



**UNIVERSIDAD CATÓLICA LOS ÁNGELES
CHIMBOTE**

**FACULTAD DE INGENIERÍA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

TITULO:

**PROPUESTA DE UNA METODOLOGÍA DE EXTRACCIÓN DE
CONOCIMIENTOS A PARTIR DE DATOS DE LAS
PRESTACIONES DEL SEGURO INTEGRAL DE SALUD EN LA
REGIÓN PIURA EN EL AÑO 2016.**

**TESIS PARA OPTAR EL GRADO ACADÉMICO DE MAESTRO
EN INGENIERÍA DE SISTEMAS, CON MENCIÓN EN
TECNOLOGÍA DE INFORMACIÓN Y COMUNICACIÓN**

AUTOR

ING. GUTIÉRREZ PACHERRES, JUAN JESÚS

ASESOR

MGTR. ING. NOÉ GREGORIO SILVA ZELADA

PIURA – PERÚ

2017

JURADO EVALUADOR Y ASESOR DE TESIS

DR. ING. VÍCTOR ÁNGEL ANCAJIMA MIÑAN

Presidente

MGTR. ING. MARLENY SERNAQUE BARRANTES

Secretaria

MGTR. ING. OSWALDO CORONADO ZULOETA

Miembro

MGTR. ING. NOE GREGORIO SILVA ZELADA

Asesor

AGRADECIMIENTO

A Dios.

Por haberme permitido llegar hasta este punto y haberme dado salud para lograr mis objetivos, además de su infinita bondad y amor.

A mi madre Olga.

Por haberme apoyado en todo momento, por sus consejos, sus valores, por la motivación constante que me ha permitido ser una persona de bien, pero más que nada, por su amor.

A mi padre Martín.

Por los ejemplos de perseverancia y constancia que lo caracterizan y que me ha infundado siempre, por el valor mostrado para salir adelante y por su amor.

A mi hermano Rodolfo y hermanos en Cristo.

Por estar ahí hermano, sabiendo que estamos juntos, una sola familia. A mis compañeros, amigos, hermanos en Cristo siempre os llevo presentes.

¡Gracias, a nuestro asesor de tesis y coordinador de maestría!

DEDICATORIA

Dedico esta tesis a DIOS, a San Martín de Porres, patrono de la justicia social y a la Virgen María, quienes inspiraron mi espíritu para la conclusión de esta tesis de maestría, en Ingeniería de Sistemas.

A mis padres Martín y Olga, hermano José Rodolfo, quienes me dieron y dan vida, educación, apoyo y consejos.

A mis compañeros de estudio, a mis maestros y amigos, quienes sin su ayuda nunca hubiera podido hacer esta tesis.

A ti Amparo quien me apoya y alienta para continuar, por nuestra vida MH.

A todos ellos se los agradezco desde el fondo de mi alma. Para todos ellos hago esta dedicatoria.

RESUMEN

El presente trabajo tiene como base la línea de investigación de la Universidad Católica Los Ángeles de Chimbote para la escuela de Ingeniería de Sistemas que menciona la Implementación de las Tecnologías de Información y Comunicaciones (TIC) para la mejora continua de la calidad en las organizaciones del Perú, de la Escuela Profesional de Ingeniería de Sistemas de la Universidad los Ángeles de Chimbote (ULADECH CATÓLICA). El objetivo principal es desarrollar una metodología de extracción de conocimientos que permita descubrir conocimientos relevantes, partiendo de datos de prestaciones del Seguro Integral de Salud en la Región Piura, En la actualidad, el uso de técnicas de minería de datos forman parte del análisis de grandes cantidades de datos da lugar a la revelación, a la adquisición de conocimiento y patrones de comportamiento útil para la toma de decisiones, es un proceso mediante el cual se realiza una elección entre las opciones o formas para resolver situaciones de la vida en diferentes contextos. En el presente trabajo de tesis se emplean el tipo de investigación descriptiva, no experimental y cuantitativa. Se determinó aplicar como instrumento, un cuestionario para obtener conocimiento sobre prestaciones, patologías, insumos o medicamentos en relación a las prestaciones brindadas a los/las afiliadas (as) al Seguro Integral de Salud (SIS) de la Región Piura en un determinado tiempo, aplicada a la Población y muestra de 08 trabajadores relacionados con el uso de las (TIC), al disponer de este conocimiento empleando técnicas de minería de datos permite contar con datos necesarios para establecer relaciones que ayude en la toma de decisiones con respecto a las prestaciones brindadas en las Instituciones Prestadoras de Servicios de Salud (IPRESS). La serie de datos utilizadas corresponden a los registros de los Formatos Únicos de Atención (FUA) del SIS de prestaciones brindadas a los/las afiliadas (as). El análisis de las prestaciones de la región con base a esta serie de datos es importante, puesto que frecuentemente solo se hacen análisis locales cuantificando datos, lo cual es adecuado, pero no brinda un horizonte, obteniéndose los siguientes resultados: el 75% trabajadores estuvo de acuerdo con la implementación de la propuesta de mejora. Un 63% opino que totalmente la metodología optimizaría el tiempo de trabajo, también 88% de trabajadores les sería totalmente útil el uso de la herramienta, Solo un 50% opino que podría el conocimiento obtenido ayudar a optimizar procesos y un 75% opino que contar con el conocimiento obtenido ayudaría a las autoridades/directivos a la toma de decisiones.

Palabras Clave: Minería de Datos, Servicio de Salud, conocimiento sobre prestaciones.

ABSTRACT

This paper is based on the research line of the Catholic University Los Angeles of Chimbote for the school of Systems engineering that mentions the implementation of information and communications technologies (ICT) for the continuous improvement of the Quality in the organizations of Peru, of the professional School of Systems Engineering of the university Los Angeles of Chimbote (Catholic ULADECH). The main objective is to develop a methodology for the extraction of knowledge that allows to discover relevant knowledge, starting from data of benefits of the Integral health insurance in the region Piura, at the present time, the use of data mining techniques are part of the analysis of large quantities of data gives place to The revelation, to the acquisition of knowledge and patterns of behavior useful for the decision making, is a process by which a choice is made between the options or ways to solve situations of life in different contexts. In this thesis work, the type of descriptive, non-experimental and quantitative research is used. It was determined to apply as an instrument, a questionnaire to obtain knowledge about benefits, pathologies, inputs or medicines in relation to the benefits provided to the members of the Integral Health Insurance (SIS) of the Region Piura in a certain time, applied to the population and sample of 08 workers related to the use of the (ICT), to have this knowledge using techniques of data mining allows to have data necessary to establish relationships that Help in decision making regarding benefits Provided in the health services Institutions (IPRESS). The series of data used correspond to the records of the unique care formats (FUA) of the SIS of benefits provided to the affiliates. The analysis of the benefits of the region based on this series of data is important, since it is often only Local analyses are made by quantifying data, which is adequate, but does not provide a horizon, obtaining the following results: 75% workers agreed with the implementation of the improvement proposal. 63% I think that totally the methodology would optimize the working time, also 88% of workers would be totally useful to use the tool, only 50% I think that could the knowledge to help optimize processes and a 75% think I think that having the knowledge obtained would help the authorities/managers to make decisions.

Key Words: Data Mining, Health Service, knowledge about benefits.

ÍNDICE

HOJA DE FIRMA DEL JURADO Y ASESOR	ii
AGRADECIMIENTO	iii
DEDICATORIA	iv
RESUMEN	v
ABSTRACT	vi
ÍNDICE DE CONTENIDO	vii
ÍNDICE DE FIGURAS	ix
ÍNDICE DE TABLAS	x
ÍNDICE DE GRÁFICOS	xii
I. INTRODUCCIÓN	1
II. MARCO TEÓRICO	4
2.1.ANTECEDENTES.....	4
2.1.1 ANTECEDENTES A NIVEL INTERNACIONAL	4
2.1.2 ANTECEDENTES A NIVEL NACIONAL	5
2.1.2 ANTECEDENTES A NIVEL LOCAL.....	7
2.2 BASES TEÓRICAS RELACIONADAS CON EL ESTUDIO.....	7
2.3 HIPÓTESIS.....	30
2.4 VARIABLES	30
III. METODOLOGIA	31
3.1 EL TIPO Y EL DISEÑO DE LA INVESTIGACIÓN	31
3.2 DISEÑO DE LA INVESTIGACIÓN.....	31
3.3 POBLACIÓN Y MUESTRA	31
3.4 DEFINICIÓN Y OPERACIONALIZACIÓN DE LAS VARIABLES Y LOS INDICADORES.....	32
3.5 TÉCNICAS E INSTRUMENTOS	33
3.6 PLAN DE ANÁLISIS	33
3.7 VALIDEZ DEL INSTRUMENTO	34
4 RESULTADOS	35
4.1 RESULTADOS.....	35
4.2 ANÁLISIS DE RESULTADOS DE ALGORITMOS	45
4.2.1.Reglas de Asociación:	45
4.2.2.Bayes Naive.....	45
4.2.3.Clústeres	45
4.2.4.Árbol de Decisión.....	45
4.3. ANÁLISIS DE RESULTADOS DE INSTRUMENTO DE EVALUACIÓN	46
4.4. PROPUESTA MEJORA	52

4.5. DISEÑO DE LA METODOLOGÍA PROPUESTA	52
4.6. CONCLUSIONES	53
4.7. RECOMENDACIONES	54
5. REFERENCIAS BIBLIOGRAFICAS	55
ANEXOS.....	58

ÍNDICE DE FIGURAS

FIGURA 1: WHERE ANALYTICS, DATA MINING, DATA SCIENCE WERE APPLIED IN 2016. / DONDE SE APLICÓ ANÁLISIS, MINERÍA DE DATOS, CIENCIA DE DATOS, EN 2016.....	16
FIGURA 2: METODOLOGÍAS UTILIZADAS EN DATA MINING ()	18
FIGURA 3: FASE DE COMPRESIÓN DEL NEGOCIO (CRISP-DM, 2000).....	20
FIGURA 4: FASE DE COMPRESIÓN DE LOS DATOS (CRISP – DM, 2000)	21
FIGURA 5: FASE DE PREPARACIÓN DE LOS DATOS (CRISP-DM, 2000).....	23
FIGURA 6: FASE DE MODELADO (CRISP-DM, 2000).....	24
FIGURA 7: FASE DE EVALUACIÓN (CRISP-DM, 2000)	26
FIGURA 8: FASE DE IMPLEMENTACIÓN (CRISP-DM, 2000)	28
FIGURA 9: BASE DE DATOS SIS – ESTRUCTURA DE LA TABLA ATENCIÓN.....	36
FIGURA 10: BASE DE DATOS SIS – ESTRUCTURA DE LA TABLA DIAGNÓSTICO.	36
FIGURA 11: BASE DE DATOS SIS – CONSULTA SQL DE AGRUPACIÓN.	37
FIGURA 12: BASE DE DATOS – MINERIA_DATOS.....	37
FIGURA 13: ESTRUCTURA DE LA TABLA PARA APLICACIÓN DE MINERÍA DE DATOS.	38
FIGURA 14: PROCESAMIENTO DE VARIABLES.	38
FIGURA 15: SOFTWARE DE APLICACIÓN MINERÍA DE DATOS: RAPIDMINER.....	39
FIGURA 16: CONSTRUCCIÓN DE MODELO – REGLAS DE ASOCIACIÓN (SOFTWARE RAPIDMINER).	39
FIGURA 17: RESULTADOS DE APLICACIÓN DE MODELO DE MINERÍA DE DATOS.....	40
FIGURA 18: CONSTRUCCIÓN DE MODELO – BAYES NAIVE (SOFTWARE RAPIDMINER).....	40
FIGURA 19: CONSTRUCCIÓN DE MODELO - CLUSTERING (SOFTWARE RAPIDMINER).	41
FIGURA 20: CONSTRUCCIÓN DE MODELO – ÁRBOL DE DECISIÓN (SOFTWARE RAPIDMINER). ..	42
FIGURA 21: EVALUACIÓN DEL MODELO DE REGLAS DE ASOCIACIÓN (SOFTWARE RAPIDMINER).	43
FIGURA 22: EVALUACIÓN DEL MODELO DE BAYES NAIVE (SOFTWARE RAPIDMINER).	43
FIGURA 23: EVALUACIÓN DEL MODELO DE CLUSTERING (SOFTWARE RAPIDMINER).	44
FIGURA 24: EVALUACIÓN DEL MODELO ÁRBOL DE DECISIÓN (SOFTWARE RAPIDMINER).....	44

ÍNDICE DE TABLAS

TABLA 1: CUADRO COMPARATIVO DE METODOLOGÍA DE MINERÍA DE DATOS.....	29
TABLA 2: TABLA RESUMEN - VALIDEZ DE INSTRUMENTO	34
TABLA 3: MATRIZ DE CONSISTENCIA	59
TABLA 4: CUADRO DE LA PROPUESTA DE LA METODOLOGÍA	60
TABLA 5: FASE 1 CRISP-DM: ENTENDIMIENTO DEL NEGOCIO.....	60
TABLA 6: FASE 2 CRISP-DM: ENTENDIMIENTO DE LOS DATOS	60
TABLA 7: FASE 3 CRISP-DM: PREPARACIÓN DE LOS DATOS	60
TABLA 8: FASE 4 CRSIP-DM: MODELADO	60
TABLA 9: FASE 5 CRISP-DM: EVALUACIÓN.....	60
TABLA 10: FASE 6 CRISP-DM: IMPLEMENTACIÓN	60
TABLA 11: FASES DE LA METODOLOGÍA PROPUESTA.....	60
TABLA 12: CUADRO DE TABULACIONES DE ENCUESTA.....	60

ÍNDICE DE GRÁFICOS

GRÁFICO 1: GRÁFICO DE LA PREGUNTA N° 01.....	46
GRÁFICO 2: GRÁFICO DE LA PREGUNTA N° 02.....	47
GRÁFICO 3: GRÁFICO DE LA PREGUNTA N° 03.....	47
GRÁFICO 4: GRÁFICO DE LA PREGUNTA N° 04.....	48
GRÁFICO 5: GRÁFICO DE LA PREGUNTA N° 05.....	49
GRÁFICO 6: GRÁFICO DE LA PREGUNTA N° 06.....	49
GRÁFICO 7: GRÁFICO DE LA PREGUNTA N° 07.....	50
GRÁFICO 8: GRÁFICO DE LA PREGUNTA N° 08.....	51

I. INTRODUCCIÓN

La Oficina de Informática del Equipo de Seguros de Salud de la Dirección Regional de Salud Piura se caracteriza siempre por la ejecución de políticas innovadoras para la coordinación de sus procesos. Los procesos que se llevan a cabo dentro de la oficina abarcan todos los aspectos del procesamiento de información, estadísticas, indicadores de gestión y análisis en consistencias y validación, por ello varían en su naturaleza y a la vez varía el tipo de equipos que intervienen como apoyo tecnológico para la producción. Esto hace que en la oficina de informática traten de buscar herramientas que, basadas en principios permitan orientar eficientemente el proceso de toma de decisiones asociado con la gestión de la producción de estadísticas e información.

Cada día se genera una gran cantidad de información, algunas veces conscientes de que lo hacemos y otras veces inconscientes de ello porque lo desconocemos. Nos damos cuenta de que generamos información cuando registramos nuestra entrada en el trabajo, cuando entramos en un servidor para ver nuestro correo, cuando pagamos con una tarjeta de crédito o cuando reservamos un boleto de avión. Otras veces no nos damos cuenta de que generamos información, como cuando conducimos por una vía donde están contabilizando el número de automóviles que pasan por minuto, cuando se sigue nuestra navegación por Internet o cuando nos sacan una fotografía del rostro al haber pasado cerca de una oficina gubernamental o un banco.

¿Con qué finalidad queremos generar información? Son muchos los motivos que nos llevan a generar información, ya que nos pueden ayudar a controlar, optimizar, administrar, examinar, investigar, planificar, predecir, negociar o tomar decisiones de cualquier ámbito según el dominio en que nos desarrollemos. La información por sí misma está considerada un bien patrimonial. De esta forma, si una empresa tiene una pérdida total o parcial de información provoca bastantes perjuicios. Es evidente que la información debe ser protegida, pero también explotada.

Actualmente todas estas ventajas nos han llevado a abusar del almacenamiento de la información en las bases de datos. Podemos decir que las empresas almacenan datos, pero no todas hacen uso de ellos. En cambio, si estos datos son leídos y analizados pueden proporcionar en conjunto un verdadero conocimiento que ayude en la toma de decisiones. Así pues, es necesario contar con tecnologías que nos ayuden a explotar el potencial de este tipo de datos.

La disciplina denominada Minería de Datos estudia métodos y algoritmos que permiten la extracción automática de información sintetizada que permite caracterizar las relaciones escondidas en la gran cantidad de datos; también se pretende que la información obtenida posea capacidad predictiva, facilitando el análisis de los datos de forma eficiente. Bajo la denominación de "minería de datos" se han agrupado diversas técnicas estadísticas y del aprendizaje automático (Inteligencia Artificial) enfocadas, principalmente, a la visualización, análisis, y modelización de información de bases de datos masivas.

Como parte del proceso de descubrimiento de conocimiento, se requiere la selección de un método y de una técnica de minería de datos apropiada. Además, se debe determinar la estrategia para implementar la técnica de tal forma que sea eficiente teórica y empíricamente.

El Seguro Integral de Salud (SIS) actualmente un organismo ejecutor que tiene como finalidad proteger la salud de los peruanos que no cuentan con un seguro de salud, priorizando la afiliación de aquella población más vulnerable y que se encuentra en situación de pobreza o pobreza extrema. Para el Aseguramiento Universal en Salud, el SIS es una IAFAS, es decir, administra los fondos destinados a cubrir distintas afiliaciones SUSALUD (1).

La Dirección Regional de Salud Piura tiene la Misión de proteger la dignidad personal, promoviendo la salud para construir una cultura de salud y de solidaridad, previniendo las enfermedades y garantizando la atención integral de salud de todos los habitantes; cumpliendo las políticas y objetivos nacionales de salud en concertación con todos los sectores públicos y privados y otros actores sociales. PORTAL PERUANO DEL ESTADO, SUSALUD (2,3).

En el presente trabajo ambos organizamos interactúan de la siguiente manera la Unidad Desconcentrada Regional (UDR) representa al Seguro Integral de Salud en la Región Piura, así mismo, la Oficina de Informática del Equipo de Seguros en la Dirección Regional de Salud se encarga de procesar la data de prestaciones que se brindan a través del Seguro Integral de Salud en los establecimientos de salud del Gobierno Regional de Piura.

En este ámbito en los últimos años, los datos ingresados a partir de las prestaciones del Seguro Integral de Salud en la base de datos han ido incrementándose

considerablemente. Esta información, de gran importancia estratégica para el ámbito sanitario, en el presente año se accede a través del uso de técnicas clásicas como son sentencias SQL y los procedimientos almacenados. Por tal motivo existe demora en la recuperación y el análisis de la información para la elaboración de informes, formularios y reportes de gestión solicitados por la alta dirección.

Así es necesaria la utilización de métodos analíticos más avanzados, como es la minería de datos para la explotación de datos con la finalidad de ayudar a los miembros de la alta dirección a la toma de decisiones.

Los registros pertenecen a las atenciones brindadas por las IPRESS (Instituciones Prestadoras de Servicios de Salud) es decir los centros y postas de salud del estado, a los pacientes afiliados al Seguro Integral de Salud (SIS), dichos registros ingresados a los aplicativos con el instrumento del FUA (Formato Único de Atención).

El presente proyecto planteó el desarrollo de una metodología de extracción de conocimientos a partir de datos de las prestaciones del Seguro Integral de Salud para caracterizar, describir o descubrir el comportamiento en los registros, los cuales se encuentran bajo la responsabilidad del Equipo de Seguros de Salud de la Dirección Regional de Salud Piura, esto nos conlleva al objetivo que la metodología mejorara el proceso para la toma de decisiones correctas en el momento adecuado.

II. MARCO TEÓRICO

2.1. ANTECEDENTES

2.1.1. ANTECEDENTES A NIVEL INTERNACIONAL

Yadira Robles Aranda; Anthony R. Sotolongo. " Integración de los Algoritmos de Minería de Datos 1r, Prism E Id3 A PostgreSQL" (2013). Revista de Gestión de la Tecnología de Sistemas de Información. Universidad de Ciencias Informáticas. La Habana – Cuba.

Resumen: El trabajo de referencia se analizaron las técnicas de minería de datos de árboles de decisión y de reglas de inducción para integrar varios de sus algoritmos al sistema gestor de base de datos (SGBD) PostgreSQL, buscando suplir las deficiencias de las herramientas libres existentes. También se propuso un mecanismo para optimizar el rendimiento de los algoritmos implementados con el objetivo de aprovechar las ventajas de PostgreSQL y se comprobó, mediante un experimento, que, al utilizar los algoritmos integrados al gestor, los tiempos de respuestas y los resultados obtenidos son superiores ROBLES, A. (4).

Correlación: Del presente trabajo, fijaremos las variables relacionadas a la minería de datos, las cuales establecen dependientes e independientes, trabajo similar con el procesamiento de la herramienta de PostgreSQL.

Padua, Luciana María. "Comparación interactiva de modelos de minería de datos utilizando técnicas de visualización" (2014)

Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires

Resumen: El trabajo de referencia presenta una metodología la cual genera de forma automática varios árboles de decisión, el cual es un algoritmo de minería de datos. El objetivo de este consiste en contribuir a la creciente disciplina de Visual Analytics, la cual es en una combinación de técnicas de Visualización de la Información y Minería de Datos con el fin de involucrar a las personas en los procesos de aprendizaje automático. El enfoque presentado se basa en herramientas disponibles en el mercado, sugiriendo la adaptación de estas para lograr los objetivos de interacción del usuario. Con el fin de demostrar la potencialidad de la metodología se presenta al finalizar un caso de uso real de una empresa de Telecomunicaciones PADUA, L. (5).

Correlación: El trabajo presenta una metodología de aplicación de minería de datos con la técnica de Árboles de decisión, así mismo, presenta una metodología para generar el algoritmo de forma automática. El autor considera PADUA, L. (5) el software SAS Enterprise Mine INC (6) para aplicar la técnica en la visualización de la aplicación de minería de datos.

Flores Cruz, José Antonio. "Data Warehouse y minería de datos como alternativas al análisis de datos forestales" (2014)

Instituto de Enseñanza e Investigación en Ciencias Agrícolas, Montecillo, estado de México.

Resumen: La Aplicación de Minería de Datos contribuyó al diseño de un almacén de datos o "Data Warehouse" para la construcción de cubos de análisis para volumen de madera, biomasa y carbono. Utilizando como software SQL Server 2008, en este trabajo se probaron cuatro modelos de minería de datos, para las diferentes clasificaciones de árboles, géneros maderables. Además, se programaron interfaces para la visualización y análisis de la información, algunas de ellas en Visual Studio, concluyendo que, si fue posible aplicar los modelos de minería de datos para la clasificación de los temas correspondientes CRUZ, J. (7).

Correlación: El siguiente trabajo presenta una metodología de aplicación de minería de datos inclusive se aplican cuatro algoritmos que demostraron para este caso que si se podía aplicar minería de datos a partir de base de datos de los años 2004-2009 de la institución que se presentó.

2.1.2 ANTECEDENTES A NIVEL NACIONAL

Rodríguez Cabanillas, Keller Gladys. "Análisis diseño e implementación de una solución de inteligencia de negocios para el área de compras y ventas de una empresa comercializadora de electrodomésticos" (2011).

Pontificia Universidad Católica del Perú. Lima, Perú

Resumen: La Aplicación de Minería de Datos, para el caso que se nos presenta como antecedente nacional, fue debido al crecimiento comercial de las empresas debido a esto se plantea como solución el uso de una herramienta de inteligencia de negocios que permita en tiempo real a los

gerentes y jefes de producto generar escenarios, como solución de Inteligencia de Negocios se diseña un Data Mart de Compras y un Data Mart de Ventas, luego se realizan los procesos de extracción, transformación y carga de datos, para finalmente explotar los datos mediante reportes que permitan hacer el análisis de la información KELLER, G. (8).

Correlación: El siguiente trabajo presenta una construcción de data Mart de ventas y compras, así mismo, la construcción de una solución de inteligencia de negocios, donde se decide usar las herramientas de software libre para base de datos y para los procesos de extracción y explotación por sus bajos costos. Así mismo, se diseña un modelo dimensional adecuado según la cantidad y profundidad de datos que posee cada Data Mart.

Alcántara Morí, Álvaro Alejandro. “Formulación de Minería de Datos para la Empresa Distribuidora de Productos Espinoza Aguilar S.A.” (2012). Universidad Tecnológica del Perú. Lima, Perú

Resumen: La Aplicación de Minería de Datos, para el caso fue un estudio de la viabilidad, adaptación y beneficios que ofreció la metodología de la minería de datos aplicado a la pequeña empresa, las cuales no cuentan con plan de proyección estructurado, de los análisis internos y externos, los cuales van cambiando durante el ciclo de vida de la empresa. Esta tesis llegó a abarcar las etapas de análisis y diseño y se utilizó la metodología RUP en combinación con UML, se diagramaron los casos de uso del negocio, los casos de uso en la aplicación de la metodología, los diagramas de estado, diagramas de secuencia, diagramas de clases y por último el prototipo de la aplicación resultante ALCANTARA, M. (9).

Correlación: El siguiente trabajo presento una formulación del sistema implantado que permitió conocer el impacto económico y en la alineación de los posesos al implantar un sistema de Minería de Datos en la logística de la empresa.

2.1.2 ANTECEDENTES A NIVEL LOCAL

Procesos de Información en relación con las prestaciones a los pacientes afiliados al Seguro Integral de Salud (SIS), a través de las Instituciones Prestadoras de Servicios de Salud (IPRESS), (2017). Dirección Regional de Salud. Piura, Perú

Para el presente estudio en la dependencia analizada de la Dirección Regional de Salud Piura lugar donde laboro profesionalmente como Ingeniero de Sistemas y la problemática planteada es la experiencia misma, de procesar información y querer profundizar a través de nuevos conocimientos la aplicación de Minería de Datos , en dicha dependencia el proceso de análisis de información era a nivel de consultas SQL tradicionales (reportes que se trabajaban en Hojas de Cálculo), cuando se propone implementar la metodología CRISP (Cross-Industry Standard Process for Data Mining - Proceso Estándar de Trans-Industria para la Minería de Datos) se obtiene acceso a grandes almacenes de datos (Base de Datos , Reportes, tablas dinámicas) y se formula cuestionarios para posteriormente verificar dicho proceso de implementación. Así mismo, con la propuesta de mejora: integrando a la metodología descrita el proceso de Elicitación de Requisitos del cual se obtuvo aceptación en la implementación de la metodología.

Cabe mencionar que se construyeron modelos de minería de datos para complementar la investigación de los cuales se procesaron los modelos de: Reglas de Asociación, Bayes Naive, Clústeres, Árbol de Decisión.

2.2 BASES TEÓRICAS RELACIONADAS CON EL ESTUDIO

A continuación, se presentan los principales conceptos relacionados al tema de investigación.

2.2.1 Minería De Datos

La minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiados datos MICROSOFT (10).

La extracción de información oculta y predecible de grandes bases de datos es una poderosa tecnología nueva con gran potencial para ayudar a las compañías a concentrarse en la información más importante de sus Bases de Información (Data Warehouse). Las herramientas de Data Mining predicen futuras tendencias y comportamientos, permitiendo en los negocios tomar decisiones proactivas y conducidas por un conocimiento acabado de la información ALCANTARA, M. (9).

La data mining surge como una tecnología que intenta ayudar a comprender el contenido de una base de datos, esta tecnología puede generar nuevas oportunidades de negocios ya que permite:

- ***Predicción automatizada de tendencias y comportamientos.*** Data mining automatiza el proceso de encontrar información predecible en grandes bases de datos. Esto requería un intenso análisis manual, ahora encontramos respuestas rápidas y directas desde los datos.
- ***Descubrimiento automatizado de modelos previamente desconocidos.*** Las Herramientas de Data mining exploran minuciosamente las bases de datos e identifican modelos previamente escondidos en un solo paso. Otros problemas de descubrimiento de modelos incluyen detectar transacciones fraudulentas de tarjetas de créditos e identificar datos anormales que pueden representar errores de digitación en la carga de datos MARTINEZ D. (11).
- Una definición tradicional es la siguiente: Un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos AAI (12).
- Desde el punto de vista empresarial, lo definimos como: La integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten un sesgo hacia la toma de decisión MOLINA, F. (13).

- El data mining es una tecnología compuesta por etapas que integra varias áreas y que no se debe confundir con un gran software. Durante el desarrollo de un proyecto de este tipo se usan diferentes aplicaciones software en cada etapa que pueden ser estadísticas, de visualización de datos o de inteligencia artificial, principalmente. Actualmente existen aplicaciones o herramientas comerciales de data mining muy poderosas que contienen un sinnúmero de utilerías que facilitan el desarrollo de un proyecto. Sin embargo, casi siempre acaban complementándose con otra herramienta MARTINEZ, D. MARTINEZ D. (11).

2.2.2 Procedo De Extracción De Conocimiento (KDD)

El descubrimiento de esta información oculta es posible gracias a la Minería de Datos (Data Mining), que entre otras sofisticadas técnicas aplica la inteligencia artificial para encontrar patrones y relaciones dentro de los datos permitiendo la creación de modelos, es decir, representaciones abstractas de la realidad, pero es el descubrimiento del conocimiento (KDD-Knowledge Discovery in Databases), por sus siglas en inglés) que se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a estos patrones encontrados RIVE, P. (14).

El proceso de KDD es interactivo e iterativo e implican pasos que se resumen así VERGARAY, A. (15):

1. ***Aprendizaje del Dominio de la aplicación:*** este paso incluye la adquisición del conocimiento previo relevante y planteo de los objetivos de la aplicación.
2. ***Creación de un conjunto de datos de destino:*** por medio de esto se escoge el conjunto de datos o se elige el subconjunto de variables o muestras de datos en los cuales el descubrimiento que va a realizar.
3. ***Limpieza de datos y preprocesamiento:*** aquí se dan las operaciones básicas como la eliminación del ruido, la recogida de información, necesaria para modelar, la determinación de estrategias para el manejo de

los campos de datos que falta, la contabilidad de la información en tiempo y secuencia de los cambios conocidos.

4. **Reducción de datos de proyección:** en este punto se lleva a cabo la búsqueda de las características útiles para representar los datos, dependiendo del objetivo y de la tarea. Para ello, se ejecuta la reducción de dimensionalidad o se aplican métodos de transformación que permitan acortar el número efectivo de las variables en estudio o encontrar las representaciones invariables para los datos.
5. **Elección de la función de minería de datos:** a través de esto se decide el propósito del modelo derivado por medio del algoritmo de minería de datos.
6. **Elección del algoritmo de minería de datos:** en esta parte se eligen los métodos que van a ser utilizados para la búsqueda de patrones en el conjunto de los datos, tales como decidir que modelos y parámetros pueden ser apropiados y establecer con que método de minería de datos pueden ser aplicados.
7. **Realización de minería de datos:** aquí se da la búsqueda de patrones de interés dentro de una forma de representación particular o el conjunto de tales de representaciones, en el cual se incluyen las reglas de clasificación o árboles, la regresión, el clustering, el modelo secuencia, la dependencia y el análisis lineal.
8. **Interpretación:** en este paso se realiza la interpretación de los patrones detectados y, en algunos casos, también se vuelve a cualquiera de los momentos anteriores. Luego de ello, se hace la visualización de los posibles patrones extraídos, la eliminación de los patrones redundantes o irrelevantes y la traducción de los patrones útiles a términos que sean comprensibles para los usuarios.
9. **Utilización del Conocimiento descubierto:** este paso implica la incorporación de los conocimientos extraídos en pro del rendimiento del sistema, tomando acciones basadas en dicho conocimiento o simplemente documentándolo e informándolo a las partes interesadas.

2.2.3 Modelos de Minería de Datos

Los modelos de minería de datos son predictivos y descriptivos. Los predictivos se tienen una variable con valor predictivo y su finalidad es determinarlo. Por otro lado, los modelos descriptivos, en los cuales no se cuenta con un resultado conocido para poder guiar los algoritmos, y por ello se conocen como modelos de aprendizaje no supervisado, donde el modelo se va ajustando de acuerdo con las observaciones o datos entregados, y se recurre según el autor a argumentos heurísticos para evaluar la calidad de los resultados VERGARAY, A. (15).

2.2.4 Métodos de Minería de Datos

1. **Árboles de decisión:** Los algoritmos de árbol de decisión consisten en organizar los datos en elecciones que compiten formando ramas de influencia después de una decisión inicial. El tronco del árbol representa la decisión inicial, y empieza con una pregunta de sí o no DAY, G. (16).
2. **Microsoft algoritmo de árboles de decisión:** es un algoritmo de clasificación y regresión para su uso en el modelado de predicción de atributos discretos y continuos. Para los atributos discretos, el algoritmo hace predicciones basándose en las relaciones entre las columnas de entrada de un conjunto de datos. Utiliza los valores, conocidos como estados, de estas columnas para predecir los estados de una columna que se designa como elemento de predicción. Específicamente, el algoritmo identifica las columnas de entrada que se correlacionan con la columna de predicción. Por ejemplo, en un escenario para predecir qué clientes van a adquirir probablemente una bicicleta, si nueve de diez clientes jóvenes compran una bicicleta, pero solo lo hacen dos de diez clientes de edad mayor, el algoritmo infiere que la edad es un buen elemento de predicción en la compra de bicicletas. El árbol de decisión realiza predicciones basándose en la tendencia hacia un resultado concreto MICROSOFT (17).
3. **Los árboles de decisión:** Los árboles de decisión proveen de una herramienta de clasificación muy potente. Su uso en el manejo de

datos la hace ganar en popularidad dadas las posibilidades que brinda y la facilidad con que son comprendidos sus resultados por cualquier usuario. El árbol en sí mismo, al ser obtenidos, determinan una regla de decisión, dicha técnica permite: BOUZA, N. (18)

- a. Segmentación: establecer que grupos son importantes para clasificar un cierto ítem.
 - b. Clasificación: asignar ítems a uno de los grupos en que está particionada una población.
 - c. Predicción: establecer reglas para hacer predicciones de ciertos eventos.
 - d. Reducción de la dimensión de los datos: Identificar que datos son los importantes para hacer modelos de un fenómeno.
 - e. Identificación-interrelación: identificar que variables y relaciones son importantes para ciertos grupos identificados a partir de analizar los datos.
 - f. Recodificación: discretizar variables o establecer criterios cualitativos perdiendo la menor cantidad posible de información relevante.
4. **El algoritmo K-means:** El algoritmo K-means se basa en el análisis de grupos. Trata de dividir los datos recogidos en "bloques" ("clusters" en inglés) separados agrupados por características comunes DAY, G. (16).
5. **Algoritmo del vecino k más cercano:** Este algoritmo reconoce patrones en la ubicación de los datos y los asocia a los datos con un identificador mayor. Por ejemplo, si quieres asignar una oficina postal a cada ubicación geográfica del hogar y tienes un conjunto de datos para cada ubicación geográfica del hogar, el algoritmo del vecino k más cercano asignará las casas a la oficina postal más cercana según su proximidad.
6. **Naive Baye:** El algoritmo Naive Baye predice la salida de una identidad basándose en los datos de observaciones conocidas. Por ejemplo, si una persona tiene una altura de 6 pies y 6 pulgadas (1,97 m) y lleva una talla 14 de zapatos, el algoritmo Naive Baye podría

predecir con una determinada probabilidad que la persona es un hombre DAY, G. (16).

El algoritmo Bayes Naive de Microsoft es un algoritmo de clasificación que proporciona Microsoft SQL Server Analysis Services para el modelado de predicción. El nombre Bayes Naive deriva del hecho de que el algoritmo usa el teorema de Bayes, pero no tiene en cuenta las dependencias que pudieran existir y, por consiguiente, se dice que sus suposiciones son ingenuas o "Naive". Desde el punto de vista computacional, el algoritmo es menos complejo que otros algoritmos de Microsoft y, por tanto, resulta útil para generar rápidamente modelos de minería de datos para descubrir relaciones entre columnas de entrada y columnas de predicción. Puede utilizar este algoritmo para realizar exploraciones iniciales de datos y, más adelante, aplicar los resultados para crear modelos de minería de datos adicionales con otros algoritmos más complejos y precisos desde el punto de vista computacional MICROSOFT (19).

7. **Algoritmo CART:** "CART" es una sigla en inglés que significa análisis de árbol regresivo y de clasificación. Al igual que los análisis de árboles de decisión, organiza los datos según opciones que compiten, como si una persona ha sobrevivido a un terremoto. Al contrario que los algoritmos de árboles de decisión, que sólo pueden clasificar una salida o una salida numérica basada en la regresión, el algoritmo CART puede usar los dos para predecir la probabilidad de un evento DAY, G. (16).

2.2.5 Herramientas de Minería de Datos

Algunas herramientas para aplicación de minería de datos:

1. **Clementine / SPSS:** Clementine 12.0 ofrece una productividad analítica incrementada, una visión más profunda sobre la información y una visualización avanzada, incluyendo una interactividad gráfica rica y mejorada, y unos informes tabulares personalizados ejecutados por una poderosa interfaz de diseño. SPSS ofrece así Clementine 12.0 como una solución completamente integrada de texto y data mining, que se integra también en la

- plataforma predictiva SPSS Predictive Enterprise Services y con Dimensions, el principal software de tratamiento de encuestas de SPSS SPAIN, B. (20).
2. **SAS Enterprise Miner / SAS:** SAS Enterprise Miner agiliza el proceso de minería de datos para crear de alta precisión modelos predictivos y descriptivos basados en el análisis de grandes cantidades de datos de toda una empresa. La minería de datos es aplicable en una variedad de industrias y ofrece metodologías para tan diversos problemas de negocio como la detección de fraudes, householding, la retención de clientes y el desgaste, la comercialización de bases de datos, segmentación de mercado, análisis de riesgos, El análisis de afinidad, la satisfacción del cliente, la predicción de la quiebra, y el análisis de la cartera SAS (21).
 3. **Keel:** KEEL es una herramienta software para la preparación de AEs para problemas de minería de datos. Permite ejecutarlos dentro del propio entorno (ejecución online) o generarlos para una ejecución posterior en distintas máquinas (ejecución off-line) DERRAT, J. (22).
 4. **RapidMiner / Yale:** RapidMiner es una herramienta de Minería de Datos ampliamente usada y probada a nivel internacional en aplicaciones empresariales, de gobierno y academia. Implementa más de 500 técnicas de preprocesamiento de datos, modelación predictiva y descriptiva, métodos de prueba de modelos, visualización de datos MICROSYSTEM (23).
 5. **Microsoft SQL Server:** SQL Server Business Intelligence ofrece una plataforma amplia que permite a las organizaciones compilar e implementar soluciones BI seguras, escalables y administrables.
 6. **Weka:** acrónimo de Waikato Environment for Knowledge Analysis, es un entorno para experimentación de análisis de datos que permite aplicar, analizar y evaluar las técnicas más relevantes de análisis de datos, principalmente las provenientes del aprendizaje automático, sobre cualquier conjunto de datos del usuario. Para ello únicamente

se requiere que los datos a analizar se almacenen con un cierto formato, conocido como ARFF (Attribute-Relation File Format) ISA (24).

2.2.6 Características de Minería de Datos

Explorar los datos se encuentran en las profundidades de las bases de datos, como los almacenes de datos, que algunas veces contiene información almacenada durante varios años.

En algunos casos, los datos se consolidan en un almacén de datos y en mercados de datos; en otros, se mantienen en servidores de Internet e Intranet.

2.2.7 Aplicación de Minería de Datos

Podemos revisar algunos casos de aplicación de minería de datos VERGARAY, A. (15):

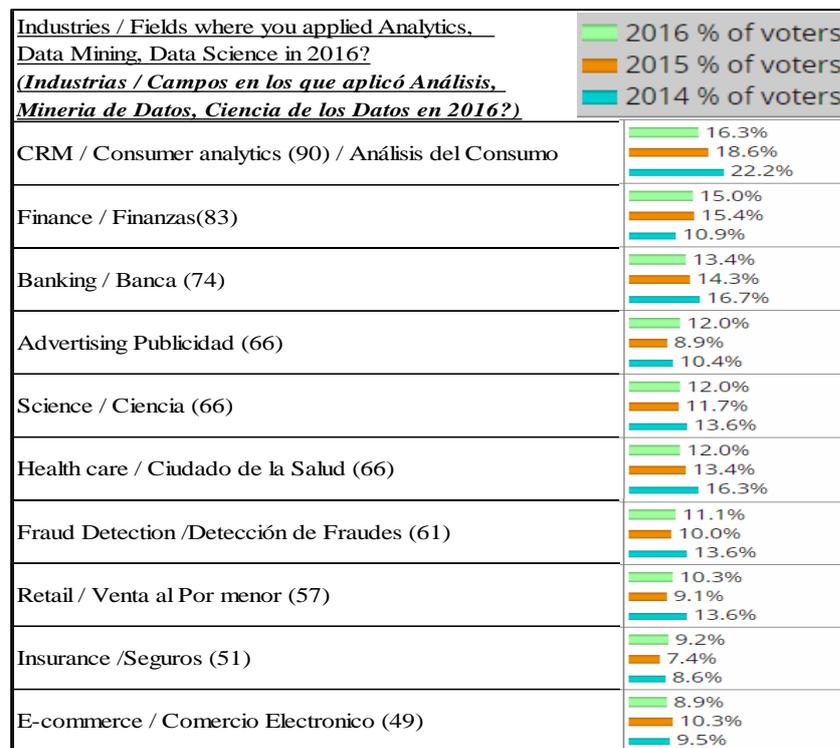
- **Comercio y Marketing:** con el objetivo de identificar patrones de compra de los clientes, buscar asociaciones entre clientes y características demográficas, predecir respuesta a campañas de e-mailing, analizar la canasta de compra.
- **Banca:** con el objetivo de detectar patrones de uso fraudulento de tarjetas de crédito, identificar a clientes leales, determinar gasto de tarjeta de créditos por grupos.
- **Seguros y Salud Privada:** con el objetivo de analizar los procedimientos médicos solicitados conjuntamente, identificar patrones de comportamiento para pacientes con riesgo, identificar terapias médicas satisfactorias para diferentes enfermedades, asociar síntomas y clasificación diferencial de patologías.

Podemos revisar algunos resultados de aplicación de minería de datos PIATETSKY, G. (25):

- La última encuesta de KDnuggets pidió a los lectores que Industrias / Campos en los que aplicó Analytics, ¿Data Mining, Data Science en 2016? Las áreas más populares fueron:

- CRM / Análisis de los Consumidores, todavía número 1 con 16,3%, pero menos con respecto al 18,6% en 2015.
 - Finanzas, 15,0% (casi lo mismo que en 2015).
 - Banca, 13,4% (ligeramente abajo)
 - Publicidad, 12,0% (+ 35% a partir de 2015))
 - Ciencia, 12,0% (casi el mismo)
 - Atención de salud, 12,0% (11% menos)
 - Nueva área añadida en la encuesta de 2016 fue Supply Chain, que recibió una respetable participación de 6,5%, (rango 20).
- En 2016, 552 personas votaron, en comparación con 350 en 2015, y 221 en 2014.

Figura 1: Where Analytics, Data Mining, Data Science were applied in 2016. / Donde se aplicó Análisis, Minería de Datos, Ciencia de Datos, en 2016.



2.2.8 Seguro Integral de Salud

El Seguro Integral de Salud (SIS), como Organismo Público Ejecutor (OPE), del Ministerio de Salud, tiene como finalidad proteger la salud de los peruanos que no cuentan con un seguro de salud, priorizando en aquellas poblacionales vulnerables que se encuentran en situación de pobreza y pobreza extrema.

De esta forma, estamos orientados a resolver la problemática del limitado acceso a los servicios de salud de nuestra población objetivo, tanto por la existencia de barreras económicas, como las diferencias culturales y geográficas. Pero el SIS también busca mejorar la eficiencia en la asignación de los recursos públicos e implementando instrumentos de identificación del usuario, priorizando el componente materno infantil SIS (26).

2.2.9 Instituciones Prestadores de Servicios de Salud (IPRESS)

Las Instituciones Prestadoras de Servicios de Salud (IPRESS), según el artículo 7 del D. L. N.º 1158, son aquellos establecimientos de salud y servicios médicos de apoyo, públicos, privados o mixtos, creados y por crearse, que realizan atención de salud con fines de prevención, promoción, diagnóstico, tratamiento y/o rehabilitación; así como aquellos servicios complementarios o auxiliares de la atención médica, que tienen por finalidad coadyuvar en la prevención, promoción, diagnóstico, tratamiento y/o rehabilitación de la salud SUSALUD (3).

2.2.10 Prestaciones de Salud

Las prestaciones de salud corresponden a las atenciones brindadas a los afiliados al Seguro Integral de Salud, estas se clasifican en preventivas y recuperativas. Dichas prestaciones son registradas por los profesionales de la salud en un Formato Único de Atención (FUA), la cual es ingresada o digitada a un Aplicativo de Registro de Formato Único de Atención FUA del Seguro Integral de Salud SIS (ARFSIS) en los puntos de digitación SIS (26).

2.2.11 Metodología utilizada

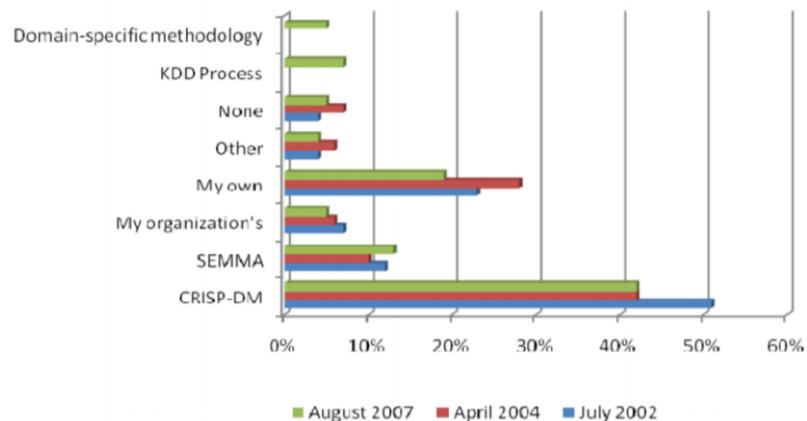
Para llevar a cabo el presente trabajo de investigación de minería de datos se tomó como base el proceso de desarrollo de minería de datos **CRISP (Cross-Industry Standard Process for Data Mining - Proceso Estándar de Trans-Industria para la Minería de Datos)**

- **CRISP-DM:** El modelo de proceso corriente para la minería de datos proporciona una descripción del ciclo de vida del proyecto de minería de datos. Este contiene las fases de un proyecto, sus tareas respectivas, y las relaciones entre estas tareas.

CRISP-DM [CRISP-DM, 2000], es la guía de referencia más ampliamente utilizada en el desarrollo de proyectos de Data Mining, como se puede constatar en la gráfica presentada en la figura 2. Esta gráfica, publicada el año 2014 por kdnuggets.com, representa el resultado obtenido en sucesivas encuestas efectuadas durante los últimos años, respecto del grado de utilización de las principales guías de desarrollo de proyectos de Data Mining. En ella se puede observar, que CRISP-DM sigue siendo la principal metodología para proyectos de minería de datos, con el mismo porcentaje que en 2007 (43% vs 42%). Sin embargo, se informa que se utiliza por menos del 50%.

Figura 2: Metodologías utilizadas en Data Mining ()

Metodologías utilizadas en Data Mining ([kdnuggets, 2007]).



- Fases de la metodología de referencia CRISP-DM
 - **Comprensión del negocio:** Esta fase inicial se enfoca en la comprensión de los objetivos de proyecto y exigencias desde una perspectiva de negocio, luego convirtiendo este conocimiento de los datos en la definición de un problema de minería de datos y en un plan preliminar diseñado para alcanzar los objetivos.

Tareas de la fase

Determinar los objetivos del negocio. Esta es la primera tarea para desarrollar y tiene como metas, determinar cuál es el problema que se desea resolver, por qué la necesidad de utilizar Data Mining y definir los criterios de éxito. Los problemas pueden ser diversos como, por ejemplo, detectar fraude en el uso de tarjetas de crédito, detección de intentos de ingreso indebido a un sistema, asegurar el éxito de una determinada campaña publicitaria, etc. En cuanto a los criterios de éxito, estos pueden ser de tipo cualitativo, en cuyo caso un experto en el área de dominio, califica el resultado del proceso de DM, o de tipo cuantitativo, por ejemplo, el número de detecciones de fraude o la respuesta de clientes ante una campaña publicitaria.

Evaluación de la situación. En esta tarea se debe calificar el estado de la situación antes de iniciar el proceso de DM, considerando aspectos tales como: ¿cuál es el conocimiento previo disponible acerca del problema?, ¿se cuenta con la cantidad de datos requerida para resolver el problema?, ¿cuál es la relación coste beneficio de la aplicación de DM?, etc. En esta fase se definen los requisitos del problema, tanto en términos de negocio como en términos de Data Mining.

Determinación de los objetivos de DM. Esta tarea tiene como objetivo representar los objetivos del negocio en términos de las metas del proyecto de DM, como, por ejemplo, si el objetivo del negocio es el desarrollo de una campaña publicitaria para incrementar la asignación de créditos hipotecarios, la meta de DM será, por ejemplo, determinar el perfil de los clientes respecto de su capacidad de endeudamiento. Producción de un plan del proyecto. Finalmente, esta última tarea de la primera fase de CRISP-DM, tiene como meta desarrollar un plan para el proyecto,

que describa los pasos a seguir y las técnicas a emplear en cada paso.

Figura 3: Fase de Comprensión del Negocio (CRISP-DM, 2000).



- **Comprensión de los datos:** La fase de entendimiento de datos comienza con la colección de datos inicial y continua con las actividades que le permiten familiarizar primero con los datos, identificar los problemas de calidad de datos, descubrir los primeros conocimientos en los datos, y/o descubrir subconjuntos interesantes para formar hipótesis en cuanto a la información oculta.

Tareas de la fase

Recolección de datos iniciales. La primera tarea en esta segunda fase del proceso de CRISP-DM, es la recolección de los datos iniciales y su adecuación para el futuro procesamiento. Esta tarea tiene como objetivo, elaborar informes con una lista de los datos adquiridos, su localización, las técnicas utilizadas en su recolección y los problemas y soluciones inherentes a este proceso.

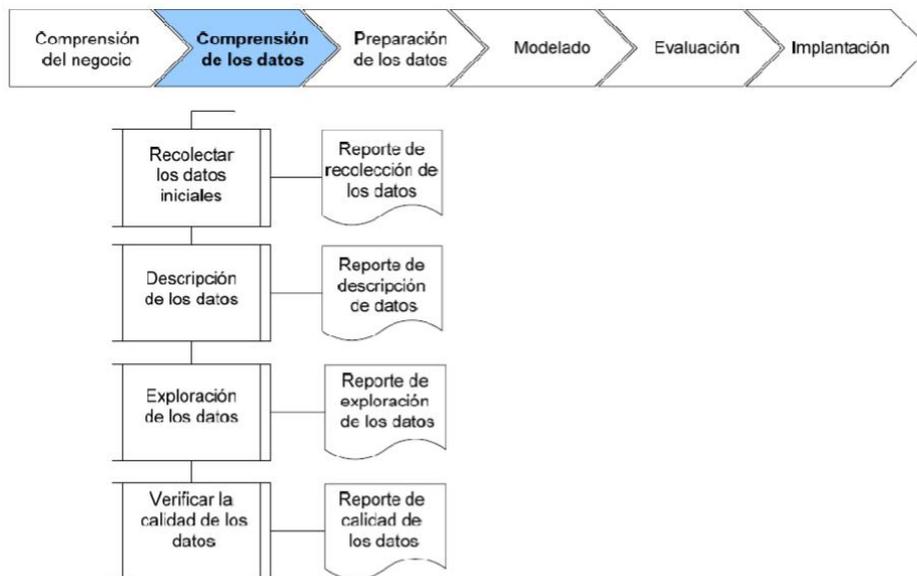
Descripción de los datos. Después de adquiridos los datos iniciales, estos deben ser descritos. Este proceso involucra establecer volúmenes de datos (número de registros y campos por registro), su

identificación, el significado de cada campo y la descripción del formato inicial.

Exploración de datos. A continuación, se procede a su exploración, cuyo fin es encontrar una estructura general para los datos. Esto involucra la aplicación de pruebas estadísticas básicas, que revelen propiedades en los datos recién adquiridos, se crean tablas de frecuencia y se construyen gráficos de distribución. La salida de esta tarea es un informe de exploración de los datos.

Verificación de la calidad de los datos. En esta tarea, se efectúan verificaciones sobre los datos, para determinar la consistencia de los valores individuales de los campos, la cantidad y distribución de los valores nulos, y para encontrar valores fuera de rango, los cuales pueden constituirse en ruido para el proceso. La idea en este punto es asegurar la completitud y corrección de los datos.

Figura 4: Fase de Comprensión de los Datos (CRISP – DM, 2000)



- **Preparación de datos:** La fase de preparación de datos cubre todas las actividades necesarias para construir el conjunto de datos final [los datos que serán provistos en las herramientas de modelado] de los datos en brutos iniciales. Las tareas de preparación de datos probablemente van a ser realizadas muchas veces y no en cualquier orden prescripto. Las tareas

incluyen la selección de tablas, registros, y atributos, así como la transformación y la limpieza de datos para las herramientas que modelan.

Tareas de la fase

Selección de Datos. En esta etapa, se selecciona un subconjunto de los datos adquiridos en la fase anterior, apoyándose en criterios previamente establecidos en las fases anteriores: calidad de los datos en cuanto a completitud y corrección de los datos y limitaciones en el volumen o en los tipos de datos que están relacionadas con las técnicas de DM seleccionadas.

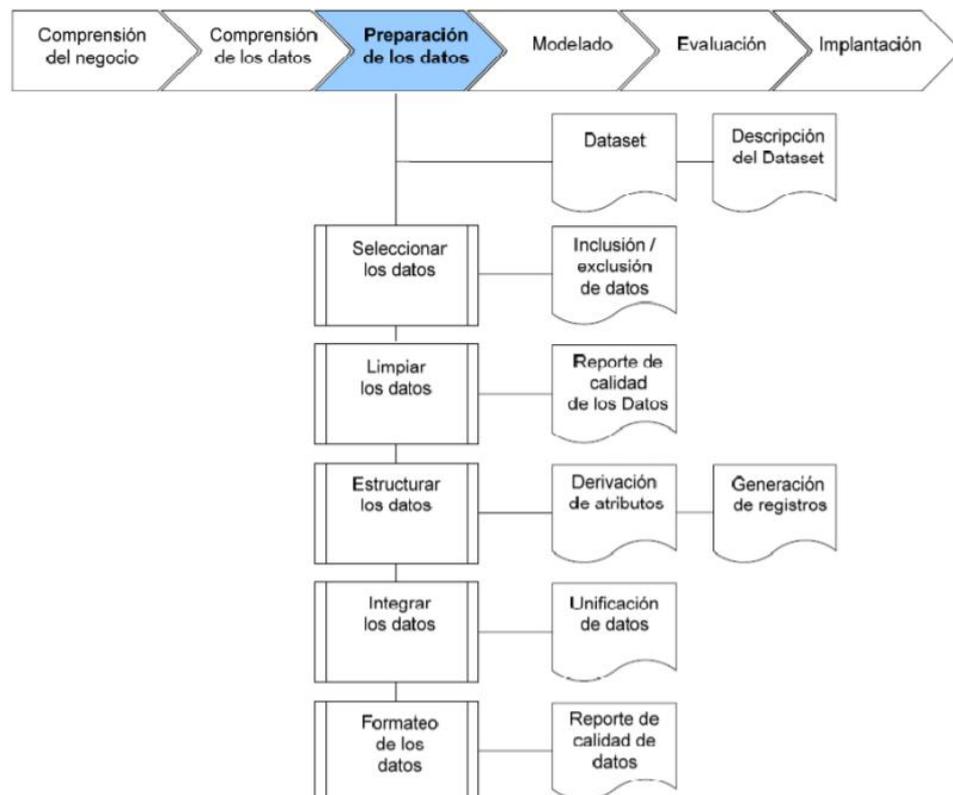
Limpieza de Datos. Esta tarea complementa a la anterior, y es una de las que más tiempo y esfuerzo consume, debido a la diversidad de técnicas que pueden aplicarse para optimizar la calidad de los datos al objeto de prepararlos para la fase de modelación. Algunas de las técnicas a utilizar para este propósito son: normalización de los datos, discretización de campos numéricos, tratamiento de valores ausentes, reducción del volumen de datos, etc.

Construcción de los datos. Esta tarea incluye las operaciones de preparación de los datos tales como la generación de nuevos atributos a partir de atributos ya existentes, integración de nuevos registros o transformación de valores para atributos existentes.

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

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

Figura 5: Fase de Preparación de los datos (CRISP-DM, 2000).



- **Modelado:** En esta fase, varias técnicas de modelado son seleccionadas y aplicadas, y sus parámetros son calibrados a valores óptimos. Típicamente hay varias técnicas para el mismo tipo de problema de minería de datos. Algunas técnicas tienen requerimientos específicos sobre la forma de datos. Por lo tanto, volver a la fase de preparación de datos es a menudo necesario.

Tareas de la fase

Selección de la técnica de modelado. Esta tarea consiste en la selección de la técnica de DM más apropiada al tipo de problema a resolver. Para esta selección, se debe considerar el objetivo principal del proyecto y la relación con las herramientas de DM existentes.

Generación del plan de prueba. Una vez construido un modelo, se debe generar un procedimiento destinado a probar la calidad y validez de este. Por ejemplo, en una tarea supervisada de DM como la clasificación, es común usar la razón de error como medida de la calidad. Entonces, típicamente se separan los datos en dos conjuntos, uno de entrenamiento y otro de prueba, para luego construir el modelo basado en el conjunto de entrenamiento y medir la calidad del modelo generado con el conjunto de prueba.

Construcción del Modelo. Después de seleccionada la técnica, se ejecuta sobre los datos previamente preparados para generar uno o más modelos. Todas las técnicas de modelado tienen un conjunto de parámetros que determinan las características del modelo a generar. La selección de los mejores parámetros es un proceso iterativo y se basa exclusivamente en los resultados generados. Estos deben ser interpretados y su rendimiento justificado.

Evaluación del modelo. En esta tarea, los ingenieros de DM interpretan los modelos de acuerdo con el conocimiento preexistente del dominio y los criterios de éxito preestablecidos. Expertos en el dominio del problema juzgan los modelos dentro del contexto del dominio y expertos en Data Mining aplican sus propios criterios (seguridad del conjunto de prueba, pérdida o ganancia de tablas, etc..).

Figura 6: Fase de Modelado (CRISP-DM, 2000)

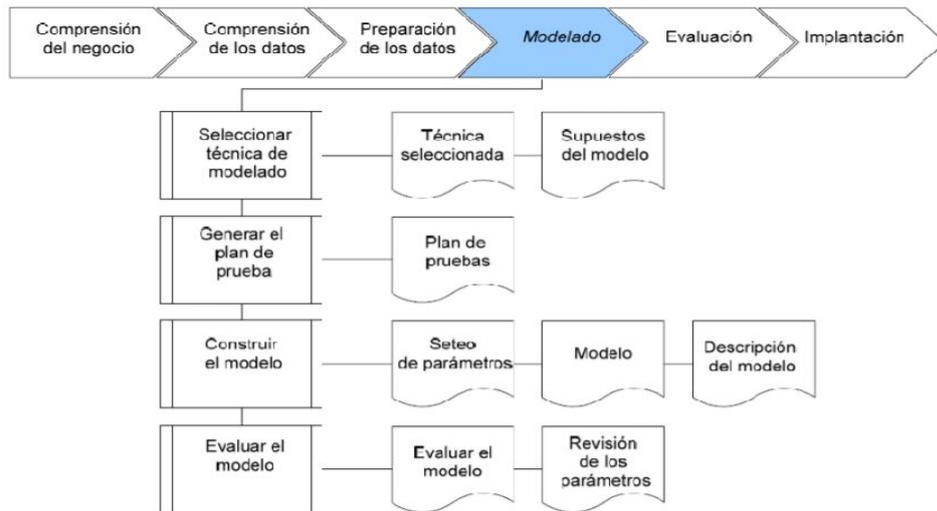


Figura No. 2.9. Fase de modelado ([CRISP-DM, 2000]).

- **Evaluación:** En esta etapa en el proyecto usted ha construido un modelo (o modelos) que parece tener la alta calidad de una perspectiva de análisis de datos. Antes de proceder al despliegue final del modelo, es importante evaluar a fondo ello y la revisión de los pasos ejecutados para crearlo, para comparar el modelo correctamente obtenido con los objetivos de negocio. Un objetivo clave es determinar si hay alguna cuestión importante de negocio que no ha sido suficientemente considerada. En el final de esta fase, una decisión en el uso de los resultados de minería de datos debería ser obtenida.

Tareas de la fase

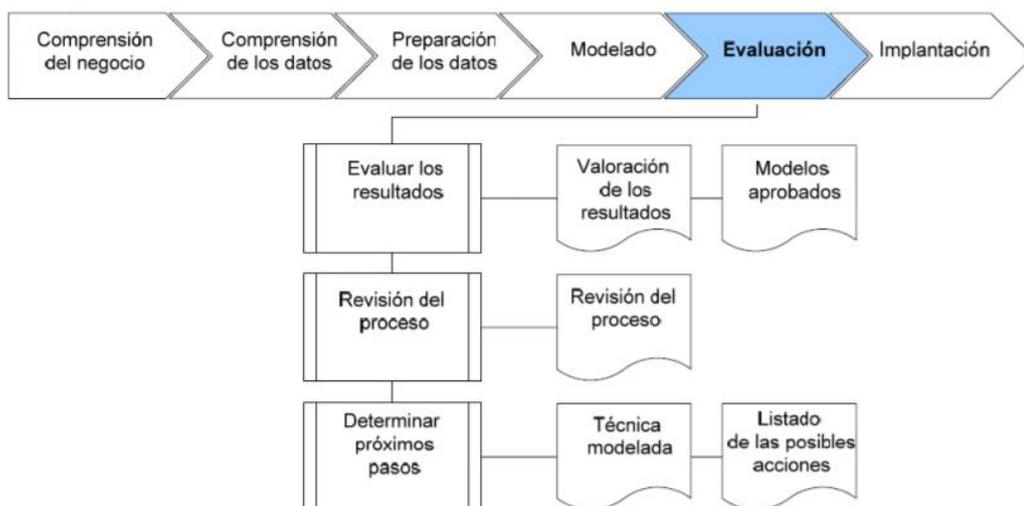
Evaluación de los resultados. En los pasos de evaluación anteriores, se trataron factores tales como la exactitud y generalidad del modelo generado. Esta tarea involucra la evaluación del modelo en relación con los objetivos del negocio y busca determinar si hay alguna razón de negocio para la cual, el modelo sea deficiente, o si es aconsejable probar el modelo, en un problema real si el tiempo y restricciones lo permiten. Además de los resultados directamente relacionados con el objetivo del proyecto, ¿es aconsejable evaluar el modelo en relación con otros objetivos

distintos a los originales?, esto podría revelar información adicional.

Proceso de revisión. El proceso de revisión se refiere a calificar al proceso entero de DM, al objeto de identificar elementos que pudieran ser mejorados.

Determinación de futuras fases. Si se ha determinado que las fases hasta este momento han generado resultados satisfactorios, podría pasarse a la fase siguiente, en caso contrario podría decidirse por otra iteración desde la fase de preparación de datos o de modelación con otros parámetros. Podría ser incluso que en esta fase se decida partir desde cero con un nuevo proyecto de DM.

Figura 7: Fase de Evaluación (CRISP-DM, 2000)



- **Desarrollo:** La creación del modelo no es generalmente el final del proyecto. Incluso si el objetivo del modelo es aumentar el conocimiento de los datos, el conocimiento ganado tendrá que ser organizado y presentado en el modo en el que el cliente pueda usarlo. Ello a menudo implica la aplicación de modelos "vivos" dentro de un proceso de toma de decisiones de una organización, por ejemplo, en tiempo real la personalización de página Web o la repetida obtención de bases de datos de mercadeo. Dependiendo de los requerimientos, la fase de desarrollo puede ser tan simple como la generación de un

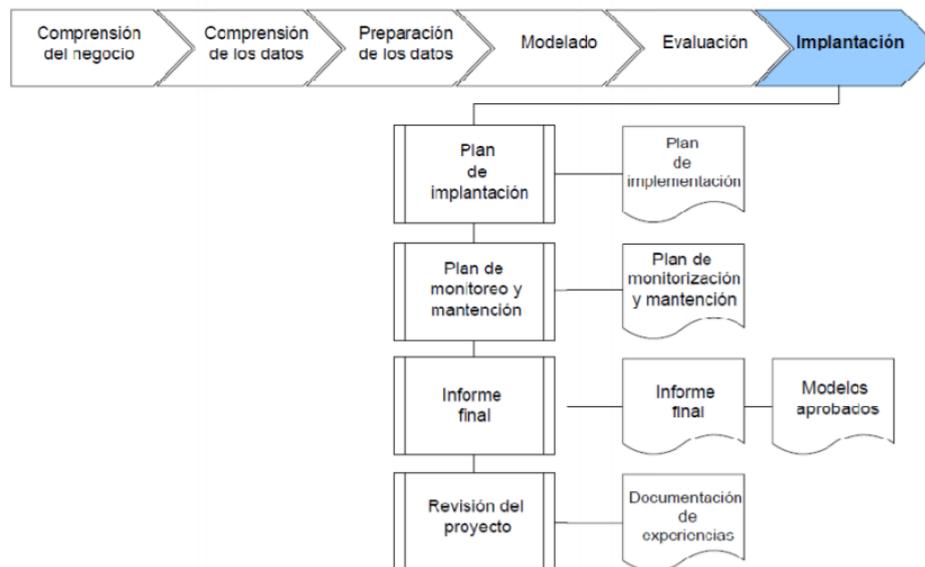
informe o tan compleja como la realización repetida de un proceso cruzado de minería de datos a través de la empresa. En muchos casos, es el cliente, no el analista de datos, quien lleva el paso de desarrollo. Sin embargo, incluso si el analista realizara el esfuerzo de despliegue, esto es importante para el cliente para entender de frente que acciones necesita para ser ejecutadas en orden para hacer uso de los modelos creados actualmente.

Tareas de la fase

- ***Plan de implementación.*** Para implementar el resultado de DM en la organización, esta tarea toma los resultados de la evaluación y concluye una estrategia para su implementación. Si un procedimiento general se ha identificado para crear el modelo, este procedimiento debe ser documentado para su posterior implementación. Monitorización y Mantenimiento. Si los modelos resultantes del proceso de Data Mining son implementados en el dominio del problema como parte de la rutina diaria, es aconsejable preparar estrategias de monitorización y mantenimiento para ser aplicadas sobre los modelos. La retroalimentación generada por la monitorización y mantenimiento pueden indicar si el modelo está siendo utilizado apropiadamente.

Informe Final. Es la conclusión del proyecto de DM realizado. Dependiendo del plan de implementación, este informe puede ser sólo un resumen de los puntos importantes del proyecto y la experiencia lograda o puede ser una presentación final que incluya y explique los resultados logrados con el proyecto. Revisión del proyecto: En este punto se evalúa qué fue lo correcto y qué lo incorrecto, qué es lo que se hizo bien y qué es lo que se requiere mejorar.

Figura 8: Fase de Implementación (CRISP-DM, 2000)



2.2.12 Comparación de Metodologías

- A continuación, se desarrolla el siguiente apartado comparando las metodologías estudiadas las cuales para el presente trabajo se describen: Proceso de Extracción de Conocimientos (**KDD**), (Cross-Industry Standard Process for Data Mining - Proceso Estándar de Trans-Industria para la Minería de Datos **CRISP - DM** y la de **SEMMA** que por sus abreviaturas significan (muestreo, exploración, modificación, modelado y evaluación).
- Según la revisión de la literatura y fuentes bibliográficas puedo referir que la metodología **KDD (1996)** da inicio al proceso de estudio para la extracción de conocimientos, por lo que podemos afirmar que SEMMA y CRISP – DM, son una implementación de la metodología KDD.
- También puedo llegar a la conclusión de que CRISP-DM es más completo que SEMMA, no por poseer una etapa o fase que es la comprensión del negocio (CRISP-DM).
- Sin embargo, analizándolo más profundamente, podemos integrar el desarrollo de una comprensión del dominio de la aplicación, el conocimiento previo relevante y los objetivos del usuario final, en la

etapa de muestreo de SEMMA, porque los datos no se pueden muestrear a menos que exista una comprensión verdadera de todos los aspectos presentados, es decir, podría estar implícito.

- En este documento, se pretende establecer un agregado a los procesos como una implementación adicional a los ya conocidos en las metodologías descritas.
- Podemos presentar de esta manera la siguiente tabla resumen:

Tabla 1: Cuadro Comparativo de Metodología de Minería de Datos.

KDD	SEMMA	CRISP – DM
Pre – KDD	-----	Comprensión del Negocio
Selección	Muestreo	Comprensión de los Datos
Preprocesamiento	Exploración	
Transformación	Modificación	Preparación de los Datos
Minería de Datos	Modelado	Modelado
Interpretación/Evaluación	Evaluación	Evaluación
Post KDD		Implementación

2.3 HIPÓTESIS

La propuesta de una metodología de extracción de conocimientos mejorará el proceso de análisis del conocimiento a partir de las prestaciones del Seguro Integral de Salud en la Región Piura en el año 2016.

2.4 VARIABLES

- Para el presente trabajo se consideran dos variables la que buscare medir (metodología) y la medida (proceso de análisis).

2.4.1 Variable Dependiente: Metodología de extracción de conocimientos, la metodología determinada por fases y tareas.

2.4.2 Variable Independiente: Proceso de Análisis, el resultado de la implementación de la metodología.

III. METODOLOGIA

3.1 EL TIPO Y EL DISEÑO DE LA INVESTIGACIÓN

- a) El presente trabajo es de tipo descriptivo, porque busca especificar las propiedades, características y los perfiles importantes que se someta a un análisis.
- b) Nivel cuantitativo: porque la intención es buscar la exactitud de mediciones o indicadores sociales con el fin de generalizar sus resultados a poblaciones o situaciones amplias. Trabajan fundamentalmente con el número, el dato cuantificable PALELLA, S. (27).

3.2 DISEÑO DE LA INVESTIGACIÓN

- 3.2.1 Diseño no experimental es el que se realiza sin manipular en forma deliberada ninguna variable. El investigador no sustituye intencionalmente las variables independientes. Se observan los hechos tal y como se presentan en su contexto real y en un tiempo determinado o no, para luego analizarlos EUMED (28).

3.3 POBLACIÓN Y MUESTRA

- 3.3.1 La población está determinada por ocho (08) trabajadores relacionados con el uso de la tecnología de la información, distribuidos: cuatro (04) de la Dirección Regional de Salud Piura y (04) del Seguro Integral de Salud.
- 3.3.2 La población muestral está definida por (08) trabajadores, es decir, el total de la población relacionada con el uso de la tecnología de la información.
- 3.3.3 Se consigna la cantidad de ocho (08) trabajadores porque son entes decisores de categoría directiva en la oficina del Equipo de Seguros de Salud de la Dirección Regional de Salud – Piura.

3.4 DEFINICIÓN Y OPERACIONALIZACIÓN DE LAS VARIABLES Y LOS INDICADORES

VARIABLE	DEFINICIÓN CONCEPTUAL	DIMENSIONES	DEFINICIÓN OPERACIONAL	INDICADORES
VD: Metodología de Extracción de Conocimientos	En un método probado para orientar trabajos de minería de datos, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas (09)	Base de Datos Análisis	SQL SERVER Modelos estadísticos, descriptivos, predicciones.	Utilización de Gestor de Base de Datos. Software para aplicación de minería de datos.
VI: Procesos de Análisis	Proceso de análisis de reportes y resultados de modelos de algoritmo de minería de datos.	Resultado de Algoritmos	Reporte de la Herramienta de Minería de Datos	Reporte que ayude a la toma de decisiones

3.5 TÉCNICAS E INSTRUMENTOS

3.5.1 Se utilizó la técnica de entrevista personal para la evaluación de la metodología empleada que se ejecutó a partir de la aplicación de los algoritmos de minería de datos y las fases que corresponden.

3.5.2 Instrumentos que se construyeron para evaluación de metodología.

3.6 PLAN DE ANÁLISIS

3.6.1 La fase de análisis de datos comprende la recolección de datos de las prestaciones del Seguro Integral de Salud, obtenidos de la base de datos del SIS, por intermedio del Equipo de Seguros de Salud de la Dirección Regional de Salud Piura.

3.6.2 Inicialmente se tiene una base de datos de prestaciones, con las tablas Atención, Medicamentos, Procedimientos, Materno Infantil y Diagnósticos, la cual ayudó a encontrar los principales indicadores, así como, aplicar correctamente los algoritmos de minería de datos, permitiendo que se valide la aplicación de estas técnicas y los resultados.

3.6.3 Se inició con la creación de almacén de datos, para llevar el proceso de extracción, transformación y carga, seleccionando los datos que sirvieron para la aplicación de minería de datos, a continuación, se llevó a cabo la limpieza y transformación de estos que nos permitió construir un modelo apropiado al objetivo que se tiene.

3.6.4 Aquí se utilizó para el proceso de minería de datos, RapidMiner Studio 7.4 a partir de ahí se procesó e implemento unas vistas preseleccionadas de los algoritmos que se aplicaron. Se incluyó el software Microsoft SQL Server.

3.6.5 Así mismo, se utilizó el programa de Office – Excel para el registro y procesamiento de respuestas del cuestionario aplicado para posterior construcción de tabulaciones y gráficos.

3.7 VALIDEZ DEL INSTRUMENTO

El instrumento fue validado a través de la validación de contenido mediante el juicio de expertos, (Anexo 11,12 y 13). En tal sentido, se sometió al juicio de tres profesionales con grado de maestría, quienes revisaron y evaluaron la pertinencia, coherencia, congruencia, suficiencia, etc. Del instrumento, de acuerdo con la ficha de validación propuesta. Los resultados para el instrumento se indican en la tabla siguiente:

Tabla 2: Tabla resumen - validez de instrumento

Instrumento	Experto 1	Experto 2	Experto 3	Promedio
O1	0,90	0,93	0,93	0,92

Tal como se observa, los tres profesionales han validado de manera favorable dicho instrumento, con un promedio de 0,92 lo cual corresponde a una validez Muy Buena, lo que significa que el instrumento está midiendo bien el concepto para el que ha sido preparado.

4 RESULTADOS

4.1 RESULTADOS

A continuación, se presentan los resultados de la metodología aplicada **CRISP (Cross-Industry Standard Process for Data Mining - Proceso Estándar de Trans-Industria para la Minería de Datos)**

A. Fases del modelo de referencia CRISP-DM

a. Comprensión del negocio:

i. Tareas de la fase

1. **Determinar los objetivos del negocio:** La organización el Equipo de Seguros de Salud pertenece a la Dirección Regional de Salud de Piura (DIRESA Piura), la cual como ente rector sanitario tiene como objetivo supervisar y monitorizar el cumplimiento de las prestaciones según las normas técnicas vigentes emitidas por el Ministerio de Salud.
2. **Evaluación de la situación:** La situación actual de la organización está destinada a supervisar y monitorear el cumplimiento de la gratuidad de las prestaciones a los pacientes afiliados al Seguro Integral de Salud en el ámbito de la jurisdicción sanitaria de la Región Piura, para ello se requiere análisis de información.
3. **Determinación de los objetivos de DM:** Para el presente trabajo de investigación se plantearon como objetivos la aplicación de diferentes algoritmos de minería de datos en el Software RapidMiner el cual se eligió por ser el de mayor aceptación.

b. Comprensión de los datos:

i. Tareas de la fase

1. **Recolección de datos iniciales:** las recolecciones de datos iniciales pertenecen a los registros obtenidos de la digitación del Formato Único de Atención (FUA) del Seguro Integral de Salud de las prestaciones brindadas por las Instituciones Prestadoras de Salud (IPRESS) a los afiliados al SIS a la Región Piura durante el año 2016, las cuales están almacenadas en base de datos de Microsoft SqlServer 2014.

2. **Descripción de los datos:** los registros obtenidos de la base de datos pertenecen a los campos de DNI y los diagnósticos definitivos de los pacientes que recibieron atención durante el año 2016 en los diferentes servicios en las IPRESS. El campo DNI es de tipo varchar y con una longitud de 8 el cual pertenece a la tabla (**dbo.I_Atencion**), así mismo se obtuvo el código del diagnóstico que es de tipo varchar y con longitud 5 el cual pertenece a la tabla (**dbo.I_AtencionDia**).

Figura 9: Base de Datos SIS – Estructura de la tabla atención.

Nombre de columna	Tipo de datos	Permitir val...
ate_TipoDocumento	varchar(1)	<input checked="" type="checkbox"/>
ate_dni	varchar(10)	<input checked="" type="checkbox"/>
ate_idCategoriaEESS	varchar(2)	<input checked="" type="checkbox"/>
ate_costoServ	decimal(18, 4)	<input checked="" type="checkbox"/>
ate_costoMedi	decimal(18, 4)	<input checked="" type="checkbox"/>
ate_costoProc	decimal(18, 4)	<input checked="" type="checkbox"/>

Figura 10: Base de Datos SIS – Estructura de la tabla diagnóstico.

Nombre de columna	Tipo de datos	Permitir val...
adia_numregate	int	<input type="checkbox"/>
adia_Idnumreg	int	<input type="checkbox"/>
adia_iNroDia	int	<input type="checkbox"/>
adia_cdx	varchar(1)	<input checked="" type="checkbox"/>
adia_CodDia	varchar(5)	<input type="checkbox"/>
adia_TipoDia	varchar(1)	<input type="checkbox"/>

3. **Exploración de datos:** para la exploración de los datos se utilizó el lenguaje Transact-SQL es el lenguaje que se utiliza para administrar instancias del SQL Server Database Engine (Motor de base de datos de SQL Server), para crear y administrar objetos de base de datos, y para insertar, recuperar, modificar y eliminar datos. Para ello se elaboró consultas SQL:

Figura 11: Base de Datos SIS – Consulta SQL de agrupación.

```

AGROPACION.sql - ...N GUTIERREZ (56)  x  SQLQuery1.sql - Sl...AN GUTIERREZ (55)
USE MINERIA_DATOS

drop table DISLIPIDEMIAS
select distinct MAX(ate_fecatencion) as Fec_Atc ,ATE_DNI,sm.asmi_codsmi,sm.asmi_numero,
CASE ate_idsexo WHEN 0 THEN 'F' else 'M' end as 'Sexo',
ate_edad as Edad,Distrito,ad.adia_coddia as Codigo into DISLIPIDEMIAS
from [SISREG2016].[dbo].[I_Atencion] join [SISREG2016].[dbo].[I_AtencionDia] ad on ate_Idnumreg=ad.adia_numregate
join [SISREG2016].[dbo].[I_AtencionSmi] sm on ate_idnumreg=sm.asmi_numregate
JOIN [MAESTROS_ODSIS_2016].[dbo].[EESS16] ON ate_ideess=codigo_sis
join [MAESTROS_ODSIS_2016].[dbo].[M_CIE10] addx on ad.adia_coddia=addx.codigo
where --ate_idservicio in ('056') AND
convert(char(6), ate_fecatencion, 112) between 201601 and 201612
---CIE-X CONTEMPLADOS PARA DISLIPIDEMIAS
and ad.adia_coddia in ('E78X','E780','E781','E782','E783','E784','E785','E786')
AND adia_tipodia = 1
AND sm.asmi_codsmi IN ('003','004')
and ATE_DNI <> ''
and ate_loteafiins+ate_numregafiins+(CAST(ate_fecatencion as varchar))
IN (SELECT ate_loteafiins+ate_numregafiins+(CAST(MAX(ate_fecatencion) as varchar))
FROM [SISREG2016].[dbo].[I_Atencion] GROUP BY ate_loteafiins, ate_numregafiins)
group by ate_dni,asmi_codsmi,asmi_numero,ate_idsexo,ate_edad,Distrito,adia_CodDia
    
```

4. **Verificación de la calidad de los datos:** para la verificación y aprobación de las cláusulas y atributos se consultó al especialista considerado por la parte técnica para la validación de la calidad de los datos.

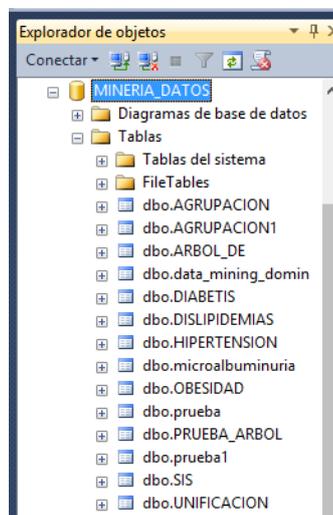
Así mismo, se seleccionó los códigos CIE X, que se procesaron para el presente modelo.

c. **Preparación de datos:**

i. **Tareas de la fase**

1. **Estructuración de los datos:** los datos pre-procesados fueron almacenados en una base de datos de nombre **Mineria_Datos** en la tabla que corresponde a reglas de asociación, para su posterior utilización.

Figura 12: Base de Datos – Minería_Datos.



2. **Integración de los datos:** con referencia a los datos fueron almacenados con las siguientes variables las cuales contienen 30,599 registros que serán utilizados en el procesamiento de minería de datos.

Figura 13: Estructura de la Tabla para aplicación de minería de datos.

Nombre de columna	Tipo de datos	Permitir val...
ATE_DNI	varchar(10)	<input checked="" type="checkbox"/>
DISLIPIDEMIAS	int	<input checked="" type="checkbox"/>
DIABETIS	int	<input checked="" type="checkbox"/>
OBESIDAD	int	<input checked="" type="checkbox"/>
HIPERTENSION	int	<input checked="" type="checkbox"/>
MICROALBU	int	<input checked="" type="checkbox"/>

3. **Formateo de los datos:** el procesamiento de datos incluye el ejecutar comandos update para los valores obtenidos en los registros:

Figura 14: Procesamiento de Variables.

```
UPDATE A SET DISLIPIDEMIAS = '0' FROM AGRUPACION A
UPDATE A SET DIABETIS = '0' FROM AGRUPACION A
UPDATE A SET OBESIDAD = '0' FROM AGRUPACION A
UPDATE A SET HIPERTENSION = '0' FROM AGRUPACION A
UPDATE A SET MICROALBU = '0' FROM AGRUPACION A

UPDATE A SET DISLIPIDEMIAS = '1' FROM AGRUPACION A JOIN DISLIPIDEMIAS D ON A.ate_dni=D.ate_dni
UPDATE A SET DIABETIS = '1' FROM AGRUPACION A JOIN DIABETIS D ON A.ate_dni=D.ate_dni
UPDATE A SET OBESIDAD = '1' FROM AGRUPACION A JOIN OBESIDAD D ON A.ate_dni=D.ate_dni
UPDATE A SET HIPERTENSION = '1' FROM AGRUPACION A JOIN HIPERTENSION D ON A.ate_dni=D.ate_dni
UPDATE A SET MICROALBU = '1' FROM AGRUPACION A JOIN microalbuminuria D ON A.ate_dni=D.ate_dni
```

d. **Modelado:**

Algoritmo: Reglas de Asociación

i. **Tareas de la fase**

1. **Selección de la técnica de modelado:** se utilizó la técnica de reglas de asociación porque nos permitió asociar los registros como variable nominal. Las reglas de asociación se crean mediante el análisis de los datos de frecuencia si / entonces los patrones y utilizando los criterios apoyan y confianza para identificar las relaciones más importantes. El apoyo es una indicación de la frecuencia con la que aparecen los elementos en la base de datos. Confianza indica el número de veces que el if / then se han encontrado para ser verdad.

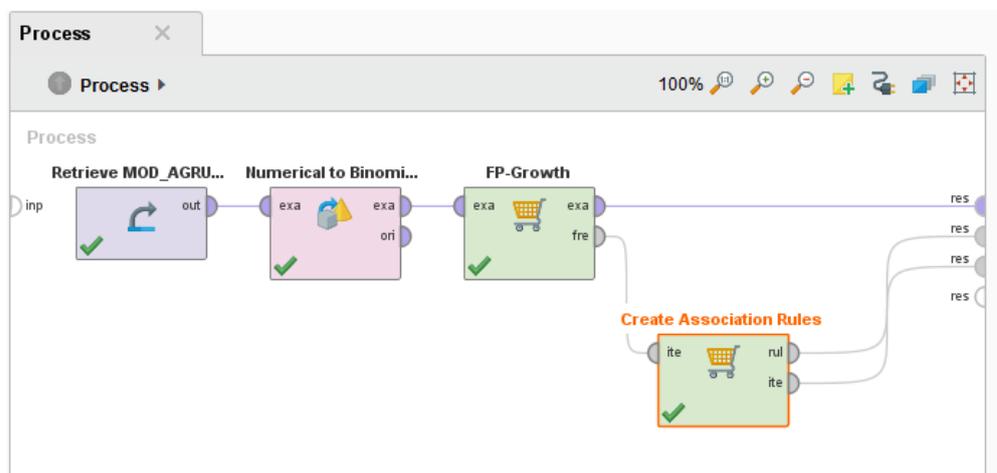
2. **Generación del plan de prueba:** se generó las pruebas necesarias en el software RapidMiner Studio 7.4

Figura 15: Software de Aplicación Minería de Datos: Rapidminer.



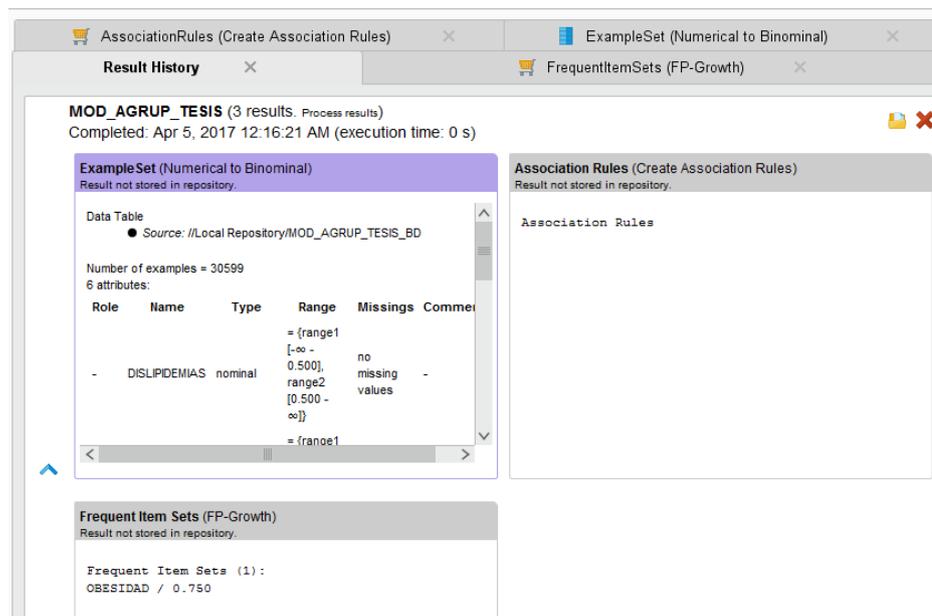
3. **Construcción del Modelo:** la construcción del modelo está en base a operadores en relación con el algoritmo de reglas de asociación:

Figura 16: Construcción de Modelo – Reglas de Asociación (Software Rapidminer).



4. **Evaluación del modelo:** la evaluación del modelo estará a cargo post aplicación del cuestionario para visualizar las variables del presente estudio, y así obtener el grado de aceptación del modelo implementado.

Figura 17: Resultados de Aplicación de modelo de minería de datos.

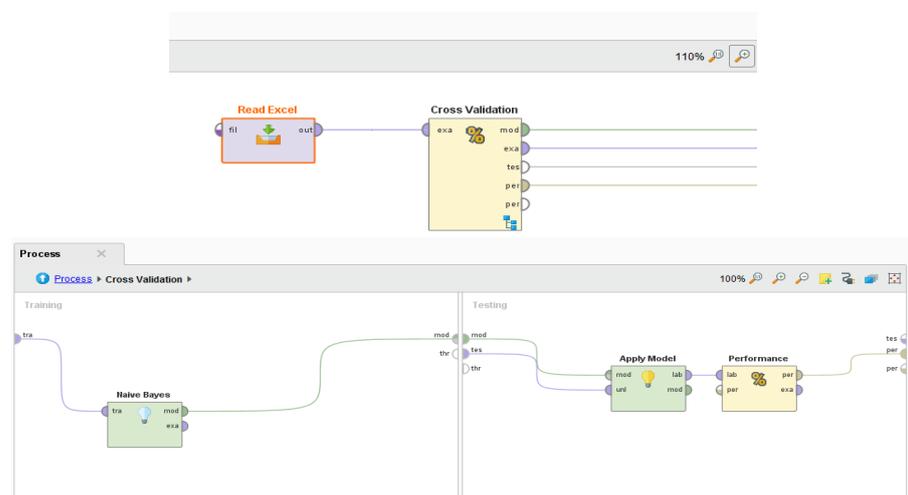


Algoritmo: Bayes Naive

i. Tareas de la fase

1. **Selección de la técnica de modelado:** se utilizó el algoritmo de clasificación Naive Bayes es un clasificador probabilístico. porque nos permitió aprender sobre relaciones de dependencia y causalidad.
2. **Generación del plan de prueba:** se generó las pruebas necesarias en el software RapidMiner Studio 7.4 (Figura 15).
3. **Construcción del Modelo:** la construcción del modelo está en base a operadores en relación con el algoritmo de Bayes Naive:

Figura 18: Construcción de Modelo – Bayes Naive (Software Rapidminer).



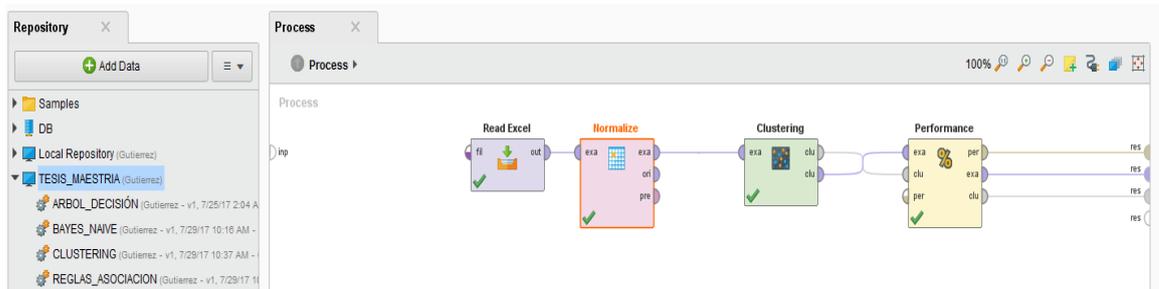
4. **Evaluación del modelo:** la evaluación del modelo estará a cargo post aplicación del cuestionario para visualizar las variables del presente estudio, y así obtener el grado de aceptación del modelo implementado.

Algoritmo: Clusters

ii. Tareas de la fase

1. **Selección de la técnica de modelado:** se utilizó la técnica de clustering porque nos permitió utilizando algoritmos matemáticos agrupar objetos. Usando la información que brindan las variables que pertenecen a cada objeto se mide la similitud entre los mismos, y una vez hecho esto se colocan en clases que son muy similares internamente (entre los miembros de la misma clase) y a la vez diferente entre los miembros de las diferentes clases.
2. **Generación del plan de prueba:** se generó las pruebas necesarias en el software RapidMiner Studio 7.4 (Figura 15).
3. **Construcción del Modelo:** la construcción del modelo está en base a operadores en relación con el algoritmo de clustering:

Figura 19: Construcción de Modelo - Clustering (Software Rapidminer).



4. **Evaluación del modelo:** la evaluación del modelo estará a cargo post aplicación del cuestionario para visualizar las variables del presente estudio, y así obtener el grado de aceptación del modelo implementado.

Algoritmo: Árboles de Decisión

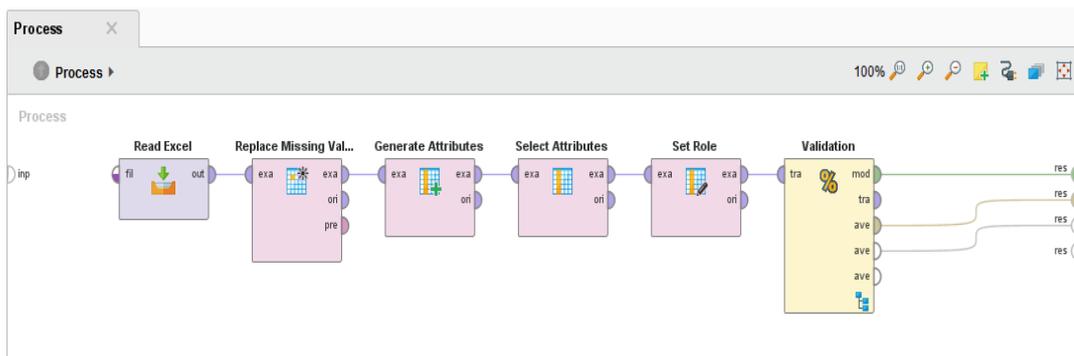
iii. Tareas de la fase

1. **Selección de la técnica de modelado:** se utilizó la técnica de árbol de decisión porque nos permitió a través de una representación esquemática de las alternativas disponible facilita la toma de mejores decisiones, especialmente cuando existen riesgos, costos, beneficios y múltiples opciones. Así mismo, Nos ayudan a tomar la decisión “más acertada”, desde un punto de vista probabilístico, ante un abanico de posibles decisiones.

Proveen una estructura sumamente efectiva dentro de la cual estimar cuales son las opciones e investigar las posibles consecuencias de seleccionar cada una de ellas. También ayudan a construir una imagen balanceada de los riesgos y recompensas asociados con cada posible curso de acción.

2. **Generación del plan de prueba:** se generó las pruebas necesarias en el software RapidMiner Studio 7.4 (Figura 15).
3. **Construcción del Modelo:** la construcción del modelo está en base a operadores en relación con el algoritmo de árbol de decisión.

Figura 20: Construcción de Modelo – Árbol de Decisión (Software Rapidminer).



4. **Evaluación del modelo:** la evaluación del modelo estará a cargo post aplicación del cuestionario para visualizar las variables del presente estudio, y así obtener el grado de aceptación del modelo implementado.

e. **Evaluación:**

i. **Tareas de la fase**

1. ***Evaluación de los resultados:*** los resultados preliminares de la ejecución del modelo se muestran a continuación:

Figura 21: Evaluación del Modelo de Reglas de Asociación (Software Rapidminer).

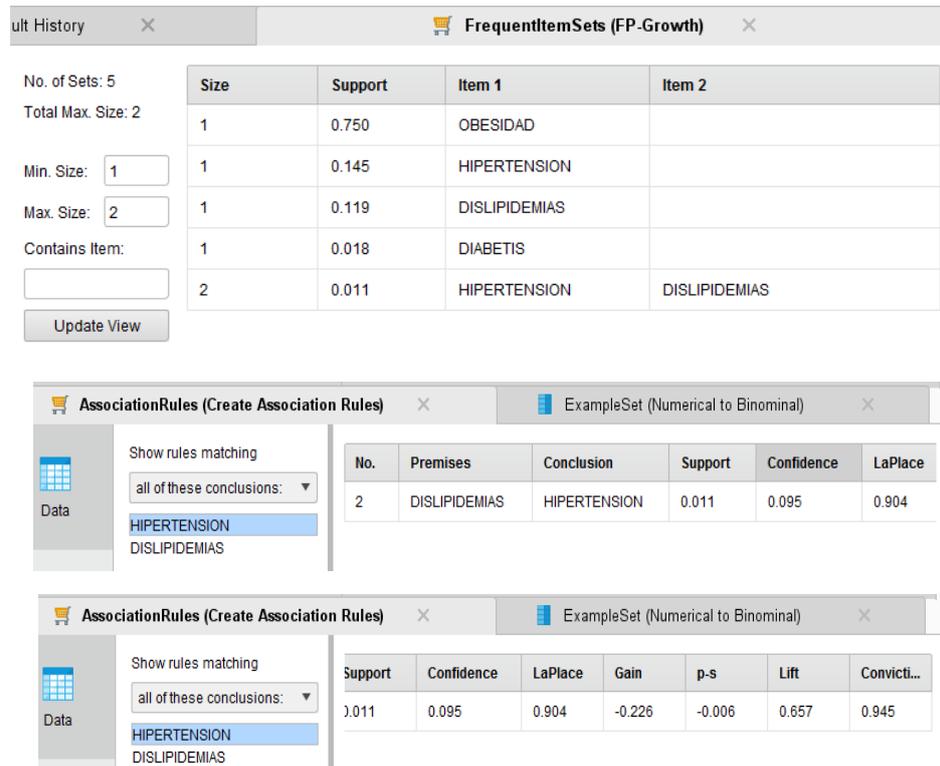


Figura 22: Evaluación del Modelo de Bayes Naive (Software Rapidminer).

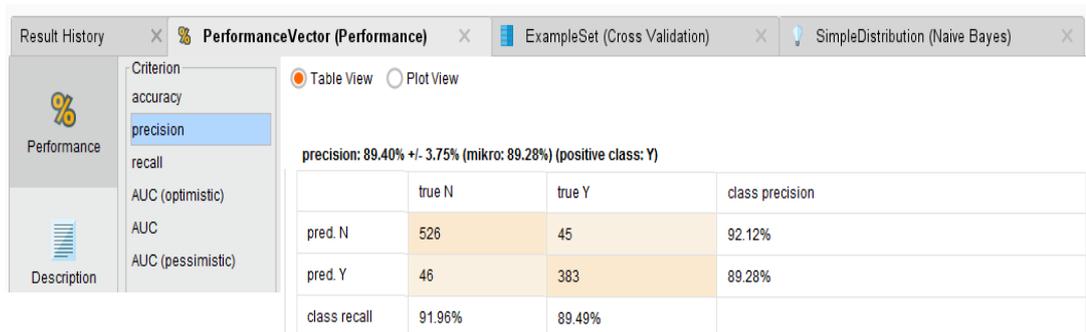


Figura 23: Evaluación del Modelo de Clustering (Software Rapidminer).

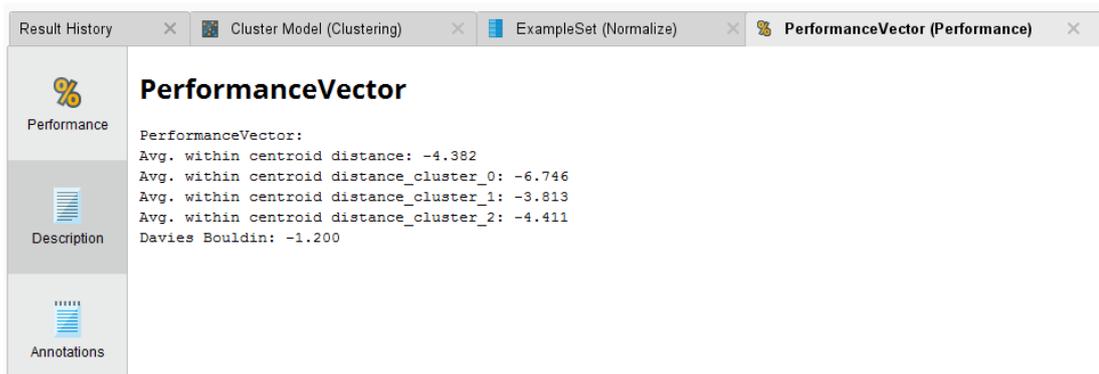
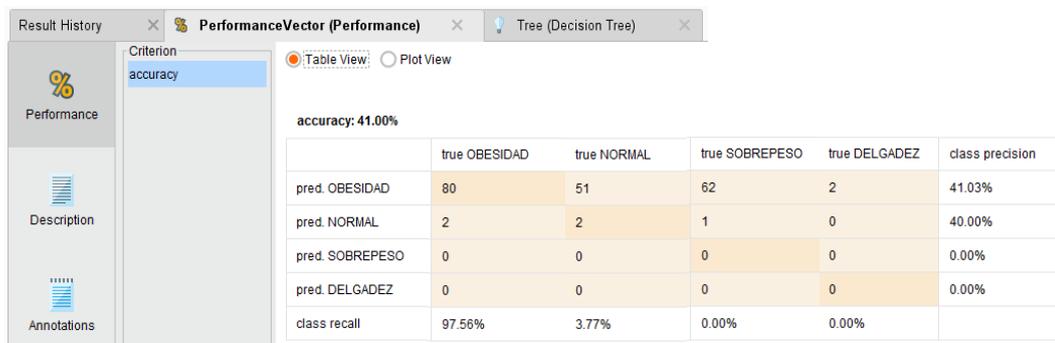


Figura 24: Evaluación del Modelo Árbol de Decisión (Software Rapidminer).



2. **Proceso de revisión:** la revisión del proceso lleva al investigador a comprender los siguientes conceptos:

- ✓ La preparación de los datos conlleva a realizar un análisis previo para la ejecución de Operaciones básicas de manipulación de datos en SQL en el proceso de extracción de datos.
- ✓ El enfoque al problema es esencial porque una vez que los datos están preparados para el análisis, la construcción de los modelos es más fácil.

3. **Determinación de futuras fases:** para el presente proyecto los modelos aplicados de minería de datos han sido satisfactorios, en la medida que se implemente un plan de análisis de la información por la institución la Dirección Regional de Salud Piura y llevar a cabo la aplicación de los resultados podemos afirmar que tanto en la práctica con la muestra y la realidad los modelos son válidos.

f. Desarrollo:

i. Tareas de la fase

1. *Plan de implementación:* para el presente trabajo se propone crear planes de desarrollo, de difusión o implementación posteriores a la aplicación de minería de datos como estrategia para la organización en este caso la Dirección Regional de Salud Piura.

2. *Informe Final:* se propone un análisis de resultados que detallara en el numeral (4.2 Análisis de Resultados), así mismo esta tarea de la metodología aterriza en la recomendación (número 1) del presente trabajo.

4.2 ANÁLISIS DE RESULTADOS DE ALGORITMOS

4.2.1. Reglas de Asociación:

El descubrimiento de asociación encuentra relaciones que aparecen conjuntamente a un acontecimiento y la secuencia la asocia al tiempo.

4.2.2. Bayes Naive

La dependencia y causalidad de las variables, que fueron Diabetes e Índice de Masa Corporal IMC, en relación con siete (07) variables. Modelo de distribución para el atributo Resultado: Clase N (0,572) 8 distribuciones y Clase Y (0,428) 8 distribuciones.

4.2.3. Clústeres

Para agrupar objetos la utilización de este algoritmo nos dio como resultado tres (03) clúster: C0 (90), C1 (399), C2 (511) de un total de 1000. Lo que significa la agrupación de mil resultados de pacientes en relación con el procedimiento de diabetes siendo el resultado positivo o negativo con relación a 08 (variables)

4.2.4. Árbol de Decisión

La decisión más acertada que brinda este algoritmo, conlleva analizar 12 variables que analizaba en relación con el Índice de Masa Corporal IMC que relación tenía según la construcción de un árbol de decisión:

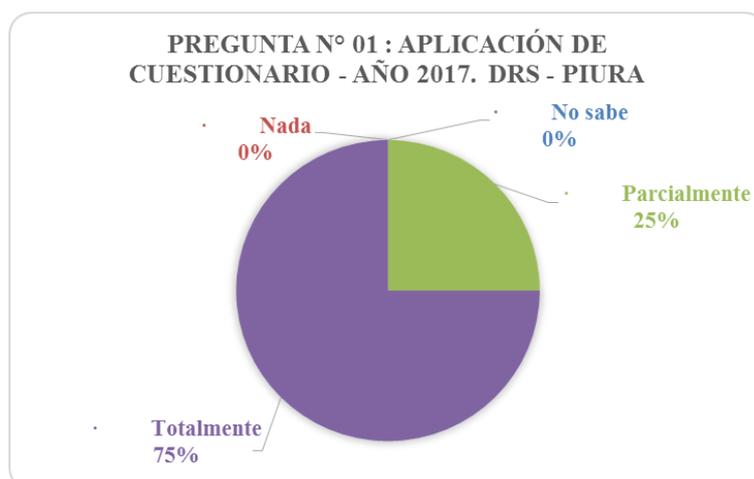
obteniendo un resultado que la variable que más influye es la edad, seguida de la educación y estatus. Así mismo, la clase de precisión fue mayor a 40% para el resultado de obesidad y normal (IMC).

4.3. ANÁLISIS DE RESULTADOS DE INSTRUMENTO DE EVALUACIÓN

4.3.1. Se aplicó una encuesta a Coordinadores, Personal Asistencial, Médicos Auditores y Personal de Sistemas (Ver Anexo N° 10).

- a. Pregunta N° 01: ¿Está de acuerdo con la implementación de una metodología para el análisis de los registros de prestaciones que permita obtener conocimiento para la toma de decisiones?

Gráfico 1: Gráfico de la pregunta N° 01.

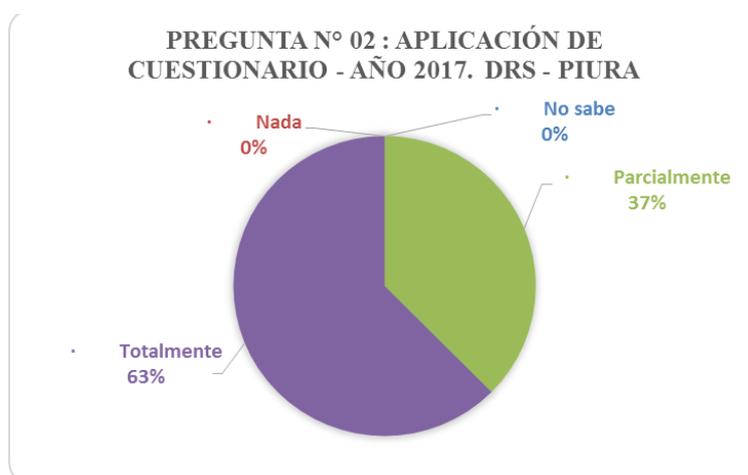


ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 75% responden que mejoraría **TOTALMENTE** es indispensable el uso de una metodología de extracción de conocimientos que analice automática y rápidamente grandes cantidades de datos para obtener información valiosa del catastro de establecimientos, el 25% responden que mejoraría **PARCIALMENTE**, mientras que el 0% responde que no mejoraría **NADA** y **NO SABE**.

- b. Pregunta N° 02: ¿Cree usted que la implementación de la metodología de extracción de conocimientos optimizaría el tiempo de trabajo para el análisis de las prestaciones de salud?

Gráfico 2: Gráfico de la pregunta N° 02.

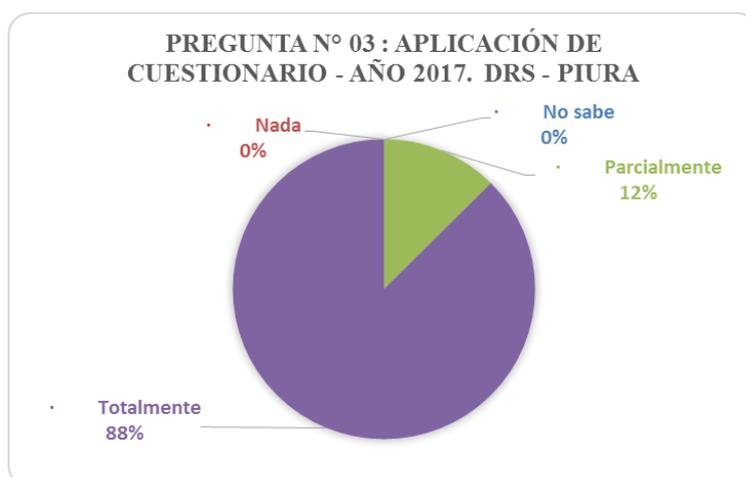


ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 63% responden que TOTALMENTE la implementación de la metodología de extracción de conocimientos optimizará el tiempo de trabajo de los entes de salud en el proceso de análisis de información de las prestaciones, el 37% responden que PARCIALMENTE, mientras que el 0% responde que NADA y NO SABE.

- c. Pregunta N° 03: ¿Considera usted útil contar con el uso de una herramienta como la metodología, que permita conocer nueva información?

Gráfico 3: Gráfico de la pregunta N° 03.

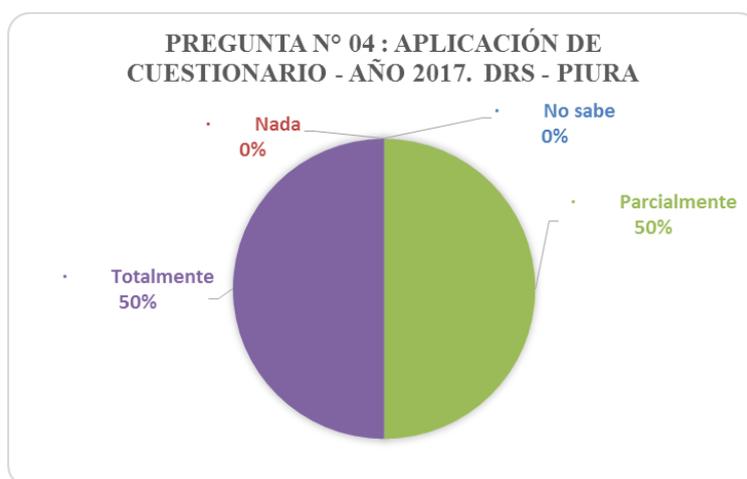


ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 88% responden que TOTALMENTE será útil contar con el uso de una herramienta como la metodología que permita conocer los posibles establecimientos que incurrirán en mora, el 12% responden que PARCIALMENTE, mientras que el 0% responde que NADA y NO SABE.

- d. Pregunta N° 04: ¿Cree usted que la información obtenida puede ayudar a optimizar los procesos y mejorar el proceso de toma de decisiones?

Gráfico 4: Gráfico de la pregunta N° 04.

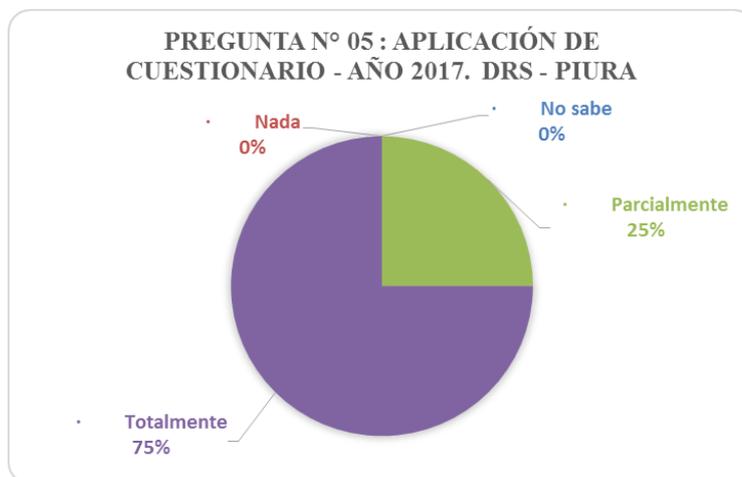


ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 50% responden que TOTALMENTE la información obtenida puede ayudar a al descubrimiento y análisis de información, el 50% responden que PARCIALMENTE, mientras que el 0% responde que NADA y NO SABE. Lo que se evidencia que debería tener implementarse o tener una adaptación de la fase de elicitación de requisitos, para que los resultados sean precisos.

- e. Pregunta N° 05: ¿Cree usted que contar con información detallada del análisis de los registros de prestaciones ayudara a los entes directivos a la toma de decisiones?

Gráfico 5: Gráfico de la pregunta N° 05.

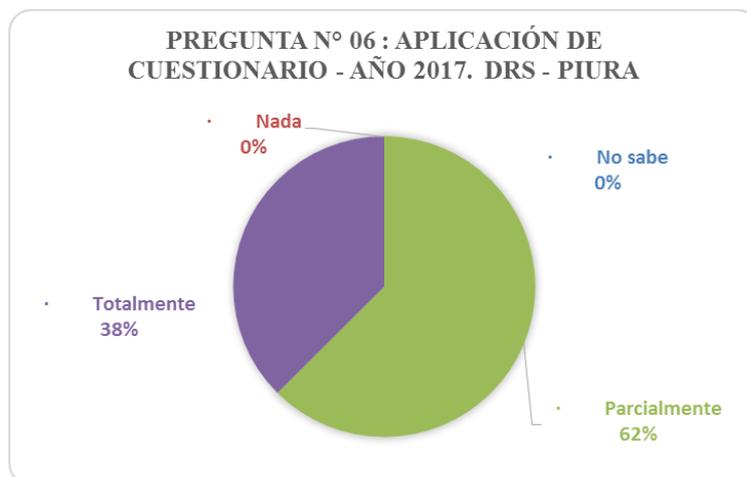


ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 75% responden que TOTALMENTE al contar con información detallada de los establecimientos de salud ayudará a las autoridades a tomar decisiones previas sobre las prestaciones de salud, el 25% responden que PARCIALMENTE, mientras que el 0% responde que NADA y NO SABE.

- f. Pregunta N° 06: ¿Cree usted que mejoraría el desempeño en el trabajo contar con un análisis profundo con el conocimiento adquirido?

Gráfico 6: Gráfico de la pregunta N° 06.

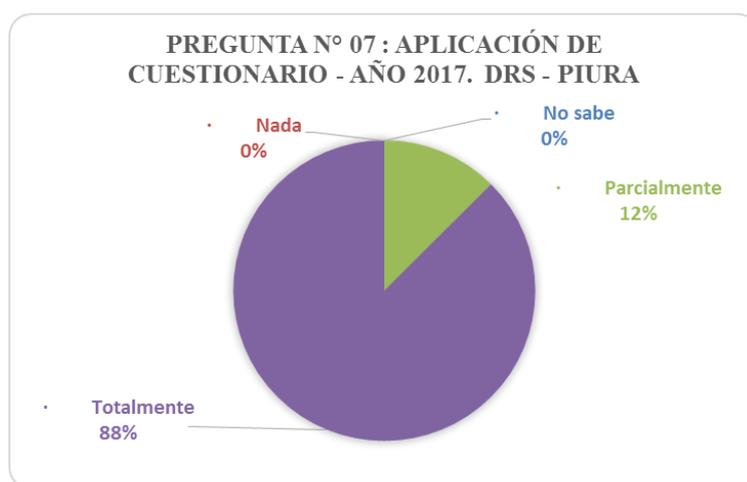


ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 38% responden que **TOTALMENTE** mejorará el desempeño en el trabajo al contar con análisis descriptivo de las prestaciones de salud, el 62% responden que **PARCIALMENTE**, mientras que el 0% responde que **NADA** y **NO SABE**.

- g. Pregunta N° 07: ¿Cree usted que la información procesada automática y ordenadamente de los registros de las prestaciones puede ayudar a realizar un mejor análisis para la toma de decisiones?

Gráfico 7: Gráfico de la pregunta N° 07.



ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 88% responden que **TOTALMENTE** la información procesada automática y ordenadamente de los establecimientos puede ayudar a realizar un mejor análisis y toma de decisiones, el 12% responden que **PARCIALMENTE**, mientras que el 0% responde que **NADA** y **NO SABE**.

- h. Pregunta N° 08: ¿Cree usted que los resultados que arroje la minería de datos ayuden a determinar políticas de control en el análisis de la nueva información?

Gráfico 8: Gráfico de la pregunta N° 08.



ANÁLISIS E INTERPRETACIÓN

Se obtiene que el 75% responden que **TOTALMENTE** los resultados que arroje el proceso de implementación de metodología y las fases incluida la minería de datos ayuden a determinar políticas de control en el análisis de la nueva información, el 25% responden que **PARCIALMENTE**, mientras que el 0% responde que **NADA** y **NO SABE**:

4.4. PROPUESTA MEJORA

Con el presente trabajo se pretende contribuir a la metodología de origen KDD, así mismo, con lo propuesta en las implementaciones que tuvieron la metodología CRISP – DM y SEMMA que se compararon en el capítulo III del presente trabajo. Para ello debo mencionar que, la propuesta es agregar a las fases ya conocidas la **Elicitación de Requisitos –ER–** que tiene que ver con el desarrollo de proyectos y tiene un impacto muy alto en el diseño y en las demás fases del ciclo de vida de un producto.

Además, la calidad de la elicitación determina la exactitud de la retroalimentación al cliente acerca de la integridad y validez de los requisitos. Debido a que esta fase es crítica y de alto impacto en el proyecto, es muy importante que la labor de elicitar se realice lo más cercano posible a la “perfección”. Teniendo en cuenta las diferentes características de los proyectos, en este trabajo se propone adicionar a la metodología la cual se describirá en el numeral 4.4 diseño de la metodología.

4.5. DISEÑO DE LA METODOLOGÍA PROPUESTA

Para el presente trabajo se propone la adaptación de la metodología utilizada (CRISP – DM) incluidas las tareas por cada fase (Ver Anexo N° 03 y 08). La investigación tiene como objetivo dar una propuesta de una metodología de extracción de conocimientos, como puede observarse en con mi aporte agregaría una fase adicional: **La Elicitación de Requisitos**, con las tareas de Entrevistas, lluvia de ideas, y la utilización de escenarios.

4.5.1. Elicitación de Requisitos (Ver Anexo N° 09):

- **Entrevista: las fases que corresponderían son:** Preparación de entrevistas, Realización de entrevistas, Análisis de las entrevistas,
- **Lluvia de Ideas: las fases que corresponderían son:** preparación, generación, consolidación, documentación.
- **Utilización de Escenarios: las fases que corresponderían son:** diagramas, relaciones y organización.

CONCLUSIONES Y RECOMENDACIONES

4.6. CONCLUSIONES

- 4.6.1. Con relación al objetivo general de este trabajo de investigación, concluyo que para realizar una metodología en minería de datos debemos tener en cuenta lo siguiente: los datos (base de datos) el insumo de entrada, determinar los modelos que se aplicaran, elegir una metodología que permita elaborar, definir y sistematizar el conjunto de técnicas, métodos y procedimientos que se deben seguir y la utilización de un Software que sirva como soporte para operar los modelos.
- 4.6.2. La propuesta de la metodología de análisis de información para extracción de conocimiento provee resultados confiables y demostró a través de los modelos, es factible la adaptación de esquemas de minería de datos.
- 4.6.3. La minería de datos aconteció como un proceso, que extrae informaciones esenciales de grandes bases de datos sin requerir de ningún conocimiento previo, para tomar decisiones y aprender sobre el fenómeno.
- 4.6.4. En el ámbito de la salud, resulta interesante e importante encontrar patrones ocultos de consumo de medicamentos, resultados en procedimientos aplicados a pacientes esto conlleva a poder explorar nuevos horizontes.
- 4.6.5. La decisión de seleccionar una solución de data mining no es una tarea simple, porque existen diferentes modelos y además se necesita consultar diferentes fuentes confiables y fidedignas del tema.
- 4.6.6. La minería de datos se presenta como la tecnología emergente, así mismo la metodología empleada abre a través de sus resultados nuevas oportunidades de negocio e implica cuidar los detalles en los resultados porque estos convergen en la toma de decisiones.
- 4.6.7. Para la aplicación de las técnicas de minería de datos realizó un proceso de depuración de la información proporcionada por la institución pública de la Dirección Regional de Salud Piura. Se procedió a refinar los datos, y así alcanzar un mejor partido de la información para identificar los patrones.

4.7. RECOMENDACIONES

- 4.7.1. Socializar los resultados teniendo en cuenta el público objetivo, es decir receptores del informe. Así mismo, podría incluir la descripción detallada del problema, procedimientos, costos, resumen de resultados y recomendaciones para futuros trabajos de investigación.
- 4.7.2. Investigar nuevos algoritmos de minería de datos con la finalidad de evaluar la posibilidad y beneficio de implementarlos en el contexto del proyecto
- 4.7.3. Implementar un Data Warehouse, para que sea un paso siguiente en la evaluación del proyecto sería la implementación de un almacén de datos a fin de contar con datos históricos que enriquezcan los procesos de minería de datos.
- 4.7.4. Para dar continuidad al trabajo aquí expuesto se propone analizar otras diferentes metodologías y técnicas de reconocimiento de patrones que puedan aportar una mejor forma de obtener conocimiento para la toma de decisiones.

5. REFERENCIAS BIBLIOGRAFICAS

1. SUSALUD. SUSALUD. [Online]. 2016 [Citado: 2016 10 01. Disponible en: <http://portales.susalud.gob.pe/web/portal/seguro-integral-de-salud-sis>
2. PERUANO PDE. PORTAL DEL ESTADO PERUANO. [Online]. 2016 [Citado: 2016 10 01. Disponible en: http://www.peru.gob.pe/directorio/pep_directorio_detalle_institucion.asp?cod_institucion=12787.
3. SUSALUD. SUSALUD. [Online]. 2016 [Citado: 2016 10 01. Disponible en: <http://portales.susalud.gob.pe/web/mundo-ipress/instituciones-prestadoras-servicios>.
4. Robles Aranda Y, Sotolongo AR. Scielo. [Online]. 2013 [Citado: 2016 10 20. Disponible en: <http://www.scielo.br/pdf/jistm/v10n2/1807-1775-jistem-10-02-00389.pdf>.
5. Padua LM. digital.bl.fcen.uba.a. [Online]. 2014 [Citado: 2016 10 02. Disponible en: http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_5612_Padua.pdf.
6. Inc. SI. [Software de Aplicación]. 2016 [Citado: 2016 10 02. Disponible en: <http://support.sas.com/software/products/miner/>.
7. Cruz Flores JA. colposdigital. [Online]. 2014 [Citado: 2016 10 02. Disponible en: http://colposdigital.colpos.mx:8080/jspui/bitstream/handle/10521/2521/Flores_Cruz_JA_MC_Computo_Aplicado_2014.pdf?sequence=1&isAllowed=y.
8. Keller Gladys RC, Angela Lucia MP. Pontificia Universidad Católica del Perú. [Online]. 2011 [Citado: 2016 10 20. Disponible en: http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/931/RODRIGUEZ_CABANILLAS_KELLER_INTELIGENCIA_NEGOCIOS_ELECTRODOMESTICOS.pdf;jsessionid=D8BBF37C464BB85A6828EE1A67DF3B19?sequence=1.
9. Alcántara Morí AA. Formulación de Minería de Datos para la Empresa Distribuidora de Productos Espinoza Aguilar S.A. [Online]. 2012 [Citado: 2016 10 05. Disponible en: <http://e.exam-10.com/ekonomika/15685/index.html>.

10. Microsoft. Conceptos de Minería de Datos. [Online]. 2015 [Citado: 2016 10 02].
Disponible en: <https://msdn.microsoft.com/es-es/library/ms174949.aspx>.
11. La Red DM, Martínez DLR. exa.unne.edu.a. [Online]. 2006 [Citado: 2016 09 30].
Disponible en:
http://exa.unne.edu.ar/informatica/SO/Mineria_Datos_Vallejos.pdf.
12. UF, GPS, PS. From Data Mining to. [Online]. 1996 [Citado: 2016 09 30].
Disponible en:
<https://www.aaai.org/ojs/index.php/aimagazine/article/viewFile/1230/1131>.
13. Félix LCM. Data mining: torturando a los datos hasta que confiesen. [Online].
[Citado: 2016 09 30. Disponible en:
<http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.pdf>.
14. Rive PEC. Universidad Técnica Paritular de Loja. [Online]. 2009 [Citado: 2016
09 30. Disponible en:
<http://dspace.utpl.edu.ec/bitstream/123456789/13055/1/Chamba%20Rivera%20Patricia%20Elvira.pdf>.
15. Vergaray AD. Data Mining - Minería de Datos. Primera Edición ed. EIRL EEM,
editor. Lima: Macro; 2016.
16. Day g. eHow. [Online]. 2014 [Citado: 2016 09 30. Disponible en:
http://www.ehowenespanol.com/mejores-10-algoritmos-mineria-datos-info_295108/.
17. Microsoft. Algoritmo de árboles de decisión de Microsoft. [Online]. 2016 [Citado:
2016 09 30. Disponible en: <https://msdn.microsoft.com/es-es/library/ms175312.aspx>.
18. Bouza N C, Santiago A. Researchgate. [Online]. 2014 [Citado: 2016 10 05].
Disponible en:
https://www.researchgate.net/profile/Carlos_Bouza/publication/268516570_LA_MINERIA_DE_DATOS_ARBOLES_DE_DECISION_Y_SU_APLICACION_EN_ESTUDIOS_MEDICOS/links/546f70760cf2d67fc03113ef.pdf?inViewer=0&pdfJsDownload=0&origin=publication_detail.

19. Microsoft. Algoritmo Bayes naive de Microsoft. [Online]. 2016 [Citado: 2016 09 30. Disponible en: [https://technet.microsoft.com/es-es/library/ms174806\(v=sql.105\).aspx](https://technet.microsoft.com/es-es/library/ms174806(v=sql.105).aspx).
20. Spain B. Bi Latino. [Online]. 2008 [Citado: 2016 10 08. Disponible en: <http://www.bi-spain.com/articulo/50774/spss/el-nuevo-software-predictivo-spss-clementine-120-combina-informacion-estructurada-y-desestructurada>.
21. SAS Institute Inc. SCD. SAS. [Online]. 2011 [Citado: 2016 10 15. Disponible en: <https://support.sas.com/documentation/cdl/en/emgsj/64144/PDF/default/emgsj.pdf>.
22. Derrac J, Alberto F, Julián L. UHU. [Online]. 2010 [Citado: 2016 10 15. Disponible en: <http://www.uhu.es/estylf2010/trabajos/SS05-05.pdf>.
23. MicroSystem. RapidMiner. [Online]. 2011 [Citado: 2016 10 10. Disponible en: <http://www.microsystem.cl/plataformas/rapidminer/>.
24. ISA. INGENIERÍA DE SISTEMAS Y AUTOMÁTICA. [Online]. 2016 [Citado: 2016 10 15. Disponible en: <http://isa.umh.es/asignaturas/crss/tutorialWEKA.pdf>.
25. Piatetsky G. Kdnuggets. [Online]. 2016 [Citado: 2017 05 15. Disponible en: <http://www.kdnuggets.com/2016/12/poll-analytics-data-mining-data-science-applied-2016.html>.
26. Salud SIId. Seguro Integral de Salud. [Online]. 2016 [Citado: 2016 09 30. Disponible en: <http://www.sis.gob.pe/nuevoPortal/index.html>.
27. Palella Stracuzzi S, FM. Metodología de la investigación cualitativa. 2nd ed. Palella Stracuzzi S, editor.: FEDUPEL; 2006.
28. Eumed. eumed. [Online].; 2014 [Citado: 2016 09 29. Disponible en: http://www.eumed.net/tesis/doctorales/2012/mirm/cualitativo_cuantitativo_mixto.html.

ANEXOS

ANEXO N° 01

Tabla 3: MATRIZ DE CONSISTENCIA

PROBLEMA	OBJETIVO	HIPÓTESIS	VARIABLES	INDICADORES	METODOLOGIA
¿En qué medida la propuesta de una metodología de extracción de conocimientos mejorará el Análisis del Conocimiento en relación con las prestaciones del Seguro Integral de Salud en la región Piura en el año 2016?	•Desarrollar una metodología de extracción de conocimientos que permita descubrir conocimientos relevantes, partiendo de datos de prestaciones del Seguro Integral de Salud en la región Piura en el año 2016.	La propuesta de una metodología de extracción de conocimientos mejorará el proceso de análisis del conocimiento a partir de las prestaciones del Seguro Integral de Salud en la Región Piura en el año 2016.	Variable 1: Metodología de Extracción de Conocimientos Variable 2: Proceso de Análisis	Base de Datos (Atención, Medicamentos) Análisis Modelos estadísticos, Cumplimiento de indicadores sanitarios	Aplicación Encuesta, Entrevista

ANEXO N° 02

INSTRUMENTO PARA LA ENCUESTA

Encuesta dirigida al personal del Equipo de Seguros de Salud y del Seguro Integral de Salud (UDR).

Objetivo: Verificar la necesidad de la implementación de una metodología de extracción de conocimientos para el análisis de datos de las prestaciones de salud del Seguro Integral de Salud en la Región Piura en el año 2016.

Señores: Estoy trabajando en un estudio que servirá para elaborar una tesis de maestría acerca de la propuesta de una metodología de extracción de conocimientos para el análisis de datos de las prestaciones de salud del Seguro Integral de Salud en la Región Piura en el año 2016. Sus respuestas serán confidenciales y anónimas.

N.	ITEM	RESPUESTAS	COD.
1	¿Está de acuerdo con la implementación de una metodología para el análisis de los registros de prestaciones que permita obtener conocimiento para la toma de decisiones?	<ul style="list-style-type: none">• No sabe• Nada• Parcialmente• Totalmente	1. () 2. () 3. () 4. ()
2	¿Cree usted que la implementación de la metodología de extracción de conocimientos optimizaría el tiempo de trabajo para el análisis de las prestaciones de salud?	<ul style="list-style-type: none">• No sabe• Nada• Parcialmente• Totalmente	1. () 2. () 3. () 4. ()
3	¿Considera usted útil contar con el uso de una herramienta como la metodología, que permita conocer nueva información?	<ul style="list-style-type: none">• No sabe• Nada• Parcialmente• Totalmente	1. () 2. () 3. () 4. ()
4	¿Cree usted que la información obtenida puede ayudar a optimizar los procesos y mejorar el proceso de toma de decisiones?	<ul style="list-style-type: none">• No sabe• Nada• Parcialmente• Totalmente	1. () 2. () 3. () 4. ()
5	¿Cree usted que contar con información detallada del análisis de los registros de prestaciones ayudara a los entes directivos a la toma de decisiones?	<ul style="list-style-type: none">• No sabe• Nada• Parcialmente• Totalmente	1. () 2. () 3. () 4. ()
6	¿Cree usted que mejoraría el desempeño en el trabajo contar con un análisis profundo con el conocimiento adquirido?	<ul style="list-style-type: none">• No sabe• Nada• Parcialmente• Totalmente	1. () 2. () 3. () 4. ()
7	¿Cree usted que la información procesada automática y ordenadamente de los registros de las prestaciones puede ayudar a realizar un mejor análisis para la toma de decisiones?	<ul style="list-style-type: none">• No sabe• Nada• Parcialmente• Totalmente	1. () 2. () 3. () 4. ()
8	¿Cree usted que los resultados que arroje la minería de datos ayuden a determinar políticas de control en el análisis de la nueva información?	<ul style="list-style-type: none">• No sabe• Nada• Parcialmente• Totalmente	1. () 2. () 3. () 4. ()

ANEXO N° 03

Tabla 4: CUADRO DE LA PROPUESTA DE LA METODOLOGÍA

Fases	PROPUESTA DE METODOLOGÍA
Elicitacion de Requisitos	Definir requisitos: Definir los productos entregables oficialmente al cliente como parte del desarrollo en fechas previamente acordadas, mientras que los no entregables son productos internos al desarrollo que no se entregan al cliente.
Análisis y Comprensión del negocio	Comprensión del negocio: en esta fase se determinan los objetivos y requerimientos del proyecto desde una perspectiva del negocio, definiendo el problema de minería y el plan de trabajo.
Selección y Separación de los datos	Comprensión de los datos: fase que consiste en la recolección de datos que se utilizarán en el proyecto y la familiarización con los mismos. En esta etapa es posible el surgimiento de las primeras hipótesis acerca de la información que podría estar oculta. Preparación de los datos: comprende aquellas actividades de tratamiento de los datos para construir la vista minable o conjunto de datos final sobre el cual se aplicarán las técnicas de minería (Tabla N° 3).
Modelado	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 (Tabla N° 4).
Evaluación	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 y si el nuevo conocimiento será implementado, es decir, si se pasará a la próxima etapa (Tabla N° 5).
Implementación	Implementación: consiste en la comunicación e implementación del nuevo conocimiento, el cual debe ser representado de forma entendible para el usuario (Tabla N° 6).

Fuente : Elaboracion propia.

ANEXO N° 04

Tabla 5: Fase 1 CRISP-DM: Entendimiento del Negocio

Tarea	Descripción	Salidas
1.1. Determinar los objetivos del negocio	Entender y establecer cuáles son los objetivos que el cliente pretende alcanzar, desde una perspectiva del negocio.	<ul style="list-style-type: none"> • Background, con la información que se conoce sobre la situación actual de la organización, incluyendo una descripción general del problema y la solución actual para el mismo (si es que existe). • Objetivos del negocio, identificando los objetivos principales del cliente. • Criterios de éxito, describiendo los resultados esperados desde una perspectiva de negocio.
1.2. Evaluar la situación	Profundizar en la evaluación de la situación actual del negocio. Analizar con mayor profundidad las restricciones y factores que se deben tener en cuenta para el proyecto.	<ul style="list-style-type: none"> • Inventario de recursos, donde deberán incluirse los recursos disponibles (como los recursos humanos, fuentes de datos, hardware y software). • Lista de requerimientos del proyecto, supuestos y restricciones que se han detectado. • Riesgos y planes de contingencia. Consiste en la identificación de los potenciales riesgos para el proyecto y la planificación de las acciones reactivas que se llevarán a cabo (planes de contingencia). • Glosario con terminología relevante para el proyecto. En el mismo deberá incluirse un glosario de terminología del negocio y otro de minería de datos. • Análisis costo-beneficio del proyecto.
1.3. Determinar los objetivos de la minería de datos	Los objetivos del negocio se describen en términos organizacionales, en cambio los objetivos de minería de datos describen los objetivos del proyecto en "términos técnicos". Es decir, si un objetivo de negocio es aumentar el volumen de ventas, el objetivo de minería de datos podría ser el "agrupamiento" de los clientes para la promoción de nuevas campañas publicitarias.	<ul style="list-style-type: none"> • Objetivos de minería de datos, describiendo los resultados previstos del proyecto que permiten el logro de los objetivos de negocio. • Definir un criterio de éxito para el proyecto de minería. Especificar las condiciones bajo las cuales se aceptarán los resultados obtenidos.
1.4. Crear un plan para el proyecto de minería de datos.	Crear una planificación para el proyecto de minería, el cual debe ser consistente con los objetivos planteados.	<ul style="list-style-type: none"> • Plan de proyecto: listar las tareas que deben ser ejecutadas, duraciones y recursos necesarios, así como sus entradas y salidas. El plan del proyecto es un documento dinámico, que debe ser revisado y ajustado al final de cada fase. • Evaluación inicial de técnicas y herramientas de minería que podrían ser utilizadas en el proyecto.

ANEXO N° 05

Tabla 6: Fase 2 CRISP-DM: Entendimiento de los datos

Tarea	Descripción	Salidas
2.1. Recolectar los datos iniciales	Recolectar todos los datos necesarios especificados en la lista de recursos del proyecto.	<ul style="list-style-type: none">• Reporte de recolección inicial de datos, donde se detalla la forma en la que han sido obtenidos los conjuntos de datos y los problemas que han surgido en el proceso.
2.2. Describir los datos	Describir en líneas generales los datos recolectados.	<ul style="list-style-type: none">• Descripción de los datos, incluyendo el formato de los mismos y su tamaño (como cantidad de registros y variables).
2.3. Explorar los datos	Realizar una exploración de los datos, observando la distribución y comportamiento de las variables con mayor relevancia. En esta fase es conveniente el uso de técnicas simples de análisis estadístico.	<ul style="list-style-type: none">• Reporte inicial de exploración de datos, donde se expongan los resultados del análisis y las hipótesis iniciales con su impacto en el proyecto.
2.4. Verificar la calidad de los datos.	Examinar la calidad de los datos, incluyendo un análisis de su completitud, de potenciales errores en los mismos y de los datos ausentes.	<ul style="list-style-type: none">• Reporte de calidad de los datos, donde se documente el análisis de calidad efectuado y se propongan potenciales soluciones a los problemas encontrados.

ANEXO N° 06

Tabla 7: Fase 3 CRISP-DM: Preparación de los datos

Tarea	Descripción	Salidas
3.1. Seleccionar los datos	Seleccionar los datos que serán utilizados para el análisis. En esta etapa se debe seleccionar con qué atributos (columnas) y con qué observaciones (filas o registros) se trabajará. La selección debe estar justificada.	<ul style="list-style-type: none"> Justificación de la selección. Documento donde se justifiquen las causas por las cuales se incluyeron y excluyeron los datos.
3.2. Limpieza de datos	Es una etapa que tiene por objetivo mejorar la calidad de los datos. En ella se deberán tomar decisiones sobre los problemas de calidad encontrados en los mismos, como datos ausentes o datos anómalos.	<ul style="list-style-type: none"> Reporte de limpieza de datos, donde se incluyan las decisiones tomadas sobre los problemas de calidad de los datos (reportados en la fase "2.4 Verificar la calidad de los datos")
3.3. Construcción de los datos	En esta fase se lleva a cabo la construcción de nuevos datos, derivados de los disponibles, que son importantes para el análisis. Estos nuevos datos pueden ser, por ejemplo, atributos calculados o atributos transformados.	<ul style="list-style-type: none"> Atributos derivados. Estos atributos se calculan a partir de otros atributos del mismo registro. Por ejemplo: $edad_cliente = fecha_venta - fecha_nacimiento$. Registros creados. Estos nuevos registros se crean cuando son necesarios en la fase posterior de modelado.
3.4. Integrar los datos	Consiste en la integración de datos provenientes de diferentes tablas o registros.	<ul style="list-style-type: none"> Datos combinados. Resultan de integrar la información de dos o más tablas que tienen diferente información de las mismas observaciones. Por ejemplo, la integración de los datos personales y los datos de las atenciones efectuadas a un paciente en un centro de salud. En esta fase se incluye el cálculo de agregaciones, donde se calculan nuevos datos resumiendo información de diferentes tablas y registros. Siguiendo con el ejemplo del centro de salud, podríamos integrar en un solo registro los datos personales del paciente, el total de atenciones efectuadas, y el promedio anual de consultas médicas realizadas.
3.5. Formatear los datos	Esta etapa se refiere al cambio que debe realizarse en el formato de los datos (pero no en su significado) por los requisitos de las técnicas de modelado elegidas. Por ejemplo, el formato de las fechas o el ordenamiento del set de datos.	<ul style="list-style-type: none"> Conjunto de datos reformateados.

ANEXO N° 07

Tabla 8: Fase 4 CRSIP-DM: Modelado

Tarea	Descripción	Salidas
4.1 Seleccionar la técnica de modelado	<p>Consiste en seleccionar qué técnica de minería de datos será utilizada. Por ejemplo, en un caso donde se ha definido un problema de agrupamiento (clustering), se puede decidir utilizar el algoritmo k-medias.</p> <p>Si se ha optado por el uso de múltiples técnicas, se debería repetir esta tarea para cada una.</p>	<ul style="list-style-type: none"> • Técnica de modelado. Documentar la técnica de modelado con la que se trabajará. • Supuestos del modelo. Algunas técnicas asumen supuestos sobre el conjunto de datos, como por ejemplo distribución normal de una variable. Documentar todos los supuestos realizados.
4.2. Diseñar las pruebas del modelo	<p>Una vez construidos los modelos, necesitaremos un mecanismo para determinar su calidad y validez. Por ejemplo, en problemas de agrupamiento se puede utilizar el coeficiente de silueta para evaluar la robustez de los grupos encontrados y en problemas de clasificación la tasa de error para estimar la capacidad del clasificador.</p> <p>En esta fase se dividirá el conjunto de datos en un grupo para entrenar el modelo (training) y otro para probarlo (test).</p>	<ul style="list-style-type: none"> • Diseño de los test. Determinar y documentar de qué forma se entrenarán y evaluarán los modelos generados. Incluir las decisiones tomadas sobre los datos que se utilizarán para entrenamiento y prueba.
4.3. Construir el modelo	<p>Aplicar la técnica seleccionada sobre el conjunto de datos para generar uno o más modelos. En esta fase el modelo será evaluado con distintos valores de parámetros. Por ejemplo, en un algoritmo de agrupamiento k-medias, se podrían generar distintos modelos para diferentes valores de "k" o grupos.</p>	<ul style="list-style-type: none"> • Parámetros seleccionados. Listar los parámetros que se han proporcionado al modelo, justificando la elección de los mismos. • Modelos producidos por las herramientas de minería. • Descripción de los modelos.
4.4. Evaluar el modelo	<p>En esta fase, el equipo de proyecto interpreta y evalúa el modelo en función de su conocimiento del dominio, los criterios de éxito definidos para el proyecto (tarea 1.3) y las pruebas diseñadas para el modelo (tarea 4.2).</p> <p>Los modelos pueden ser valorados y rankeados.</p>	<ul style="list-style-type: none"> • Evaluación de los modelos. Generar un reporte de evaluación de los modelos obtenidos, describiendo sus características y un ranking para los mismos. • Evaluación de los parámetros. En función de la evaluación anterior, revisar los parámetros y ajustar los mismos para volver a la fase de construcción del modelo (tarea 4.3). Repetir las etapas 4.3 y 4.4 hasta asegurarse de que se han encontrado los "mejores" modelos.

ANEXO N° 08

Tabla 9: Fase 5 CRISP-DM: Evaluación

Tarea	Descripción	Salidas
5.1. Evaluar los resultados	En esta etapa se evalúa el modelo en función de los objetivos del negocio, determinando su validez de acuerdo a los intereses organizacionales. Además del modelo, puede haber surgido como parte del proceso nueva información relevante y futuras líneas de investigación.	<ul style="list-style-type: none"> • Evaluación de los resultados de la minería de datos con respecto a los criterios de éxito y objetivos de negocio. • Modelos evaluados y aprobados.
5.2. Revisión del proceso	Realizar una revisión completa del proceso efectuado en búsqueda de posibles errores u omisiones.	<ul style="list-style-type: none"> • Revisión del proceso, documentando un resumen del mismo. Incluir las actividades omitidas o bien aquellas que deberían ser repetidas.
5.3. Determinar las próximas etapas	En función de la evaluación de resultados y la revisión del proceso, se debe decidir cómo continua el proyecto: si se pasa a la próxima fase (implementación) o bien si se retorna a una fase anterior.	<ul style="list-style-type: none"> • Lista de posibles acciones. • Descripción de la decisión tomada.

Tabla 10: Fase 6 CRISP-DM: Implementación

Tarea	Descripción	Salidas
6.1. Planificar la implementación	En esta etapa se genera el plan de implementación de los resultados obtenidos mediante la minería de datos.	<ul style="list-style-type: none"> • Plan de implementación, incluyendo las etapas y cómo llevarlas a cabo.
6.2. Planificar el monitoreo y el mantenimiento	El monitoreo y mantenimiento es de gran importancia si los resultados de la minería formarán parte del trabajo diario del negocio y su entorno.	<ul style="list-style-type: none"> • Plan de mantenimiento y monitoreo.
6.3. Crear un reporte final	Generar un reporte final, que podría resumir el desarrollo del proyecto o bien mostrar un análisis comprensivo de los resultados obtenidos en el proceso de minería.	<ul style="list-style-type: none"> • Reporte final del proyecto. • Presentación final al cliente, incluyendo resultados y conclusiones.
6.4. Revisión del proyecto	Consiste en identificar y analizar los puntos que fueron bien realizados, los que fueron mal realizados, y los que podrían mejorarse.	<ul style="list-style-type: none"> • Documentación de la experiencia adquirida durante el desarrollo del proyecto.

ANEXO N° 09

Tabla 11: Fases de la Metodología propuesta.

Fases	PROPUESTA DE METODOLOGÍA - TAREAS
Elicitacion de Requisitos	1.- Proceso de Entrevista 2.- Lluvia de Ideas. 3.- Utilización de escenarios.
Entendimiento del Negocio	1.- Determinar los objetivos del negocio. 2.- Evaluar la situación. 3.- Determinar los objetivos de minería de datos. 4.- Crear un plan para el proyecto de minería de datos.
Entendimiento de los Datos	1.- Recolectar los datos iniciales. 2.- Describir los datos 3.- Explorar los Datos 4.- Verificar la calidad de los datos.
Preparación de los Datos	1.- Seleccionar los datos. 2.- Limpieza de los datos. 3.- Construcción de los Datos. 4.- Integrar los datos. 5.- Formatear los Datos
Modelado	1.- Seleccionar la técnica de modelado. 2.- Diseñar las pruebas del modelo. 3.- Construir el modelo. 4.- Evaluar el modelo.
Evaluación	1.- Evaluar los resultados. 2.- Revisión del Proceso. 3.- Determinar las próximas etapas.
Implementación	1.- Planificar la implementación. 2.- Planificar Monitoreo. 3.- Crear reporte o informe final. 4.- Revisión del Proyecto.

Fuente : Elaboracion propia.

ANEXO N° 10

Tabla 12: Cuadro de Tabulaciones de encuesta

Preg. N° 1.- Opc.	Frecuencia	(%)	Preg. N° 2.- Opc.	Frecuencia	(%)
- No sabe	0	0%	- No sabe	0	0%
- Nada	0	0%	- Nada	0	0%
- Parcialmente	2	25%	- Parcialmente	3	38%
- Totalmente	6	75%	- Totalmente	5	63%
TOTAL	8	100%	TOTAL	8	100%

Preg. N° 3.- Opc.	Frecuencia	(%)	Preg. N° 4.- Opc.	Frecuencia	(%)
- No sabe	0	0%	- No sabe	0	0%
- Nada	0	0%	- Nada	0	0%
- Parcialmente	1	13%	- Parcialmente	4	50%
- Totalmente	7	88%	- Totalmente	4	50%
TOTAL	8	100%	TOTAL	8	100%

Preg. N° 5.- Opc.	Frecuencia	(%)	Preg. N° 6.- Opc.	Frecuencia	(%)
- No sabe	0	0%	- No sabe	0	0%
- Nada	0	0%	- Nada	0	0%
- Parcialmente	2	25%	- Parcialmente	5	63%
- Totalmente	6	75%	- Totalmente	3	38%
TOTAL	8	100%	TOTAL	8	100%

Preg. N° 7.- Opc.	Frecuencia	(%)	Preg. N° 8.- Opc.	Frecuencia	(%)
- No sabe	0	0%	- No sabe	0	0%
- Nada	0	0%	- Nada	0	0%
- Parcialmente	1	13%	- Parcialmente	2	25%
- Totalmente	7	88%	- Totalmente	6	75%
TOTAL	8	100%	TOTAL	8	100%

ANEXO N° 11

**UNIVERSIDAD CATOLICA LOS ANGELES DE CHIMBOTE
ESCUELA DE POSGRADO**

**FICHA DE VALIDACIÓN
DEL INSTRUMENTO**

1.1 Nombres y apellidos del validador : Franklin Daniel Mena Nizama
 1.2 Cargo e institución donde labora : Gobierno Regional Piura - Analista sistema-Auditor
 1.3 Nombre del instrumento evaluado : ENCUESTA PARA VALIDAR METODOLOGÍA DE EXTRACCIÓN DE CONOCIMIENTOS
 1.4 Autor del instrumento : GUTIERREZ PACHERRES, JUAN JESÚS

II. ASPECTOS DE VALIDACIÓN

Revisar cada uno de los ítems del instrumento y marcar con un aspa dentro del recuadro (X), según la calificación que asigna a cada uno de los indicadores.

1. Deficiente (Si menos del 30% de los ítems cumplen con el indicador).
2. Regular (Si entre el 31% y 70% de los ítems cumplen con el indicador).
3. Buena (Si más del 70% de los ítems cumplen con el indicador).

Aspectos de validación del instrumento		1	2	3	Observaciones Sugerencias
Criterios	Indicadores	D	R	B	
• PERTINENCIA	Los ítems miden lo previsto en los objetivos de investigación.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• COHERENCIA	Los ítems responden a lo que se debe medir en la variable y sus dimensiones.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• CONGRUENCIA	Los ítems son congruentes entre sí y con el concepto que mide.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
• SUFICIENCIA	Los ítems son suficientes en cantidad para medir la variable.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• OBJETIVIDAD	Los ítems se expresan en comportamientos y acciones observables.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
• CONSISTENCIA	Los ítems se han formulado en concordancia a los fundamentos teóricos de la variable.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• ORGANIZACIÓN	Los ítems están secuenciados y distribuidos de acuerdo a dimensiones e indicadores.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• CLARIDAD	Los ítems están redactados en un lenguaje entendible para los sujetos a evaluar.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• FORMATO	Los ítems están escritos respetando aspectos técnicos (tamaño de letra, espaciado, interlineado, nitidez).	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• ESTRUCTURA	El instrumento cuenta con instrucciones, consignas, opciones de respuesta bien definidas.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
CONTEO TOTAL (Realizar el conteo de acuerdo a puntuaciones asignadas a cada indicador)		0	4	24	28
		C	B	A	Total

Coefficiente de validez :

$$\frac{A+B+C}{30} = 0.93$$

III. CALIFICACIÓN GLOBAL

Ubicar el coeficiente de validez obtenido en el intervalo respectivo y escriba sobre el espacio el resultado.

Validez Muy Buena

Piura, Octubre del 2017

Intervalos	Resultado
0,00 – 0,49	• Validez nula
0,50 – 0,59	• Validez muy baja
0,60 – 0,69	• Validez baja
0,70 – 0,79	• Validez aceptable
0,80 – 0,89	• Validez buena
0,90 – 1,00	• Validez muy buena

[Firma]
 Mg. Ing. Franklin Daniel Mena Nizama
 INGENIERO INFORMÁTICO
 CIP 102607

ANEXO N° 12

**UNIVERSIDAD CATOLICA LOS ANGELES DE CHIMBOTE
ESCUELA DE POSGRADO**

**FICHA DE VALIDACIÓN
DEL INSTRUMENTO**

1.1 Nombres y apellidos del validador : Richard Javier Yovera Risco
 1.2 Cargo e institución donde labora : Analista - Gobierno Regional Piura
 1.3 Nombre del instrumento evaluado : ENCUESTA PARA VALIDAR METODOLOGÍA DE EXTRACCIÓN DE CONOCIMIENTOS
 1.4 Autor del instrumento : GUTIERREZ PACHERRES, JUAN JESÚS

II. ASPECTOS DE VALIDACIÓN

Revisar cada uno de los ítems del instrumento y marcar con un aspa dentro del recuadro (X), según la calificación que asigna a cada uno de los indicadores.

1. Deficiente (Si menos del 30% de los ítems cumplen con el indicador).
2. Regular (Si entre el 31% y 70% de los ítems cumplen con el indicador).
3. Buena (Si más del 70% de los ítems cumplen con el indicador).

Aspectos de validación del instrumento		1	2	3	Observaciones Sugerencias
Criterios	Indicadores	D	R	B	
• PERTINENCIA	Los ítems miden lo previsto en los objetivos de investigación.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• COHERENCIA	Los ítems responden a lo que se debe medir en la variable y sus dimensiones.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• CONGRUENCIA	Los ítems son congruentes entre sí y con el concepto que mide.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
• SUFICIENCIA	Los ítems son suficientes en cantidad para medir la variable.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• OBJETIVIDAD	Los ítems se expresan en comportamientos y acciones observables.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
• CONSISTENCIA	Los ítems se han formulado en concordancia a los fundamentos teóricos de la variable.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• ORGANIZACIÓN	Los ítems están secuenciados y distribuidos de acuerdo a dimensiones e indicadores.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• CLARIDAD	Los ítems están redactados en un lenguaje entendible para los sujetos a evaluar.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• FORMATO	Los ítems están escritos respetando aspectos técnicos (tamaño de letra, espaciado, interlineado, nitidez).	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• ESTRUCTURA	El instrumento cuenta con instrucciones, consignas, opciones de respuesta bien definidas.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
CONTEO TOTAL		<u>0</u>	<u>4</u>	<u>24</u>	<u>28</u>
(Realizar el conteo de acuerdo a puntuaciones asignadas a cada indicador)		C	B	A	Total

Coefficiente de validez :

$$\frac{A + B + C}{30} = 0.93$$

III. CALIFICACIÓN GLOBAL

Ubicar el coeficiente de validez obtenido en el intervalo respectivo y escriba sobre el espacio el resultado.

Validez Muy Buena

Intervalos	Resultado
0,00 – 0,49	• Validez nula
0,50 – 0,59	• Validez muy baja
0,60 – 0,69	• Validez baja
0,70 – 0,79	• Validez aceptable
0,80 – 0,89	• Validez buena
0,90 – 1,00	• Validez muy buena

Richard Javier Yovera Risco
INGENIERO DE SISTEMAS
REGISTRO COP N° 102783

Piura, Octubre del 2017

ANEXO N° 13

**UNIVERSIDAD CATOLICA LOS ANGELES DE CHIMBOTE
ESCUELA DE POSGRADO**

**FICHA DE VALIDACIÓN
DEL INSTRUMENTO**

1.1 Nombres y apellidos del validador : Javier Eduardo Aquino Moreno
 1.2 Cargo e institución donde labora : Informático del VDR - SW.
 1.3 Nombre del instrumento evaluado : ENCUESTA PARA VALIDAR METODOLOGÍA DE EXTRACCIÓN DE CONOCIMIENTOS
 1.4 Autor del instrumento : GUTIERREZ PACHERRES, JUAN JESÚS

II. ASPECTOS DE VALIDACIÓN

Revisar cada uno de los ítems del instrumento y marcar con un aspa dentro del recuadro (X), según la calificación que asigna a cada uno de los indicadores.

1. Deficiente (Si menos del 30% de los ítems cumplen con el indicador).
2. Regular (Si entre el 31% y 70% de los ítems cumplen con el indicador).
3. Buena (Si más del 70% de los ítems cumplen con el indicador).

Aspectos de validación del instrumento		1	2	3	Observaciones Sugerencias
Criterios	Indicadores	D	R	B	
• PERTINENCIA	Los ítems miden lo previsto en los objetivos de investigación.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• COHERENCIA	Los ítems responden a lo que se debe medir en la variable y sus dimensiones.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• CONGRUENCIA	Los ítems son congruentes entre sí y con el concepto que mide.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• SUFICIENCIA	Los ítems son suficientes en cantidad para medir la variable.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• OBJETIVIDAD	Los ítems se expresan en comportamientos y acciones observables.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• CONSISTENCIA	Los ítems se han formulado en concordancia a los fundamentos teóricos de la variable.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
• ORGANIZACIÓN	Los ítems están secuenciados y distribuidos de acuerdo a dimensiones e indicadores.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• CLARIDAD	Los ítems están redactados en un lenguaje entendible para los sujetos a evaluar.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
• FORMATO	Los ítems están escritos respetando aspectos técnicos (tamaño de letra, espaciado, interlineado, nitidez).	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
• ESTRUCTURA	El instrumento cuenta con instrucciones, consignas, opciones de respuesta bien definidas.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
CONTEO TOTAL		0	6	21	
(Realizar el conteo de acuerdo a puntuaciones asignadas a cada indicador)		C	B	A	Total

Coefficiente de validez :

$$\frac{A+B+C}{30} = 0.90$$

III. CALIFICACIÓN GLOBAL

Ubicar el coeficiente de validez obtenido en el intervalo respectivo y escriba sobre el espacio el resultado.

Validez Muy buena ✓

Piura, Octubre del 2017

Intervalos	Resultado
0,00 – 0,49	• Validez nula
0,50 – 0,59	• Validez muy baja
0,60 – 0,69	• Validez baja
0,70 – 0,79	• Validez aceptable
0,80 – 0,89	• Validez buena
0,90 – 1,00	• Validez muy buena


ING. JAVIER E. AQUINO MORENO
 INGENIERO INFORMÁTICO
 CIP. N° 85888