquiera sea la modalidad de trabajo que realicen. Elección del sistema: Voluntaria por parte de trabajadores dependientes e independientes.	
Tasa Aporte	13%. Recaudación va un fondo común.	10% aporte obligatorio. Comisión AFP (% variable) Prima de seguro de invalidez o sobrevivencia (% variable)
Prestaciones	Jubilación, invalidez, viudez, orfandad, ascendencia y capital.	Jubilación, invalidez, sobrevivencia y gastos de sepelio.
Definición Pensión	En función a los años de aportación y remuneración promedio. Si no se alcanza cuando menos 20 años no existe derecho a pensión.	En función al saldo, producto y grupo familiar. Siempre se otorga una pensión con lo que hubiera acumulado en la cuenta.
Productos	Renta Vitalicia	Retiro Programado, Rentas Vitalicias con diferentes opciones de protección, duración y moneda.
Posibilidad de Herencia	No	Sí, respecto del saldo que quedara en la cuenta si es que no tiene beneficiarios.

Elaboración: los autores.

Como se aprecia en el cuadro comparativo, la elección del sistema a elegir está ligado a muchas variables que van desde lo económico, profesional y social. Un afiliado al SNP puede cambiarse al SPP, pero no en el sentido contrario.



CAPÍTULO II METODOLOGÍA

En toda investigación, tecnológica o científica, es absolutamente fundamental la prueba, hasta un nivel prefijado de confianza (Palacios Quichiz, 2012). En el presente capítulo se indican y describen los costos de los materiales, los recursos y la rentabilidad esperada del proyecto; así también, se muestran las diversas metodologías que son aplicables para el desarrollo de un Data Mart y Data Mining, comparándolas entre ellas, y hallar la más adecuada para la elaboración de este proyecto.

2.1 Materiales

2.1.1 Recursos humanos

En la presente tesis, participarán dos (2) personas que desempeñarán los roles de Gestor de Proyecto, Analista funcional y Analista de Base de Datos.

2.1.2 Hardware

En la tabla 2.1, se detallan los recursos de hardware que requieren para la implementación del proyecto.

Tabla 2.1: Recursos de hardware

Equipo	Detalle	Cantidad
Servidor Data Mart	Intel Xeon E5-2640 , 16 GB de RAM y 1TB disco	1
Laptop	Intel core i3, 8 GB de RAM	1
PC estación	AMD PHENOM 2 X4, 6GB de RAM	1

Elaboración: los autores.

2.1.3 Software

Con respecto al software a utilizarse en la tabla 2.2, se especifican la cantidad de licencias y tipos de software requeridos.

Tabla 2.2: Recursos de software.

Software	Tipo	Licencias
IBM SPSS 22	Software estadístico	1
SQL Server 2008 R2	Sistema Gestor de BD	1
Microsoft Office 2013	Software de Oficina	2

Elaboración: los autores.

2.1.4 Presupuesto

De los recursos mencionados en los puntos anteriores, se debe considerar que el hardware y las licencias de software son proporcionados por la empresa donde se desarrolla la presente tesis, que serán mencionados para el cálculo de la rentabilidad del proyecto.

De acuerdo con el cronograma del proyecto se estima una duración de 66 días para su realización. (Ver Anexo 2: Cronograma del Proyecto).

2.1.4.1 Costos de Personal

Para la ejecución del proyecto, se necesitará de los roles que desempeñara del personal humano. En la tabla 2.3, se detallan los roles y costos para el proyecto.

Tabla 2.3: Costo de los recursos humanos.

Recurso	Horas dedicadas	Costo por Hora	Costo Total
Gestor de Proyecto	90	S/. 80.00	S/. 7,200.00
Analista Funcional	110	S/. 60.00	S/. 6,600.00
Analista de Base de Datos	110	S/. 60.00	S/. 6,600.00
TOTAL			S/. 20,400.00

Elaboración: los autores.

2.1.4.2 Costos directos

Dentro de los costos directos son aquellos necesarios para la implementación del proyecto y que participan directamente en el desarrollo de la actividad. Se muestran en la siguiente tabla 2.4.

Tabla 2.4: Costos directos.

Nombre	Tipo	Cantidad	Costo Total
Servidor Data Mart	HW	1	S/. 20,000.00
Laptop	HW	1	S/. 2,000.00
PC estación	HW	1	S/. 4,000.00
IBM SPSS 22	Licencia SW	1	S/. 16,500.00
SQL Server 2008 R2	Licencia SW	1	S/. 24,000.00
Microsoft Office 2013	Licencia SW	2	S/. 500.00
TOTAL			S/. 67,000.00

Elaboración: los autores.

2.1.4.3 Costos indirectos

Están conformados por los servicios que se utilizan, indirectamente, y que permiten el desarrollo del proyecto. Se detallan en la tabla 2.5.

Tabla 2.5: Costos indirectos.

Materiales y Servicios	Costo Mensual	Total Anual
Internet	S/. 140	S/. 1,680.00
Útiles de oficina	S/. 50	S/. 600.00
Alquiler Oficina	S/. 600	S/. 7,200.00
Otros	S/. 40	S/. 4,000.00
TOTAL		S/. 9,960.00

Elaboración: los autores.

2.1.4.4 Costos de Operación

Como se muestra en la tabla 2.6, estos costos están conformados por las licencias de software y el mantenimiento de los equipos.

Tabla 2.6: Costos de operación.

Importe	Costo Anual
Licencia SQL Server 2008 R2	S/. 24,000
Licencia IBM SPSS	S/. 16,500
Mantenimiento	S/. 8,000
TOTAL	S/. 48,500

Elaboración: los autores.

2.1.5 Ingresos

Los ingresos serán estimados de acuerdo con el ratio de rentabilidad que obtiene la empresa sobre sus inversiones, que actualmente oscila entre 9% a 10% de acuerdo a (Class & Asociados, 2015).

Para el cálculo del ingreso por rentas vitalicias anuales del proyecto se realizó una estimación de 10 ventas en un mes, con un CIC en

valor promedio de S/. 200,000 debido a que representa un valor significativo para que la empresa realice una oferta ver tabla 2.7.

Tabla 2.7: Estimación de las primas de seguros anuales.

Cantidad Vendida	CIC Promedio	Ingreso por RV Mensual	Ingreso por RV Anual
10	S/. 200,000	S/ 2,000,000	S/ 24,000,000

Elaboración: los autores.

El margen de contribución del proyecto fue obtenido calculando el 9% del retorno sobre la inversión realizada, para esto se tomó el monto proveniente de los CIC ganados al cerrar una venta, ya que ellos serán utilizados para que la empresa realice inversiones durante el ejercicio ver tabla 2.8.

Tabla 2.8: Estimación del margen de contribución del proyecto.

Margen de Contribución Mensual	Margen de Contribución Anual
S/. 180,000	S/. 2,160,000

Elaboración: los autores.

En cuanto la estimación del VAN, se tiene el siguiente cuadro graficado en la Tabla 2.9:

Tabla 2.9: Estimación del VAN.

EVALUACIÓN DEL BENEFICIO DE LA IMPLANTACIÓN DEL PROYECTO PARA LA PREDICCIÓN DE LA TASA DE VENTAS EN RENTAS VITALICIAS
(EN S/.) (*)

	Año 0	Año 1	Año 2	Año 3	Año 4	Año 5	Año 6	Año 7	Año 8	Año 9	Año 10
Ingresos											
Margen de Contribución		S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000
Total		S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000	S/. 2,160,000
Costo de la Operación											
Inversiones	S/. 87,400										
Operación		S/. 48,500	S/. 48,500	S/. 48,500	S/. 48,500	S/. 48,500	S/. 48,500	S/. 48,500	S/. 48,500	S/. 48,500	S/. 48,500
Gastos	S/. 9,960	S/. 9,960	S/. 9,960	S/. 9,960	S/. 9,960	S/. 9,960	S/. 9,960	S/. 9,960	S/. 9,960	S/. 9,960	S/. 9,960
Total	S/. 97,360	S/. 58,460	S/. 58,460	S/. 58,460	S/. 58,460	S/. 58,460	S/. 58,460	S/. 58,460	S/. 58,460	S/. 58,460	S/. 58,460
Flujo Neto	-S/. 97,360	S/. 2,101,540	S/. 2,101,540	S/. 2,101,540	S/. 2,101,540	S/. 2,101,540	S/. 2,101,540	S/. 2,101,540	S/. 2,101,540	S/. 2,101,540	S/. 2,101,540
TMAR		10%									
VAN 5 AÑOS		S/. 7,869,130									
VAN 8 AÑOS		S/. 11,114,201									
VAN 10 AÑOS		S/. 12,815,694									

(*) A una estimación de rentabilidad sobre la inversión, por el proyecto, del 9% anual.

Elaboración: los autores.

2.2 Métodos

Para el desarrollo del proyecto se explican las metodologías seleccionadas, así como los criterios de selección para cada una ellas.

2.2.1 Metodología para el Desarrollo del Data Mart

Para la elaboración del Data Mart, se utilizará la metodología Kimball, que se basa en el ciclo de Vida Dimensional del Negocio.

Tiene cuatro principios básicos:

- Centrarse en el negocio
- Construir una infraestructura de información adecuada
- Realizar entregas en incrementos significativos
- Ofrecer la solución completa

Contiene cuatro etapas definidas, como se muestra en la Figura 2.1:



Figura 2.1: Etapas de la Metodología Kimball.

Fuente: Kimball et al, 2008.

- Etapas de la metodología de Kimball

- a) Planificación

Determina el propósito del proyecto de BI, sus objetivos específicos y el alcance del mismo, los principales riesgos y una aproximación inicial a las necesidades de información.

- b) Análisis

Define los requerimientos del negocio, sienta las bases para identificación de los procesos y las dimensiones a implementar.

- c) Diseño

Crea un modelo dimensional e iterativo a partir de los procesos identificados en la etapa anterior.

- ✓ Elegir el proceso de negocio.
- ✓ Establecer el nivel de granularidad.
- ✓ Elegir las dimensiones.
- ✓ Identificar medidas y las tablas de hechos.

- d) Implementación

Selecciona las herramientas, tanto software como hardware, y la solución de Inteligencia de Negocios para la extracción de la información.

Para la selección de la metodología de Kimball se realizó un cuadro comparativo como se muestra en la tabla 2.10, entre ellos y luego se procedió a realizar la ponderación como se detalla en la tabla 2.11, y que permite determinar la más adecuada para el proyecto.

Tabla 2.10: Comparación de metodologías para Data Mart.

Metodología	Kimball	Inmon
Enfoque general	Bottom-up	Top-down
Estructura de la Arquitectura	Data Marts por proceso de negocio.	Toda la empresa (atómico)
Complejidad	Mediana	Alta
Orientación de datos	Orientado a procesos	Orientado a datos
Accesibilidad del usuario final	Alta	Baja
Objetivo	Solución que facilita la consulta de datos por los usuarios.	Solución técnica basada en métodos y tecnología comprobada.

Elaboración: los autores.

La ponderación realizada fue a base de la información y enfoque que se tiene del negocio, y tener en cuenta que se sitúa en un área específica.

Tabla 2.11: Ponderación por metodología Data Mart.

Metodología	Kimball	Inmon
Enfoque general	4	3
Estructura de la Arquitectura	4	3
Complejidad	3	4
Orientación de datos	4	3
Accesibilidad del usuario final	5	3
Objetivo	4	3
Total	24	19

Elaboración: los autores.

La metodología que mejor se adecúa a las necesidades del proyecto es Kimball, según los resultados que se muestran en la Tabla 2.10 y Tabla 2.11

2.2.2 Metodología de Data Mining

Para el desarrollo de proyectos que utilizan técnicas de Data Mining, se tienen varias metodologías reconocidas. Para este proyecto se utilizará la metodología CRISP-DM.

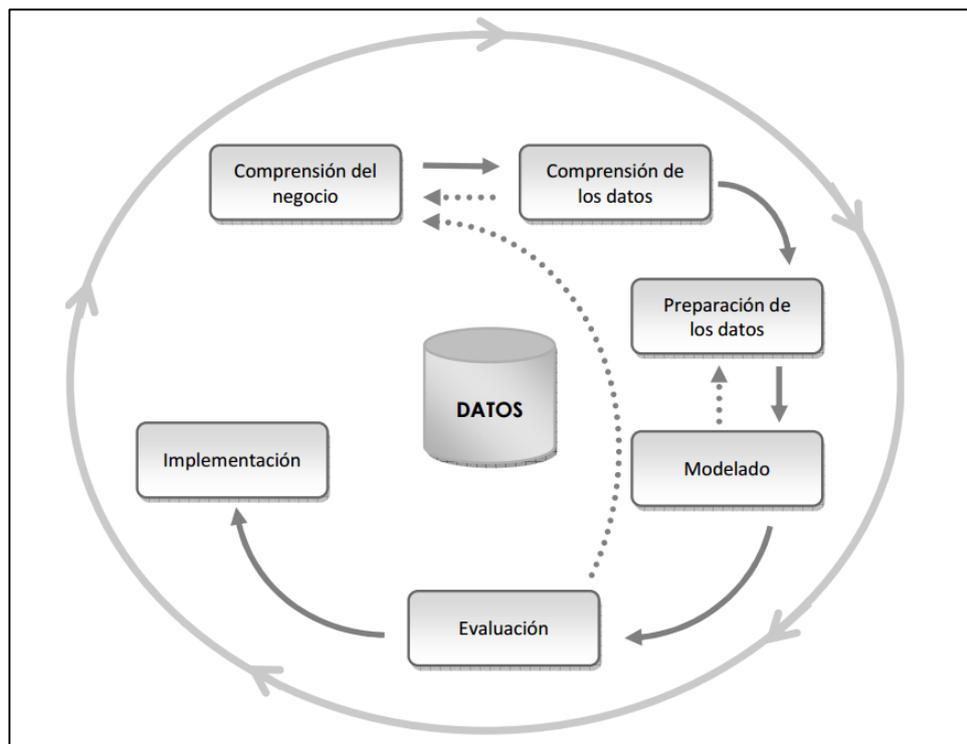


Figura 2.2: Fases del modelo de referencia CRISP-DM.

Fuente: Chapman et al, 2000.

Como se muestra en la Figura 2.2, las seis fases que componen el proceso de minería de datos para esta metodología son:

- **Comprensión del negocio:** En esta fase, se determinan los objetivos y requerimientos del proyecto desde una perspectiva del negocio, que definen el problema de minería y el plan de trabajo.
- **Comprensión de los datos:** Fase que consiste en la recolección de datos que se utilizará en el proyecto y la familiarización con los mismos

- Preparación de los datos: Comprende aquellas actividades de tratamiento de los datos para construir la vista que se puede explotar o conjunto de datos finales sobre los cuales se aplicarán las técnicas de minería.
- Modelado: En esta etapa, se aplican las diversas técnicas y algoritmos de minería sobre el conjunto de datos para obtener la información oculta y los patrones implícitos en ellos.
- Evaluación: Fase en la que se analizan los patrones obtenidos en función de los objetivos organizacionales. En esta etapa, se debería determinar si se ha omitido algún objetivo importante del negocio.
- Implementación: Consiste en la comunicación e implementación del nuevo conocimiento, que debe ser representado de forma comprensible para el usuario.

Para la selección de la metodología CRISP-DM, se realizó un cuadro comparativo, como se muestra en la tabla 2.12, luego se procedió a realizar la ponderación entre ellos como se detalla en la tabla 2.13, y que permite precisar la más adecuada para el proyecto.

Tabla 2.12: Comparación de metodologías para Data Mining.

PASOS	KDD	SEMMA	CRISP-DM
Planificación	Pre KDD	-	Determinar los objetivos y requerimientos del proyecto
Selección de Datos	Creación de un conjunto de datos de destino	Realizar el muestreo de datos mediante extracción de información significativa	Recolección de datos a utilizar en el proyecto para facilitar su comprensión
Limpieza de Datos	Limpieza de Datos para obtener datos consistentes	Explorar los datos mediante la búsqueda de tendencias y/o anomalías	
Transformación	Utilizar métodos de reducción de dimensión de datos	Modificar los datos mediante la transformación de variables	Comprende las actividades de tratamiento de datos sobre el cual se aplicara las técnicas de DM
Modelado	Búsqueda de patrones dependiendo del objetivo	Búsqueda de combinaciones de datos a predecir	Se aplican técnicas y algoritmos de DM sobre el conjunto de datos buscando patrones implícitos
Evaluación	Interpretación y evaluación de los patrones extraídos	Evaluación de los datos por utilidad y confiabilidad de los resultados	Se analizan los patrones obtenidos en función de los objetivos planteados.
Implementación	Post KDD	-	Implementación del nuevo conocimiento

Elaboración: los autores.

Tabla 2.13: Ponderación por metodología.

PASOS	KDD	SEMMA	CRISP-DM
Planificación	3	0	5
Selección de Datos	4	3	4
Limpieza de Datos	4	4	4
Transformación	4	3	4
Modelado	3	3	5
Evaluación	4	4	5
Implementación	3	0	4
Total	25	17	31

Elaboración: los autores.



CAPÍTULO III

DESARROLLO DEL PROYECTO

Para el desarrollo del proyecto, debido al alcance y a la funcionalidad del mismo, se han definido dos etapas:

- Desarrollo de la Evaluación predictiva (Predictor)
- Desarrollo del Data Mart

Cabe mencionar que, si bien es cierto, la funcionalidad de ambas partes del proyecto es paralela, estas herramientas se integran para cubrir todas las necesidades desde el punto de vista de la información y de método, para así, alcanzar la mayor eficacia del cálculo de la Tasa de Venta.

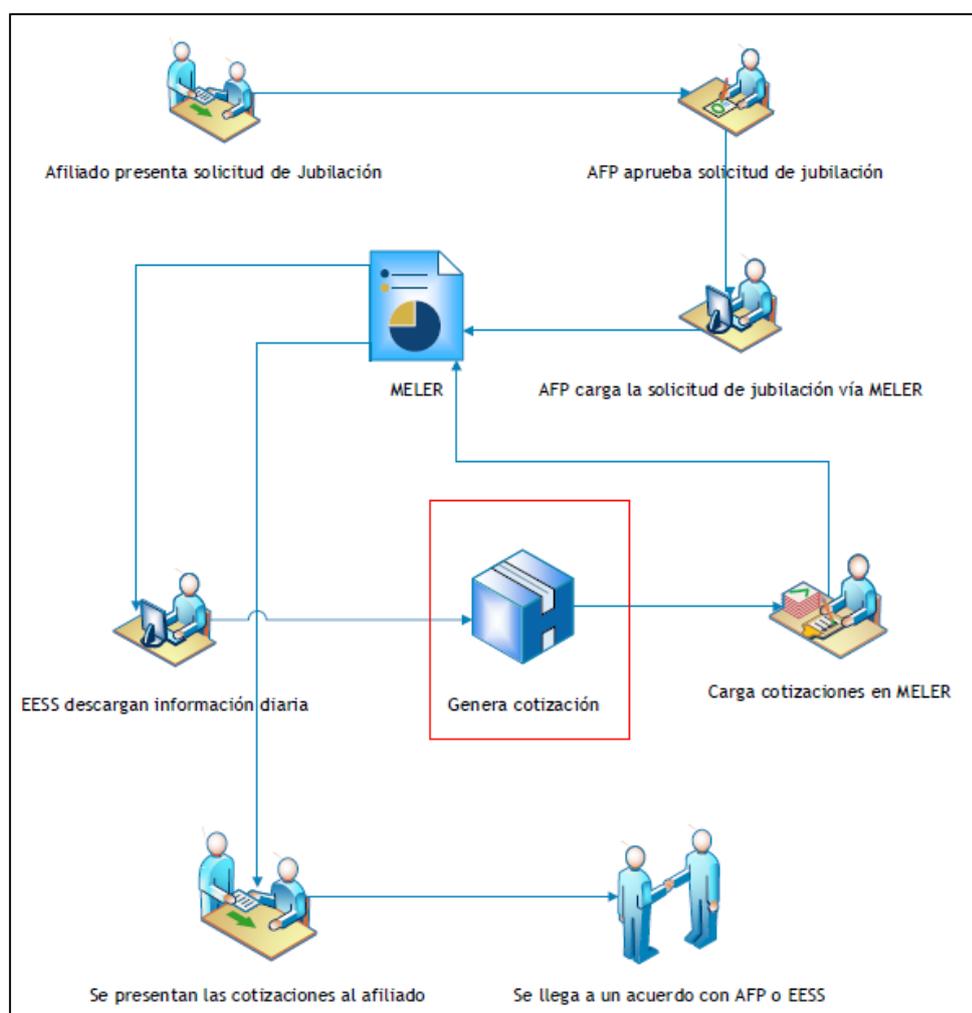
3.1 Desarrollo de la evaluación predictiva (PREDICTOR)

En la presente tesis, se propondrá la creación de un método (algoritmo junto con las reglas del negocio de la empresa aseguradora) que permita la predicción de la mejor tasa de venta para las cotizaciones en Rentas Vitalicias, mejorando la generación de cotizaciones para los futuros pensionistas. Para ello se detallarán las etapas de la metodología CRISP-DM aplicadas al proyecto.

3.1.1 Comprensión del Negocio

El proceso de venta de una renta vitalicia se inicia desde que el afiliado solicita su jubilación a su AFP, independientemente, de la modalidad y el tipo que sea. Una vez realizada la solicitud, por norma dictada por la SBS, la AFP notifica la jubilación a las empresas aseguradoras y se comienza el proceso detallado en la Figura 3.1:

Figura 3.1: Flujo de venta de renta vitalicia.



Elaboración: los autores.

Como se aprecia en la Figura 3.1, el proceso de venta de una renta vitalicia comienza cuando el afiliado a alguna AFP solicita su jubilación a dicha entidad, esta última procede con la aprobación de dicha solicitud y la carga al MELER - SBS (Mercado Electrónico de Rentas y Retiros - SBS) que juega un papel de comunicación entre la SBS y las empresas aseguradoras.

La data, con la información de las solicitudes de jubilación, son extraídas de forma diaria por todas las empresas aseguradoras para generar las cotizaciones (proceso al cual apunta este proyecto) a ofrecer a los futuros pensionistas; una vez generada las cotizaciones, estas son enviadas nuevamente (vía MELER) a las AFP para la presentación ante la persona, y ella elija en qué compañía obtener su jubilación y la modalidad de pago de su pensión.

3.1.2 Comprensión de los datos

En esta parte de la metodología, se realizará la extracción de los datos que se encuentran en las bases de datos transaccionales de la empresa, y se pueden clasificar de la siguiente manera según la unidad de negocio a la que pertenecen:

- Rentas Vitalicias
 - RviAdm, MercadoRV, RV, Cotizaciones
- Inteligencia Comercial
 - Exp_Serv, Contact, LasLomas, Comercial

Luego de la extracción de los datos, se considerarán a aquellas pólizas que cerraron en el último mes móvil, es decir, 28 días atrás al día anterior en que se extrae la data, así como las cotizaciones ofrecidas por las compañías de seguro en el mismo rango de tiempo. Esto nos servirá para la obtención de la tendencia de la tasa de venta en el último mes y dependerá de los diferentes perfiles en los que se clasifican los asegurados (jubilados).

3.1.2.1 Variables independientes

Son aquellos datos procedentes de las bases de datos transaccionales y que sirven para distribuir la base de muestra, y posibilitar la identificación del valor pronosticado en la variable dependiente. Algunas de las variables independientes más representativas son:

- Moneda, compañía, categoría, modalidad, porcentaje de periodo garantizado, edad, CIC, zona (Lima, Provincia), departamento, ubicación geográfica (Clasificación Arellano Marketing).

3.1.2.2 Variables dependientes

Es el dato a pronosticarse, en este caso en particular, la tasa de venta. Es resultado del árbol de decisión, y es altamente cambiante según las variables independientes que compongan el modelo predictivo. Generalmente, es un valor numérico cuya medida tendencial es la media aritmética.

3.1.3 Preparación de los datos

En esta etapa, se presentan los pasos para realizar la extracción y limpieza de los datos.

Como se muestra en la Figura 3.2, el primer paso es extraer los datos que van a conformar la Base Predictiva o Base de Muestra (contiene los registros con las cotizaciones de las distintas compañías de seguros de los últimos 28 días o llamado también mes móvil, descrita en el punto anterior). El resultado de esta extracción de datos, se muestra en la Figura 3.3.

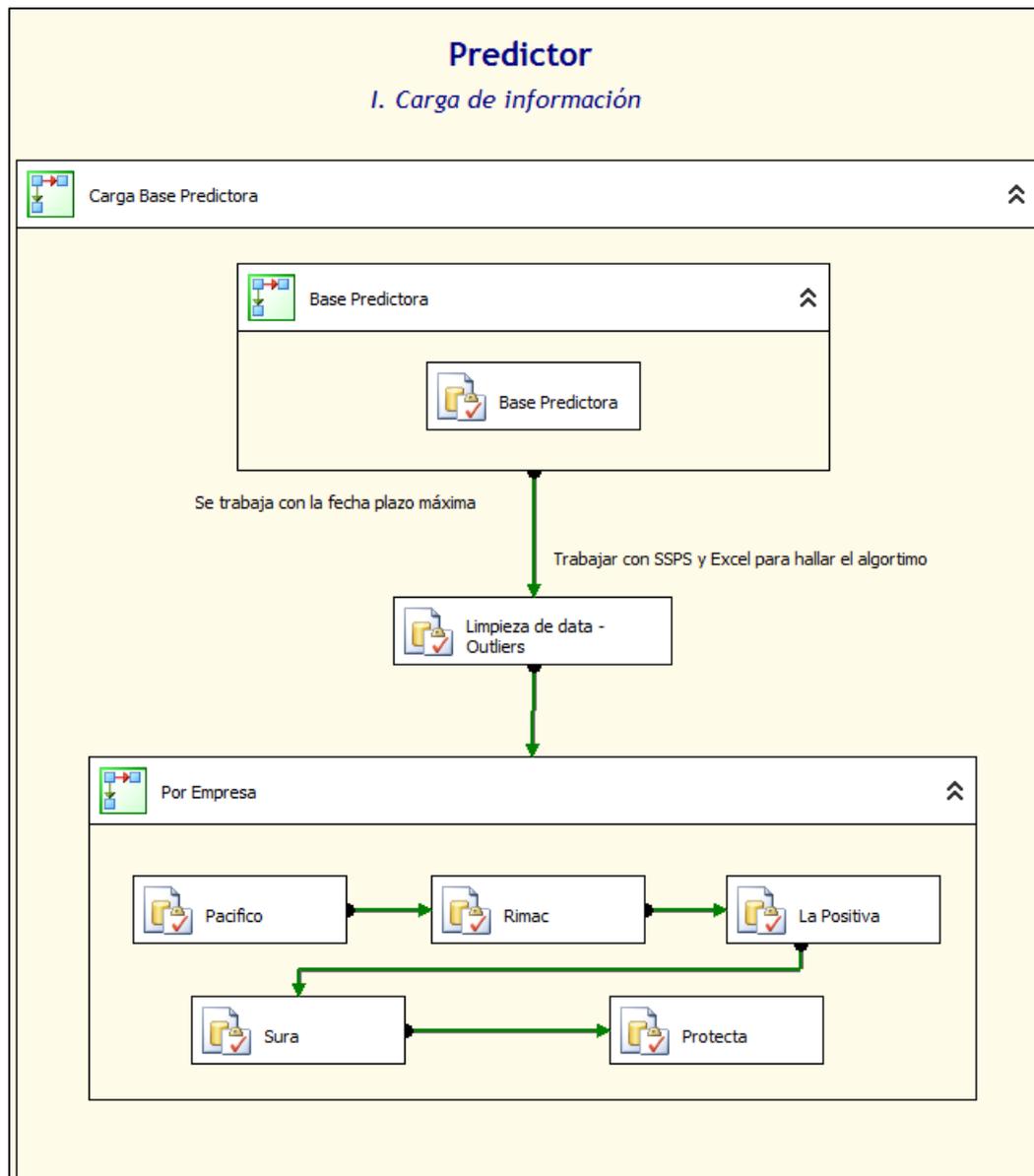


Figura 3.2: Proceso de generación de base de muestra.

Elaboración: los autores.

	num_operacion	num_cuspp	compañía	moneda	categoria	cod_modalidad	periodo_temporal	pje_pension	val_periodo_garantizado
1	107309	466901VTATU0	INTERSEGURO S.A.	S/.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
2	107309	466901VTATU0	LA POSITIVA VIDA	S/.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
3	107309	466901VTATU0	PACIFICO	S/.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
4	107309	466901VTATU0	SURA	S/.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
5	107310	528711RMNRR8	INTERSEGURO S.A.	S/.	INVALIDEZ	RTVD	2,00	50,00	10,00
6	107310	528711RMNRR8	LA POSITIVA VIDA	S/.	INVALIDEZ	RTVD	2,00	50,00	10,00
7	107310	528711RMNRR8	PACIFICO	S/.	INVALIDEZ	RTVD	2,00	50,00	10,00
8	107310	528711RMNRR8	SURA	S/.	INVALIDEZ	RTVD	2,00	50,00	10,00
9	107311	220271ETACE2	INTERSEGURO S.A.	S/.Aj.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
10	107311	220271ETACE2	LA POSITIVA VIDA	S/.Aj.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
11	107311	220271ETACE2	PACIFICO	S/.Aj.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
12	107311	220271ETACE2	PROTECTA	S/.Aj.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
13	107311	220271ETACE2	RIMAC	S/.Aj.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
14	107311	220271ETACE2	SURA	S/.Aj.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
15	107312	617901LMPHE1	INTERSEGURO S.A.	S/.Aj.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
16	107312	617901LMPHE1	LA POSITIVA VIDA	S/.Aj.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
17	107312	617901LMPHE1	PACIFICO	S/.Aj.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
18	107312	617901LMPHE1	PROTECTA	S/.Aj.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
19	107312	617901LMPHE1	RIMAC	S/.Aj.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
20	107312	617901LMPHE1	SURA	S/.Aj.	SOBREVIVENCIA	RTVD	2,00	50,00	15,00
21	107313	489791RBBIA6	INTERSEGURO S.A.	S/.Aj.	INVALIDEZ	RTVD	2,00	50,00	15,00
22	107313	489791RBBIA6	LA POSITIVA VIDA	S/.Aj.	INVALIDEZ	RTVD	2,00	50,00	15,00

Figura 3.3: Base muestra – Base predictiva.

Elaboración: los autores.

Luego de la extracción de la base de muestra, se procede a la limpieza de los datos, con el fin de evitar que la dispersión de los datos (que hacen 'ruido') altere la medida tendencial, en este caso, la media o promedio. Para ello, se utilizará la técnica denominada "Identificación de Outliers" (César Pérez López, 2009). La técnica consiste en identificar los outliers (valores atípicos), mediante el cálculo de cuartiles (que se hallan mediante el software IBM SPSS), como se muestra en la Figura 3.4 y Figura 3.5; y que nos ayudará a encontrar el RIC (Rango Intercuartílico). Una vez encontrado el RIC, se obtendrá el límite superior y el inferior de la muestra. Las fórmulas para el cálculo de lo mencionado anteriormente, son las siguientes:

- RIC:

$$IQR = Q_3 - Q_1$$

- Límites:

$$LI = Q_1 - 1,5 \cdot IQR$$
$$LS = Q_3 + 1,5 \cdot IQR$$

Donde,

Q_1 = Cuartil 1 (Percentil 25)

Q_3 = Cuartil 3 (Percentil 75)

IQR = Rango Intercuartílico (RIC)

LI = Límite Inferior

LS = Límite Superior

Hallados los límites necesarios para la consideración de la muestra (LI y LS), estos se ajustan a la base de muestra dependiendo de las variables utilizadas para el cálculo, se seleccionarán así a los registros que ingresaran en el árbol de decisión. La selección de los datos aptos para el árbol de clasificación, se obtiene de un algoritmo obtenido luego de emplear la técnica descrita anteriormente, tal como se muestra en la Figura 3.6.

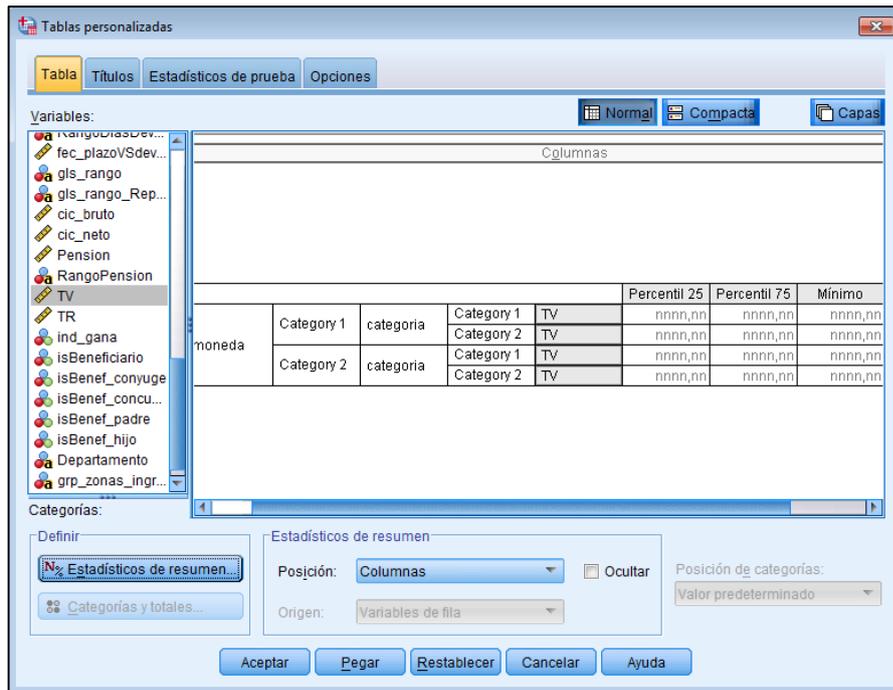


Figura 3.4: Proceso de generación de cuartiles en IBM SPSS.
Elaboración: los autores.

→ **Tablas personalizadas**

[Conjunto_de_datos1]

					Percentil 25	Percentil 75	Mínimo	Máximo
moneda	S/.	categoria	EDAD LEGAL	TV	2,70	3,26	,70	3,65
			INVALIDEZ	TV	2,10	3,13	,48	3,38
			JUB ANT ALTO RIESGO	TV	3,45	4,08	3,34	4,18
			REJA	TV	2,84	3,27	2,07	3,58
			SOBREVIVENCIA	TV	1,98	3,01	,44	3,35
	S/Aj.	categoria	EDAD LEGAL	TV	5,57	6,14	2,65	6,64
			INVALIDEZ	TV	5,47	5,89	3,21	6,49
			JUB ANT ALTO RIESGO	TV	6,26	6,76	5,79	9,43
			JUB ANT ORDINARIA	TV	5,68	6,11	3,42	6,42
			REJA	TV	5,87	6,17	4,78	6,69
			SOBREVIVENCIA	TV	5,37	5,83	3,00	6,27
US\$	categoria	SOBREVIVENCIA	TV	3,00	3,54	,76	4,33	
US\$Aj.	categoria	EDAD LEGAL	TV	4,09	4,75	1,88	5,15	
		INVALIDEZ	TV	3,97	4,35	3,85	4,56	
		REJA	TV	4,28	4,48	4,06	4,86	
		SOBREVIVENCIA	TV	4,12	4,35	4,06	4,46	

Figura 3.5: Resultado de la generación de cuartiles en IBM SPSS.
Elaboración: los autores.

Proceso de Limpieza de Muestra (Outliers)

					Percentil 25	Percentil 75	Mínimo	Máximo	RIC	LI	LS	Logica de Limpieza
moneda	S/.	categoria	EDAD LEGAL	TV	2.73	3.22	.70	3.65	0.50	1.98	3.65	When Moneda = 'S/.' And Categoria = 'EDAD LEGAL' And TV >= 1.97956091922293 And TV <= 3.65264119945652 Then 'OK'
			INVALIDEZ	TV	2.09	3.00	.44	3.37	0.91	0.73	3.37	When Moneda = 'S/.' And Categoria = 'INVALIDEZ' And TV >= 0.730642304555066 And TV <= 3.37031776422528 Then 'OK'
			JUB ANT ORDINARIA	TV	2.28	3.17	2.04	3.19	0.89	2.04	3.19	When Moneda = 'S/.' And Categoria = 'JUB ANT ORDINARIA' And TV >= 2.03770399297398 And TV <= 3.18765422648072 Then 'OK'
			REJA	TV	2.97	3.34	2.05	3.80	0.37	2.42	3.80	When Moneda = 'S/.' And Categoria = 'REJA' And TV >= 2.42293545287251 And TV <= 3.80480196707451 Then 'OK'
			SOBREVIVENCIA	TV	1.93	2.92	.47	3.35	0.99	0.47	3.35	When Moneda = 'S/.' And Categoria = 'SOBREVIVENCIA' And TV >= 0.470648706429969 And TV <= 3.35312214084006 Then 'OK'
	S/Aj.	categoria	EDAD LEGAL	TV	5.56	6.15	2.76	6.74	0.58	4.69	6.74	When Moneda = 'S/Aj.' And Categoria = 'EDAD LEGAL' And TV >= 4.69269099183263 And TV <= 6.73582141029884 Then 'OK'
			INVALIDEZ	TV	5.53	5.92	3.52	6.26	0.38	4.96	6.26	When Moneda = 'S/Aj.' And Categoria = 'INVALIDEZ' And TV >= 4.9598249516688 And TV <= 6.25640476748139 Then 'OK'
			JUB ANT ALTO RIESGO	TV	6.16	6.76	3.93	7.35	0.60	5.25	7.35	When Moneda = 'S/Aj.' And Categoria = 'JUB ANT ALTO RIESGO' And TV >= 5.25133410993833 And TV <= 7.35192256783514 Then 'OK'
			JUB ANT ORDINARIA	TV	5.78	6.12	3.50	6.42	0.33	5.29	6.42	When Moneda = 'S/Aj.' And Categoria = 'JUB ANT ORDINARIA' And TV >= 5.28531820110517 And TV <= 6.41724095390221 Then 'OK'
			REJA	TV	5.86	6.15	4.19	6.57	0.29	5.42	6.57	When Moneda = 'S/Aj.' And Categoria = 'REJA' And TV >= 5.41686947999181 And TV <= 6.5670675104962 Then 'OK'
	US\$	categoria	SOBREVIVENCIA	TV	5.36	5.83	1.29	6.27	0.48	4.64	6.27	When Moneda = 'US\$.' And Categoria = 'SOBREVIVENCIA' And TV >= 4.63927979113664 And TV <= 6.27432607741748 Then 'OK'
			SOBREVIVENCIA	TV	3.39	3.87	2.14	4.24	0.48	2.66	4.24	When Moneda = 'US\$.' And Categoria = 'SOBREVIVENCIA' And TV >= 2.66329494571286 And TV <= 4.23940705559405 Then 'OK'
	US\$Aj.	categoria	EDAD LEGAL	TV	4.03	4.69	1.88	4.98	0.66	3.04	4.98	When Moneda = 'US\$Aj.' And Categoria = 'EDAD LEGAL' And TV >= 3.04377899809576 And TV <= 4.98356827101769 Then 'OK'
			INVALIDEZ	TV	4.14	4.51	4.04	4.56	0.37	4.04	4.56	When Moneda = 'US\$Aj.' And Categoria = 'INVALIDEZ' And TV >= 4.03538724811092 And TV <= 4.56203500265111 Then 'OK'
			REJA	TV	4.37	4.53	4.01	4.88	0.16	4.13	4.77	When Moneda = 'US\$Aj.' And Categoria = 'REJA' And TV >= 4.12846464145267 And TV <= 4.76866434805125 Then 'OK'
			SOBREVIVENCIA	TV	4.12	4.35	4.06	4.46	0.23	4.06	4.46	When Moneda = 'US\$Aj.' And Categoria = 'SOBREVIVENCIA' And TV >= 4.06343115899998 And TV <= 4.45670500843353 Then 'OK'

Figura 3.6: Proceso de limpieza de muestra – Identificación de Outliers.

Elaboración: los autores.

3.1.4 Modelado

Luego de la limpieza de los datos, la base muestral se clasificará con la técnica de árboles de decisión. Se trabaja con 12 variables independientes de la base muestral, las que son: Moneda, categoría, zona, CIC, sexo, AFP, rango_CIC, % de periodo garantizado, modalidad, % de modalidad, departamento y zona geográfica.

Estas variables tendrán las siguientes características: serán validadas por el método “Validación Cruzada”, que consiste en seccionar la base a evaluar en partes iguales y calcular la media aritmética de cada una de partes; método de crecimiento CHAID exhaustivo, y seleccionar la división más significativa a medida que se va analizando la base de muestra (César Pérez López, 2007); y con una profundidad máxima de 15 niveles (cantidad de ramas en la cual el árbol puede crecer) Las características mencionadas se encuentran provistas por el software estadístico IBM SPSS 22.

Luego de estas especificaciones, y como se muestra en la Figura 3.7 y la Figura 3.8, el software SPSS evaluará las distintas variables y mostrará la media de la Tasa de Venta de aquellas que muestran una distribución considerable, además del algoritmo a considerar para la asignación de la tasa de venta por tipo de perfil como se muestra en la Figura 3.9:

Especificaciones	Método de crecimiento	EXHAUSTIVE CHAID	
	Variable dependiente	TV	
	Variables independientes	moneda, categoria, cod_modalidad, periodo_temporal, pje_pension, val_periodo_garantizado, afp, zona, sexo, tit_edad, gls_rango, cic_bruto, Departamento, grp_zonas_geografica	
	Validación	Validación cruzada	
	Máxima profundidad del árbol		15
Resultados	Casos mínimos en nodo padre		100
	Casos mínimos en nodo hijo		50
	Variables independientes incluidas	categoria, cic_bruto	
	Número de nodos		5
	Número de nodos terminales		3
	Profundidad		2

Figura 3.7: Resumen de método predictivo.

Elaboración: los autores.

Como se muestra en la Figura 3.7, el árbol de decisión, luego de su análisis con las características y variables independientes mencionadas anteriormente, dio como resultado que las variables más significativas son la Categoría o producto y el CIC de la persona que desea jubilarse. También nos muestra que el árbol de decisión tiene cinco nodos, de los cuales, tres de ellos son terminales; haciendo que tenga dos de profundidad o ramas.

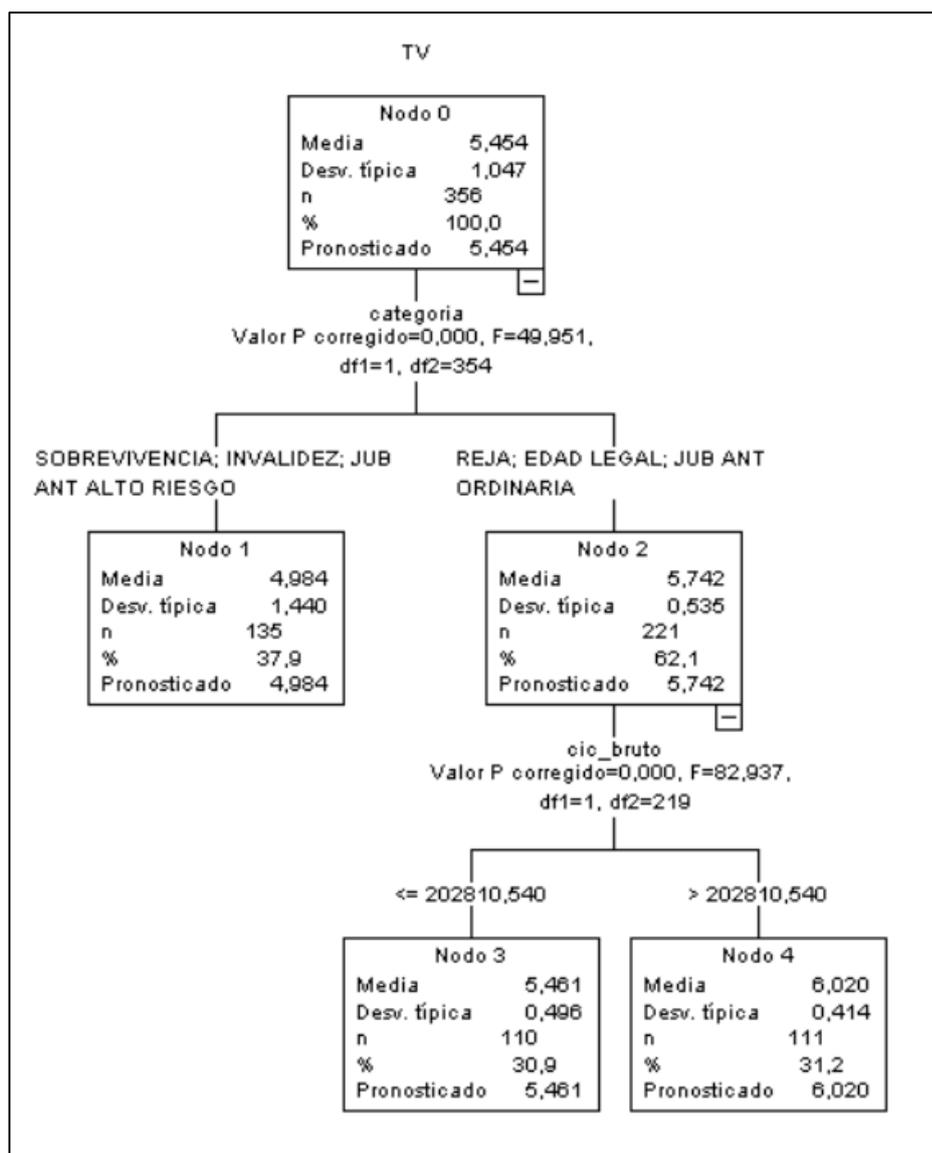


Figura 3.8: Representación de Árbol de Decisión.

Elaboración: los autores.

La Figura 3.7 descrita, anteriormente, es el resultado del árbol de decisión de manera numérica. El software IBM SPSS 22 también nos muestra el resultado de manera gráfica como se muestra en la Figura 3.8. Nos muestra información por nodo de evaluación y la probabilidad que se obtuvo en dicha evaluación. Cabe indicar que, el algoritmo resultante del árbol de decisión tiene lectura descendente que va desde el Nodo 0 hasta el Nodo terminal, según corresponda.

Por ejemplo: Para la base evaluada, las personas que hayan optado por una categoría de Edad Legal y tengan un CIC mayor a

S/.202,810.540 brutos, tendrán una Tasa de Venta preliminar de 6.02%, por el contrario, aquellos clientes que hayan optado por una categoría de Edad Legal y tengan un CIC menor a S/.202,810.540 brutos, tendrán una Tasa de Venta Preliminar de 5.46%.

Como se ha mencionado anteriormente, el proceso genera un algoritmo de clasificación y este puede ser generado en lenguaje SQL, que es necesario para la aplicación de nuestro proyecto. El script que genera el software estadístico como se muestra en la Figura 3.9, se actualiza semanalmente en el procedimiento almacenado o “stored procedure” utilizado para predicción de las diferentes cotizaciones ingresadas. Antes de ello, y por políticas de la empresa aseguradora, el procedimiento verifica automáticamente que la tasa a pronosticar tenga un 90% de confianza como mínimo para los diferentes perfiles. Si la tasa pronosticada cumple con esta validación, el perfil correspondiente será considerado en el proceso de predicción, de lo contrario, el proceso continuará con el flujo normal sin la tasa pronosticada.

```

/* PACIFICO */

UPDATE [RV].dbo.RV_Predictor_BasePredictiva_Temporal_Pacifico
    SET TV_Arbol = 4.98404545628405
WHERE (categoria = 'SOBREVIENCIA' OR categoria = 'INVALIDEZ' OR categoria = 'JUB ANT ALTO RIESGO');

UPDATE [RV].dbo.RV_Predictor_BasePredictiva_Temporal_Pacifico
    SET TV_Arbol = 5.46062477546798
WHERE ((categoria IS NULL) OR categoria <> 'SOBREVIENCIA' AND categoria <> 'INVALIDEZ' AND categoria <> 'JUB ANT ALTO RIESGO');

UPDATE [RV].dbo.RV_Predictor_BasePredictiva_Temporal_Pacifico
    SET TV_Arbol = 6.02009239461832
WHERE ((categoria IS NULL) OR categoria <> 'SOBREVIENCIA' AND categoria <> 'INVALIDEZ' AND categoria <> 'JUB ANT ALTO RIESGO');

|
Select Cast(Moneda As Varchar(15)) Moneda,
Cast(Categoria As Varchar(25)) Categoria,
Cast(Zona As Varchar(1)) Zona,
Cast(gls_Rango_Reporte As Varchar(15)) gls_Rango,
Cast('PACIFICO' As Varchar(25)) Compañia
Into [MercadoRV].[dbo].[RV_Predictor_NivelConfianza]
From [RV].dbo.RV_Predictor_BasePredictiva_Temporal_Pacifico
Group by Moneda, Categoria, Zona, gls_Rango_Reporte
Having Case When Abs(Avg(TV_Arbol)-Avg(TV)) <= 0.10 Then 'VA' Else 'NO VA' End = 'VA'

```

Figura 3.9: Algoritmo predictivo y validación de perfiles.

Elaboración: los autores.

3.1.5 Evaluación

En esta etapa, se realizan las primeras pruebas del método predictivo desarrollado. En coordinación con la gerencia de Rentas Vitalicias, se acordó seleccionar casos potenciales y reales diariamente, en un formato dado, y de los cuales aquellos casos que cumplan con las condiciones de evaluación mencionadas anteriormente, con excepción de aquellos que sean Cartera Blanca (casos con tratamiento especial por ser de personas con CIC muy elevados), tal como se muestra en la Figura 3.10.

1	Nro Solicitud	Compañía	Moneda	Categoría	CIC Bruto	AFP	ACOM	DCOM	TRA	Tasa C/cam	Tasa S / Cambio	Jefe	Columna
2	108584_01	SURA	S/.	REJA	1683985.93	PRIMA	0	75	1.5	3.18	3.07	magali	
3	108576_03	PACIFICO	S/Aj.	EDAD LEGAL	S/ 286,601.00	PRIMA	1.2	0	-1.5	5.74	5.91	cesar	
4	108550_01	LA POSITIVA VIDA	S/Aj.	REJA	S/ 160,778.52	INTEGRA	50	1	1.5	5.72	5.77	cesar	CARTERA BLANCA
5	108636_01	PACIFICO	S/Aj.	EDAD LEGAL	S/ 674,980.00	INTEGRA	0	100	-2	5.99	5.82	RECOTIZACION	CARTERA BLANCA
6													

Figura 3.10: Registros reales para prueba.

Elaboración: los autores.

Para agilizar el proceso de evaluación del método predictivo, se consideró crear una herramienta en Microsoft Access llamada "Predictor", y aprovechar las facilidades de su integración con SQL Server y con Microsoft Excel, en donde, con un botón, se conecta con la base de datos y envía la información ingresada en la herramienta en Microsoft Access, la que será procesada mediante procedimientos almacenados llamados: `dbo.sp_RV_PredictorArbol` (contiene el algoritmo de árbol de decisión) y `dbo.sp_RV_LogicaPredictor` (contiene las reglas de negocio); luego de ejecutarse, se genera la Tasa de Venta pronosticada (TV Predictor) para aquellos casos seleccionados como se muestra en la Figura 3.11:

Figura 3.11: Aplicación para utilización de la predicción en los casos de prueba.
Elaboración: los autores.

3.2 Desarrollo del Data Mart

Adicionalmente, al método predictivo, esta tesis contempla crear un Data Mart, que tendrá como función principal, la integración de la información proveniente de los sistemas transaccionales, y que será utilizada para el método predictivo. Además de ello, servirá para alimentar de información el tablero de control con los resultados de las tasas predichas (Predictor).

Objetivo del Data Mart:

“Proporcionar la fuente de datos integrados y estructurados, para la ejecución del método de predicción de la tasa de venta en Rentas Vitalicias.”

Como se ha dicho anteriormente, la metodología elegida para la implementación del Data Mart es Kimball, que tiene las siguientes fases:

3.2.1 Fase de Planificación

En esta etapa, se describe el objetivo principal para el desarrollo del Data Mart, de acuerdo en las necesidades observadas en el área de Inteligencia Comercial.

Las fuentes de datos serán identificadas según las tablas que se utilizan en las consultas y requerimientos necesarios para la predicción de la tasa de venta en las cotizaciones de Rentas Vitalicias. Se identificarán las bases de datos, luego las tablas consultadas y finalmente los campos necesarios en considerar.

Con la identificación de las tablas fuentes se obtendrá un diseño lógico en el que tendremos relacionadas tanto las tablas de hechos (valores de las medidas del negocio, indicadores) como las tablas de dimensiones (parámetros de los que dependen otros datos).

El alcance del Data Mart se rige desde la consolidación, transformación y limpieza de la información proveniente de las bases de datos transaccionales hasta la integración de los datos, de manera que sirvan de utilidad para la predicción de la tasa de venta en el proceso de cotizaciones.

3.2.2 Fase de Análisis

En esta etapa, se realiza la captura de los requerimientos que debe cubrir el Data Mart. Para ello, se realizó una entrevista al Coordinador de Inteligencia Comercial, quien describió el siguiente proceso.

Proceso de Generación de Cotización de Rentas Vitalicias:

Dentro de las actividades realizadas en la venta de Rentas Vitalicias, se encuentra la generación de cotizaciones, esto de acuerdo con las solicitudes de jubilación recibidas de manera diaria. Es, en este proceso donde se hace uso de la tasa de venta, que sirve para aumentar de manera porcentual, el CIC de la persona que desea jubilarse y la cual es calculada de manera mecánica sin tener un análisis detrás. Esta varía en el tiempo, acorde con las ofrecidas por las empresas competidoras o depender del perfil de la persona.

Actualmente, el cálculo de la tasa de venta se hace de manera mecánica, como se ha dicho anteriormente.

Luego de la entrevista con el coordinador de Inteligencia Comercial, se definieron, además de la función principal de este Data Mart, los siguientes requerimientos funcionales (Reportes) que debe cumplir, para el proceso de predicción de la tasa de ventas:

- Cantidad prospectos creados (Cotizaciones)
- Porcentaje de cotizaciones con tasa de ventas pronosticada, cotizaciones sin tasa de ventas pronosticada
- Cantidad de cierres por tiempo (semanalmente)
- Cantidad de cierres ganados por la empresa
- Monto total de CIC neto ganados por la empresa
- Cantidad (%) ganada de “*Market Share*”

3.2.2.1 Análisis dimensional

Luego de haber identificado los requerimientos funcionales, solicitados por los usuarios finales, se definieron las dimensiones dentro del modelo lógico del Data Mart. En esta etapa, se desarrollará el documento “Análisis Dimensional” (Ver Anexo 3: Análisis Dimensional del Data Mart)

Las dimensiones que conforman el Modelo Lógico para el Data Mart Comercial son las siguientes, tal como se muestra en la Tabla 3.1:

Tabla 3.1: Dimensiones del Data Mart.

N°	Dimensión
1	D_Ubigeo
2	D_Compañía
3	D_Tiempo
4	D_Predictor
5	D_Cierres
6	D_CierreCotizacion
7	D_CotizacionesPredictor
8	D_Persona

Elaboración: los autores.

Luego del análisis de dimensiones, se identificó la tabla de hechos, que cumple los requerimientos funcionales como se muestra en la Tabla 3.2:

Tabla 3.2: Tablas de Hechos (facts) del Data Mart.

N°	Tema
1	F_ResultadoPredictor

Elaboración: los autores.

En la Figura 3.12, se muestra el modelo lógico resultante del análisis descrito anteriormente.

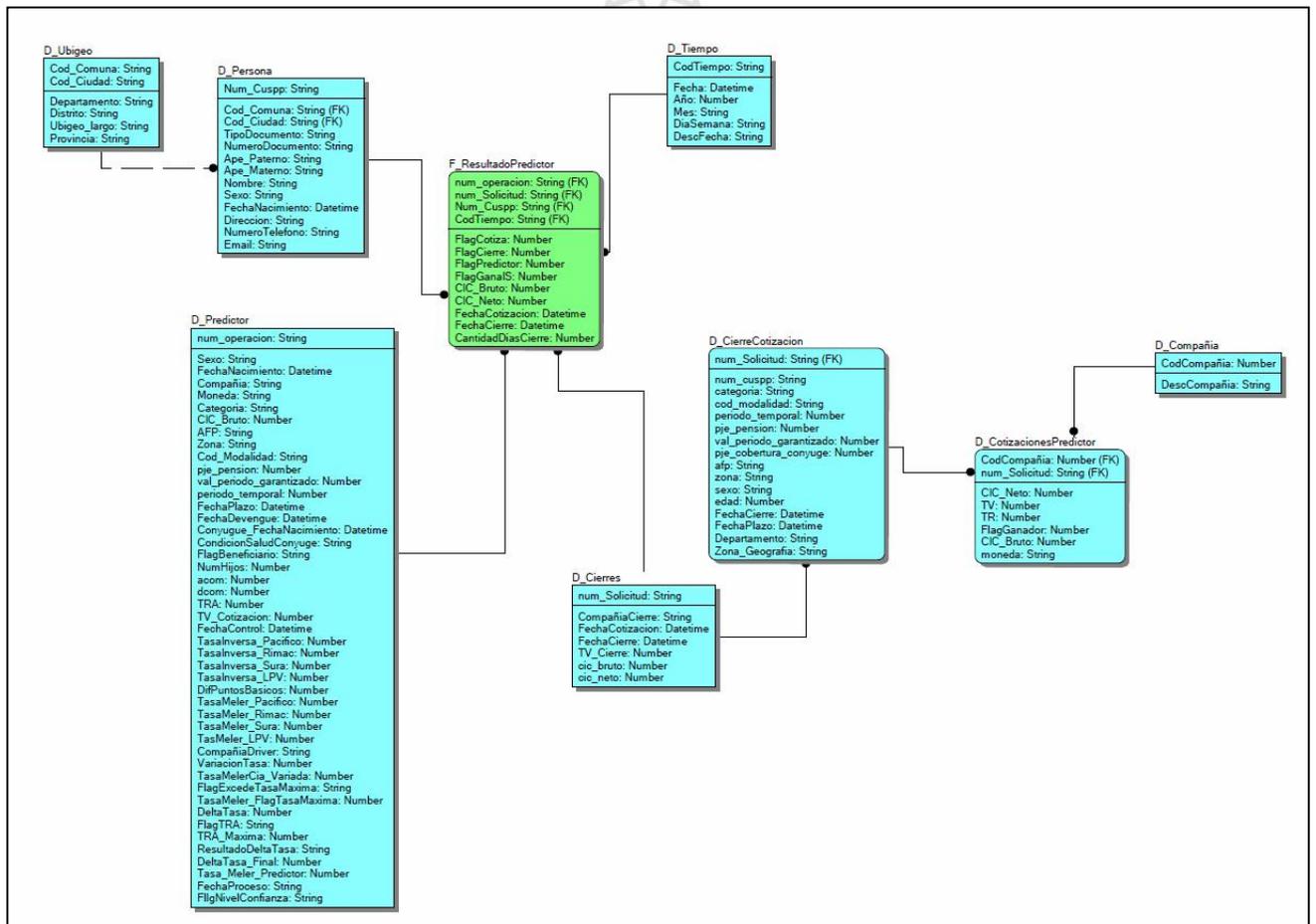


Figura 3.12: Modelo Lógico del Data Mart.

Elaboración: los autores.

3.2.3 Fase de Diseño

En esta etapa de la metodología, se mostrará la arquitectura planteada para la solución, así como el proceso de explotación de las DB seleccionadas a base de los requerimientos identificados en la etapa anterior.

3.2.3.1. Arquitectura de la solución

Con el fin de entender los componentes que se interrelacionarán en la arquitectura del Data Mart, a continuación se mencionarán los más importantes:

- Componente Fuentes de Información: Está formado por la base de datos que contiene la información transaccional del área Comercial de la empresa aseguradora.
- Componente ETL: Implementa los procesos necesarios para extraer la información de los datos fuentes, transformándolos y cargándolos en la base de datos de Data Mart.
- Componente Data Mart: Representa a la base de datos multidimensional.

En la Figura 3.13, se muestra la arquitectura del Data Mart, que utilizará SQL Server 2008 como motor de Base de Datos. Asimismo, en la tabla 3.3 se muestra las bases de datos que servirán como fuentes de información para poblar la estructura del Data Mart.

Tabla 3.3: Bases de datos transaccionales del proceso de Rentas Vitalicias

N°	Nombre BD	Descripción
1	RV	Contiene información del área de Rentas Vitalicias, tales como los clientes y/o prospectos a jubilarse.
2	RVIAdmin	Contiene información del área de Rentas Vitalicias, tales como los clientes y/o prospectos a jubilarse.
3	MercadoRV	Contiene información del mercado de jubilaciones en sus distintos tipos.
4	LasLomas	Contiene información de la zonificación y clasificación de las provincias o lugares geográficos.
5	Exp_Serv	Contiene información de los sistemas transaccionales de la empresa.
6	Cotizaciones	Contiene información de las cotizaciones generadas por la empresa en todos sus productos aseguradores.
7	Contact	Contiene información de las personas que son clientes de la empresa además de sus datos de contactabilidad.
8	Comercial	Contiene información generada por Inteligencia Comerciales en cuando análisis o prospectos generados para los diferentes productos de la empresa.

Elaboración: los autores.

Para realizar la carga de los datos, se utilizará la herramienta SQL Server Integration Services. El esquema del Data Mart propuesto será copo de nieve debido a que la base muestra a utilizar debe eliminar la redundancia de los datos y ser normalizada para facilitar su explotación.

La explotación de la información se realizará mediante la lógica del Predictor, que permitirán la explotación de la información y así cumplirá con los requerimientos propuestos y facilitará la comprensión de la información obtenida para los usuarios de la herramienta.

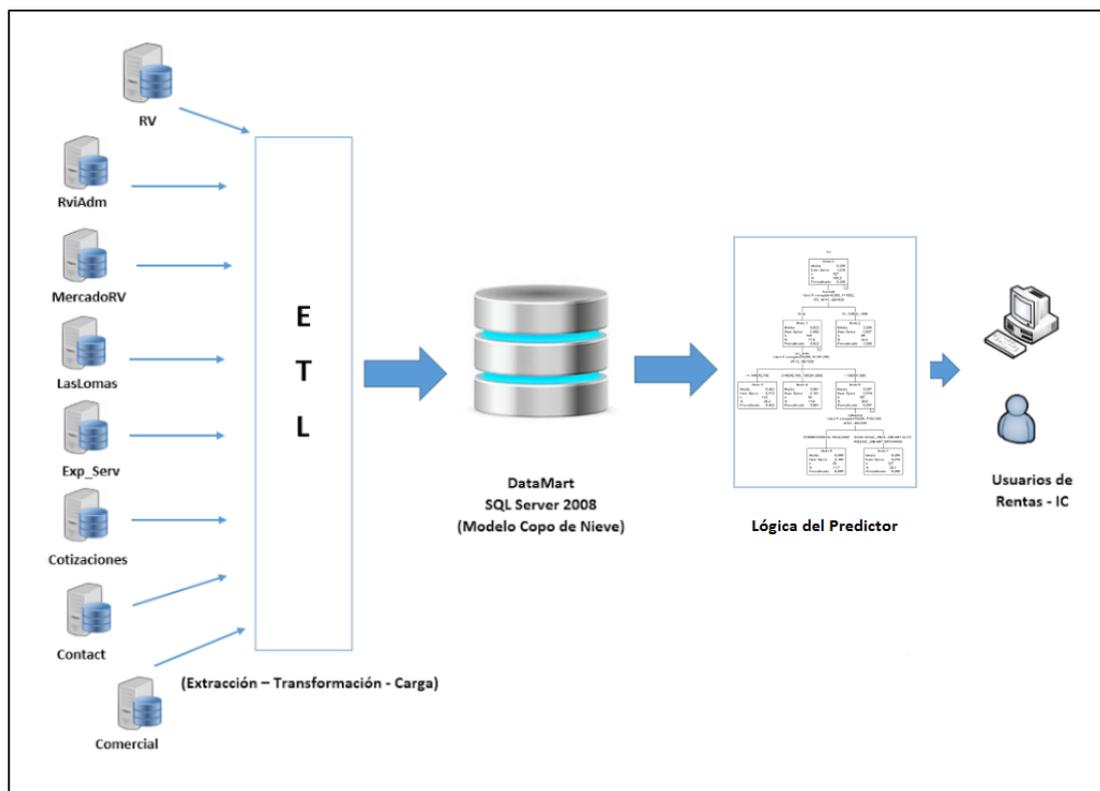


Figura 3.13: Arquitectura del Data Mart.

Elaboración: los autores.

Para mayor detalle del diseño y creación del Data Mart, ver Anexo 4: Diseño del Data Mart.

3.2.4 Fase de implementación

Una de las últimas fases de la metodología, es la fase de Implementación del Data Mart. A continuación, se mostrarán los criterios y aspectos tomados en cuenta en la etapa de la implementación.

3.2.4.1. Construcción

La fase de Implementación del Data Mart está definida principalmente por la implementación de los procesos ETL, debido a que los modelos de bases de datos fuentes y del Data Mart ya están definidos en los capítulos anteriores.

3.2.4.2. Proceso ETL

El proceso de Extracción, Transformación y Carga de la información, consiste en la ejecución de sentencias SQL para conseguir la información en la estructura que tiene contemplada el Data Mart.

Para esta tesis, se utilizará la herramienta que brinda Visual Studio 2008, en su versión Integration Services, la cual nos brinda la posibilidad de crear DTS, en su función de ETL para generar la información antes mencionada y así, poder alimentar y/o poblar el Data Mart con la información correspondiente.

Para mayor detalle de los procesos de ETL y construcción del Data Mart, ver Anexo 5: Implementación del Data Mart.

3.3 Factores de riesgo

Para detallar los factores que afectan el proyecto de tesis se realizará un análisis en los objetivos propuestos, así como el cumplimiento de los mismos. Durante la realización del proyecto, se encontraron los siguientes:

3.3.1. Utilización de información desactualizada

En el marco de la construcción del Data Mart, se tiene como principio fundamental ubicar el proceso de negocio crítico, donde se llevará a cabo la explotación de la información. Para ello es clave tener actualizados los datos con los que se ejecutará el proceso, así como las fuentes de datos a utilizar.

Ante esta situación, se diseñó el proceso de limpieza y carga de datos, programado para recolectar la información de las fuentes de datos seleccionadas para poblar el Data Mart, actualizando su contenido en cada ejecución del proceso.

3.3.2. Seguridad de la Información

Este factor es crítico debido a que la información con la que se trabaja es sensible, y del que depende la efectividad del cálculo predictivo a realizarse para determinar la tasa de venta más óptima del mercado.

Ante esta situación, se utilizó una interfaz diseñada en Microsoft Access para facilitar la interacción de la asistente de Rentas Vitalicias con el método predictivo, asegurando la integridad y efectividad del método desarrollado.

3.3.3. Efectividad del nuevo método utilizado

El uso del método predictivo desarrollado en la presente tesis debe ser comparado con la anterior forma de trabajo en la empresa aseguradora, evaluando el impacto sobre el negocio y validando su efectividad.

Para facilitar la interpretación de los resultados obtenidos, se desarrolló un cuadro de resultados, donde se muestran las comparativas de la nueva forma de cálculo de la tasa de venta con el método propuesto así como el rendimiento obtenido.



CAPÍTULO IV

PRUEBAS Y RESULTADOS

Como se ha dicho en puntos anteriores, las pruebas se han realizado de forma diaria con muestra enviada por la gerencia de Rentas Vitalicias.

Según lo coordinado entre las áreas participantes en el proceso (área de Inteligencia Comercial – gerencia de Rentas Vitalicias), se tomaría como prueba diaria los casos seleccionados y enviados por la Asistente de Rentas Vitalicias, asimismo, Inteligencia Comercial devolvería aquellos casos a los cuales se les haya generado tasa de venta pronosticada y que, a su vez, cumplan con el nivel de confianza del 90% en la evaluación del árbol de clasificación.

4.1 Pruebas

En la presente sección, se describirá el procedimiento que se ha seguido durante cuatro semanas para las pruebas del método de predicción de la tasa de ventas para rentas vitalicias.

Se debe tener en cuenta que el proceso consta de dos fases, la fase de generación del algoritmo por el árbol de clasificación (Semanalmente) y las pruebas para casos reales (Diariamente).

- Generación del algoritmo (árbol de decisión)

Cada semana, específicamente, los días jueves, se generó los algoritmos resultantes de la evaluación del árbol de decisión. El proceso consistió en:

- En una primera instancia, se ejecuta el DTS que contiene el Data Mart; con la finalidad de actualizar las dimensiones que contienen la información de los cierres y así, obtener el mes móvil para el cálculo del algoritmo.
 - Una vez ejecutado el Data Mart, y por ende, con las dimensiones actualizadas, se obtiene la base muestral (o base móvil), la cual es importada en el software SPSS para el cálculo de los valores atípicos (con la finalidad de ser extraídos para no distorsionar el resultado) Luego de ello, se separa la “base limpia” por compañía competidora, a las cuales se les generará un árbol de decisión para cada una de ellas. Al finalizar la evaluación, el software nos proveerá de un algoritmo en lenguaje SQL, el cual se reemplazará semanalmente, en los procedimientos almacenados en el motor de base de datos SQL Server 2008.
- Pruebas diarias (Casos reales)

Como se mencionó en la introducción del presente capítulo, la asistente de rentas vitalicias enviaría diariamente una selección de solicitudes para ser procesadas y obtener su tasa de ventas pronosticada (tasa predictor) Para ello, se siguen las siguientes tareas:

- La gerencia de Rentas vitalicias, en la persona de su asistente, envía diariamente una selección de casos reales, cumpliendo con las indicaciones de Inteligencia comercial y el formato establecido.
- Inteligencia comercial recibe el formato con la información diaria y procede a ingresarlos a la base de datos mediante la herramienta “Predictor”, desarrollado en Microsoft Access. Seguidamente, procede a generar la tasa predictor.
 - El proceso comienza con el procedimiento almacenado llamado `dbo.sp_RV_PredictorArbol`, que contiene el algoritmo generado, semanalmente, descrito en el punto anterior. A su vez, el proceso ejecuta el procedimiento almacenado `dbo.sp_RV_LogicaPredictor`, con las instrucciones codificadas en lenguaje SQL y que tiene como función el ajuste de la tasa

pronosticada con las reglas de negocio. El proceso solo generará la tasa predictor a aquellos perfiles a los cuales, en la evaluación semana de la base móvil, obtuvieron un 90% de confianza entre la tasa pronosticada por el árbol de decisión y la tasa real de cierre de la base.

- Luego del proceso y la asignación de la tasa predictor para las solicitudes que cumplieron con las validaciones y reglas de negocio, Inteligencia comercial regresa el formato enviado con la información adicional generada a la gerencia de Rentas Vitalicias.

Para un mayor entendimiento de los procesos mencionados anteriormente, se muestra la siguiente Figura 4.1

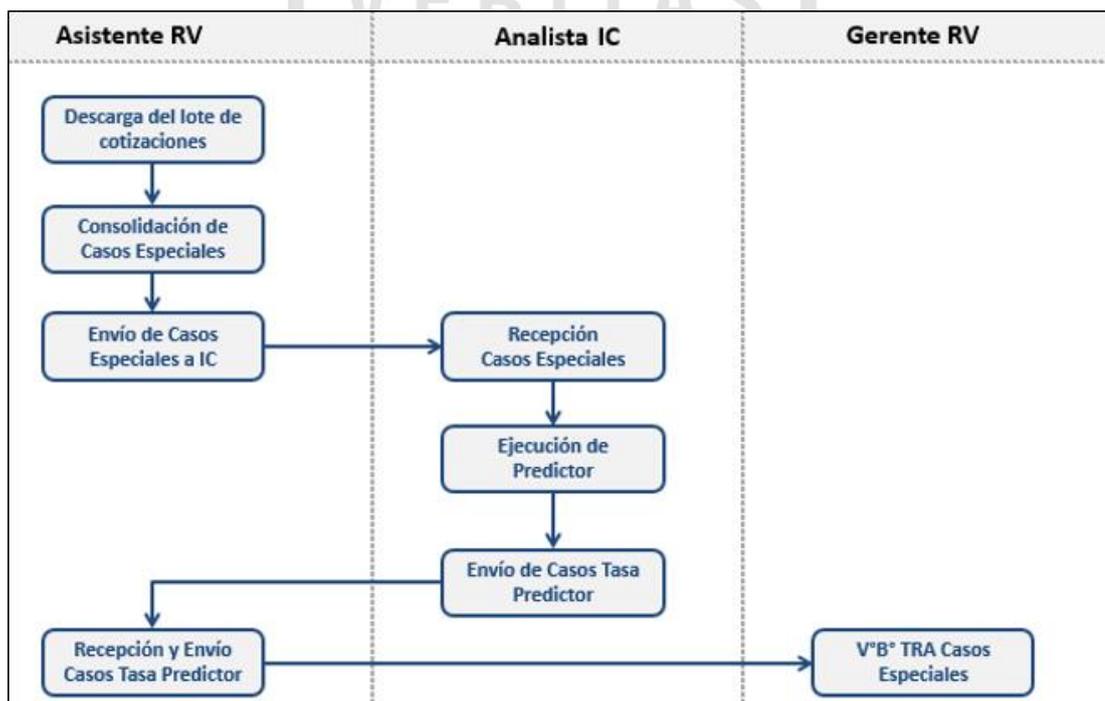


Figura 4.1: Proceso de pruebas Predictor.

Elaboración: los autores.

4.2 Resultados

Las pruebas se realizaron diariamente por cuatro semanas, tal como se muestra en la tabla 4.1:

Tabla 4.1: Cantidad de solicitudes enviadas a pruebas.

Semana	Predictor	No Predictor	Total general
3. Sem. 04/05 - 08/05	12	25	37
4. Sem. 11/05 - 15/05	13	25	38
5. Sem. 18/05 - 22/05	20	40	60
6. Sem. 25/05 - 29/05	30	41	71
Total general	75	131	206

Elaboración: los autores.

Como se comentó anteriormente, solo las solicitudes que cumplan con las validaciones y reglas negocio, tendrían generada la tasa predictor para su solicitud. En este caso, vemos que en las cuatro semanas de prueba, y de un total de 206 solicitudes enviadas, el 36% de las solicitudes han recibido la tasa predictor.

Cabe indicar que para que una cotización logre cerrarse con alguna compañía de seguros, tiene un plazo máximo de tres días a seis días luego del envío de la cotización. En este escenario, se esperó una semana luego del envío de las cotizaciones para ver los resultados de las pruebas la cual obtuvo un resultado satisfactorio para las pruebas enviadas ya que, como se puede ver en la Tabla 4.2, a pesar de que la cantidad de casos enviados con tasa predictor fue en menor cantidad, estos obtuvieron un mejor *Market Share* en las cuatro semanas de prueba, superando inclusive, los casos que no fueron considerados en el proceso de generación de la tasa predictor (no enviados por la gerencia de rentas vitalicias).

Tabla 4.2: Resultados de cierres – Rentas Vitalicias.

Fecha de Envío	Predictor							
	Tasa Predictor (Nivel Confianza $\geq 90\%$)				No Tasa Predictor (Nivel Confianza $< 90\%$)			
	Cotizacion	Cierre	Ganados		Cotizacion	Cierre	Ganados	
Cantidad	Cantidad	Cantidad	Market Share (S/.)	Cantidad	Cantidad	Cantidad	Market Share (S/.)	
3. Sem. 04/05 - 08/05	12	7	5	72.52%	25	17	7	67.07%
4. Sem. 11/05 - 15/05	13	8	3	24.62%	25	22	8	27.41%
5. Sem. 18/05 - 22/05	20	16	7	43.49%	40	25	8	23.91%
6. Sem. 25/05 - 29/05	30	9	6	77.14%	41	17	8	56.43%

Elaboración: los autores.



Como se muestra en el Tabla 4.2, se obtuvo más y mejores ventas en las solicitudes enviadas con la tasa predictor que las solicitudes que tuvieron un flujo normal.

Adicional a los casos o cotizaciones ganadas por la empresa aseguradora, también es importante tomar en cuenta los resultados de las variaciones, tanto en el TRA (Tasa de Rentabilidad Ajustada) que es utilizada para el ejercicio de venta, es decir, este indicador nos indica cuanto (en puntos básicos) puede subir la oferta. Para ello, tenemos como resultado lo que se muestra en la Tabla 4.3:

Tabla 4.3: Variación de TRA.

	TRA (Casos con Tasa Predictor)					
	Cant. x Variación TRA			Δ Prom. Variación TRA		
	Aumentó	Disminuyó	Se Mantuvo	Aumentó	Disminuyó	Se Mantuvo
3. Sem. 04/05 - 08/05	2	4	6	0.35	-0.95	-
4. Sem. 11/05 - 15/05	0	10	3	-	-0.52	-
5. Sem. 18/05 - 22/05	6	6	8	0.52	-0.73	-
6. Sem. 25/05 - 29/05	5	16	9	0.40	-0.89	-

Elaboración: los autores.

En la Tabla 4.3, podemos notar que, dado que el TRA es un indicador para subir la oferta de venta, se obtuvo disminución de este último debido al ajuste que el proceso realiza bajo las reglas de negocio establecidas, es decir, se consiguió un mejor ajuste en cesión de puntos básicos de oferta, pero se obtuvo una tasa de venta más eficiente, avalada por los resultados del Tabla 4.2.

Adicional a la evaluación del TRA, es importante saber cómo se comportó la variación de las tasas pronosticadas con respecto a las tasas emitidas por el Sistema Cotizador. En la Tabla 4.4, podemos notar que las tasas de venta pronosticadas, en su totalidad, ofrecieron una mejor oferta que las tasas de venta calculadas por el Sistema Cotizador.

Tabla 4.4: Variación de Tasa de Venta.

	Tasa Predictor					
	Cant. x Variación TV			Δ Prom. Variación TV (.pb)		
	Aumentó	Disminuyó	Se Mantuvo	Aumentó	Disminuyó	Se Mantuvo
3. Sem. 04/05 - 08/05	11	0	1	0.15	-	-
4. Sem. 11/05 - 15/05	13	0	0	0.15	-	-
5. Sem. 18/05 - 22/05	20	0	0	0.19	-	-
6. Sem. 25/05 - 29/05	29	1	0	0.16	-0.15	-

Elaboración: los autores.

Es necesario complementar que, si bien es cierto, las tasas de ventas pronosticadas aumentaron, el TRA disminuyó; esto hace notar que el algoritmo con el ajuste de reglas de negocio, hacen un impacta directamente en la generación de la tasa de ventas.



CAPÍTULO V DISCUSIÓN Y APLICACIÓN

5.1 Discusión

Se presenta el resultado de la evaluación del Predictor en la gerencia de Rentas Vitalicias de la empresa.

5.1.1 Situación sin predictor

Para realizar una cotización, los agentes y asistentes de la gerencia Rentas Vitalicias hacen uso del sistema “Cotizador”, donde se ingresan los datos de la solicitud a cotizar en el sistema y se genera la tasa de venta a ofrecer, producto del cálculo de las variables ingresadas como son la categoría, moneda, AFP, zona y CIC bruto del afiliado.

Categoría	Moneda	AFP	Zona	CIC	TV Cotizador
Edad Legal	S/. Aj.	Prima	Lima	250,000	5.90
Edad Legal	S/. Aj.	Integra	Lima	500,000	6.10

Figura 5.1: Cotización actual.

Elaboración: los autores.

En la figura 5.1, se observa la forma de cálculo actual obtenida por el Sistema cotizador, donde el cálculo solo se basa en los valores ingresados actuales. La veracidad y valor competitivo de esta tasa no puede ser medido

con la ofrecida por el mercado siendo, en muchos casos, mayor a lo que pudo ofrecer la competencia que generó, de este modo, un mayor desembolso de dinero para la empresa ya que se necesita un mayor capital para rentabilizar este nuevo ingreso.

5.1.2 Situación con predictor

Para el cálculo de la tasa bajo el análisis de medida tendencial, se propone el método predictivo “Predictor” que, a diferencia del Sistema Cotizador actual, realiza un análisis por compañía y perfil, calculando la tasa ofrecida por compañía en el transcurso del mes móvil (base de muestra con las cotizaciones en un lapso de un mes) como se muestra en la figura 5.2.

Categoría	Moneda	AFP	Zona	CIC	TV Cotizador	TV pred RIMAC	TV pred PACIFICO	TV pred SURA	TV pred POSITIVA
Edad Legal	S/. Aj.	Prima	Lima	250,000	5.90	6.00	6.03	5.98	6.10
Edad Legal	S/. Aj.	Integra	Lima	500,000	6.10	6.03	6.05	6.08	6.12

Figura 5.2: Calculo de la tasa de venta ofrecida por compañía.

Elaboración: los autores.

Este proceso genera las tasas ofrecidas en el mes móvil y de acuerdo con ello se selecciona la compañía competidora “driver” (competencia directa, por perfil, para la empresa aseguradora) para los perfiles a evaluar. Lo siguiente es aplicar una variación en puntos básicos por regla de negocio a la tasa seleccionada de acuerdo con el perfil, para este ejemplo se debe aplicar +5 puntos básicos como se muestra en la figura 5.3, por ser la compañía RÍMAC la competidora directa para este perfil.

Categoría	Moneda	AFP	Zona	CIC	TV Cotizador	TV pred RIMAC	Regla de Negocio	TV' pred RIMAC
Edad Legal	S/. Aj.	Prima	Lima	250,000	5.90	6.00	+0.05	→ 6.05
Edad Legal	S/. Aj.	Integra	Lima	500,000	6.10	6.02	+0.05	→ 6.07

Figura 5.3: Selección de la compañía driver.

Elaboración: los autores.

Una vez que se tiene la tasa predicha se realiza el ajuste de la tasa conocido como TRA (Tasa de Rentabilidad Ajustada) por regla de negocio. Para este ejemplo, se encontró que el perfil estaba sujeto a un TRA de 1.50 que permite una variación máxima de 13 puntos básicos respecto de lo ofrecido en primera instancia por el Sistema Cotizador.

Categoría	Moneda	AFP	Zona	CIC	TV Cotizador	TV' pred RIMAC	TRA Máximo	Δ (TV' pred - TV Cot)	TV'' pred RIMAC
Edad Legal	S/. Aj.	Prima	Lima	250,000	5.90	6.05	1.50	0.13	6.03
Edad Legal	S/. Aj.	Integra	Lima	500,000	6.10	6.07	--	→ --	6.07

Figura 5.4: Aplicación de la Tasa de Rentabilidad Ajustada.

Elaboración: los autores.

Como se muestra en la Figura 5.4, existe una variación entre la tasa predicha y la tasa generada por el Sistema Cotizador de 15 puntos básicos, valor que sobrepasa el ajuste del TRA de 13 puntos básicos como máximo; por regla de negocio, si el ajuste del TRA es superado se toma la tasa generada por el Sistema Cotizador más los puntos de ajuste del TRA, y así generar, de esta forma, la tasa final a cotizar como se muestra en la figura 5.5.

Categoría	Moneda	AFP	Zona	CIC	TV final
Edad Legal	S/. Aj.	Prima	Lima	250,000	6.03
Edad Legal	S/. Aj.	Integra	Lima	500,000	6.07

Figura 5.5: Tasa de venta final a cotizar.

Elaboración: los autores.

Las tasas de venta final, en la Figura 5.5, muestran una variación a las primeras tasas generadas por el Sistema Cotizador, la interpretación para estas diferencias es la siguiente:

- En el primer perfil, el cotizador indicó una tasa de 5.90, luego de realizar la evaluación tendencial del mercado con el método predictivo se llegó a una tasa de 6.03, por tanto, se determina que la empresa debe ofrecer esta última tasa ya que representa un valor de competencia directa en el mercado.

- En el segundo perfil, el cotizador indicó una tasa de 6.10, luego de realizar la evaluación tendencial del mercado con el método predictivo se llegó a una tasa de 6.07, por tanto, se determina que la empresa debe ofrecer esta última tasa ya que representa el valor suficiente para cerrar esa venta y que simboliza un ahorro en tres puntos básicos.

5.2 Aplicaciones

Las aplicaciones consideradas para el método predictivo se muestran en la Figura 5.6.



Figura 5.6: Aplicaciones del Método Predictivo.

Elaboración: los autores.

5.2.1 Integración al Sistema cotizador actual

Como se observa en el punto anterior, el Sistema Cotizador solo realiza el cálculo de la tasa de manera mecánica con los valores ingresados al momento del registro de la solicitud. Una de las aplicaciones del método predictivo es la automatización del Sistema Cotizador para que realice un análisis estadístico a través del método de predicción y, al momento de generar la tasa de venta, contemple todas las consideraciones, tanto estadísticas como las reglas de negocio. De esta forma, se lograría la automatización en esta parte del proceso de generación de cotizaciones.

5.2.2 Integración al CRM como dato informativo

Debido a las tendencias de mercado mostrado por el método predictivo se pensó en la disponibilidad de mostrar esta información en el CRM con que cuenta la empresa aseguradora; dirigido hacia a la Fuerza de Ventas, representada por los agentes de ventas que tiene la empresa, ya que la información podría ser utilizada para canalizar los esfuerzos de marketing y despliegues de campañas hacia clientes que representen una fuente de ingresos potenciales.

5.2.3 Como aplicativo para casos especiales

5.2.3.1 Inteligencia comercial

Para esta aplicación, Inteligencia comercial tendría la responsabilidad de ejecutar el método predictivo para aquellos casos que simbolicen una complejidad en el cálculo de la tasa de venta y sean seleccionados por la gerencia de Rentas Vitalicias, ya que puedan representar un CIC superior al promedio, un tema de recotización o sobre los casos de TRA Especiales.

5.2.3.2 Gerencia de Rentas vitalicias

Debido a la mejor comprensión y manejo de las reglas de negocio aplicadas al proceso, la gerencia de Rentas Vitalicias realizará la ejecución del método predictivo mediante una herramienta desarrollada por Inteligencia Comercial que facilite la explotación de la información y que sea de fácil uso, siguiendo los mismos pasos que se ha descrito en el punto anterior.

CONCLUSIONES

1. Se logró la integración de la información de las BD transaccionales participantes en la generación de cotizaciones y el cierre de las mismas, donde, además de ello, con este Data Mart se redujo el tiempo de ejecución del método predictivo debido a que realizaba las consultas a una sola fuente información.
2. Se elaboró de un método predictivo capaz de asignar la mejor tasa de venta a una cotización; para ello se emplearon las técnicas de árboles de decisión e identificación de *outliers*, optimizando la selección de la muestra, con un nivel de confianza superior al 90%, y así mejorar el cálculo de la Tasa de venta para la generación de cotizaciones en Rentas vitalicias.
3. Se logró ajustar la calidad de las cotizaciones sustentándolas con la información estadística producto del análisis dimensional y de la identificación de las variables que participan directamente en el proceso. Asimismo, se automatizó el manejo y aplicación de las reglas del negocio al proceso de cotizaciones.

RECOMENDACIONES

1. Debido a la complejidad del método de predicción, se recomienda capacitar a todos los involucrados que sean afectados por la implementación del mismo directa e indirectamente, desde la gerencia de Rentas Vitalicias (incluida la fuerza de ventas) hasta los empleados del área de Inteligencia comercial, a fin de evitar los llamados “cuellos de botella” por desinformación sobre el uso de la herramienta.
2. El método de predicción utiliza hasta cinco software entre estadísticos, motor de base de datos, entorno de desarrollo y Microsoft Office, se recomienda crear un sistema computarizado que automatice la explotación y utilización de las aplicaciones antes mencionadas en donde se obtenga un menor tiempo de ejecución del proceso, como la automatización del mismo.
3. El proyecto realizado en esta tesis, tiene un campo de expansión con más oportunidades de mejora; una de las mejoras identificadas es la utilización de otras técnicas estadísticas a fin de mejorar el enfoque en los procesos previos al método estadístico como por ejemplo, la utilización de la regresión logística, que nos dará una probabilidad, que, bien utilizada, puede mejorar el enfoque del método predictivo. Adicional a ello, se propone una investigación más a fondo de la utilización de técnicas de predicción para el cálculo de la tasa de venta, con la utilización de más variables independientes o nuevas técnicas como por ejemplo, la regresión

lineal utilizando variables ficticias “dummies”, técnicas estadísticas que evalúen la base de forma personalizada y no agrupando por nodos de evaluación.



FUENTES DE INFORMACIÓN

Bibliográficas

1. Arenas, Maria. Rodriguez, Paula (2012) Credit Scoring: Evaluación del riesgo crediticio de la cartera de microcréditos en una institución financiera en Uruguay (Tesis de Pregrado). Universidad de la Republica. Montevideo, Uruguay.
2. Bissantz, N., & Hagedorn, J. (2009) Data Mining. Business & Information Systems Engineering, 1(1), 118-122.
3. Chang, J. F. (2006) Bussiness Process Management System. Auerbach.
4. Inmon, W. H. (2005) Building the data warehouse. Segunda edición, Nueva York, Estados Unidos: Wiley.
5. Kimball et al. (2008). The Data Warehouse Lifecycle Toolkit. Segunda Edición. Nueva York, Estados Unidos: Wiley.
6. Pérez, César (2007) Minería de datos: técnicas y herramientas. Madrid - España: Paraninfo.
7. Pérez, César (2009) Técnicas de Análisis de Datos con SPSS 15. Universidad Computense de Madrid: Pearson Educación S.A.
8. Pérez, C. (2011) Técnicas de segmentación. Conceptos, herramientas y aplicaciones. Madrid: Gaceta Grupo Editorial.

9. Pérez, C. & Santín, D. (2007) Minería de Datos: Técnicas y Herramientas. Madrid: Ediciones Paraninfo, S.A.

Hemerográficas

1. Chapman, Pete et al. (2000), CRISP-DM 1.0: "Step-by-step data mining guide". CRISP-DM consortium.
2. Berlanga, Rubio y Vilá (2013) Cómo aplicar árboles de decisión en SPSS. REIRE, Revista d'Innovació i Recerca en Educació, Publicado en enero del 2013, N° 6, p. 79.
3. Fayyad, U. & Piatetsky-Shapiro, G. & Smythet, P. (1996) From Data Mining to Knowledge Discovery in Databases. AI MAGAZINE, volumen 17 N°3 pp 37-54. Recuperado de: <http://dx.doi.org/10.1609/aimag.v17i3.1230>
4. Goebel, M., & Gruenwald, L. (1999) A survey of data mining and knowledge discovery software tools. ACM SIGKDD Explorations Newsletter, 1(1), 20-33.
5. Interseguro (2014) Boletín Mensual – Noviembre. Publicado en noviembre del 2014, N° 11, p. 13. "Proceso de Venta de Rentas Vitalicias". Perú.
6. Kamath, C. (2006) Mining science data. Institute of Physics Publishing, N° 46, pp. 500-504. Recuperado de: http://iopscience.iop.org/1742-6596/46/1/068/pdf/1742-6596_46_1_068.pdf
7. Kritzer Barbara, kay Stephen & Sinha Tapen (2011) Next Generation of Individual Account Pension Reforms in Latin America. (Social Security Bulletin, Vol. 71, No. 1, pp. 35-76, 2011). Recuperado de SSRN:<http://ssrn.com/abstract=1769350>

Electrónicas

1. Class & Asociados (2014) Fundamentos de Clasificación de Riesgo. Interseguro Compañía de Seguros S.A. Recuperado de: <<http://www.classrating.com/Interseg.pdf>>
2. Chang, C. (2014, Enero 29). El 56% del fondo de pensión del afiliado corresponde a la rentabilidad, afirma AFP Integra [Video]. Publicado el 29 de enero del 2014, obtenido el 20 de Marzo del 2015, desde: <http://gestion.pe/tu-dinero/modalidades-pension-jubilacion-renta-vitalicia-retiro-programado-afp-spp-2087550>
3. De Los Ángeles, M. (2005). IBM DB2 OLAP Server. Publicado 2005. Obtenido el 10 Abril del 2015, desde: <http://exa.unne.edu.ar/informatica/SO/OLAP.pdf>
4. Diario Gestión. (2015) Rentas vitalicias de aseguradoras habrían crecido 24.9%. Publicado el 28 de enero del 2015, obtenido el 20 de Marzo del 2015, desde: <http://gestion.pe/tu-dinero/rentas-jubilacion-aseguradoras-habrian-crecido-249-sumando-casi-s-2000-mllns-2014-2121837>
5. Fort, M. (2015). ¿AFP o aseguradora? Averigua dónde conviene recibir una pensión [Video]. Publicado el 11 de febrero del 2015, obtenido el 06 de Abril del 2015, desde: <http://gestion.pe/tu-dinero/afp-aseguradora-averigua-donde-conviene-recibir-pension-2123124>
6. Gartner. (2012). Gartner Identifies the Top 10 Strategic Technologies for 2012. Publicado el 18 Octubre del 2012, obtenido el 13 de Abril del 2015, desde: <http://www.gartner.com/newsroom/id/1826214>
7. López, Leonardo (2013) Predicción de fuga de clientes desde un enfoque de competencia. Universidad de Chile. Santiago de Chile, Chile
8. Méndez, Marcela del Pilar (2013). Desarrollo de un Modelo de Recomendación de Compra para Clientes de una Empresa de Seguros (Tesis de Pregrado). Universidad de Chile, Chile.

9. Palacios Quichiz, Luis (2012). Modelo de RNA para predecir la morosidad de Microcrédito en la Banca Estatal Peruana (Tesis de maestría). Universidad Nacional Mayor de San Marcos, Lima - Perú.
10. Pérez, Alex (2014) Modelo de Activación de Tarjetas de Crédito en el mercado crediticio ecuatoriano. Escuela Politécnica Nacional. Quito, Ecuador.
11. Ranjan, Jayanthi (2009) Business Intelligence: Concepts, components, techniques and benefits. Journal of Theoretical and Applied Information Technology. Institute of Management Technology. India.
12. Romano, R. & Davino, C. & Naes, T. (2014) Classification trees in consumer studies for combining both product attributes and consumer preferences with additional consumer characteristics, Food Quality and Preference. Elsevier, volumen 33 (Abril 2014), pp. 27-36. Recuperado de: <http://dx.doi.org/10.1016/j.foodqual.2013.11.006>
13. SAS Enterprise Miner (2015) – SEMMA. SAS Institute. Publicado en Mayo 2008, obtenido el 23 de Marzo del 2015, desde: http://www.sas.com/en_us/software/analytics/enterprise-miner.html
14. SBS (2013) Estadísticas de cotizaciones en Rentas Vitalicias Diciembre del 2014. Publicado el 31 de enero de 2015, obtenido el 15 de Marzo del 2015, desde: <http://www.sbs.gob.pe/>
15. SBS (2015). Carpeta de Información del Sistema Privado de Pensiones. Superintendencia de Banca y Seguros. Febrero del 2015. Recuperado de: <https://intranet1.sbs.gob.pe/estadistica/financiera/2015/Febrero/FP-1339-fe2015.ZIP>
16. So Young Sohn & Ji Won Kim. (2012) Decision tree-based technology credit scoring for start-up firms: Korean case, Expert Systems with Applications. Elsevier, volumen 39 (Marzo 2012), pp. 4007-4012. Recuperado de: <http://dx.doi.org/10.1016/j.eswa.2011.09.075>

17. Valcárcel, Violeta (2004) Data Mining y el Descubrimiento del Conocimiento. Revista de la Facultad de Ingeniería Industrial [en línea]. Noviembre de 2004, vol.7, pp. 83-86. Recuperado de: http://sisbib.unmsm.edu.pe/bibvirtualdata/publicaciones/indata/Vol7_n2/Pdf/a13.pdf
18. Ventura, M. (2014). ¿Qué debo tomar en cuenta si me jubilaré por renta vitalicia? [Video]. Publicado el 18 de Abril del 2014, obtenido el 15 de Marzo del 2015 desde: <http://gestion.pe/tu-dinero/tu-dinero-que-debo-tomar-cuenta-si-me-jubilare-renta-vitalicia-2094829>



ANEXOS

Anexo 1: Entrevistas a expertos	82
Anexo 2: Cronograma del proyecto	86
Anexo 3: Análisis dimensional del Data Mart	88
Anexo 4: Diseño del Data Mart	100
Anexo 5: Implementación del Data Mart	106



ANEXO 1: Entrevistas a expertos

Entrevista N° 1: Gerente Central de Operaciones de una AFP

Se presenta la entrevista (Diario Gestión, 2014) realizada a César Chang, gerente central de operaciones de AFP Integra.

¿Cuál es la pensión histórica promedio en ambos sistemas de pensiones?

La pensión promedio histórica en el Sistema Privado de Pensiones (SPP) supera los S/. 1,000 hasta la fecha. En el caso de la ONP, ese monto se reduce a S/. 800. Menciona que la constancia de los afiliados a las AFP en sus aportes y la rentabilidad que estos generan en el tiempo explica el referido dato.

¿Si se realiza un análisis sobre la rentabilidad que produce la AFP sobre el fondo aportado por el afiliado que porcentaje se estimaría?

Actualmente el 56% del fondo de pensión del afiliado es rentabilidad pura, entonces conforme pasen los años ese porcentaje va a ir aumentando sobre el total. Esto dependerá de la cantidad de aportes que realice el afiliado y la constancia en ello para obtener una buena pensión.

¿Cuáles son las modalidades de jubilación más solicitadas?

El afiliado al SPP puede acogerse a una de las siguientes alternativas:

Jubilación por edad legal. La persona puede acceder a ella a partir de los 65 años de edad.

Régimen especial de jubilación anticipada por desempleo. Uno de los requisitos es estar desempleado durante los últimos doce meses, las mujeres pueden acceder a partir de los 50 años y los hombres a partir de los 55 años.

La pensión calculada en el Sistema Privado de Pensiones tendrá que resultar igual o superior al valor de la Remuneración Mínima Vital. Si al hacerse los cálculos correspondientes la pensión no llega a los S/.750 mensuales, la persona tiene automáticamente derecho a solicitar la devolución del 50% del fondo.

Pensión mínima. El Estado le brinda al trabajador esa garantía bajo una serie de condiciones para que reciba la pensión mínima que obtendría en la ONP.

Una vez que la persona califica para percibir una pensión de jubilación, puede acceder a distintas modalidades. El retiro programado y la renta vitalicia son los más utilizados, lo contrario ocurre con la renta mixta y la renta bimoneda.

¿Por qué la elección de una renta Vitalicia?

Para explicar la elección de la renta vitalicia se mencionaron las siguientes características con respecto al retiro programado que es la modalidad ofrecida por la AFP.

Con el retiro programado el afiliado se mantiene como propietario del fondo, que es directamente administrado por la AFP. La entidad ofrece una pensión en función a la rentabilidad que se va generando, por lo que el monto de la pensión varía. Por tanto, “en periodos en que pudiera haber una baja rentabilidad probablemente el monto de pensión caiga, esa variabilidad podría afectarle al afiliado si solo depende de ese ingreso”, advirtió. De otro lado, “en caso la persona fallezca y no tenga beneficiarios directos, el fondo es heredable”.

Señaló que, por su parte, la renta vitalicia requiere que el afiliado entregue su fondo a una compañía de seguros. La diferencia central respecto al retiro programado consiste en que la persona siempre recibirá una pensión fija mientras viva. En ese caso, el afiliado se olvida del riesgo de longevidad.

Otra diferencia clave, añadió, es que “el fondo es un producto no heredable; a menos que se pacte con la compañía de seguros alguna característica de ‘periodos garantizados’ que le permitan a la familia recibir una pensión, a pesar de que el afiliado fallezca”.

“El proceso de elección de pensión de jubilación requiere una asesoría personal para que cada afiliado entienda mejor, en función a sus características, qué le conviene más”.

Entrevista N° 2: Gerente General de una Aseguradora

Se presenta la entrevista (Diario Gestión, 2014) realizada a Mario Ventura, gerente general de Seguros Sura.

De acuerdo al entrevistado a diferencia de la jubilación bajo retiro programado (la otra opción en el mercado propuesta por las AFP), el riesgo de vivir más de lo esperado lo asume la Aseguradora, pues no le pagarán a partir de un fondo que pueda agotarse, sino de los términos del contrato con la aseguradora

¿Qué otros aspectos se deben tomar en cuenta?

Bajo la modalidad de renta vitalicia, el jubilado transfiere su fondo a una empresa de seguros que pagará la pensión dependiendo de la modalidad elegida, pactando distintos aspectos como el tiempo máximo que recibirá la mensualidad. Incluso, puede asegurar a su cónyuge en caso de fallecimiento.

1. Moneda

Lo primero es preguntarse qué moneda necesitará en el futuro. Normalmente son nuevos soles, pero quizás algunos planeen viajar o vivir en el extranjero.

2. Informarse

Lo segundo es informarse bien sobre las empresas, visitar y preguntar. “No se queden con ninguna duda porque es una decisión irrevocable. La idea es que lo clientes formen una opinión propia”, subrayó.

Cabe resaltar que hay decenas de opciones el mercado: Bimoneda, vitalicia diferida, inmediata, entre otras.

3. Respaldo

Si bien la SBS dispone que todas las empresas de seguros ofrezcan los mismos productos de renta vitalicia, Ventura recalcó que es necesario que el aportante compare el respaldo que tiene cada institución, así como el manejo de los recursos.

“Las empresas de seguros necesitan operadores solventes, idealmente con experiencia internacional en otros mercados y acceso a mejores inversiones, con métodos de control de riesgos y de gobierno corporativo alineados a los intereses del clientes”, acotó. Es decir, es mejor si puede comprobar la sostenibilidad de la empresa.

Entrevista N° 3: Vicepresidente comercial de una AFP

Se presenta la entrevista (Diario Gestión, 2015) realizada a Marcel Fort, vicepresidente comercial de AFP Integra.

¿Dónde me conviene recibir una pensión: AFP o aseguradora?

Según Fort, la diferencia fundamental entre los dos es el tipo de riesgo que uno asume.

“Si uno aporta bajo retiro programado (AFP), lo puede tomar en nuevos soles, mientras que en una aseguradora se pueda hacer en nuevos soles y dólares”, declaró. Fort dijo que esta última opción permite mitigar el riesgo cambiario.

El segundo riesgo, según el experto, es la expectativa de vida. En un retiro programado, la pensión comienza a reducirse si vives más de lo estimado. Por otro lado, en una renta vitalicia (aseguradora), puedes seguir recibiendo el mismo monto, si vives más que la probabilidad de vida. “La aseguradora cubre el riesgo de vida o muerte”, añadió.

Fort detalló que, el tercer riesgo es la forma en que las compañías asumen variaciones en los fondos. En una AFP, los fondos están sujetos a cambios, y eso podría afectar el recalcule del monto.

En cambio, en una aseguradora, recibirás el mismo monto, independientemente si sube o baja el fondo.

La conclusión es simple: si deseas asumir riesgos, tu opción de jubilación sería una AFP (retiro programado). Sin embargo, si no estás dispuesto a soportar dichos riesgos, tu opción sería una compañía de seguros, donde pagas una prima por dicho riesgo.

Otro dato es que si la rentabilidad del fondo es muy grande, los clientes prefieren quedarse en una AFP; si pasa lo contrario, y los rendimientos no son tan atractivos, los pensionistas optan por tomar una renta vitalicia en una aseguradora.

ANEXO 2: Cronograma del Proyecto

En el presente anexo se describen las actividades a realizar en el proyecto, cabe mencionar que, las metodologías a considerar son Kimball y CRISP-DM, para el Data Mart y el método predictivo, respectivamente.

El proyecto tiene una duración de 66 días, dando 35 días para el diseño y la construcción del Data Mart, 17 días para la implementación del método predictivo y 8 días para la integración de ambas herramientas. Cabe mencionar que, ambas fases son totalmente independientes pudiendo construirse simultáneamente, inclusive.

1. Cronograma



▣ Proyecto "PREDICCIÓN DE LA TASA DE VENTA PARA RENTAS VITALICIAS EN UNA EMPRESA ASEGURADORA"	66 días	lun 09/03/15	sáb 06/06/15
Definición de Requerimientos del Negocio	5 días	lun 09/03/15	vie 13/03/15
▣ DATAMART	35 días	sáb 14/03/15	jue 30/04/15
Definición del Objetivo	4 días	lun 16/03/15	jue 19/03/15
▣ Análisis	5 días	vie 20/03/15	jue 26/03/15
Modelo Dimensional	5 días	vie 20/03/15	jue 26/03/15
▣ Diseño	6 días	mar 07/04/15	mar 14/04/15
Arquitectura de Data Mart	3 días	mar 07/04/15	jue 09/04/15
Diseño y Extracción del Data Mart	4 días	jue 09/04/15	mar 14/04/15
Implementación	2 días	mié 15/04/15	jue 16/04/15
▣ MODELO PREDICTIVO	17 días	vie 01/05/15	lun 25/05/15
Selección y Limpieza de datos	4 días	vie 01/05/15	mié 06/05/15
Construcción de datos	4 días	vie 01/05/15	mié 06/05/15
Selección de técnica de modelado	3 días	jue 07/05/15	lun 11/05/15
Diseño de pruebas de modelo	4 días	mar 12/05/15	vie 15/05/15
Construcción de modelo	3 días	lun 18/05/15	mié 20/05/15
Prueba de Modelo	4 días	jue 21/05/15	mar 26/05/15
▣ Automatización de Proceso	8 días	lun 25/05/15	mié 03/06/15
Prueba de Reglas de Negocio	3 días	mar 26/05/15	jue 28/05/15
Prueba de Resultado Esperado	2 días	vie 29/05/15	dom 31/05/15
Pruebas finales	3 días	lun 01/06/15	mié 03/06/15
Presentación Final	3 días	jue 04/06/15	sáb 06/06/15

Cronograma del Proyecto.

Elaboración: los autores.

ANEXO 3: Análisis Dimensional del Data Mart

En el presente anexo se describen las dimensiones y tabla de hechos que componen el Data Mart, así como la función que cumple cada una de ellas para satisfacer los requerimientos funcionales.

1. Dimensiones

Las dimensiones que conforman el esquema copo de nieve propuesto para el desarrollo de la presente tesis son las siguientes:

N°	Dimensión
1	D_Compañía
2	D_CotizacionesPredictor
3	D_CierreCotización
4	D_Cierres
5	D_Persona
6	D_Ubigeo
7	D_Tiempo
8	D_Predictor

1.1. D_Compañía

Esta dimensión contiene los datos de las compañías aseguradoras que componen el sistema asegurador, las cuales envían sus cotizaciones ante la recepción de nuevas solicitudes de jubilación.

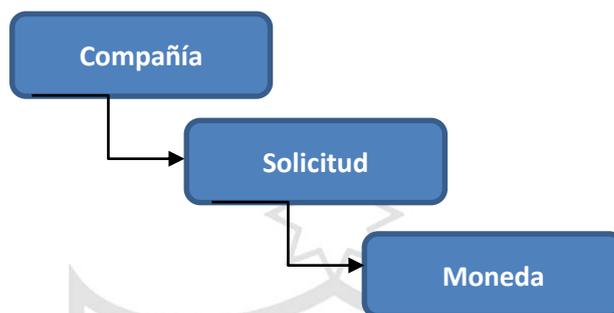
1.1.1. Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
CodCompañía	Identificador llave de la compañía.	Numérico	0
DescCompañía	Descripción de la Compañía aseguradora.	A-Z: Cualquier valor entre A y Z	Ninguno

1.2. D_CotizacionesPredictor

Esta dimensión contiene los datos de las cotizaciones hechas por las diferentes compañías del mercado asegurador hacia una persona que solicitó jubilación.

1.2.1. Jerarquías



1.2.2. Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
CodCompañía	Código de la Compañía que realizó la solicitud	Numérico	0
Num_Solicitud	Numero de solicitud de jubilación	Alfanumérico	Ninguno
TV	Tasa de Venta ofrecida por la compañía	Numérico	0
TR	Factor de rentabilidad de la CIC	Numérico	0
FlagGanador	Indica que compañía ganó la solicitud. Indica con el número 1	Numérico	0
CIC_Bruto	Indica la totalidad del CIC que cuenta el afiliado.	Numérico	0
CIC_Neto	Indica el total, luego de cobros, de CIC que el afiliado contaría.	Numérico	0

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
Moneda	Indica el tipo de moneda en la cual fue cotizada la solicitud.	Alfabético	Ninguno

1.3. D_CierreCotización

Esta dimensión contiene el detalle de las cotizaciones generadas hacia un cliente, considerando datos que servirán como variables para el árbol de decisión.

1.3.1. Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
Num_Solicitud	Numero de solicitud de jubilación	Alfanumérico	Ninguno
num_cuspp	Código Único en el Sistema Privado de Pensiones	Alfanumérico	Ninguno
Categoría	Tipo de producto de jubilación a ofrecer	Alfabético	Ninguno
Cod_modalidad	Modalidad de pago de pensión	Alfabético	Ninguno
Periodo_temporal	Periodo de renta temporal que el cliente elija, si lo hiciere	Numérico	0
Pje_pension	% de pensión en Renta Temporal, si lo hiciere	Numérico	0
Val_periodo_garantizado	Cantidad de años que el cliente elija como seguro.	Numérico	0
Pje_cobertura_conyuge	% de cobertura hacia el cónyuge del asegurado	Numérico	0
Afp	Empresa AFP en la que estuvo afiliado la persona	Alfabético	Ninguno

Zona	Ubicación geográfica, Lima o Provincia, donde se encuentra la persona	Alfabético	Ninguno
Sexo	Género de la persona	Alfabético	Ninguno
Edad	Edad de la persona	Numérico	Ninguno
FechaCierre	Fecha en la que se concretó la jubilación	Fecha	Ninguno
FechaPlazo	Fecha en la que se generó la cotización	Fecha	Ninguno
Departamento	Departamento en el cual reside la persona	Alfabético	Ninguno
Zona_Geografica	Zona geográfica, según Arellano Marketing, en la que reside la persona	Alfabético	Ninguno

1.4. D_Cierres

Esta dimensión contiene la información sobre el cierre de las solicitudes en su etapa final.

1.4.1. Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
Num_Solicitud	Numero de solicitud de jubilación	Alfanumérico	Ninguno
CompañiaCierre	Compañía ganadora de la solicitud de la cotización	Alfabético	Ninguno
FechaCotizacion	Fecha en la que se generó la cotización	Fecha	Ninguno
FechaCierre	Fecha en la que se concretó la jubilación	Fecha	Ninguno
TV_Cierre	Tasa de venta (Meler) con la que se cerró la jubilación	Numérico	Ninguno
cic_bruto	Indica la totalidad del CIC que cuenta el afiliado.	Numérico	0

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
cic-neto	Indica el total, luego de cobros, de CIC que el afiliado contaría.	Numérico	0

1.5. D_Persona

Esta dimensión contienen los datos personales y de contacto de las personas que accedieron a una jubilación.

1.5.1. Atributos

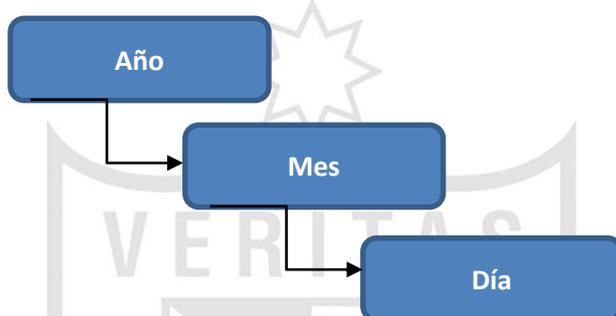
Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
Num_Cussp	Código Único en el Sistema Privado de Pensiones	Alfanumérico	Ninguno
Cod_Comuna	Código de Comuna (Distrito) en la que reside la persona	Alfabético	-
Cod_Ciudad	Código de Ciudad en la que reside la persona	Alfabético	-
TipoDocumento	Tipo de documento de identificación	Alfabético	Ninguno
NumeroDocumento	Numero de documento de identificación	Alfabético	Ninguno
Ape_Paterno	Apellido Paterno	Alfabético	Ninguno
Ape_Materno	Apellido Materno	Alfabético	Ninguno
Nombre	Nombre	Alfabético	Ninguno
Sexo	Género de la persona	Alfabético	Ninguno
FechaNacimiento	Fecha de Nacimiento	Fecha	Ninguno
Direccion	Dirección de residencia	Alfabético	Ninguno

NumeroTelefono	Número telefónico de contacto	Alfabético	Ninguno
Email	Correo Electrónico de contacto	Alfabético	Ninguno

1.6. D_Tiempo

Esta dimensión contiene los niveles en tiempo en el cual se harán los filtros de consulta.

1.6.1. Jerarquía



1.6.2. Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
CodTiempo	Identificado de tiempo	Alfabético	Ninguno
Fecha	Fecha Exacta	Fecha	Ninguno
Año	Año en el que se realizó la solicitud	Numérico	Ninguno
Mes	Mes en el que se realizó la solicitud	Alfabético	Ninguno
DiaSemana	Día de la semana en la que se realizó la solicitud	Alfabético	Ninguno
DescFecha	Descripción de fecha en la que se realizó la solicitud	Alfabético	Ninguno

1.7. D_Predictor

Esta dimensión contiene los registros, que son seleccionados por la gerencia de Rentas Vitalicias, para el cálculo de la tasa de venta por medio de la predicción.

1.7.1. Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
Num_operacion	Número de identificación de la solicitud	Alfabético	Ninguno
Sexo	Género de la persona	Alfabético	Ninguno
FechaNacimiento	Fecha de Nacimiento	Fecha	Ninguno
Compañía	Compañía Competidora	Alfabético	Ninguno
Moneda	Tipo de moneda a cotizar	Alfabético	Ninguno
Categoría	Tipo de Jubilación a cotizar	Alfabético	Ninguno
AFP	Empresa AFP de la persona	Alfabético	Ninguno
Zona	Ubicación: Lima o Provincia de la persona	Alfabético	Ninguno
Cod_Modalidad	Modalidad de pago de pensión	Alfabético	Ninguno
pje_pension	% de pensión en Renta Temporal, si lo hiciere	Numérico	0
val_periodo_garantizado	Cantidad de años que el cliente elija como seguro.	Numérico	0
periodo_temporal	Cantidad de años en Renta Temporal, si lo hiciere	Numérico	0
FechaPlazo	Fecha en la que se realizó la cotización	Fecha	Ninguno
FechaDevengue	Ultima fecha en la que se hizo devengue a la AFP	Fecha	Ninguno
Conyugue_FechaNacimiento	Fecha de Nacimiento de Cónyuge	Fecha	Ninguno
CondicionSaludConyugue	Condición de salud de Cónyuge	Alfabético	Sano

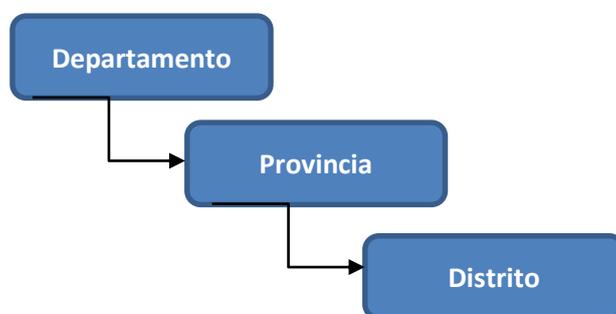
Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
FlagBeneficiario	Indicador si tiene beneficiarios	Numérico	0
NumHijos	Número de hijos	Numérico	0
Acom	Numero de ACOM	Numérico	0
Dcom	Numero de DCOM	Numérico	0
TRA	Factor TRA	Numérico	0
TV_Cotizacion	Tasa de Venta generada por Cotizador	Numérico	0
FechaControl	Fecha Actual	Alfabético	Ninguno
TasalInversa_Pacifico	Tasa de Venta inversa para Pacífico	Numérico	0
TasalInversa_Rimac	Tasa de Venta inversa para Rímac	Numérico	0
TasalInversa_Sura	Tasa de Venta inversa para Sura	Numérico	0
TasalInversa_LPV	Tasa de Venta inversa para La Positiva	Numérico	0
DifPuntosBasicos	Cantidad de puntos básicos para conversión a Tasa Meler	Numérico	0
TasaMeler_Pacifico	Tasa Meler para Pacífico	Numérico	0
TasaMeler_Rimac	Tasa Meler para Rimac	Numérico	0
TasaMeler_Sura	Tasa Meler para Sura	Numérico	0
TasMeler_LPV	Tasa Meler para La Positiva	Numérico	0
CompañíaDriver	Compañía a comparar por perfil	Alfabético	Ninguno
VariacionTasa	Incremento/Decremento en puntos básicos perfil	Numérico	0
TasaMelerCia_Variada	Tasa Predictor con variación de tasa	Numérico	0

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
FlagExcedeTasaMaxima	Indicación de límite de valor tasa	Alfabético	Ninguno
TasaMeler_FlagTasaMaxima	Tasa Predictor ajustada a limite	Numérico	0
DeltaTasa	Diferencia entre Tasa Predictor y Tasa Cotizador	Numérico	0
FlagTRA	Resultado (<, > o =) entre tasas	Alfabético	Ninguno
TRA_Maxima	TRA máximo por perfil	Alfabético	Ninguno
ResultadoDeltaTasa	Resultado de evaluación de tasa vs TRA	Alfabético	Ninguno
DeltaTasa_Final	Tasa ajustada a condiciones TRA	Numérico	0
Tasa_Meler_Predictor	Tasa Predictor	Numérico	0
FechaProceso	Fecha Actual	Alfabético	Ninguno
FlgNivelConfianza	Grado de Confianza por perfil	Alfabético	Ninguno

1.8. D_Ubigeo

Esta dimensión contiene información sobre la ubicación geográfica en la cual una persona afiliada se encuentra. Lugar de residencia de la persona.

1.8.1. Jerarquía



1.8.2. Atributos

Nombre del Atributo	Contenido		
	Descripción	Formato	Valor por Defecto
Cod_Comuna	Código de Comuna (Distrito) en la que reside la persona	Alfabético	-
Cod_Ciudad	Código de Ciudad en la que reside la persona	Alfabético	-
Departamento	Descrip. Departamento	Alfabético	Ninguno
Provincia	Descrip. Provincia	Alfabético	Ninguno
Distrito	Descrip. Distrito	Alfabético	Ninguno
Ubigeo_Largo	Descrip. Departamento / Provincia / Distrito	Alfabético	Ninguno

2. Tabla de Hechos

La tabla de hechos está conformada por las llaves foráneas (que en su conjunto forman la llave primaria) de las dimensiones existentes en el esquema.

2.1. F_ResultadoPredictor

Permite mostrar los resultados de los cierres de los prospectos a los cuales se les generó tasa predictor y a los que no se les generó; haciendo su respectiva comparación.

Granularidad

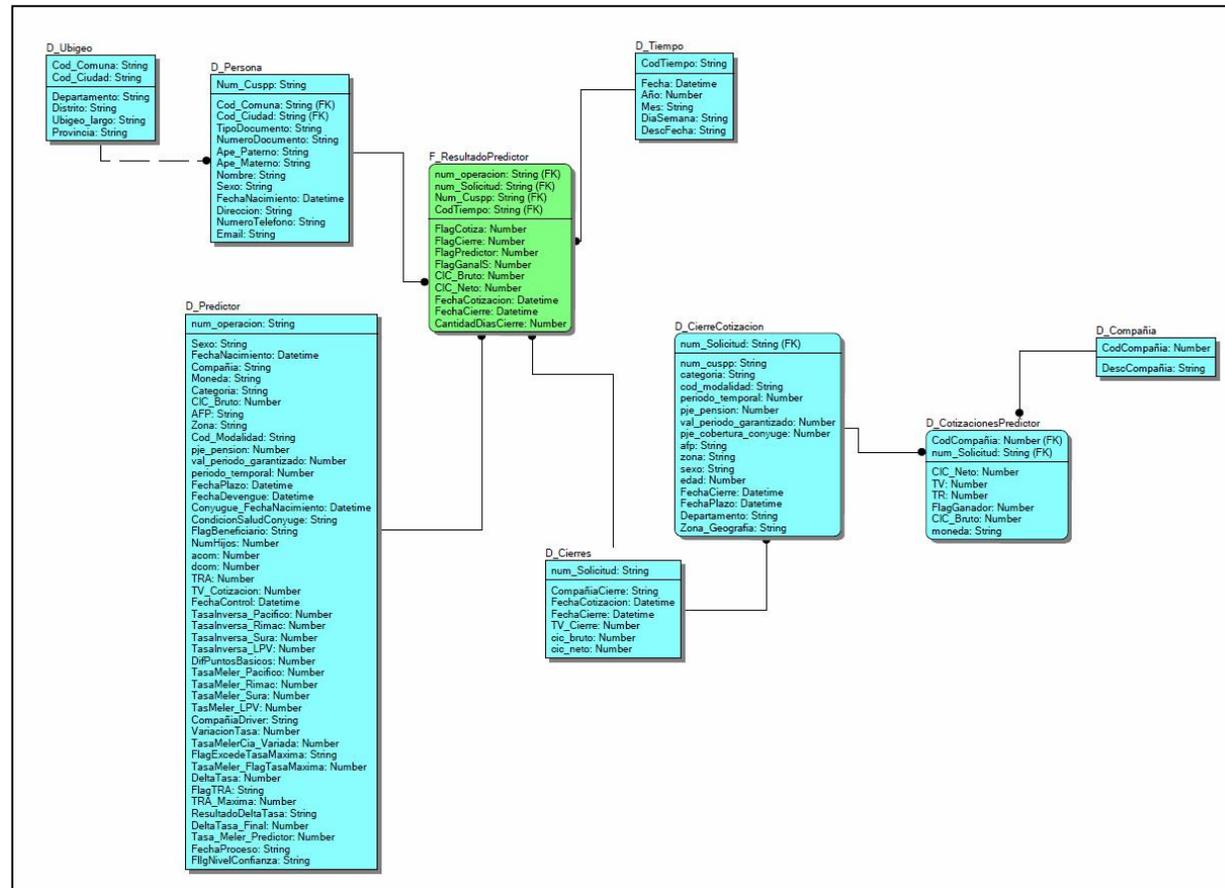
Nombre del Atributo	Descripción	
	Descripción	Formato
CodTiempo	Representa el periodo de tiempo a evaluar	Forma parte de la llave primaria
Num_Cuspp	Representa al código SPP que cuenta la persona	Forma parte de la llave primaria

Nombre del Atributo	Descripción	Formato
num_operacion	Representa el número de operación asignada a la solicitud y enviada a evaluación Predictor	Forma parte de la llave primaria
num_Solicitud	Representa el número de solicitud que cerró ante la SBS la petición de jubilación	Forma parte de la llave primaria

Medidas

Nombre del Atributo	Descripción
FechaCotizacion	Fecha en la que se generó la cotización para la solicitud de jubilación
FechaCierre	Fecha en la que se cerró la solicitud de jubilación con alguna empresa aseguradora o AFP.
FlagCotiza	Indicador de registro cotizado
FlagCierre	Identificador de cierre de la solicitud
FlagPredictor	Identificador de solicitud generada con tasa predictor
FlagGanaIS	Identificador de solicitud ganada por la empresa aseguradora
CIC_Bruto	Monto de CIC Bruto
CIC_Neto	Monto de CIC Neto (luego de cobros e impuestos)
CantidadDiasCierre	Cantidad de días transcurridos entre la generación de cotización y cierre.

3. Modelo Lógico



Modelo Lógico del Data Mart.

Elaboración: los autores.

ANEXO 4: Diseño del Data Mart

En el presente anexo se describe la forma de cómo fue creado el Data Mart, así como la forma de cómo se llena con información. También se presenta el modelo físico de la solución propuesta, así como el script de este último.

1. Diseño y Creación del Data Mart

Luego del análisis realizado, que tuvo como resultado el modelo lógico del Data Mart, se utiliza la herramienta SQL Server 2008 R2 para la creación del mismo. Para ello se utiliza las siguientes sentencias SQL:

```
CREATE TABLE D_Compañia
(
CodCompañia          integer NOT NULL ,
DescCompañia        varchar(20) NULL
)
go
ALTER TABLE D_Compañia
ADD CONSTRAINT XPKD_Compañia PRIMARY KEY CLUSTERED (CodCompañia
ASC)
go
CREATE TABLE D_CotizacionesPredictor
(
CodCompañia          integer NOT NULL ,
CIC_Neto             decimal(15,2) NULL ,
TV                  decimal(5,3) NULL ,
TR                  decimal(5,3) NULL ,
FlagGanador         integer NULL ,
num_Solicitud        varchar(13) NOT NULL ,
CIC_Bruto           decimal(15,2) NULL ,
moneda              varchar(10) NULL
)
go
ALTER TABLE D_CotizacionesPredictor
ADD CONSTRAINT XPKD_CotizacionesPredictor PRIMARY KEY CLUSTERED
(CodCompañia ASC,num_Solicitud ASC)
go
CREATE TABLE D_Persona
(
Cod_Comuna          varchar(10) NOT NULL ,
Cod_Ciudad          varchar(10) NOT NULL ,
Num_Cuspp           varchar(12) NOT NULL ,
TipoDocumento       varchar(2) NULL ,
NumeroDocumento     varchar(12) NULL ,
```

```

Ape_Paterno          varchar(50) NULL ,
Ape_Materno          varchar(50) NULL ,
Nombre               varchar(60) NULL ,
Sexo                  char(1) NULL ,
FechaNacimiento      datetime NULL ,
Direccion             varchar(120) NULL ,
NumeroTelefono       varchar(20) NULL ,
Email                 varchar(100) NULL
)
go
ALTER TABLE D_Persona
ADD CONSTRAINT XPKD_Persona PRIMARY KEY CLUSTERED (Num_Cuspp ASC)
Go

CREATE TABLE D_Predictor
(
Sexo                  char(1) NULL ,
FechaNacimiento      datetime NULL ,
Compañia              varchar(20) NULL ,
Moneda                varchar(10) NULL ,
Categoria             varchar(20) NULL ,
CIC_Bruto             decimal(15,2) NULL ,
AFP                   varchar(20) NULL ,
Zona                  char(1) NULL ,
Cod_Modalidad         varchar(10) NULL ,
pje_pension           integer NULL ,
val_periodo_garantizado integer NULL ,
periodo_temporal      integer NULL ,
FechaPlazo            datetime NULL ,
FechaDevengue         datetime NULL ,
Conyugue_FechaNacimiento datetime NULL ,
CondicionSaludConyuge varchar(10) NULL ,
FlagBeneficiario     varchar(2) NULL ,
NumHijos              integer NULL ,
acom                  decimal(10,2) NULL ,
dcom                  decimal(10,2) NULL ,
TRA                   decimal(5,2) NULL ,
TV_Cotizacion         decimal(5,3) NULL ,
FechaControl          datetime NULL ,
TasaInversa_Pacifico decimal(5,3) NULL ,
TasaInversa_Rimac    decimal(5,3) NULL ,
TasaInversa_Sura     decimal(5,3) NULL ,
TasaInversa_LPV      decimal(5,3) NULL ,
DifPuntosBasicos     decimal(5,3) NULL ,
TasaMeler_Pacifico   decimal(5,3) NULL ,
TasaMeler_Rimac      decimal(5,3) NULL ,

```

```

TasaMeler_Sura          decimal(5,3) NULL ,
TasMeler_LPV           decimal(5,3) NULL ,
CompañiaDriver         varchar(30) NULL ,
VariacionTasa          decimal(5,3) NULL ,
TasaMelerCia_Variada   decimal(5,3) NULL ,
FlagExcedeTasaMaxima   varchar(20) NULL ,
TasaMeler_FlagTasaMaxima decimal(5,3) NULL ,
DeltaTasa              decimal(5,3) NULL ,
FlagTRA                varchar(40) NULL ,
TRA_Maxima             decimal(5,3) NULL ,
ResultadoDeltaTasa     varchar(30) NULL ,
DeltaTasa_Final        decimal(5,3) NULL ,
Tasa_Meler_Predictor   decimal(5,3) NULL ,
FechaProceso           varchar(8) NULL ,
FllgNivelConfianza     varchar(10) NULL ,
num_operacion          varchar(13) NOT NULL
)
go
ALTER TABLE D_Predictor
ADD CONSTRAINT XPKD_Predictor PRIMARY KEY CLUSTERED
(num_operacion ASC)
Go
CREATE TABLE D_Tiempo
(
CodTiempo              varchar(8) NOT NULL ,
Fecha                  datetime NULL ,
Año                    integer NULL ,
Mes                    varchar(20) NULL ,
DiaSemana              varchar(20) NULL ,
DescFecha              varchar(100) NULL
)
go

ALTER TABLE D_Tiempo
ADD CONSTRAINT XPKD_Tiempo PRIMARY KEY CLUSTERED (CodTiempo ASC)
go
CREATE TABLE D_Ubigeo
(
Cod_Comuna             varchar(10) NOT NULL ,
Cod_Ciudad             varchar(10) NOT NULL ,
Departamento          varchar(50) NULL ,
Provincia              varchar(50) NULL ,
Distrito               varchar(50) NULL ,
Ubigeo_largo           varchar(200) NULL
)
go

```

```

ALTER TABLE D_Ubigeo
ADD CONSTRAINT XPKD_Ubigeo PRIMARY KEY CLUSTERED (Cod_Ciudad
ASC,Cod_Comuna ASC)
go
CREATE TABLE F_ResultadoPredictor
(
CodTiempo          varchar(8) NOT NULL ,
FlagCotiza         integer NULL ,
FlagCierre         integer NULL ,
FlagPredictor      char(18) NULL ,
FlagGanaIS         integer NULL ,
CIC_Bruto          decimal(15,2) NULL ,
CIC_Neto           decimal(15,2) NULL ,
FechaCotizacion    datetime NULL ,
FechaCierre        datetime NULL ,
CantidadDiasCierre integer NULL ,
Num_Cuspp          varchar(12) NOT NULL ,
num_operacion      varchar(13) NOT NULL ,
num_Solicitud      varchar(13) NOT NULL
)
go
ALTER TABLE F_ResultadoPredictor
ADD CONSTRAINT XPKF_ResultadoPredictor PRIMARY KEY CLUSTERED
(CodTiempo ASC,Num_Cuspp ASC,num_operacion ASC,num_Solicitud ASC)
go
ALTER TABLE D_CierreCotizacion
ADD CONSTRAINT R_10 FOREIGN KEY (num_Solicitud) REFERENCES
D_Cierres(num_Solicitud)
ON DELETE NO ACTION
ON UPDATE NO ACTION
go
ALTER TABLE D_CotizacionesPredictor
ADD CONSTRAINT R_9 FOREIGN KEY (CodCompañía) REFERENCES
D_Compañía(CodCompañía)
ON DELETE NO ACTION
ON UPDATE NO ACTION
go
ALTER TABLE D_CotizacionesPredictor
ADD CONSTRAINT R_8 FOREIGN KEY (num_Solicitud) REFERENCES
D_CierreCotizacion(num_Solicitud)
ON DELETE NO ACTION
ON UPDATE NO ACTION
go
ALTER TABLE D_Persona
ADD CONSTRAINT R_2 FOREIGN KEY (Cod_Ciudad,Cod_Comuna) REFERENCES
D_Ubigeo(Cod_Ciudad,Cod_Comuna)

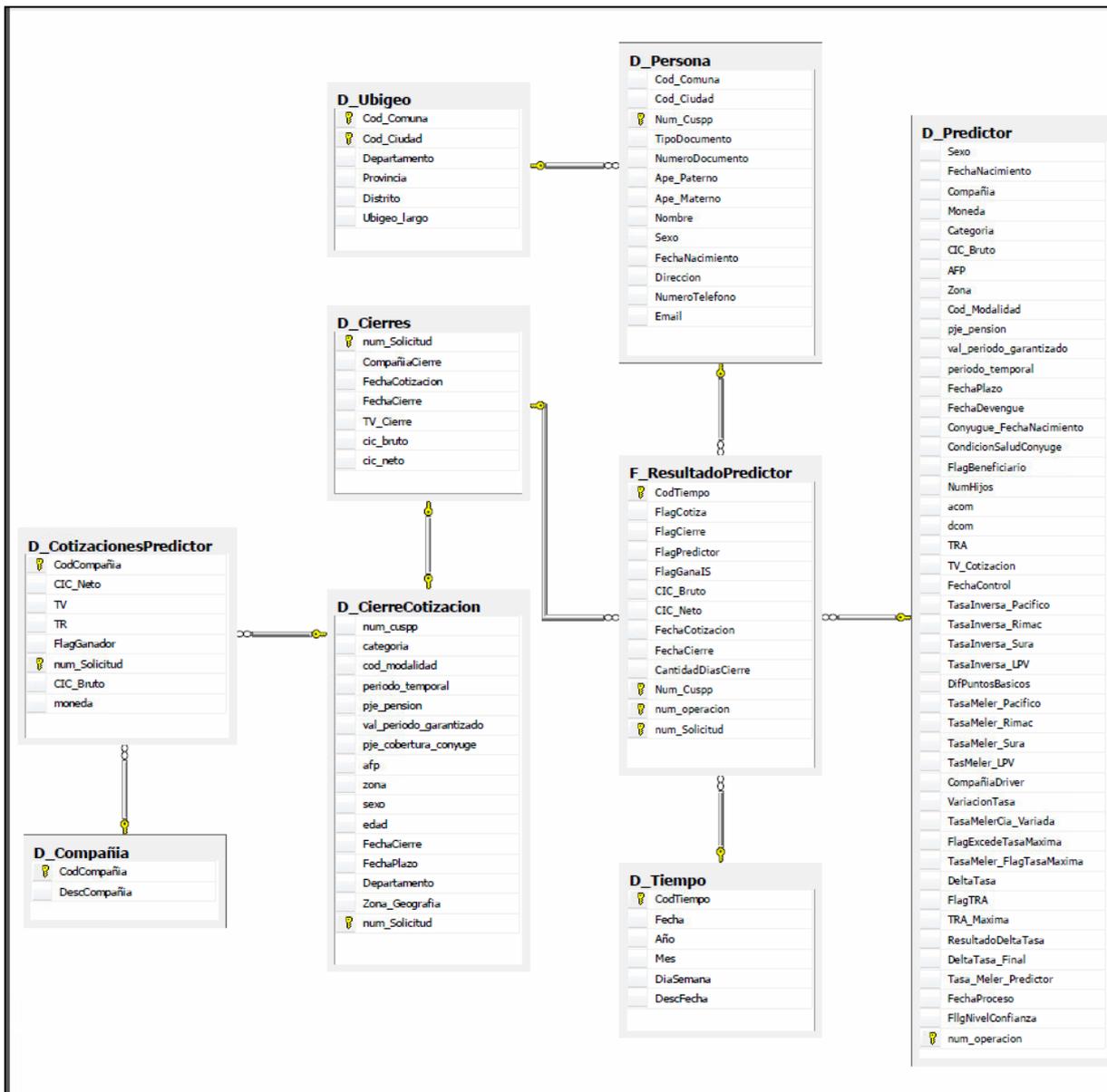
```

```

ON DELETE NO ACTION
ON UPDATE NO ACTION
go
ALTER TABLE F_ResultadoPredictor
ADD CONSTRAINT R_1 FOREIGN KEY (CodTiempo) REFERENCES
D_Tiempo (CodTiempo)
ON DELETE NO ACTION
ON UPDATE NO ACTION
go
ALTER TABLE F_ResultadoPredictor
ADD CONSTRAINT R_3 FOREIGN KEY (Num_Cuspp) REFERENCES
D_Persona (Num_Cuspp)
ON DELETE NO ACTION
ON UPDATE NO ACTION
go
ALTER TABLE F_ResultadoPredictor
ADD CONSTRAINT R_5 FOREIGN KEY (num_operacion) REFERENCES
D_Predictor (num_operacion)
ON DELETE NO ACTION
ON UPDATE NO ACTION
go
ALTER TABLE F_ResultadoPredictor
ADD CONSTRAINT R_11 FOREIGN KEY (num_Solicitud) REFERENCES
D_Cierres (num_Solicitud)
ON DELETE NO ACTION
ON UPDATE NO ACTION

```

Luego de la ejecución del script de creación del Data Mart, se puede generar el diagrama físico del mismo:



Modelo Físico del Data Mart.

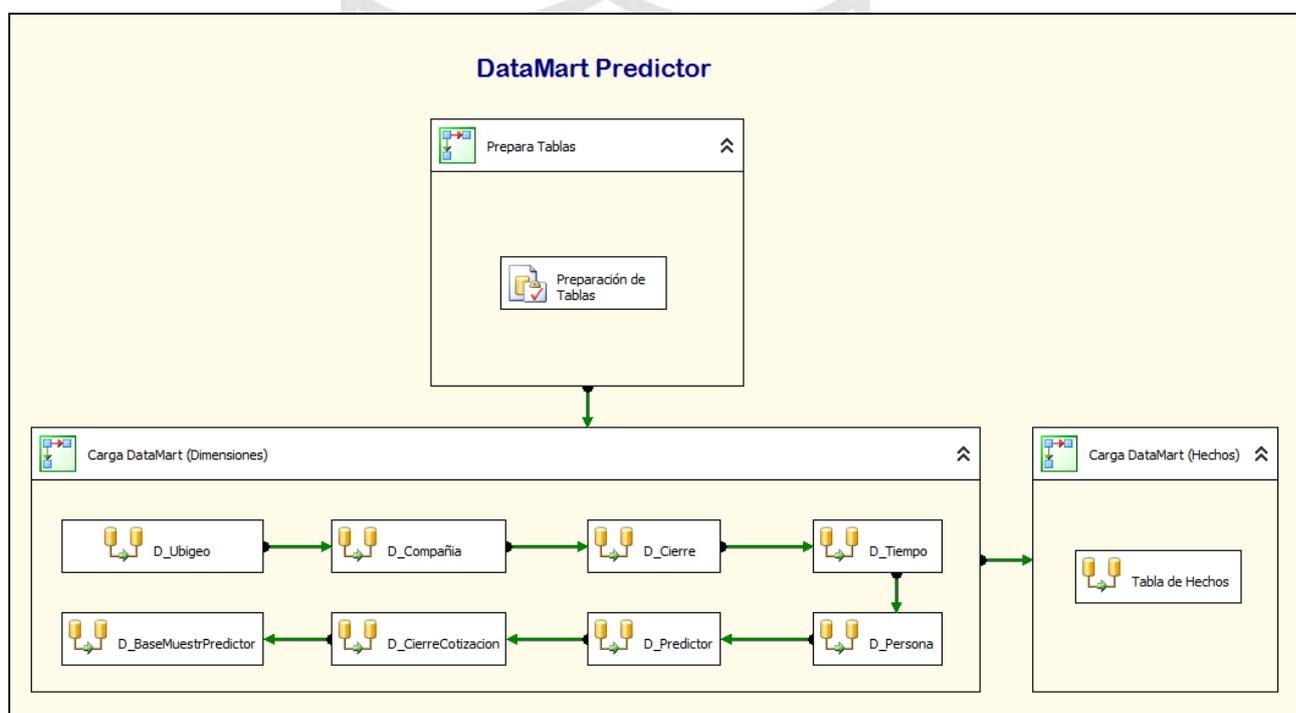
Elaboración: los autores.

ANEXO 5: Implementación del Data Mart

En el presente anexo se describe la forma de implementación del Data Mart, es decir, la herramienta utilizada para la extracción de información, como las sentencias SQL utilizadas para este fin.

8. Extracción de Información e Implementación del Data Mart

Una vez creado el Data Mart, se procede al llenado de las dimensiones en una primera instancia, para luego proceder con el llenado de la tabla de hechos. Para ello, se creará un DTS (herramienta de Integration Services), que nos proporcionará toda la facilidad y dinamismo para el ETL, tal como si muestra en la siguiente imagen:



Herramienta para el ETL del Data Mart.

Elaboración: los autores.

La herramienta que nos permite realizar el ETL, en sincronía con la generación del algoritmo resultando del árbol de decisión, se ejecuta semanalmente, siendo una actividad previa a la generación de los árboles de decisión anteriormente mencionados. Se ha dividido en 3

partes: *Preparación de tablas*, donde se acondiciona las dimensiones para albergar la carga de datos; *Carga Data Mart (Dimensiones)*, se llena de información proveniente de las bases de datos transaccionales y *Carga Data Mart (Hechos)*, genera la tabla de hechos resultante de la información que brindan las dimensiones, así como el cálculo de los indicadores para el proceso.

1.1. D_Tiempo

Contiene los niveles en tiempo, en referencia a las cotizaciones generadas. Servirán de filtro para los análisis a realizar y generar.

Sentencia SQL de Carga: Datos extraídos de la tabla que alberga la información de los casos a cotizar bajo con la tasa Predictor. Genera su código correlativo, así como una descripción del mismo en diferentes maneras.

```

Select Distinct
Cast(Cast(Cast(FechaEnvio As Int) As Varchar(8)) As Varchar(8)) CodTiempo,
CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120) Fecha,
Cast(SUBSTRING(Cast(Cast(FechaEnvio As Int) As Varchar(8)),1,4) As Int) Año,
DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) Mes,
DATENAME(weekday, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) Dia,
Case
  When DATENAME(weekday, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'MONDAY' Then 'Lunes'
  When DATENAME(weekday, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'TUESDAY' Then 'Martes'
  When DATENAME(weekday, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'WEDNESDAY' Then 'Miércoles'
  When DATENAME(weekday, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'THURSDAY' Then 'Jueves'
  When DATENAME(weekday, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'FRIDAY' Then 'Viernes'
  When DATENAME(weekday, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'SATURDAY' Then 'Sábado'
  When DATENAME(weekday, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'SUNDAY' Then 'Domingo'
End+', '+
Cast(DATENAME(Day, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) As Varchar(2))+ ' '+
Case
  When DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'JANUARY' Then 'Enero'
  When DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'FEBRUARY' Then 'Febrero'
  When DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'MARCH' Then 'Marzo'
  When DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'APRIL' Then 'Abril'
  When DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'MAY' Then 'Mayo'
  When DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'JUNE' Then 'Junio'
  When DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'JULY' Then 'Julio'
  When DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'AUGUST' Then 'Agosto'
  When DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'SEPTEMBER' Then 'Septiembre'
  When DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'OCTOBER' Then 'Octubre'
  When DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'NOVEMBER' Then 'Noviembre'
  When DATENAME(Month, CONVERT(DateTime, Cast(Cast(FechaEnvio As Int) As Varchar(8)),120)) = 'DECEMBER' Then 'Diciembre'
End+' de '+SUBSTRING(Cast(Cast(FechaEnvio As Int) As Varchar(8)),1,4) DescFecha
From MercadoRV.dbo.RV_CasosEnviadosPredictor
Where IsNull(Cast(Cast(FechaEnvio As Int) As Varchar(8)), '') <> ''

```

1.2. D_Cierres

Contiene la información sobre el cierre de las solicitudes en su etapa final.

Sentencia SQL de Carga: Se extrae de la base transaccional todos los cierres que se han obtenido en lo que va el año en curso.

Contempla datos importantes como compañía que ganó la solicitud de jubilación, la tasa de venta ganadora, la fecha de cierre, etc.

```
Declare @AnioActual Varchar(4)
Set @AnioActual = SUBSTRING(Convert(Varchar(8),GetDate()),112),1,4)

Select
nro_solicitud,
compañia,
fecha_cotizacion,
fecha_cierre,
tasa_soles,
cic_bruto,
cic_netto
From Comercial.dbo.vMELER_cierresXcompañia_corto
Where Substring(Convert(Varchar(8),fecha_cierre,112),1,4) = @AnioActual
```

1.3. D_Ubigeo

Contiene información sobre la ubicación geográfica en la cual una persona afiliada se encuentra. Lugar de residencia de la persona. Se conecta con la dimensión Persona, brindándole la información antes mencionada.

Sentencia SQL de Carga: Extrae información de la tabla de ubicación geográfica (ubigeo) de la base de datos transaccional para enlazarse con la dimensión Persona y brindar mayor descripción de la ubicación geográfica de la persona.

```
Select
Distinct
cod_comuna,
cod_ciudad,
gls_departamento,
gls_provincia,
gls_distrito,
gls_larga_ubigeo
From exp_serv.dbo.t_ubigeo
Where cod_comuna is not null
And cod_ciudad is not null
```

1.4. D_Persona

Contiene los datos personales y de contacto de las personas que accedieron a una jubilación, con información obtenida de la dimensión de Cierres, creada anteriormente.

Sentencia SQL de Carga: De las personas que concretaron un cierre, se obtiene sus datos personales como son CUSPP, documento de identificación, nombres, dirección, teléfono, correo electrónico, entre otros.

```
Select Distinct
Persona.num_cuissp Num_Cuspp,
Persona.cod_tipo_documento TipoDocumento,
Persona.rut_persona NumeroDocumento,
ape_paterno,
ape_materno,
nom_nombre Nombre,
cod_sexo Sexo,
fec_nacimiento FechaNacimiento,
gls_direccion Direccion,
Case When Ubigeo.cod_comuna IS null then '-'
else Ubigeo.cod_comuna
End Cod_Comuna,
Case When Ubigeo.cod_ciudad IS null then '-'
else Ubigeo.cod_ciudad
End Cod_Ciudad,
num_telefono,
gls_mail Email
From exp_serv.dbo.rvi_bd_ini Persona
Inner Join (Select Cierres.num_cuspp
From MercadoRV.dbo.Predictor_Access Access
Left Outer Join Comercial.dbo.vMELEL_cierresXcompañia_corto Cierres
On Ltrim(Rtrim(Cierres.Nro_Solicitud)) = Ltrim(Rtrim(Substring(Access.num_operacion,1,6)))
Where FechaProceso <> '' ) Predictor
On LTRIM(RTRIM(Predictor.num_cuspp)) = LTRIM(RTRIM(Persona.num_cuissp))
Left Outer Join Datamart_Predictor.dbo.D_Ubigeo Ubigeo
On Persona.cod_comuna = Ubigeo.cod_comuna
And Persona.cod_ciudad = Ubigeo.cod_ciudad
Where LEN(num_cuspp) = 12
And IsNull(num_cuissp,'') <> ''
And Persona.num_cuissp <> '000000000000'
```

1.5. D_Compañía

Contiene los datos de las compañías aseguradoras que componen el sistema asegurador, las cuales envían sus cotizaciones ante la recepción de nuevas solicitudes de jubilación.

Sentencia SQL de Carga: La sentencia toma como referencia la tabla de cotizaciones para obtener todas las empresas que han generado cotizaciones en el último mes móvil, asignándole un código correlativo.

```
SELECT ROW_NUMBER()  
      OVER (ORDER BY x.Compañia Desc) As CodCompañia,  
      x.compañia As DescCompañia  
FROM (SELECT DISTINCT compañia  
      FROM RV.DBO.vRV_Predictor_Tasas) x
```

1.6. D_CierreCotización

Contiene el detalle de las cotizaciones generadas hacia un cliente (afiliado), estos datos son necesarios para la evaluación a realizar por los árboles de decisión.

Sentencia SQL de Carga: Considera toda la información necesaria para saber qué producto y tipo de modalidad de pensión desea el afiliado, además de ello, muestra información importante y necesaria para el árbol de decisión. Por política de la empresa aseguradora, sólo se puede ofrecer una venta de Renta Vitalicia a personas que cuentan con un CIC bruto mayor o igual a S/. 60,000.

```

SELECT DISTINCT
prop.num_operacion, prop.num_cuspp,
CASE prop.cod_categoria WHEN 'A' THEN 'EDAD LEGAL'
WHEN 'N' THEN 'REJA'
WHEN 'C' THEN 'INVALIDEZ'
WHEN 'D' THEN 'SOBREVIVENCIA'
WHEN 'F' THEN 'JUB ANT ALTO RIESGO'
WHEN 'G' THEN 'JUB ANT ALTO RIESGO'
ELSE 'JUB ANT ORDINARIA' END AS categoria,
prop.cod_modalidad,
prop.[Periodo Temporal] AS periodo_temporal,
prop.[Porcentaje Pension] AS pje_pension,
prop.val_periodo_garantizado,
prop.pje_cobertura_conyuge,
prop.solici_gls_afp AS afp,
prop.ClienteGeo_lp AS zona,
solici.cod_genero AS sexo,
CASE WHEN prop.fec_nacimiento IS NULL THEN NULL
ELSE DATEDIFF([month], CASE WHEN prop.fec_nacimiento < getdate() THEN prop.fec_nacimiento
ELSE getdate()
END, DATEADD([day], DATEDIFF([day], 0, GETDATE()), 0)) / 12
END AS tit_edad,
prop.FechaCierre AS fec_cierre,
prop.[Fecha de plazo] AS fec_plazo,
Ubigeo.gls_departamento AS Departamento,
Zonificacion.laslomas_zonificacion AS grp_zonas_geografica
FROM Cotizaciones.dbo.CotizaCierre_cia_promax AS prop
LEFT OUTER JOIN exp_serv.dbo.sbs_solici_meler AS solici
ON prop.num_operacion = solici.num_operacion COLLATE Latin1_General_CI_AS
LEFT OUTER JOIN exp_serv.dbo.t_ubigeo AS Ubigeo
ON prop.cod_ubigeo = Ubigeo.cod_ubigeo COLLATE Latin1_General_CI_AS
LEFT OUTER JOIN LasLomas.dbo.LasLomas_Zonificacion AS Zonificacion
ON prop.cod_ubigeo = Zonificacion.cod_ubigeo COLLATE Latin1_General_CI_AS
LEFT OUTER JOIN Datamart_Predictor.dbo.D_Cierres AS Cierre
ON prop.num_operacion = Cierre.num_Solicitud
WHERE (prop.CIC >= 60000) AND (prop.IV_prod_FondoTx / prop.FondoTx > 0)
And Cierre.num_Solicitud Is Not Null

```

1.7. D_CotizacionesPredictor

Contiene los datos de las cotizaciones hechas por las diferentes compañías del mercado asegurador hacía una persona que solicitó jubilación.

Sentencia SQL de Carga: Se consideran tablas del Data Mart llenas anteriormente, como tablas de bases de datos transaccionales. Genera la información de todas las cotizaciones que ha recibido un afiliado por parte de las empresas aseguradoras, donde se especifica la tasa de venta ofrecida, el CIC neto a obtener, entre otros datos importantes.

```

Select
Compañía.CodCompañía,
prop.num_operacion,
prop.TV_prod_FondoTx / prop.FondoTx AS TV,
prop.TR_prod_FondoTx / prop.FondoTx AS TR,
prop.ind_gana,
prop.CIC AS cic_bruto,
prop.FondoTx AS cic_netto,
prop.moneda
FROM      Cotizaciones.dbo.CotizaCierre_cia_promax AS prop
          LEFT OUTER JOIN exp_serv.dbo.sbs_solici_meler AS solici
            ON prop.num_operacion = solici.num_operacion COLLATE Latin1_General_CI_AS
          LEFT OUTER JOIN exp_serv.dbo.t_ubigeo AS Ubigeo
            ON prop.cod_ubigeo = Ubigeo.cod_ubigeo COLLATE Latin1_General_CI_AS
          LEFT OUTER JOIN LasLomas.dbo.LasLomas_Zonificacion AS Zonificacion
            ON prop.cod_ubigeo = Zonificacion.cod_ubigeo COLLATE Latin1_General_CI_AS
          LEFT OUTER JOIN Datamart_Predictor.dbo.D_CierreCotizacion AS Cierre
            ON prop.num_operacion = Cierre.num_Solicitud
          LEFT OUTER JOIN Datamart_Predictor.dbo.D_Compañía AS Compañía
            ON prop.Entidad = Compañía.DescCompañía
WHERE     (prop.CIC >= 60000) AND (prop.TV_prod_FondoTx / prop.FondoTx > 0)
          And Cierre.num_Solicitud Is Not Null
          And Compañía.DescCompañía Is Not Null

```

1.8. D_Predictor

Contiene los registros seleccionados para el cálculo de la tasa de venta por medio de la predicción. Es utilizada para la medición del proceso.

Sentencia SQL de Carga: Extrae la información de la tabla que guarda los casos enviados por la gerencia de Rentas Vitalicias, así como de la tabla que guarda los datos del proceso de generación de tasa predictor. Es utilizada como dimensión base para el llenado de la tabla de hechos.

```

Select Distinct
Access.Sexo,
Access.tit_fec_nacimiento,
Casos.Compañia,
Casos.Moneda,
Casos.Categoria,
Casos.afp,
Access.Zona,
Access.cod_modalidad,
Access.pje_pension,
Access.val_periodo_garantizado,
Access.periodo_temporal,
Access.fec_plazo,
Access.fec_devengue,
Access.conyuge_fec_nacimiento,
Access.Condicion_Salud_Conyuge,
Access.isBeneficiario,
Access.hijos_q,
Casos.ACOM,
Casos.DCOM,|
Casos.TRA,
Casos.IV_Cotizador,
Access.fec_control,
Access.TasaInversa_Pacifico,
Access.TasaInversa_Rimac,
Access.TasaInversa_Sura,
Access.TasaInversa_LPV,
Access.DifPuntosBasicos,
Access.TasaMELER_Pacifico,
Access.TasaMELER_Rimac,
Access.TasaMELER_Sura,
Access.TasaMELER_LPV,
Access.CompañiaComparar,
Access.VariacionTasa,
Access.TasaMelerCia_Variada,
Access.FlagExcedeTasaMaxima,
Access.TasaMeler_FlagTasaMaxima,
Access.DeltaTasa,
Access.FlagTra,
Access.TRA_Maximo,
Access.ResultadoDeltaTasa,
Access.DeltaTasa_Final,
Access.Tasa_Meler_Predictor,
Cast(Cast( Cast(Casos.FechaEnvio As Int) As Varchar(8)) As Varchar(8)) FechaProceso,
Access.FlqNivelConfianza,
Casos.NroSolicitud Num_Operacion
From MercadoRV.dbo.RV_CasosEnviadosPredictor Casos
Left Outer Join MercadoRV.dbo.Predictor_Access Access
On Casos.NroSolicitud = Access.num_operacion
Where Casos.FechaEnvio is not null

```

1.9. F_ResultadoPredictor

Permite mostrar los resultados de los cierres de los prospectos a los cuales se les generó tasa predictor y a los que no se les generó; haciendo su respectiva comparación y medición.

Sentencia SQL de Carga: Extrae la información proveniente de las dimensiones ya llenadas de información previamente, de esta información, genera los indicadores y cantidades necesarias para la realización de los distintos reportes que cubren los requerimientos funcionales.

```

Select
Predictor.num_operacion num_operacion,
Tiempo.CodTiempo,
IsNull(Persona.Num_Cuspp,'00000000000') Num_Cuspp,
Case When Cierre.num_Solicitud Is Null Then 0
Else Cierre.num_Solicitud
End num_Solicitud ,
Cast('1' As Int) FlagCotiza,
Case When Cierre.num_Solicitud is null Then 0
Else 1
End FlagCierre,
Case When Predictor.Tasa_Meler_Predictor Is Null Then 0
Else 1
End FlagPredictor,
Case When Cierre.num_Solicitud Is Not Null And CompañiaCierre = 'INTERSEGURO S.A.' Then 1
Else 0
End FlagGanaIS,
Case When Cierre.cic_bruto Is Null Then 0
Else Cierre.cic_bruto
End cic_bruto,
Case When Cierre.cic_neto Is Null Then 0
Else Cierre.cic_neto
End cic_neto,
Case When Cierre.FechaCotizacion Is Null Then 0
Else Cierre.FechaCotizacion
End FechaCotizacion,
Case When Cierre.FechaCierre Is Null Then 0
Else Cierre.FechaCierre
End FechaCierre,
Case When Cierre.FechaCierre Is Null Then 0
Else DateDiff(day,Cierre.FechaCotizacion, Cierre.FechaCierre)
End CantidadDiasCierre
From Datamart_Predictor.dbo.D_Predictor Predictor
Left Outer Join Datamart_Predictor.dbo.D_Tiempo Tiempo
On Ltrim(Rtrim(Predictor.FechaProceso)) = Ltrim(Rtrim(Tiempo.CodTiempo))
Left Outer Join (Select a.num_Solicitud, b.Num_Cuspp
From Datamart_Predictor.dbo.D_CierreCotizacion a
Inner Join Datamart_Predictor.dbo.D_Persona b
On a.num_cuspp = b.Num_Cuspp) Persona
On Ltrim(Rtrim(Persona.Num_Cuspp)) = LTRIM(Rtrim(SUBSTRING(Predictor.num_operacion,1,6)))
Left Outer Join Datamart_Predictor.dbo.D_Cierres Cierre
On Ltrim(Rtrim(Cierre.num_Solicitud)) = LTRIM(Rtrim(SUBSTRING(Predictor.num_operacion,1,6)))

```

