

**UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN**

**FACULTAD DE INGENIERÍA**

**ESCUELA DE FORMACIÓN PROFESIONAL DE INGENIERÍA DE**

**SISTEMAS Y COMPUTACIÓN**



**T E S I S**

**Aplicación de un modelo predictivo basado en minería de datos para  
mejorar los procesos de atención médica ambulatoria del Policlínico  
Metropolitano Huancayo**

**Para optar el título profesional de:**

**Ingeniero de Sistemas y Computación**

**Autores:**

**Bach. Constantino GUZMAN MORALES**

**Bach. Erik Anibal TRAVEZAÑO SANTIAGO**

**Asesor:**

**Msc. Hebert Carlos CASTILLO PAREDES**

**Cerro de Pasco – Perú - 2024**

**UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN**

**FACULTAD DE INGENIERÍA**

**ESCUELA DE FORMACIÓN PROFESIONAL DE INGENIERÍA DE**

**SISTEMAS Y COMPUTACIÓN**



**T E S I S**

**Aplicación de un modelo predictivo basado en minería de datos para  
mejorar los procesos de atención médica ambulatoria del Policlínico  
Metropolitano Huancayo**

**Sustentada y aprobada ante los miembros del jurado:**

---

Msc. Melquiades Arturo TRINIDAD MALPARTIDA  
**PRESIDENTE**

---

Mg. Pit Frank ALANIA RICALDI  
**MIEMBRO**

---

Mg. Jose Luis SOSA SANCHEZ

**MIEMBRO**



**Universidad Nacional Daniel Alcides Carrión**

**Facultad de Ingeniería**

**Unidad de Investigación**

**INFORME DE ORIGINALIDAD N° 087-2024-UNDAC/UIFI**

La Unidad de Investigación de la Facultad de Ingeniería de la Universidad Nacional Daniel Alcides Carrión en mérito al artículo 23° del Reglamento General de Grados Académicos y Títulos Profesionales aprobado en Consejo Universitario del 21 de abril del 2022, La Tesis ha sido evaluado por el software antiplagio Turnitin Similarity, que a continuación se detalla:

Tesis:

**Aplicación de un Modelo predictivo basado en minería de datos para mejorar los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo**

Apellidos y nombres de los tesistas:

**Bach. Constantino GUZMAN MORALES**

**Bach. Erik Anibal TRAVEZAÑO SANTIAGO**

Apellidos y nombres del Asesor:

**Msc. Hebert Carlos CASTILLO PAREDES**

Escuela de Formación Profesional

**Ingeniería de Sistemas y Computación**


Índice de Similitud

**23 %**

**APROBADO**

Se informa el Reporte de evaluación del software similitud para los fines pertinentes:

Cerro de Pasco, 05 de marzo del 2024

  
UNDA UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN  
FACULTAD DE INGENIERÍA  
UNIDAD DE INVESTIGACIÓN  
**Luis Villa Reguis Carbajal**  
DOCTOR EN CIENCIAS - DIRECTOR

## **DEDICATORIA.**

A mi papá y mamá por su apoyo  
incondicional.

Constantino Guzman

A mis padres y hermanos por su  
comprensión y apoyo incondicional para  
forjar mi vida profesional.

Erik Travezaño

## **AGRADECIMIENTO**

A los docentes de la Universidad Nacional Daniel Alcides Carrión y Escuela de Sistemas y Computación por su enseñanza en las aulas universitarias.

A nuestro asesor, el maestro Hebert Carlos Castillo Paredes por ser nuestro guía para el desarrollo de este trabajo de investigación.

Al gerente y personal administrativo del Policlínico Metropolitano Huancayo por brindarnos su apoyo y asistencia durante la recopilación de datos para esta investigación.

## RESUMEN

El trabajo de investigación que realice se titula: “Aplicación de un Modelo predictivo basado en minería de datos para mejorar los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo”. Tuvo como objetivo aplicar un Modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo. La investigación de acuerdo a las características del problema, objetivo y la hipótesis es de tipo experimental, de diseño pre experimental, con pre-test y post-test, los diseños pre experimentales no presentan grupo control, que incluye la atención a los pacientes registrados en el periodo 2022 de 97323 citas consideradas. Los métodos de recopilación de datos son el conjunto de actividades y herramientas utilizadas para recopilar, examinar y analizar la información necesaria para lograr los objetivos de la investigación. Se utilizó la técnica de regresión lineal. Se aplicaron las fases de la metodología CRISP–DM. En nuestro estudio se obtuvo de la Tabla 4 se tiene que la cantidad de datos históricos que se tiene actualmente sin el modelo son de 6 indicadores hospitalarios ya que la información que se almacena es en formatos preoperatorios como se explicó y con el modelo se llega a obtener información de los 12 indicadores hospitalarios ya que la información se almacena en un sistema de información automatizado. Aplicando la fórmula de análisis predictivo se tiene que con el modelo se mejora en un 100%. En conclusión se logró analizar el modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo, con los datos del hospital y los datos del modelo para lo cual se aplicaron las ecuaciones obtenidas con la técnica de regresión lineal, obteniéndose un 100% de precisión, con lo cual se concluye que el modelo es válido

**Palabras Clave:** Modelo predictivo, atención médica.

## ABSTRACT

The research work I carried out is titled: "Application of a predictive model based on data mining to improve the outpatient medical care processes of the Huancayo Metropolitan Polyclinic." Its objective was to apply a predictive model based on data mining to improve the outpatient medical care processes of the Huancayo Metropolitan Polyclinic. The research according to the characteristics of the problem, objective and hypothesis is experimental, pre-experimental in design, with pre-test and post-test, pre-experimental designs do not present a control group, which includes care for registered patients. In the 2022 period of 97,323 appointments considered. Data collection methods are the set of activities and tools used to collect, examine and analyze the information necessary to achieve the research objectives. The linear regression technique was used. The phases of the CRISP-DM methodology were applied. In our study, it was obtained from Table 4 that the amount of historical data currently available without the model is 6 hospital indicators since the information that is stored is in preoperative formats as explained and with the model we reach obtain information on the 12 hospital indicators since the information is stored in an automated information system. By applying the predictive analysis formula, the model is improved by 100%. In conclusion, it was possible to analyze the predictive model based on data mining that improves the outpatient medical care processes of the Huancayo Metropolitan Polyclinic, with the hospital data and the model data for which the equations obtained with the linear regression technique were applied. obtaining 100% precision, which concludes that the model is valid

**Keywords:** Predictive model, medical care.

## INTRODUCCIÓN

Internamente el Policlínico atiende varias especialidades divididas en departamentos en el que a continuación se mencionan los más importantes: Medicina, Pediatría, Cirugía, Emergencia, Enfermería, Gineco Obstetricia, Patología, Estadística. Cada uno de estos Departamentos con procedimientos internos de atención, son atendidos por el departamento de estadística e informática, para el procesamiento de sus atenciones internas, control de sus indicadores y gestión de su toma de decisión.

Con la minería de datos se busca procesar analizar y predecir datos para mejorar la atención médica.

Esta tesis se encuentra constituida por cuatro capítulos, todos ellos relacionados de manera que haya una coherencia entre las distintas partes y sea fácil su comprensión para los lectores de la presente investigación.

La sección inicial del capítulo comprende: Identificar e identificar el problema, definir la investigación, desarrollar el tema, establecer objetivos razones límites de la investigación.

En el segundo capítulo hay secciones sobre teorías científicas, terminología, hipótesis y variables, y una explicación práctica.

El Capítulo 3 cubre una descripción general completa de los métodos de investigación, diseño de muestras, técnicas de recolección y análisis de datos (como simulaciones por computadora), procesamiento de datos estadísticos, diseño/uso de estudios y pautas éticas para la investigación en general.

En el Capítulo 4 se discuten los resultados y el tratamiento, con descripciones detalladas del trabajo, análisis de los hallazgos, prueba de hipótesis y análisis de las mismas. Finalmente se presentan conclusiones, sugerencias, referencias y apéndices.

**Los autores.**



## ÍNDICE

**DEDICATORIA**

**AGRADECIMIENTO**

**RESUMEN**

**ABSTRACT**

**INTRODUCCIÓN**

**ÍNDICE**

### CAPITULO I

#### PROBLEMA DE INVESTIGACIÓN

1.1. Identificación y determinación del problema .....	1
1.2. Delimitación de la investigación.....	2
1.3. Formulación del problema .....	2
1.3.1. Problema general: .....	2
1.3.2. Problemas específicos: .....	2
1.4. Formulación de objetivos.....	2
1.4.1. Objetivo general.....	2
1.4.2. Objetivos específicos.....	2
1.5. Justificación de la investigación.....	2
1.6. Limitaciones de la investigación .....	3

### CAPITULO II

#### MARCO TEÓRICO

2.1. Antecedentes de estudio. ....	4
2.2. Bases teóricas – científicas.....	8
2.3. Definición de términos básicos.....	24

2.4. Formulación de hipótesis .....	25
2.4.1. Hipótesis general .....	25
2.4.2. Hipótesis específicas .....	25
2.5. Identificación de variables .....	25
2.6. Definición operacional de variables e indicadores .....	26

### CAPITULO III

#### METODOLOGÍA Y TÉCNICAS DE INVESTIGACIÓN

3.1. Tipo de investigación .....	27
3.2. Nivel de investigación .....	27
3.3. Métodos de investigación .....	27
3.4. Diseño de investigación.....	27
3.5. Población y muestra .....	28
3.5.1. Población.....	28
3.5.2. Muestra .....	28
3.6. Técnicas e instrumentos de recolección de datos.....	28
3.7. Selección, validación y confiabilidad de los instrumentos de investigación.....	29
3.8. Técnicas de procesamiento y análisis de datos.....	29
3.9. Tratamiento estadístico. ....	29
3.10. Orientación ética filosófica y epistémica. ....	30

### CAPITULO IV

#### RESULTADOS Y DISCUSIÓN

4.1. Descripción del trabajo de campo.....	31
4.2. Presentación, análisis e interpretación de resultados .....	56
4.3. Prueba de hipótesis.....	58

4.4. Discusión de resultados..... 59

CONCLUSIONES

RECOMENDACIONES

REFERENCIAS BIBLIOGRÁFICAS

ANEXOS

Instrumento de Investigación

Matriz de Consistencia

## ÍNDICE DE TABLAS

Tabla 1.Operacionalización de variables.....	26
Tabla 2. Tabla de juicio de expertos .....	30
Tabla 3. Descripción de los datos .....	33
Tabla 4. Índice de ingreso .....	59

## ÍNDICE DE FIGURAS

<i>Figura 1. Métodos del Modelo Descriptivo</i> .....	10
<i>Figura 2. Métodos del Modelo Descriptivo</i> .....	11
<i>Figura 3. Representación del Método de Correlación</i> .....	11
<i>Figura 4. Fases de la Metodología CRISP-DM</i> .....	14
<i>Figura 5. Ubicación del Policlínico Metropolitano Huancayo</i> .....	32
<i>Figura 6. Estableciendo la conexión ETL</i> .....	36
<i>Figura 7. Diseño de Datamart</i> .....	37
<i>Figura 8. Flujo de datos de la tabla tiempo</i> .....	38
<i>Figura 9. Conversión de tipos de datos de campos de la tabla Admisión</i> .....	41
<i>Figura 10. Sentencias para la limpieza del Datamart</i> .....	42
<i>Figura 11. Flujo de control del ETL – Poblando Hechos</i> .....	43
<i>Figura 12. Gráfico de Precisión de Minería de Datos - Estructura_Con</i> .....	43
<i>Figura 13. Dashboard de Ingresados</i> .....	44
<i>Figura 14. Vista de origen de datos</i> .....	46
<i>Figura 15. Estructura de Minería de Datos – Estructura_Con</i> .....	46
<i>Figura 16. Ramificaciones del árbol de decisión – Actividad Atención Medica Ambulatoria</i> .....	47
<i>Figura 17. Variable más influyente – Actividad Atención Medica Ambulatoria</i> .....	48
<i>Figura 18. Detalle del nodo más cercano o influyente – Actividad Atención Medica Ambulatoria</i> .....	48
<i>Figura 19. Elaboración de la consulta DMX – Motivo Enfermedad Súbita</i> .....	49
<i>Figura 20. Flujo de Control - Tareas de Consulta de minería de datos</i> .....	50
<i>Figura 21. Selección del modelo de minería de datos – Actividad consulta Ambulatoria</i> .....	50
<i>Figura 22. Consulta DMX – Motivo Enfermedad Súbita</i> .....	52
<i>Figura 23. Salida del resultado de la consulta DMX – Motivo Enfermedad Súbita</i> .....	53
<i>Figura 24. Flujo de Control Final</i> .....	54
<i>Figura 25. Hoja de cálculo – Monitoreo Consulta Ambulatoria</i> .....	57

## **CAPITULO I**

### **PROBLEMA DE INVESTIGACIÓN**

#### **1.1. Identificación y determinación del problema**

Todas las organizaciones de todas las industrias deben alcanzar un alto nivel de competencia, lo que obliga a sus responsables a basar sus decisiones en información directa, auténtica, fáctica y demostrable y que oriente la gestión de cada miembro de la organización (MINSA, 2019)

Los indicadores de gestión en el sector salud, son herramientas que permiten medir la administración que buscan optimizar la oferta hospitalaria a una demanda de necesidades de atención de salud que les permita tomar decisiones para la gestión administrativa eficiente y oportuna (MINSA, 2019)

Bajo este esquema el Policlínico Metropolitano Huancayo que recibe una demanda de pacientes en el 2023 se registraron en promedio 8230 casos atendidos.

Para la adecuada atención en el policlínico es necesario que internamente cada departamento: medicina, pediatría, cirugía, emergencia, enfermería, gineco obstetricia, patología, tenga un procesamiento de datos, control de sus indicadores y estos sirvan de mejora en los procesos de atención médica ambulatoria. Debido a que los datos obtenidos en los procesos de atención tienen un déficit de orden.

## **1.2. Delimitación de la investigación.**

### **1.2.1. Espacial**

Con la presente investigación se analizó, aplico modelo predictivo basado en minería de datos para mejorar los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo.

### **1.2.2. Temporal**

La investigación se llevó a cabo con los datos del 2022.

### **1.2.3. Conceptual**

Concepto del modelo predictivo, minería de datos y los procesos de atención médica.

## **1.3. Formulación del problema**

### **1.3.1. Problema general**

¿De qué manera la aplicación de un Modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo?

### **1.3.2. Problemas específicos**

¿Cómo se analizará el modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo?

## **1.4. Formulación de objetivos**

### **1.4.1. Objetivo general**

Aplicar un Modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo.

### **1.4.2. Objetivos específicos**

Analizar el modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo.

## **1.5. Justificación de la investigación**

### **1.5.1. Teórica**

El personal del policlínico que atiende ambulatoriamente no tiene la información adecuada para gestionar los recursos internos del policlínico, con la investigación se les brinda datos del departamento de medicina para optimizar la administración de los servicios a nivel operativo, utilizando como medio la minería de datos la cual se fundamenta en la busca de patrones dentro de grandes bases de datos utilizando métodos predictivos, haciendo uso de recursos informáticos y tecnológicos.

La aplicación de un modelo predictivo basado en minería de datos permite analizar, procesar y predecir los indicadores que sirven para mejorar la atención médica ya que con ello se tiene el acceso inmediato a los datos.

#### **1.6. Limitaciones de la investigación**

Recursos económicos limitados para la recopilación de requerimientos, información y validación de la investigación.



## **CAPITULO II**

### **MARCO TEÓRICO**

#### **2.1. Antecedentes de estudio**

##### **2.1.1. A nivel Internacional**

- Según Reinoso (2023) en su trabajo de investigación titulado “Desarrollo de un modelo para predecir el rendimiento académico de estudiantes de la EPN en base a su nivel de acceso a TICS y factores socioeconómicos” implementó un modelo para predecir el rendimiento académico de los estudiantes, investigó la influencia que tiene los factores socioeconómicos y los factores de nivel de acceso a la tecnología y comunicación, sobre el rendimiento académico. Se determinó cuáles son los factores más influyentes, encontrándose qué edad de inicio de estudios, número de miembros en el entorno familiar, propiedades que cuenta el estudiante, si tiene vehículo entre otros. Se aplicó la metodología de minería de datos CRISP-DM. El algoritmo que mejores resultados ha mostrado es Random Forest con una precisión del 83.7%. Los resultados han servido mucho para las autoridades de la institución, para generar estrategias de ayuda a los estudiantes con bajo rendimiento.
- Según Mejía (2023) en su trabajo de investigación titulado “Modelo matemático para predecir el grado de deserción de los estudiantes en el Instituto Superior

Tecnológico Bolívar” implementó un modelo de predicción basada en algoritmos de Machine Learning. Su objetivo fue predecir la deserción estudiantil a través de algoritmos de aprendizaje automático, el cual tuvo como entrada 1036 estudiantes. Los datos a usar fueron las variables socioeconómicas de los estudiantes como tipo de bachillerato, estado civil, edad, carga familiar, entre otros a través de una encuesta aplicada. La investigación se desarrolló bajo un nivel de investigación descriptivo y con un enfoque cuantitativo. Se aplicó los algoritmos de Máquinas de Soporte Vectorial y Redes Neuronales Artificiales, donde el modelo más eficiente se encontró usando el algoritmo de Redes Neuronales Artificiales donde su métrica de F1 Score fue de 82.2%. El espacio y libertad de aplicar una encuesta para recoger los datos, ha permitido ser cuidadosos en elegir aquellos datos que estén relacionados con la deserción.

- Según Riestra (2022) en su trabajo de investigación titulado “Efectos de los patrones de aprendizaje en línea sobre el rendimiento académico desde una perspectiva de la minería de datos” desarrolló un modelo de predicción de rendimiento académico utilizando datos en línea obtenidos de la plataforma Moodle, con el objetivo de detectar estudiantes en riesgo de desaprobar y estudiantes excelentes de manera temprana. Se han utilizado algoritmos de aprendizaje automáticos supervisados para la predicción, además de buscar las variables más determinantes en el proceso de predicción. Los algoritmos utilizados son arboles de decisión, naive bayes, regresión logística, perceptron multicapa y máquinas de soporte vectorial. Los algoritmos que mostraron mejores resultados son el árbol de decisión y perceptron multicapa llegando hasta una precisión del 93%. Una fortaleza de esta investigación es que los datos van apareciendo mientras los estudiantes avanzan en el curso y el modelo tiene más información y actualizada para lograr una mejor precisión en los resultados.

- Según García (2019) en su trabajo de investigación titulada “Aplicación de técnicas de minería de datos para la predicción de la deserción en estudiantes de pregrado de la Universidad Pontificia Bolivariana, Sede Central Medellín” desarrolló un modelo predictivo utilizando minería de datos, el cual ha permitido anticiparse a la deserción en estudiantes de pregrado, basado en un enfoque cuantitativo, aunque también se aplicó técnicas cualitativas. Se aplicó la metodología CRISP-DM. Los datos para cuantificación se recolectaron de los sistemas de información, donde se aplicó una encuesta integral entre los años 2015 y 2016. De todos los algoritmos de clasificación, Random Forest presentó el mejor desempeño al predecir la deserción, de entre cuatro algoritmos aplicados. La encuesta inicialmente tuvo 277 variables, pero solo una minoría de 22 variables tenían influencia directa en la deserción, es por eso que para futuros trabajos se deben guardar datos específicos relacionados al tema. El aspecto que se puede analizar en forma más amplia, es experimentar con otros algoritmos para poder ver si otras variables que no fueron consideradas tienen influencia.

### **2.1.2. A nivel Nacional**

- Según Ccopa y Chavez (2015) en su trabajo de investigación titulado “Modelo Predictivo Basado en Minería de Datos Para la Mejora en la Toma de decisiones del departamento de Cirugía del Hospital Regional Manuel Núñez Butrón” tuvo por objetivo general desarrollar un Modelo Predictivo basado en Minería de Datos para la mejora en la toma de decisiones del Departamento Cirugía del Hospital Regional Manuel Núñez Butrón. El problema detectado es la carencia de información requerida para lograr una adecuada toma de decisiones, además de los procesos que involucran obtener dicha información. En esta investigación se pone como caso de estudio el departamento de cirugía, el cual además de no poder obtener una predicción de sus indicadores hospitalarios, tiene el gran problema de registrar sus datos escritos a mano y

por lo tanto existen datos ilegibles. En esta investigación se determinó que el criterio de éxito es el control de los indicadores hospitalarios, los cuales son base para una adecuada toma de decisiones, para ello se aplicó la técnica de regresión lineal del método de regresión del modelo predictivo de minería de datos aplicando la metodología CRISP-DM, utilizando el software Weka y para la recolección de datos se desarrolló un sistema de información. Finalmente, se realizó la prueba de la hipótesis que se demuestra empíricamente, el desarrollo de un “Modelo predictivo basado en minería de datos para la mejora en la toma de decisiones es aceptable ya que mejora en un 91.97%.

- Según Montes y Elmenthaler (2023) en su trabajo de investigación titulado “Sistema asistido por voz para la reserva predictiva de citas médicas en un hospital de categoría III-1 utilizando redes bayesianas” este proyecto propone una aplicación móvil que permite al paciente programar una atención médica de manera ágil, directa y precisa por medio de un chatbot en el cual el paciente indica sus síntomas de malestar a través de un prediagnóstico basado en redes bayesianas que deriva al paciente a la especialidad médica correspondiente. Como resultado la satisfacción del paciente ha aumentado a un 75%, luego de haber realizado una encuesta tras el uso del software. Se concluye que la solución trae beneficios al paciente y al hospital debido a que permite reducir los tiempos para programar una cita médica, gracias al sistema basado en redes bayesianas. Ello reduce los reclamos y evita las sanciones otorgadas por SUSALUD, así como la recategorización, el cierre temporal y la disminución de los ingresos económicos.
- Según Saldaña (2015) en su trabajo de investigación titulado “Modelo predictivo de minería de datos de apoyo a la gestión hospitalaria sobre la morbilidad de pacientes hospitalizados” propuso aplicar un marco estándar de actividades de minería de datos, creando un modelo predictivo, que sirva de apoyo a la Gestión Hospitalaria sobre la morbilidad con pacientes

hospitalizados, basado en el algoritmo de análisis de serie de tiempo, Modelo ARIMA (AutoRegresive Integrated Moving Average) de Box y Jenkins (Box G.E.P. & Jenkins, 1973), con información histórica de los últimos 7 años de los pacientes del Hospital Víctor Ramos Guardia. En la investigación, se tomó como referencia la metodología CRISP-DM (Cross Industry Standard Process For Data Mining), que consiste en la comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Por lo tanto se realizó la extracción de los datos, transformación de los datos, carga de datos, limpieza de datos, diseño del datamart "HEALTHMINING", la selección y creación de variables que sirvieron como datos de entrada para mi modelo, para posteriormente crear un modelo de pronósticos, que me permitió conocer los casos de morbilidad en pacientes hospitalizados del hospital VRG para los próximos tres años.

- Según Saire (2023) en su trabajo de investigación titulado "Predicción de la ruta de rendimiento académico con algoritmos de clasificación" Los resultados muestran que cada uno de los modelos alcanza una precisión entre un 87% y 93% en la predicción, de manera temprana de cómo será el rendimiento académico de los estudiantes ingresantes, obteniendo como conclusión que los algoritmos de clasificación permiten predecir el rendimiento académico. Se busca que la información obtenida en la investigación sea útil para que las instituciones educativas implementen programas de tutoría y seguimiento a los posibles estudiantes que desaprueben los cursos y más aún que muestren señales que abandonen la carrera.

### **2.1.3. A nivel Local**

- En cuanto al repositorio local no se encontró definiciones que tengan relación con el trabajo de investigación.

## **2.2. Bases teóricas – científicas**

### **2.2.1. Aplicación de un Modelo predictivo.**

- **Minería de datos**

Gabriel Amorocho define que: “Es el proceso de recolectar y analizar grandes cantidades de datos (típicamente históricos) desde diferentes perspectivas para identificar patrones, correlaciones o tendencias ocultas entre muchas variables para posteriormente inferir las razones subyacentes para estos comportamientos y proyectarlos al futuro” (Amorocho, 2015).

Otra definición de Palomino Mariana Garrido consiste en: “La aplicación de técnicas en grandes volúmenes de datos para descubrir información útil, aplicable y no trivial”. Esta definición, aplicada a un entorno más empresarial podría reconstruirse como “el conjunto de métodos que, junto con un profundo conocimiento del negocio, están orientados a identificar, en grandes volúmenes de datos, relaciones y tendencias ocultas hasta el momento” (Garrido, 2012).

Carlos Márquez Vera compara la minería de datos con la estadística: “La minería de datos está orientada a trabajar con cantidades muy grandes de datos (millones y billones de datos). En cambio, la estadística no suele funcionar tan bien en bases de datos de tan gran tamaño y alta dimensionalidad” (Carlos Marquéz Vera, 2012).

Molina Felix resalta que minería de datos: “Es un conjunto de áreas (Estadística, inteligencia artificial, reconocimiento de patrones, computación gráfica y bases de datos) que tienen como propósito la identificación de conocimiento obtenido a partir de las bases de datos que aporten un sesgo o apoyo hacia la toma de decisión” (Molina Felix,2014).

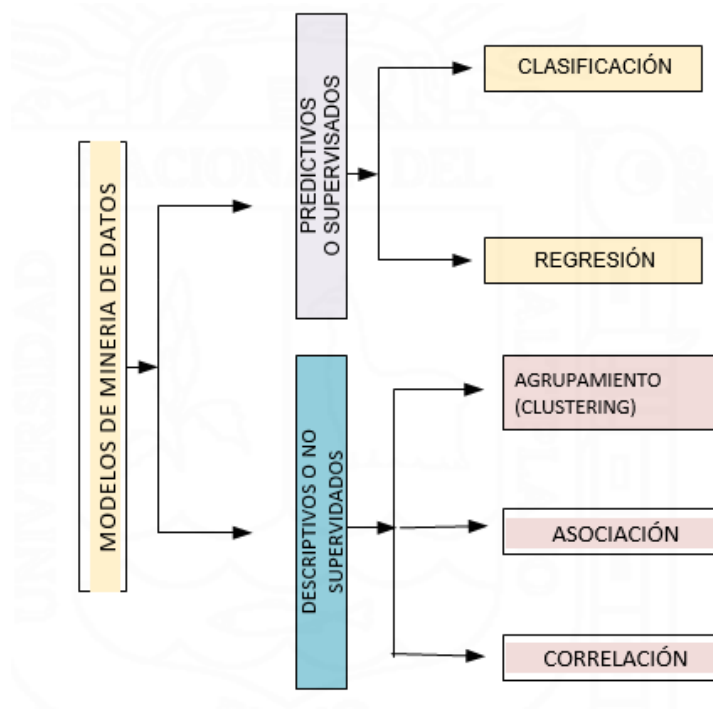
- **Modelo de Minería de Datos**

En la práctica, los modelos de minería de datos pueden ser de dos tipos: predictivos y descriptivos, otros autores los dividen en técnicas de minería de datos. Santiago Zapata menciona que: “se clasifican en dos grandes categorías: supervisadas o predictivas y no supervisadas o descriptivas” (Zapata, 2011).

- **Modelo Descriptivo**

Zapata, Santiago rescata que: “Los modelos descriptivos se clasifican en los métodos de: agrupamiento, asociación y correlación como se muestra en la Figura 1. Los modelos descriptivos también son llamados aprendizajes no supervisados ya que los modelos actúan solos y tratan de encontrar cosas en los datos” (Zapata, 2011).

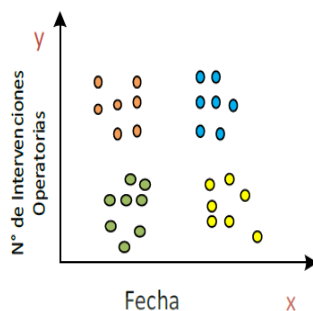
*Figura 1.*  
*Métodos del Modelo Descriptivo*



**a. Agrupamiento:**

Daniela Godoy resalta agrupamiento: “Como un método no supervisado para dividir ejemplos en grupos cuando no existen categorías predefinidas. El aprendizaje no supervisado es un método descriptivo para interpretar un conjunto de datos, es el proceso de buscar un agrupamiento natural en un conjunto de datos en base a su similitud. Ver Figura 2. El objetivo de agrupamiento es encontrar agrupaciones que estén lo más cercanos posibles malo es que se trabajan todos contra todos” (Godoy,2015).

**Figura 2.**  
*Métodos del Modelo Descriptivo*



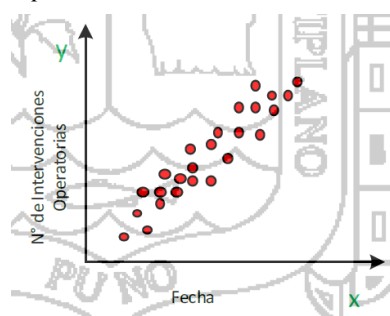
### **b. Asociación**

Según Haydeé Gommez y María de los Angeles Cerón resaltan que: “El método de asociación detecta eventos que ocurren de manera simultánea” (Gommez Díaz, Haydeé; Cerón Reyes, María de los Angeles, 2010).

### **c. Correlación**

José Hernández, María Ramírez y Cesar Ferri definen: “Las correlaciones como una tarea descriptiva que se usan para determinar el grado de similitud de los valores de dos variables numéricas. Un mecanismo estándar para medir la correlación, para un caso llamado R, el cual es un valor real comprendido entre -1 y 1. Así, si R es 1 las variables están totalmente correlacionadas; si R es -1 las variables están totalmente relacionadas negativamente; y si R es cero no existe correlación. Por consiguiente, cuando R es positivo, las variables tienen un comportamiento similar y cuando R es negativo una variable crece y la otra decrece” (Hernandez, Ramirez, Ferri, 2004).

**Figura 3.**  
*Representación del Método de Correlación*





- **Métodos del Modelo Predictivo**

- **Clasificación**

Oldemar Rodríguez Rojas también destaca que: “el objetivo de la Clasificación que es encontrar un modelo (una función o algoritmo) para predecir la clase a la que pertenecía cada registro, esta asignación de una clase se debe hacer con la mayor precisión posible. Un conjunto de prueba (tabla de testing) se utiliza para determinar la precisión del modelo. Por lo general, el conjunto de datos dado se divide en dos conjuntos al azar: el de entrenamiento y el de prueba” (Rodríguez, 2015).

- **Regresión**

Héctor Rodríguez, destaca que: “la Regresión ayuda a descubrir la dependencia del valor de un atributo con respecto a otros atributos dentro de la misma entidad u objeto. La regresión es similar a los árboles de decisión en cuanto a su contribución para clasificar datos, pero predice atributos continuos, en lugar de separados” (Rodríguez, 2009).

- **Técnicas del Modelo Predictivo**

- Redes Neuronales.
- Árboles de Decisión (ID3,C4.5.C5.0).
- Árboles de Decisión (CART).
- Otros Árboles de Decisión.
- Regresión Lineal y Logarítmica.
- Regresión Logística.
- Naive Bayes.
- Vecinos Más Próximos.
- Algoritmos Genéticos y Evolutivos.
- Máquinas de Vectores de Soporte.
- CN2 Reglas.
- Análisis Discriminante Multivariante

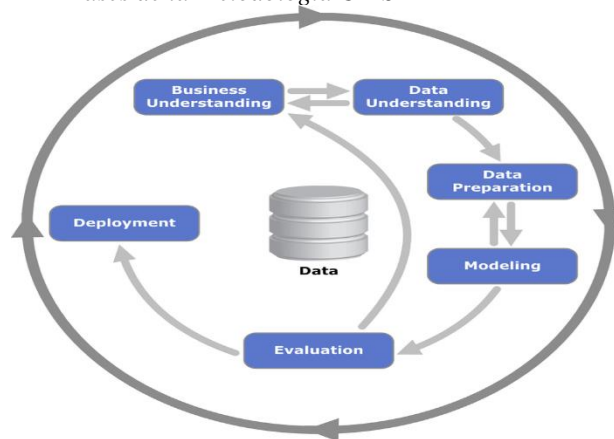
- **Metodología CRISP-DM**

Juan Miguel Moine, Ana Silvia Haedo y Silvia Gordillo resaltan que: “CRISP-DM fue creada por el grupo de empresas: SPSS, NCR y Daimler Chrysler en el año 2000, es actualmente la guía de referencia más utilizada en el desarrollo de proyectos de Data Mining. Estructura el proceso en seis fases: Comprensión del negocio, Comprensión de los datos, Procesamiento de los datos, Modelado, Evaluación y Despliegue. La sucesión de fases, no es necesariamente rígida. Cada fase es descompuesta en varias tareas generales de segundo nivel. Las tareas generales se proyectan a tareas específicas, pero en ningún momento se propone como realizarlas. Es decir, CRISP-DM establece un conjunto de tareas y actividades para cada fase del proyecto pero no especifica cómo llevarlas a cabo” (Moine , Haedo , & Gordillo, 2012).

IBM resalta que: “es un método probado para orientar trabajos de minería de datos. Y Como metodología, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas. Y como modelo de proceso, CRISP-DM ofrece un resumen del ciclo vital de minería de datos.

CRISP-DM organiza el desarrollo de un proyecto de Minería de Datos en una serie de fases o etapas, con tareas generales y específicas que permitan cumplir con los objetivos del trabajo de Investigación. Estas fases funcionan de manera cíclica e iterativa, pudiendo regresar desde alguna fase a otra anterior. En la Figura 4 se presenta las fases propuestas por esta metodología” (IBM, 2012)

**Figura 4.**  
Fases de la Metodología CRISP-DM



- **Fases de la Metodología CRISP-DM**

Juan Manuel Moine a continuación define brevemente las fases y las subfases de cada fase (Moine J. M., 2013):

- a. **Comprensión del Negocio:**

Marcelo Barrios resalta que: “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” (Barrios, 2010).

- **Determinación de Objetivos de Negocio:**

CEI rescata que: “el primer objetivo para un contexto es entender, desde una perspectiva de negocio, lo que el cliente realmente quiere lograr. A menudo el cliente tiene muchos objetivos que compiten y restricciones que deben ser correctamente equilibrados. En el objetivo se debe mostrar factores importantes.

En esta tarea se debe entender y establecer cuáles son los objetivos que el cliente pretende alcanzar, desde una perspectiva del negocio. Lo que se debe obtener es:

- Background, con la información que se conoce sobre la situación actual de la organización, incluyendo una descripción general del problema y la solución actual del 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” (CEI, 2010).

- **Evaluación de la Situación:**

Xavier Roegiers y Alexia Peyser prevalecen que: “Esta tarea implica la investigación más detallada sobre todos los recursos, restricciones, presunciones, y otros factores que deberían ser considerados en la determinación del objetivo de análisis de datos y el plan de proyecto. En la tarea anterior, su objetivo es para ponerse rápidamente al porqué de la situación.

En esta tarea se debe 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. Lo que se debe obtener es lo siguiente:

- 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” (Roegiers & Peyser, 2010).

- **Determinación de los Objetivos de la Minería de Datos:**

Paola Britos resalta que: “un objetivo de negocio declara objetivos en la terminología de negocio. Un objetivo de minería de datos declara objetivos de proyecto en términos técnicos. Por ejemplo, el objetivo de negocio podría ser “Aumentar catálogos de ventas a clientes existentes.” Un objetivo de minería de datos podría ser “Predecir cuantas baratijas un cliente comprará, obteniendo datos de sus compras de tres años pasados, información demográfica (edad,

suelo, ciudad y otros.), y el precio del artículo.” Lo que se obtiene con esta tarea es lo siguiente:

- 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” (Britos, 2010).

- **Crear un Plan para el Proyecto de Minería de Datos:**

Paola Britos rescata que: “Describir el plan intencionado para alcanzar los objetivos de minería de datos y así alcanzar los objetivos de negocio.

El plan debe especificar los pasos para ser realizados durante el resto del proyecto, incluyendo la selección inicial de herramientas y técnicas. El crear una planificación para el proyecto de minería debe ser consistente con los objetivos planteados. En esta tarea se obtiene lo siguiente:

- Plan de proyecto: listar las tareas que deben ser ejecutadas, duraciones y recursos necesarios, así como sus entradas y sus salidas.
- El plan de proyecto es un documento dinámico que debe ser ajustado y revisado en cada fase.
- Evaluación inicial de técnicas y herramientas de minería que podrían ser utilizadas en el proyecto” (Britos, 2010).

- a. **Comprensión de los Datos**

Pilar Folgueiras Bertomeu resalta que: “esta fase 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 y a continuación se detallan las tareas que se deben cumplir:” (Folgueiras Bertomeu, 2010).

- **Recolección de Datos Iniciales:**

CEI resalta que: “Esta recolección inicial incluye carga de datos, si es necesario para la comprensión de los datos. En esta tarea se recolecta todos los datos necesarios especificados en la lista de recursos del proyecto lo que se debe obtener con esta tarea es:

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 (CEI, 2010).

- **Describir los Datos:**

Barrios, Marcelo rescata de esta tarea lo siguiente: “se describe en líneas generales los datos recolectados y se obtiene lo siguiente:

Descripción de los datos, incluyendo el formato de los mismos y su tamaño (como cantidad de registros y variables)” (Barrios, 2010).

- **Explorar los Datos:**

Barrios, Marcelo resalta que: “Se realiza 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 en el que se obtiene lo siguiente:

Reporte inicial de exploración de datos, donde se expongan los resultados del análisis y las hipótesis iniciales con su impacto en el proyecto (Barrios, 2010).

- **Verificar la Calidad de los Datos:**

Pilar Folgueiras Bertomeu resalta que: “Examinar la calidad de los datos, incluyendo un análisis de su completitud, de potenciales errores en los mismos y de los datos ausentes. Se examina la calidad de los datos, dirigiendo preguntas como: ¿Los datos están completos? (¿Esto cubre todos los casos requeridos)? ¿Son correctos, o estos contienen errores y, si hay errores, que tan comunes son estos? ¿Hay valores omitidos en los datos? Si es así, ¿cómo se representan

estos, donde ocurre esto, y que tan comunes son estos? Lo que se obtiene en esta tarea:

Reporte de calidad de los datos, donde se documenta el análisis de calidad efectuado y propongan potenciales soluciones a los problemas encontrados” (Folgueiras Bertomeu, 2010).

#### **a. Procesamiento de Datos**

Rigeiro rescata que procesamiento 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.

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 y se definen a continuación” (Rigeiro, 2012).

- **Selección de Datos:**

UNED rescata que: “En esta etapa se debe seleccionar con qué atributos (columnas) y con qué observaciones (filas o registros) se trabajará. La selección debe estar justificada lo que se obtiene:

Documento donde se justifiquen las causas por las cuales se incluyeron o excluyeron los datos” (UNED, 2012).

- **Limpieza de Datos:**

UNED define que: “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. En esta tarea la salida sería la siguiente:

Reporte de limpieza de datos, donde se incluyan las decisiones tomadas sobre los problemas de calidad de los datos (reportados en la fase Verificar la calidad de los datos)” (UNED, 2012).

- **Construir Datos:**

José Luis Llavona Arregui rescata que: “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. Lo se debe obtener es lo siguiente:

- Atributos derivados. Estos atributos se calculan a partir de otros atributos del mismo registro. Por ejemplo:  $edad\_paciente = fecha\_atencion - fecha\_nacimiento$ .
- Registros creados. Estos nuevos registros se crean cuando son necesarios en la fase posterior del modelado (Llavona Arregui, 2010).

- **Integrar Datos:**

José Luis Llavona Arregui rescata que: “Esta tarea consiste en la integración de datos provenientes d diferentes tablas o registros. La salida de esta tarea es la siguiente:

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 el departamento de cirugía. 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 departamento de cirugía, podríamos integrar en un solo registro los datos personales del paciente, el total de atenciones efectuadas, y el promedio anual de operaciones realizadas” (Llavona Arregui, 2010).

- **Formatear Datos:**



UNED rescata que: “Esta etapa se refiere al cambio que debe realizarse e en 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 la salida es lo siguiente: Conjunto de fechas reformateados” (UNED, 2012).

#### **a. Implementación del Modelo**

William Meredith resalta que: “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 se cumplen las tareas siguientes en que también se definen cada un de ellos a continuación:” (Meredith, 2015).

- **Selección de la Técnica de Modelado:**

Guillermo Gilberto Molero castillo Resalta que: “esta tarea consiste en seleccionar qué técnica de minería de datos será utilizada. Por ejemplo, en una la investigación se utilizó la técnica de regresión lineal. Si se ha optado por el uso de múltiples técnicas, se debe repetir esta tarea para cada una. Las salidas de esta tarea son: La 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” (Molero Castillo, 2008).

- **Diseñar las Pruebas del Modelo:**

Regeiro rescata que: “Una vez contruidos 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. En esta fase se dividirá el conjunto de datos

en un grupo para entrenar el modelo (training) y otro para probarlo (test). La salida de esta tarea es:

El 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 utilizaran para entrenamiento y prueba” (Rigeiro, 2012).

- **Construir el Modelo:**

Guillermo Gilberto Molero Castillo rescata que: “en esta subfase se debe aplicar la técnica seleccionada sobre el conjunto de datos para generar uno o más modelos, también el modelo será evaluado con distintos valores de parámetros. Las salidas de la siguiente tarea son:

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” (Molero Castillo, 2008).

- **Evaluación del Modelo:**

Pablo Romeau Guallart rescata que: “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 determinar los objetivos de minería de datos) y las pruebas diseñadas para el modelo (tarea diseñar las pruebas del modelo). Los modelos pueden ser valorados y renqueados” (Romeau Guallart, 2010).

- a. Evaluación**

William Meredith resalta que: “es la 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 y a continuación se definen las siguientes tareas que se deben cumplir:” (Meredith, 2015).

- **Evaluación de los Resultados**

Daycy Elizabeth Peralta Cochancela resalta que: “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.

- Evaluación de los resultados de la minería de datos con respecto a los criterios de éxito y objetivos de negocio.
- Modelos evaluados y aprobados” (Peralta Cochancela, 2009).

- **Revisión del Proceso:**

Federico Peralta rescata que: “en esta etapa realizar una revisión completa del proceso efectuado en búsqueda de posibles errores u omisiones. La salida de esta tarea es:

La revisión del proceso, documentando un resumen del mismo. Incluir las actividades omitidas o bien aquellas que deberían ser repetidas (Peralta, 2014).

- **Determinación de los Próximos Pasos:**

UNED rescata que: “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. Las salidas de esta tarea son:

- Lista de posibles acciones.
- Descripción de la decisión tomada” (UNED, 2012).

- a. Distribución**

José Hernández Orallo y César Ferri Ramírez resaltan que: “Consiste en la comunicación e implementación del nuevo conocimiento, el cual debe ser representado de forma entendible las fases que se deben cumplir son:

Planificar la implementación, planificar el monitoreo y el manteniendo, crear un reporte final, revisar el proyecto a continuación se define cada una de las tareas.” (Hernández Orallo & Ferri Ramírez, 2013).

- **Planificar la implementación**

Juan Miguel Moine rescata que: “En esta etapa se genera el plan de implementación de los resultados obtenidos mediante la minería de datos. La salida de esta tarea es:

Plan de implementación, incluyendo las etapas y como llevarlas a cabo (Moine J. M., Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo, 2013).

- **Planificar el Monitoreo y el Mantenimiento**

Juan Miguel Moine rescata que: “El monitoreo y mantenimiento es de gran importancia si los resultados de la minería formarán parte del trabajo diario del negocio y su entorno. Su salida es:

El plan de mantenimiento y monitoreo” (Moine J. M., Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo, 2013).

- **Crear un reporte final**

Pilar Folgueiras Bertomeu resalta que: “en esta etapa 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.

- **Revisión del Proyecto**

Meredith, William rescata que: “esta subfase consiste en identificar y analizar los puntos que fueron bien realizados, los que fueron mal realizados, y los que podrían mejorarse.

Documentación de la experiencia adquirida durante el desarrollo del proyecto” (Meredith, 2015).

### 2.3. Definición de términos básicos

- **Árboles de Decisión:**

Es una técnica útil para problemas en los que se presentan decisiones secuenciales. Aunque esta técnica es de mayor utilidad para situaciones en que el riesgo está presente también es empleada en condiciones de certeza (Liz, 2012).

- **Casos de Uso**

Es una descripción de la secuencia de interacciones que se producen entre un actor y el sistema, cuando el actor usa el sistema para llevar a cabo una tarea específica. (Ferré Grau, Xavier; Sánchez Segura, María Isabel, 2011).

- **Dato**

Es la unidad mínima de la información. Es un elemento aislado, recabado para un cierto fin, pero que no ha pasado por un proceso que lo interrelacione con otros de manera funcional para el fin previsto. Los datos recogen un conjunto de hechos (Molina López, 2006).

- **Información**

Es un conjunto de mecanismos que permiten a un individuo estructurarlos de una manera determinada. Los patrones, asociaciones o relaciones entre los datos proporcionan información (Calderón Méndez, 2006).

- **Método**

Es una serie de pasos sucesivos, conducen a una meta. El objetivo del profesionalista es llegar a tomar las decisiones y una teoría que permita generalizar y resolver de la misma forma problemas semejantes en el futuro (Eduardo, 2015).

- **Modelo Entidad Relación**

Se basa en una percepción de un mundo real que consiste en un conjunto de objetos básicos llamados entidades y de relaciones entre estos objetos. Se

desarrolló para facilitar el diseño de bases de datos permitiendo especificar un esquema empresarial. Este esquema representa la estructura lógica general de la base de datos (López Gaona, 2011)

- **Modelo de Minería de Datos**

Es algo más que un algoritmo o un contenedor de metadatos: es un conjunto de datos, estadísticas y patrones que se pueden aplicar a los nuevos datos para generar predicciones y deducir relaciones (Vitt, 2002)

- **Predicción**

Es un pronóstico razonable y verificable, sobre un hecho o acontecimiento nuevo o desconocido. Se caracteriza por anticipar lo que va a ocurrir; la inferencia, en cambio, explica o interpreta lo que ya ha ocurrido (Yankovic, 2011)

- **Técnicas**

Estos son programas y herramientas para adquirir conocimientos (Morone, 2013)

## **2.4. Formulación de hipótesis**

### **2.4.1. Hipótesis general**

La aplicación de un Modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo.

### **2.4.2. Hipótesis específicas**

El análisis del modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo.

## **2.5. Identificación de variables**

### **2.5.1. Variables independientes**

Aplicación de un Modelo predictivo.

### **2.5.2. Variables dependientes**

Procesos de atención médica.

## 2.6. Definición operacional de variables e indicadores

*Tabla 1.*

*Operacionalización de variables*

<b>VARIABLE</b>	<b>DIMENSIONES</b>
INDEPENDIENTE	
Aplicación de un Modelo predictivo	- Análisis
DEPENDIENTE	- Atención médica
Procesos de atención médica.	

## **CAPITULO III**

### **METODOLOGÍA Y TECNICAS DE INVESTIGACIÓN**

#### **3.1. Tipo de investigación**

La presente investigación es de tipo básica pues se utilizaron conocimientos de las ciencias aplicadas a fin de detectar patrones de deserción estudiantil mediante aplicación de árboles de decisión.

#### **3.2. Nivel de investigación**

“La presente investigación es de nivel explicativo van más allá de la descripción. Están dirigidos a responder a las causas de los eventos físicos o sociales. Se centran en explicar por qué ocurre un fenómeno y en qué condiciones se da éste. Las investigaciones explicativas son más estructuradas” (Hernández Sampieri, 2014)

#### **3.3. Métodos de investigación**

El método que se usara es deductivo.(Hernández Sampieri & Mendoza Torres, 2018)

#### **3.4. Diseño de investigación**

La investigación de acuerdo a las características del problema, objetivo y la hipótesis es de tipo experimental, de diseño pre experimental, con pre-test y post-test, los diseños pre experimentales no presentan grupo control.



El grupo experimental estuvo conformado por la atención médica ambulatoria de los pacientes registrados en el periodo 2022 de 97323 citas consideradas, del Policlínico Metropolitano Huancayo.

La representación gráfica es la siguiente:

*G1: 01 X 02*

Dónde:

*G1: Grupo experimental.*

*X: Tratamiento con el modelo de minería de datos.*

*01: Test antes del experimento.*

*02: Test después del experimento.*

### **3.5. Población y muestra**

#### **3.5.1. Población**

Según (Arias, 2006) se refiere a la población como “un conjunto finito o infinito de elementos que cuentan con características comunes para los cuales serán extensivas las conclusiones de la investigación.

Pacientes registrados en el periodo 2022 de 97323 citas consideradas.

#### **3.5.2. Muestra**

Se uso como muestra de estudio toda la población para el análisis de modelo predictivo.

### **3.6. Técnicas e instrumentos de recolección de datos**

Los métodos de recopilación de datos son el conjunto de actividades y herramientas utilizadas para recopilar, examinar y analizar la información necesaria para lograr los objetivos de la investigación.

#### **Técnica**

##### **Encuestas:**

Se aplicó encuestas a la Unidad de Estadística con el fin de obtener información requerida por el Policlínico Metropolitano Huancayo.

## **Instrumento**

### **Fichas de observación:**

Se elaboró una ficha de observación para registrar los datos del experimento con modelo.

### **3.7. Selección, validación y confiabilidad de los instrumentos de investigación.**

#### **Selección de datos.**

*Tabla 2.*

*Tabla de juicio de expertos*

Experto	Grado	Calificación
1	Ingeniero	Aceptable
2	Ingeniero	Aceptable
3	Ingeniero	Aceptable

Los expertos confirmaron la “claridad, pertinencia y pertinencia” de los ítems correspondientes a diferentes aspectos de las variables de investigación. En ambos casos, los expertos coincidieron con la valoración y expresaron la siguiente opinión sobre la aplicabilidad: “Aplicable”.

### **3.8. Técnicas de procesamiento y análisis de datos.**

Se utilizó la técnica de regresión lineal.

Se aplicaron las fases de la metodología CRISP–DM:

- Comprensión del negocio.
- Comprensión de los datos.
- Preparación de datos.
- Modelado.
- Evaluación.

### **3.9. Tratamiento estadístico.**

Los datos se procesaron de forma cuantitativa.

Para los datos cuantitativos se utilizaron tablas y gráficos mediante un programa de hoja de cálculo.

Para la prueba de hipótesis se utilizó las comparaciones entre los datos reales y los que el modelo devuelve con lo cual se halló el promedio según las siguientes formulas:

**a. Predicción:**

$$\text{Precisión} = \frac{\text{N}^\circ \text{ de predicciones correctas}}{\text{N}^\circ \text{ total de predicciones}} \times 100$$

**b. Datos históricos:**

$$\text{Datos históricos} = \frac{\sum(\text{indicadores con modelo}) - \sum(\text{indicadores sin modelo})}{\sum(\text{indicadores sin modelo})} \times 100$$

**c. Mejora en la toma de decisiones:**

De las fórmulas a y b se obtiene el porcentaje de mejora para la toma de decisiones, como se muestra:

$$\text{Mejora en la toma de decisiones} = \frac{\text{Precisión} + \text{Datos Históricos}}{2}$$

### 3.10. Orientación ética filosófica y epistémica.

Se llevará a cabo la confidencialidad y se garantizará el anonimato de la ficha de registro, por lo que se omitirá sus datos personales.

## **CAPITULO IV**

### **RESULTADOS Y DISCUSIÓN**

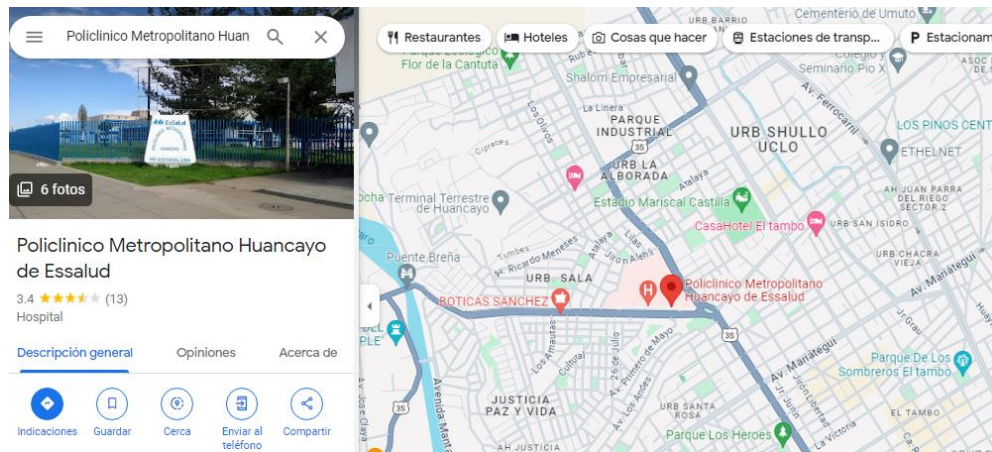
#### **4.1. Descripción del trabajo de campo**

Con el objetivo de brindar una mejor y oportuna atención, el Policlínico Metropolitano de Huancayo de la Red Asistencial Junín de EsSalud amplió su horario de atención hasta los días sábado, domingo y feriados.

Según el Dr. Carlos Cabrera Navarro, director del citado Policlínico, esta iniciativa permitirá mejorar la oferta de los servicios a fin de lograr la satisfacción de los asegurados, además agregó que gracias al Régimen de Prestaciones Complementarias de Trabajo (RPCT) promovido por la Presidencia Ejecutiva de EsSalud, se puede efectuar el pago de los profesionales asistenciales que laboren en estas fechas.

Cabrera Navarro, indicó que la nueva programación se pone en marcha desde hoy (miércoles) y estará vigente durante todo el año, en consultorios externos, laboratorio, farmacia, urgencias y apoyo al diagnóstico (ecografías, rayos x).

**Figura 5.**  
*Ubicación del Policlínico Metropolitano Huancayo*



#### 4.1.1. Aplicación de un Modelo predictivo

##### Fase I: Comprensión del negocio.

Recolectar Datos Iniciales

Se trabajarán con las tablas de AtenMed2022 en los consultorios del Hospital, estos están englobados en archivos de Excel que se exporto a una base de datos para su fácil manejo

##### Fase II: Comprensión de los datos.

##### DESCRIPCIÓN DE LOS DATOS

*Tabla 3.*

*Descripción de los datos*

Hoja2\$	Lista todas las actividades					
Total, de registros	97367 – 40.0 MB					
Nombre del campo	Tipo	Longitud	Decimal	Permite Nulls	Es Autonumérico	Descripción campo
CENTRO	Varchar	255	0	SI	NO	Centro de EESS
PERIODO	Varchar	255	0	SI	NO	Periodo
COD_SERVICIO	Varchar	255	0	SI	NO	Código de servicio

SERVICIO	Varchar	255	0	SI	NO	Servicio
COD_ACTIVIDAD	Varchar	255	0	SI	NO	Código de actividad
ACTIVIDAD	Varchar	255	0	SI	NO	Actividad
COD_SUBACTIVIDAD	Varchar	255	0	SI	NO	Código de subactividad
SUBACTIVIDAD	Varchar	255	0	SI	NO	Subactividad
DOC_MEDICO	Varchar	255	0	SI	NO	Documento del medico
APENOMB_MEDICO	Varchar	255	0	SI	NO	Apellidos y nombres de médicos
AUTOGENERADO	Varchar	255	0	SI	NO	Autogenerado
DOC_PACIENTE	Varchar	255	0	SI	NO	Documento de paciente
PACIENTE	Varchar	255	0	SI	NO	Paciente
ANNOS	Varchar	255	0	SI	NO	Años edad
MESES	Varchar	255	0	SI	NO	Meses edad
DIAS	Varchar	255	0	SI	NO	Días edad
SEXO	Varchar	255	0	SI	NO	Sexo
DIRECCION	Varchar	255	0	SI	NO	Dirección
H_C	Varchar	255	0	SI	NO	
TIPO_SEGURO	Varchar	255	0	SI	NO	Tipo de seguro
TIPO_PACIENTE	Varchar	255	0	SI	NO	Tipo de paciente
DNI_DIGITADOR	Varchar	255	0	SI	NO	DNI digitador
NOMBREDIGIT	Varchar	255	0	SI	NO	Nombre del digitador
FECHA_SOLIC	Datetime	8	0	SI	NO	Fecha de la solicitud
FECHA_CITA	Datetime	8	0	SI	NO	Fecha de cita
FECHA_ATENCION	Datetime	8	0	SI	NO	Fecha de atención

COND_CITA	Varchar	255	0	SI	NO	Condición cita
ACTO_MED	Varchar	255	0	SI	NO	Acto medico
TIPO_CONSULTA	Varchar	255	0	SI	NO	Tipo de consulta
DIAGNOSTICO	Varchar	255	0	SI	NO	Diagnostico
DES_DIAGNOSTICO	Varchar	255	0	SI	NO	Descripción de diagnostico
TIPODIAG	Varchar	255	0	SI	NO	Tipo de diagnostico
CASODIAG	Varchar	255	0	SI	NO	Caso diagnostico
RESULT_ATENCION	Varchar	255	0	SI	NO	Resultado de atención
FECHA_REG	Datetime	8	0	SI	NO	Fecha de registro
HORA_REG	Varchar	255	0	SI	NO	Hora de registro
TIPO_GRAVIDEZ	Varchar	255	0	SI	NO	Tipo de gravedad
N_R_C_SER	Varchar	255	0	SI	NO	NRC SER
N_R_C_EST	Varchar	255	0	SI	NO	NRC EST
COD_PRECEDENCIA	Varchar	255	0	SI	NO	Código de precedencia
CAS_ADSCRIPCION	Varchar	255	0	SI	NO	Caso adscripción
COD_CONSULTORIO	Varchar	255	0	SI	NO	Código consultorio
TIP_PROGRAMACION	Varchar	255	0	SI	NO	Tipo programación
FECPROGRAMA	Varchar	255	0	SI	NO	Fecha programación
HORAINI	Varchar	255	0	SI	NO	Hora inicio
HORAFIN	Varchar	255	0	SI	NO	Hora fin
TOTHORAS	Float	8	0	SI	NO	Total, horas

TOTCUPOSNORMALES	Varchar	255	0	SI	NO	Total, cupos normales
TOTCUPOSADICIONALES	Varchar	255	0	SI	NO	Total, cupos adicionales
UBIGEO_DOMICILIO	Varchar	255	0	SI	NO	Ubigeo domicilio
TELEF_FIJO	Varchar	255	0	SI	NO	Teléfono fijo
TELEF_MOVIL	Varchar	255	0	SI	NO	Teléfono móvil
TIPO_PACIENTE	Varchar	255	0	SI	NO	Tipo de paciente
TURNNO	Varchar	255	0	SI	NO	Turno
CMP	Varchar	255	0	SI	NO	CMP
CODTIP_SEGURO	Varchar	255	0	SI	NO	Código de tipo de seguro
HORA_CITA	Varchar	255	0	SI	NO	Hora de cita
NUM_RUC	Varchar	255	0	SI	NO	Numero de ruc
PERTENECE	Varchar	255	0	SI	NO	Pertenece
TIPDOC_PACIENTE	Varchar	255	0	SI	NO	Tipo de documento del paciente
COORDENADA_X	Varchar	255	0	SI	NO	Direcciones coordenadas en x
COORDENADA_Y	Varchar	255	0	SI	NO	Direcciones coordenadas en y
REFERENDIRECC	Varchar	255	0	SI	NO	Dirección referencia
FECHA_MODIF	Datetime	8	0	SI	NO	Fecha de modificación
HORA_MODIF	Varchar	255	0	SI	NO	Hora modificada
USUA_MODIF	Varchar	255	0	SI	NO	Usuario modificado



FECNACIMPACIENTE	Varchar	255	0	SI	NO	Fecha de nacimiento del paciente.
USU_REG	Varchar	255	0	SI	NO	Usuario registro
ESTADO_CIVIL	Varchar	255	0	SI	NO	Estado Civil
HIPERTENSION	Varchar	255	0	SI	NO	Hipertensión
CARDIOPATIA	Varchar	255	0	SI	NO	Cardiopatía
TIPO_VIVIENDA	Varchar	255	0	SI	NO	Tipo de vivienda
CONDICFUMADOR	Varchar	255	0	SI	NO	Condición de fumador

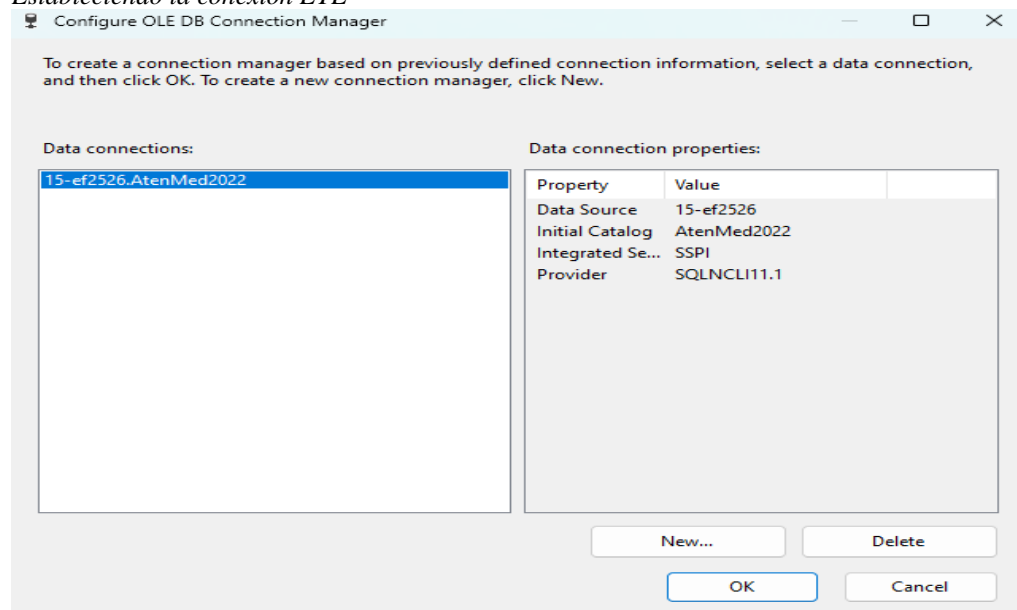
### Fase III: Preparación de datos.

Las tareas de selección, limpieza, estructuración, integración y formateo de datos componen esta fase.

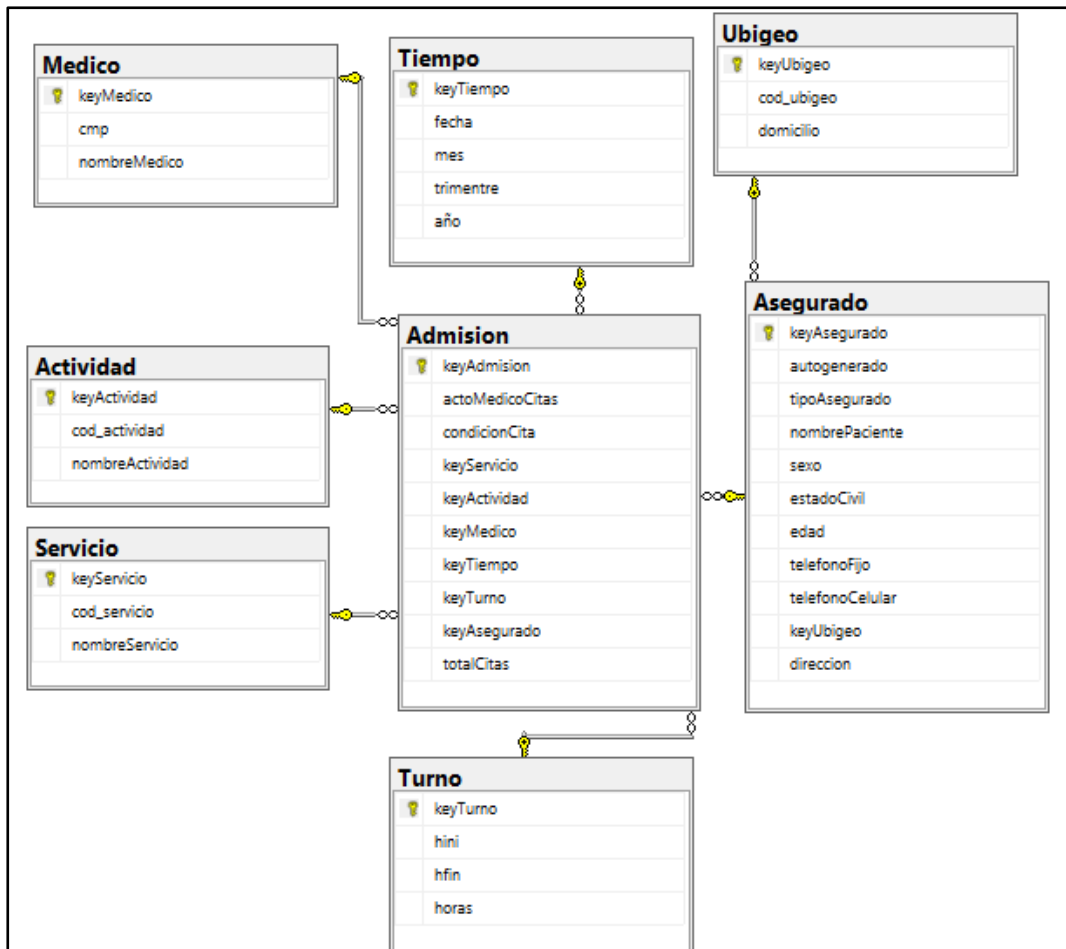
#### Limpieza, estructurado, integración y formateo de los datos

Para realizar la limpieza de los datos se creó una conexión OLE DB, con la finalidad de conectarnos con las tablas DBF a utilizar; dicha conexión tiene por nombre: ConexionAtanMed, y ha servido para crear nuestro Datamart.

**Figura 6.**  
*Estableciendo la conexión ETL*



**Figura 7.**  
Diseño de Datamart



En la herramienta SQL Server Business Intelligence Development Studio, se creó el Proyecto de Integration Services llamado ETL y seguidamente un Origen de Datos llamado “Conexión Hospital”.

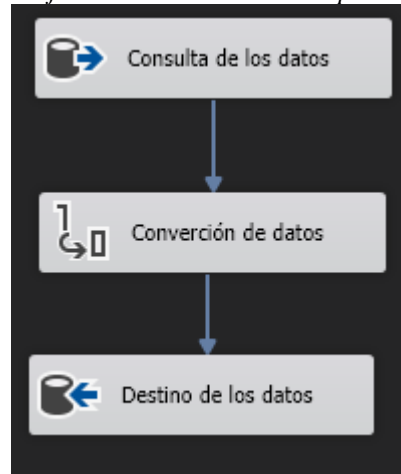
Luego se ha creado un Paquete de SQL Server Integration Services (SSIS), llamado “AtenMedETL”, donde se realizará el proceso, valga la redundancia, de ETL (Extracción, transformación y carga). Es aquí donde se realiza el poblado de las Dimensiones y los Hechos, cumpliendo con las tareas de esta fase. Las dimensiones y hechos fueron poblados mediante las herramientas que nos brinda el SSIS.

#### Dimensión Medico

Esta dimensión se pobló a través de los registros obtenidos de las citas del hospital.

```
SELECT DISTINCT CMP, APENOMB_MEDICO AS nombreMedico  
FROM dbo.Hoja2$
```

**Figura 8.**  
*Flujo de datos de la tabla tiempo.*



Para el desarrollo de esta tarea se creó primero una conexión OLE DB para obtener los datos, luego se procedió a hacer una conversión de estos datos ya que diferían de los datos de destino y finalmente se pobló las tablas dimensiones y hechos.

Cabe señalar que, para el poblado de las tablas hechos, antes se deben poblar las dimensiones.

### **Dimensión Servicio**

De la misma manera esta dimensión se pobló a través de los registros obtenidos de las citas del hospital.

```
SELECT DISTINCT COD_SERVICIO AS cod_servicio, SERVICIO AS  
nombreServicio
```

```
FROM dbo.Hoja2$
```

Este proceso se realizó de la misma manera que las otras dimensiones.

### **Dimensión Actividad**

De la misma manera esta dimensión se pobló a través de los registros obtenidos de las citas del hospital.

```

select DISTINCT COD_ACTIVIDAD AS cod_actividad, ACTIVIDAD AS
nombreActividad
from dbo.Hoja2$
WHERE COD_ACTIVIDAD is not null

```

Este proceso se realizó de la misma manera que las otras dimensiones, cabe añadir aquí que se eliminaron los datos de tipo null de cod\_actividad, ya que este lo requería.

### **Dimensión Ubigeo**

De la misma manera esta dimensión se pobló a través de los registros obtenidos de las citas del hospital.

```

SELECT DISTINCT UBIGEO_DOMIC AS cod_ubigeo,
UBIGEO_DOMICILIO AS DOMICILIO
FROM dbo.Hoja2$
WHERE UBIGEO_DOMIC is not null

```

Este proceso se realizó de la misma manera que las otras dimensiones, cabe añadir aquí que se eliminaron los datos de tipo null de cod\_ubigeo, ya que este lo requería.

### **Dimensión Tiempo**

De la misma manera esta dimensión se pobló a través de los registros obtenidos de las citas del hospital.

```

SELECT DISTINCT FECHA_ATENCION AS fecha,
MONTH(FECHA_ATENCION) AS mes, YEAR(FECHA_ATENCION) AS año,
DATEPART(quarter, FECHA_ATENCION) AS timestre
FROM dbo.Hoja2$

```

Para poblar esta dimensión se obtuvo la fecha solo el campo fecha de las tablas, los demás registros se crearon a través de este campo.

### **Dimensión Turno**

De la misma manera esta dimensión se pobló a través de los registros obtenidos de las citas del hospital.

```
SELECT DISTINCT HORAINI AS hini, HORAFIN AS hfin, TOTHORAS as horas
```

```
FROM dbo.Hoja2$
```

Este proceso se realizó de la misma manera que las otras dimensiones.

### **Dimensión Asegurado**

De la misma manera esta dimensión se pobló a través de los registros obtenidos de las citas del hospital.

```
select ubiD.keyUbigeo AS keyUbigeo, AUTOGENERADO AS autogenerado, TIPO_SEGURO AS tipoAsegurado, PACIENTE AS nombrePaciente, SEXO AS sexo, ANNOS AS edad, DIRECCION AS direccion, ESTADO_CIVIL AS estadoCivil, TELEF_FIJO AS telefonoFijo, TELEF_MOVIL AS telefonoCelular
```

```
from AtenMed2022.dbo.Hoja2$ hosData inner join BD_ATENMEDETL.dbo.Ubigeo ubiD on (ubiD.domicilio = hosData.UBIGEO_DOMICILIO)
```

Este proceso se realizó de la misma manera que las otras dimensiones, nada más que aquí se creó una conexión que se requerían para esta dimensión la cual es Ubigeo.

### **Hecho Admisión**

En esta tabla se desarrollaron las conexiones con las demás dimensiones previamente generadas.

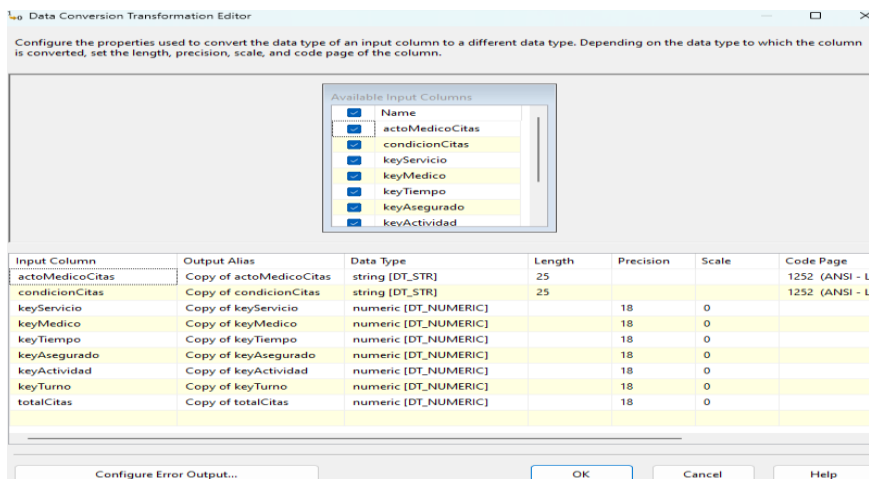
```
select ACTO_MED AS actoMedicoCitas, COND_CITA AS condicionCitas, serD.keyServicio As keyServicio, meD.keyMedico AS keyMedico, tiD.keyTiempo AS keyTiempo, asD.keyAsegurado AS keyAsegurado, acD.keyActividad AS keyActividad, turD.keyTurno AS keyTurno, ROW_NUMBER() OVER(ORDER BY hosData.CMP) AS totalCitas
```

```

from AtenMed2022.dbo.Hoja2$ hosData inner join
BD_ATENMEDETL.dbo.Servicio serD on (serD.cod_servicio =
hosData.COD_SERVICIO)
inner join
BD_ATENMEDETL.dbo.Medico meD on (meD.cmp = hosData.CMP)
inner join
BD_ATENMEDETL.dbo.Tiempo tiD on (tiD.fecha =
hosData.FECHA_ATENCION)
inner join
BD_ATENMEDETL.dbo.Asegurado asD on (asD.autogenerado =
hosData.AUTOGENERADO)
inner join
BD_ATENMEDETL.dbo.Actividad acD on (acD.cod_actividad =
hosData.COD_ACTIVIDAD)
inner join
BD_ATENMEDETL.dbo.Turno turD on (turD.hini = hosData.HORAINI)

```

**Figura 9.**  
*Conversión de tipos de datos de campos de la tabla Admisión*

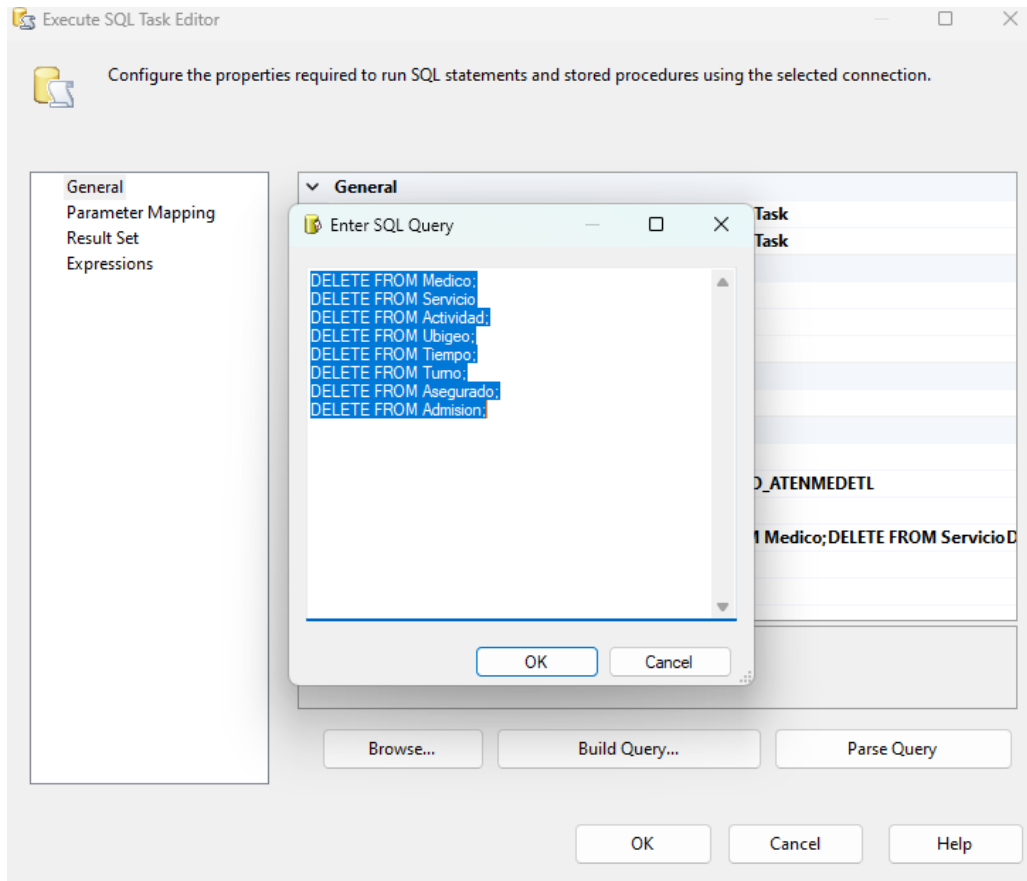


Conversión de datos; convierte columnas a distintos tipos de datos y agrega las columnas al conjunto de datos. Como se puede observar en la figura,

se utilizó esta herramienta, para convertir el tipo de datos de las columnas, esto claro se realizó con todas las tablas.

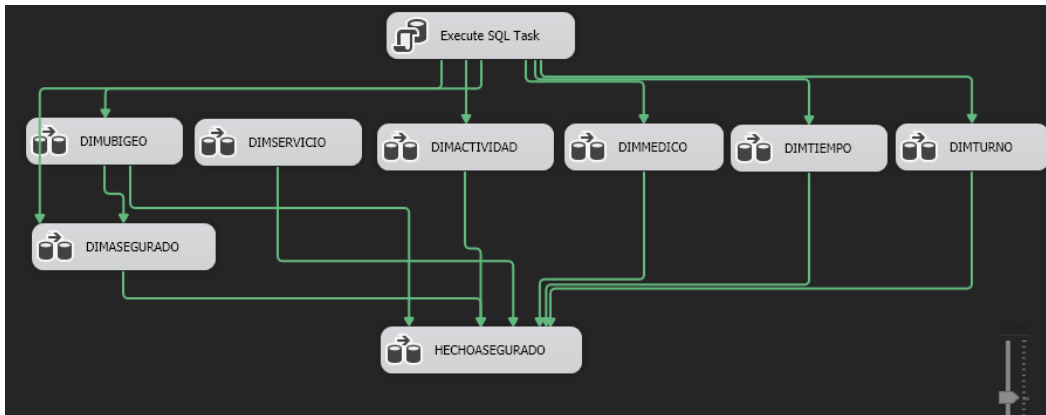
Antes de realizar el proceso de poblado de las tablas se realiza una limpieza del Datamart.

**Figura 10.**  
*Sentencias para la limpieza del Datamart.*



En la figura siguiente muestra el flujo de control que contiene las relaciones que existe entre las dimensiones y los hechos.

**Figura 11.**  
Flujo de control del ETL – Poblando Hechos



**Fase IV: Modelado.**

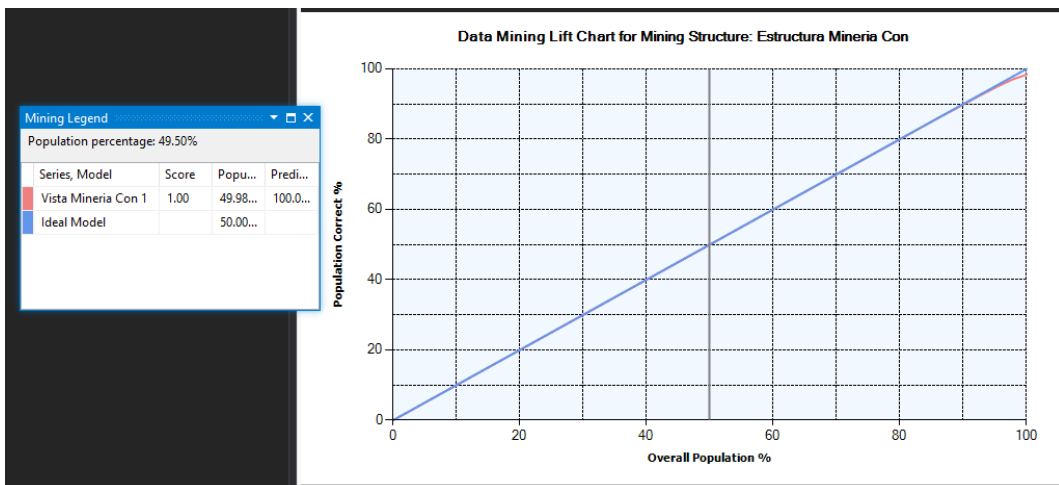
Seleccionar técnica de modelado

Aquí, se consideró 2 técnicas de modelado:

- Árboles de Decisión
- Clustering

Al final se tomó la decisión de elegir la técnica de Árboles de Decisión, pues es la que más se acerca al modelo ideal con un 100% de la Actividad de Procedimientos (Estructura\_Con).

**Figura 12.**  
Gráfico de Precisión de Minería de Datos - Estructura\_Con



**Generar el plan de prueba**



La data histórica se remonta a los meses de enero a diciembre del año 2022, donde para realiza el plan de prueba.

El resultado de este plan de prueba fue el siguiente:

- Actividad consulta ambulatoria De un total de 97323 (100%) citas registradas en el periodo del grupo de prueba.

**Figura 13.**  
*Dashboard de Ingresados*

## DASHBOARD PORCENTUAL

Total de ingresados por Actividad		
Actividad	Ingresados	Porcentajes
ATENCION MEDICA AMBULATORIA	97323	100.00%
Total	97323	100%

Total de ingresados por tipo asegurado		
Tipo seguro	Ingresados	Porcentajes
OBLIGATORIO	55101	56.617%
PENSIONISTA	30472	31.310%
S.REGULARÁ Á D.LEG.1057(CAS)	6643	6.826%
POTESTATIVO	29	0.030%
AGRARIO DEPENDIENTE	16	0.016%
TRABAJADOR DEL HOGAR	146	0.150%
SEG.SALUD TRABAJAD.INDEPENDIENTES (PEAS)	33	0.034%
AGRARIO INDEPENDIENTE	8	0.008%
BENEFICIARIO LEY 30425	3894	4.001%
CONSTRUCCION CIVIL	715	0.735%
TRABAJADOR PESQUERO-LEY 28320	1	0.001%
SALUD SEGURO POTESTATIVO	265	0.272%
TOTAL	97323	100%

### Construir el modelo

Para construir el modelo de minería de datos, se realizaron las siguientes acciones:

En la herramienta SQL Server Management Studio, se creó, dentro del Datamart, una vista llamada:

- **Vista\_Mineria\_Con**

Almacena el histórico de citas de la Actividad Médica Ambulatoria; se generó la siguiente consulta SQL:

```
CREATE VIEW Vista_Mineria_Con
AS
```

```

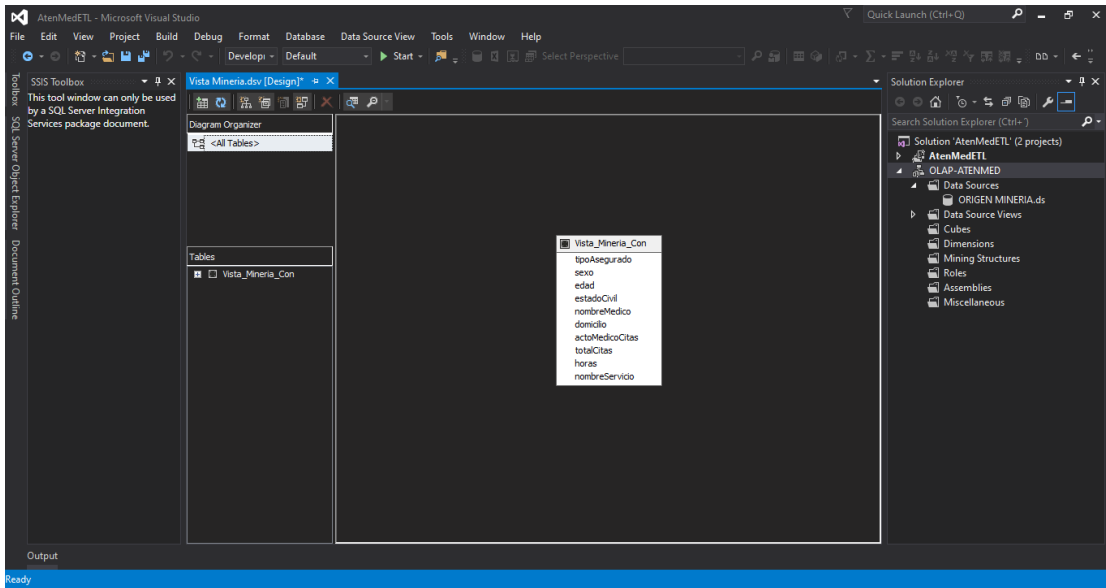
SELECT  A.tipoAsegurado,  AC.nombreActividad,  A.sexo,  A.edad,
A.estadoCivil,
        M.nombreMedico, U.domicilio,
        H.actoMedicoCitas, H.totalCitas, TU.horas,
        S.nombreServicio
FROM  Asegurado A INNER JOIN
        Ubigeo U ON A.keyUbigeo = U.keyUbigeo INNER
        JOIN Admision H ON A.keyAsegurado =
        H.keyAsegurado INNER JOIN Actividad AC ON H.keyActividad =
        AC.keyActividad INNER JOIN Turno TU ON H.keyTurno =
        TU.keyTurno INNER JOIN
        Servicio S ON H.keyServicio = S.keyServicio INNER
        JOIN Tiempo T ON H.keyTiempo = T.keyTiempo
        INNER JOIN Medico M ON H.keyMedico = M.keyMedico
WHERE  (A.tipoAsegurado = 'OBLIGATORIO') AND (AC.cod_actividad =
'91') AND (YEAR(T.año) = 2022)

```

Mediante el uso de la plataforma de SQL Server Analysis Services; se agregó un proyecto de analysis services llamado “OLAP-ATENMED”, posteriormente se creó un Origen de Datos llamado “Origen Minería”.

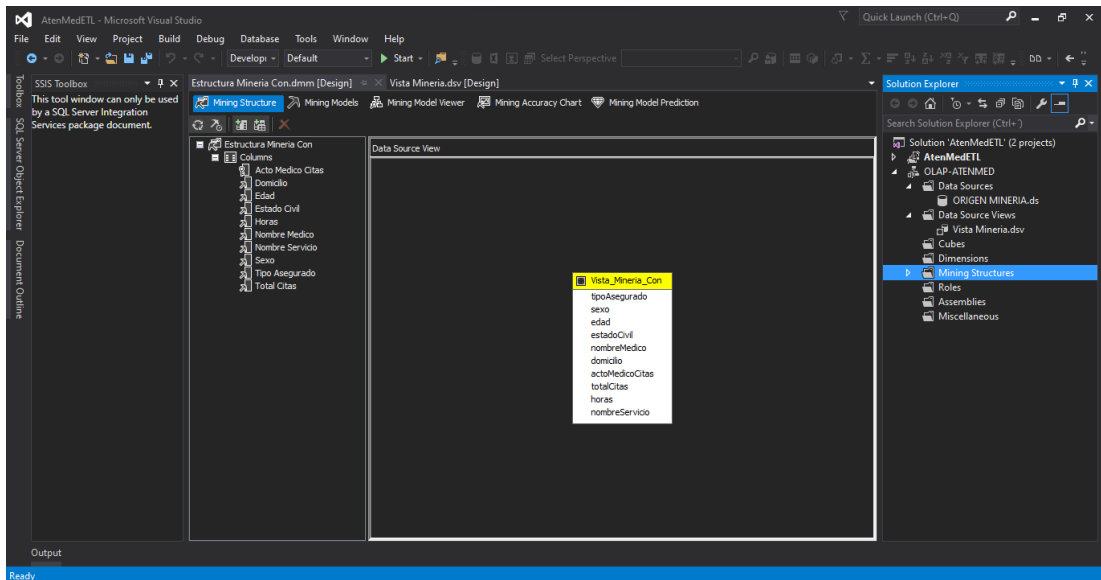
Establecida la conexión y creadas las vistas, se procedió a crear una Vista de Origen de Datos llamada “Vista Minería”, la cual estará compuesta por la vista creada en el SQL Server Management Studio.

**Figura 14.**  
Vista de origen de datos.



Luego de crear la vista, se creó las estructuras de minería de datos, para Atención Medica Ambulatoria.

**Figura 15.**  
Estructura de Minería de Datos – Estructura\_Con



En la estructura de minería de datos, podemos visualizar el modelo de minería de datos a través del Visor de modelos de minería de datos, desde 2 tipos de visores.

- Estructura de Actividad Atención Medica Ambulatoria

Estructura\_Con – Visor de árboles de Microsoft

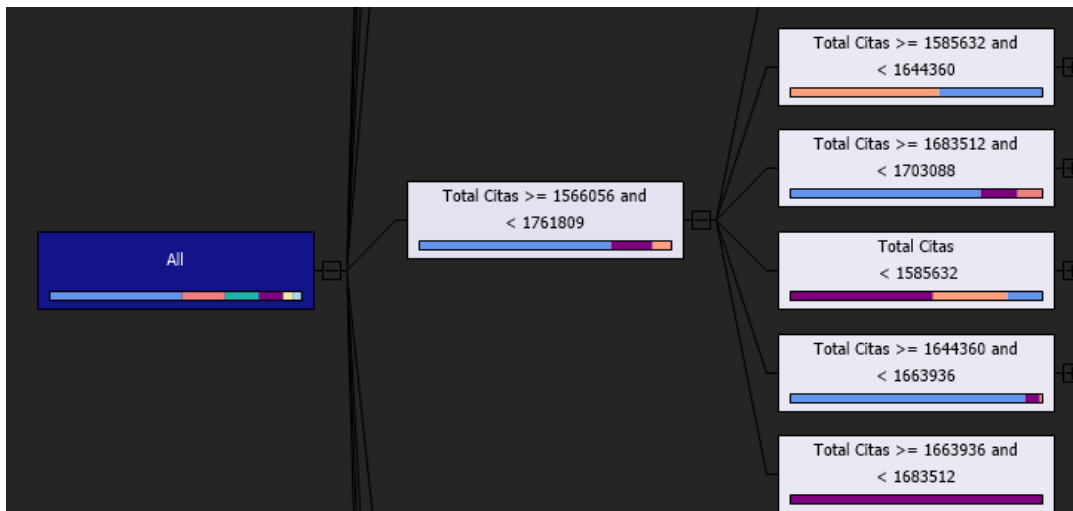
- En este visor podemos encontrar 2 pestañas:

Árbol de Decisión

En este tipo de vista, podemos visualizar la estructura del árbol, la cual tiene 5 niveles, nodos o ramificaciones.

**Figura 16.**

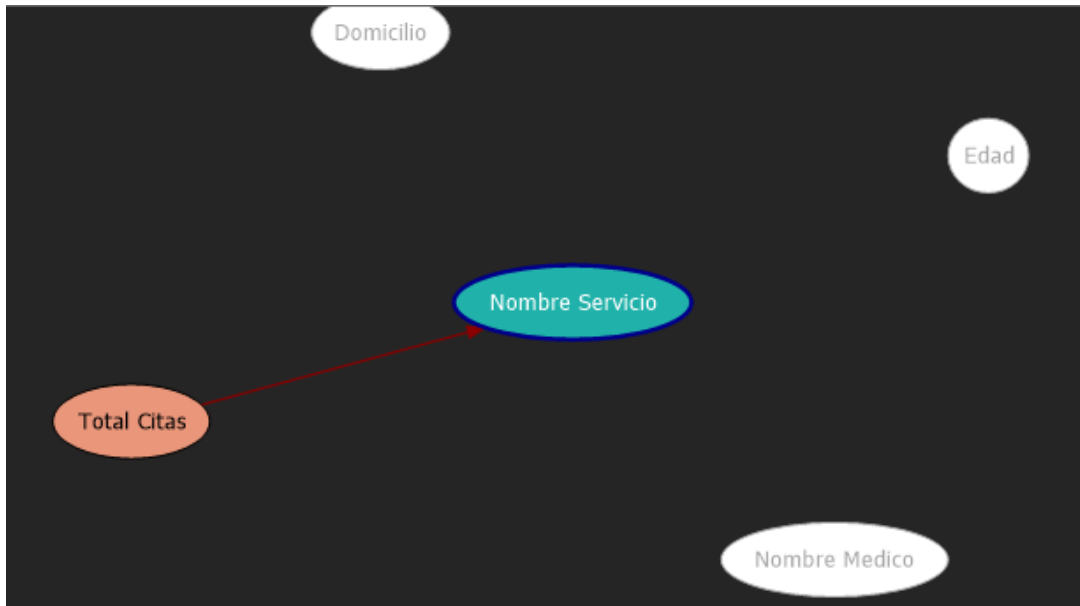
*Ramificaciones del árbol de decisión – Actividad Atención Medica Ambulatoria*



- Red de Dependencia

En esta pestaña podemos ver la influencia de las variables en relación con el motivo de ingreso; se puede apreciar que la variable más influyente es el Total de citas.

**Figura 17.**  
Variable más influyente – Actividad Atención Medica Ambulatoria



- Visor de árbol de contenido genérico de Microsoft

En este tipo de vista podemos ver de manera diferente lo mostrado en las figuras anteriores, donde la más influyente en el modelo de minería es total citas < 195785.

**Figura 18.**  
Detalle del nodo más cercano o influyente – Actividad Atención Medica Ambulatoria

Node Caption (Unique ID)	Node Details
All (000000006)	MODEL_CATALOG OLAP-ATENMED
- Total Citas < 195785 (0000000600)	MODEL_SCHEMA
- Nombre Medico = 'MENDEZ MOYED	MODEL_NAME Modelo Minería Con
- Nombre Medico = 'PERALTA DIAZ	ATTRIBUTE_NAME
- Nombre Medico = 'ROSALES URBA	NODE_NAME 0
- Total Citas >= 195785 and < 391538	NODE_UNIQUE_NAME 0
- Nombre Medico = 'BERALIN HUAPA	NODE_TYPE 1 (Model)
- Nombre Medico = 'CASTRO MANY	NODE_GUID
- Nombre Medico = 'ROSALES URBA	NODE_CAPTION
- Nombre Medico = 'VERANO PORTA	CHILDREN_CARDINALITY 1
- Total Citas >= 391538 and < 587291	PARENT_UNIQUE_NAME
- Nombre Medico = 'ALARCON TOR	NODE_DESCRIPTION
- Nombre Medico = 'BEJARANO MIR	NODE_RULE
- Nombre Medico = 'PANDO ROJAS	MARGINAL_RULE
- Nombre Medico = 'PORRAS MONT	NODE_PROBABILITY 0
- Nombre Medico = 'VERANO PORTA	MARGINAL_PROBABILITY 0
- Nombre Medico = 'ZAPATA GONZA	NODE_DISTRIBUTION ATTRIBUTE_NAME ATTRIBUTE_VALUE SUPPORT PROBA
- Total Citas >= 587291 and < 783044	NODE_SUPPORT 0
- Nombre Medico = 'SANCHEZ JAIME	MSOLAP_MODEL_COLUMN
- Nombre Medico = 'TORRES PAULI	MSOLAP_NODE_SCORE 0
- Nombre Medico = 'VILLARRUEL QU	MSOLAP_NODE_SHORT_CAPTION
- Nombre Medico = 'ZAPATA GONZA	
- Total Citas >= 783044 and < 978797	
- Nombre Medico = 'CAMPUZANO V	
- Nombre Medico = 'HUAMAN SAUÃ	
- Nombre Medico = 'VILLARRUEL QU	
- Nombre Medico = 'ZAMORA DELGA	
- Total Citas >= 978797 and < 117455	

Creada las estructuras de minería de datos, se procede a realizar la consulta DMX (Referencia de Extensiones de minería de datos), en la pestaña Predicción de modelo de minería de datos, con la finalidad de obtener los pacientes con probabilidad de deserción.

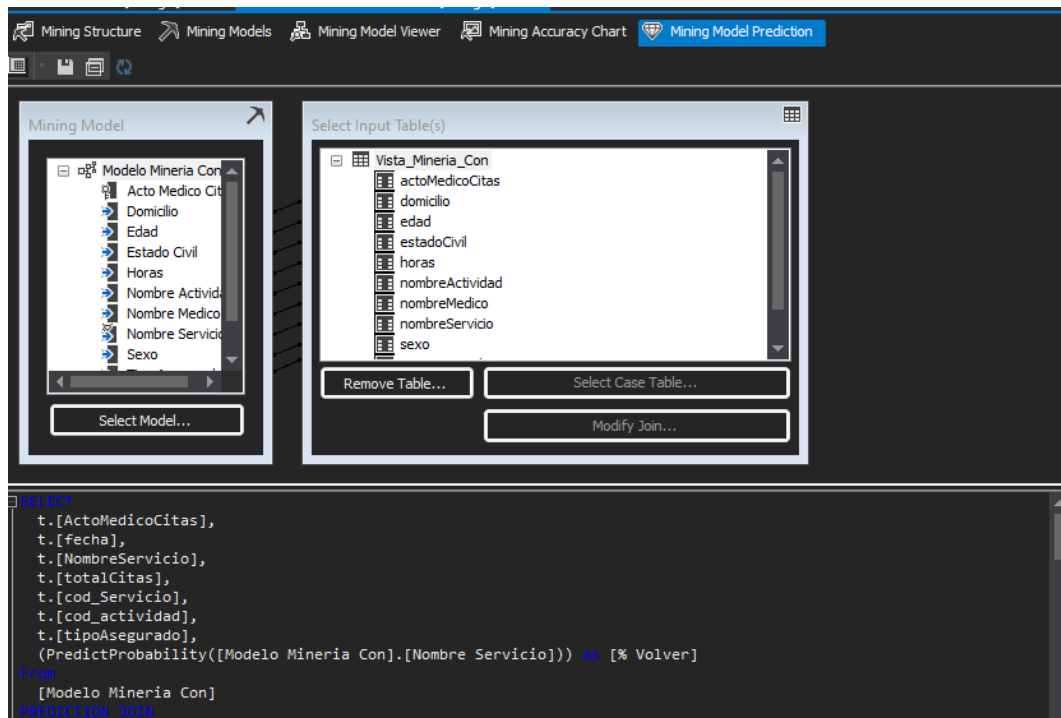
Para esto se cuenta, a su vez, con 3 tipos de vista para elaborar la consulta DMX: diseño, consulta y resultado; una vez creado la consulta, ya sea en la vista de diseño o consulta, la relación de pacientes con probabilidad de deserción se muestra en la pestaña resultado.

- Atención Medica Ambulatoria

En el cuadro “Seleccionar tabla(s) de entrada”, seleccionamos la vista Vista\_Datos\_Citas\_Con.

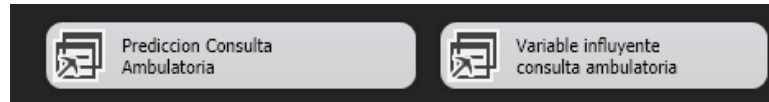
En este tipo de predicción se filtra o selecciona a aquellos pacientes que tenga una probabilidad de volver mayor al 50%.

**Figura 19.**  
*Elaboración de la consulta DMX – Motivo Enfermedad Súbita*



Una vez definida la consulta DMX a utilizar, en el entorno del ETL, se agrega 2 controles de Tarea de Consulta de minería de datos.

**Figura 20.**  
*Flujo de Control - Tareas de Consulta de minería de datos*

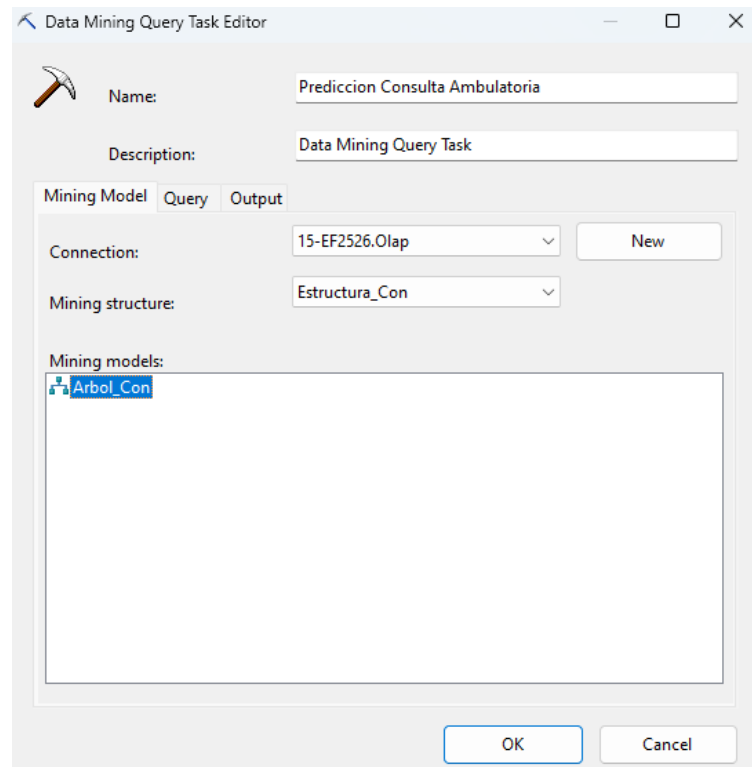


- Predicción Enfermedad común

Mediante este flujo de control procedemos a poblar una tabla temporal llamada Temporal\_Con; con los datos de los posibles pacientes desertores.

Haciendo doble clic en el control, nos aparece la siguiente ventana:

**Figura 21.**  
*Selección del modelo de minería de datos – Actividad consulta Ambulatoria*



- En la pestaña Modelo de minería de datos, elegimos nuestra conexión de Analysis Services, la estructura de minería de datos, Estructura\_Con, y nuestro modelo de minería de datos: Arbol\_Con.

SELECT

```
t.[ActoMedicoCitas],
t.[fecha],
t.[NombreServicio],
t.[totalCitas],
t.[cod_Servicio],
t.[cod_actividad],
t.[tipoAsegurado],
(PredictProbability([Modelo Minería Con].[Nombre Servicio])) as [% Volver]
```

From

```
[Modelo Minería Con]
```

PREDICTION JOIN

```
OPENQUERY([Origen Minería],
```

```
'SELECT
```

```
[ActoMedicoCitas],
```

```
[fecha],
```

```
[NombreServicio],
```

```
[totalCitas],
```

```
[cod_Servicio],
```

```
[cod_actividad],
```

```
[TipoAsegurado],
```

```
[Sexo],
```

```
[Edad],
```

```
[Domicilio],
```

```
[NombreMedico]
```

```
FROM
```

```
[dbo].[Vista_Minería_Con]
```

```
') AS t
```

ON

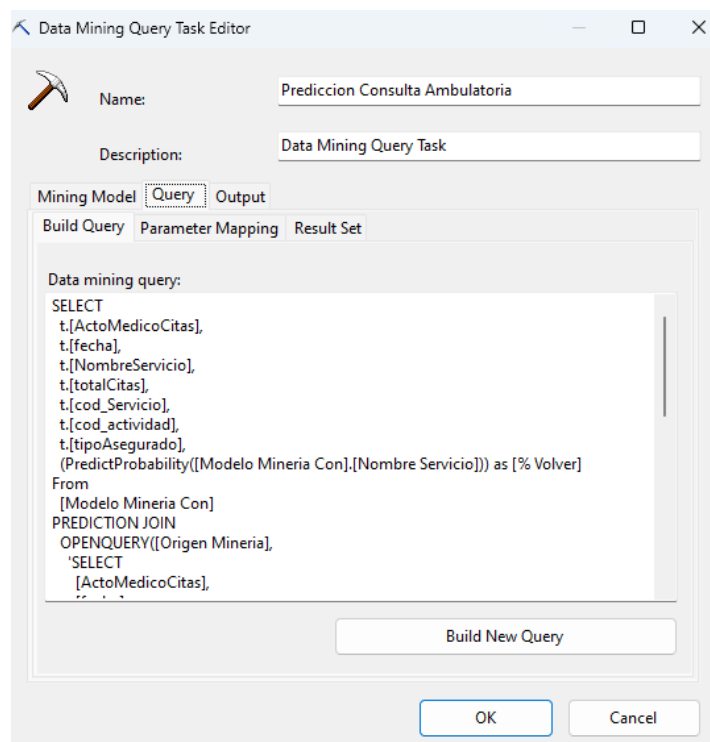


```

[Modelo Minería Con].[Tipo Asegurado] = t.[TipoAsegurado] AND
[Modelo Minería Con].[Sexo] = t.[Sexo] AND
[Modelo Minería Con].[Edad] = t.[Edad] AND
[Modelo Minería Con].[Domicilio] = t.[Domicilio] AND
[Modelo Minería Con].[Total Citas] = t.[totalCitas] AND
[Modelo Minería Con].[Nombre Servicio] = t.[NombreServicio] AND
[Modelo Minería Con].[Nombre Medico] = t.[NombreMedico]
where (PredictProbability([Modelo Minería Con].[Nombre Servicio])*100)>50
ORDER BY PredictProbability([Modelo Minería Con].[Nombre Servicio])
desc

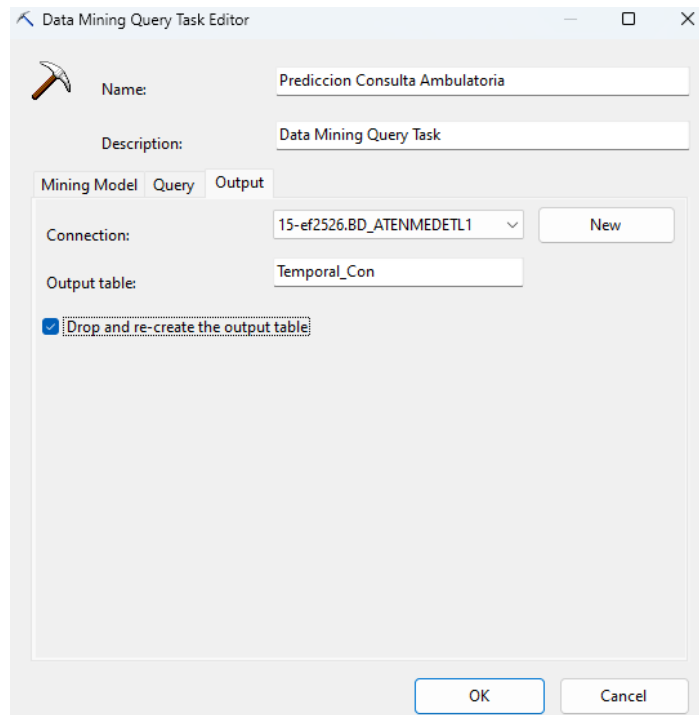
```

**Figura 22.**  
*Consulta DMX – Motivo Enfermedad Súbita*



- En la pestaña Salida, establecemos nuestra conexión de nuestro datamart “DB\_ATENMEDETL”, ingresamos el nombre de la tabla de salida (si no existiese la tabla, en este paso la crea), y, por último, marcamos la casilla que dice “Quitar y volver a crear la tabla de salida”, pues es una tabla temporal; es por eso que la llamamos Temporal\_Con.

**Figura 23.**  
Salida del resultado de la consulta DMX – Motivo Enfermedad Súbita



- Variable Actividad Consulta Ambulatoria

Mediante este flujo de control procedemos a poblar una tabla temporal llamada Influyente\_Con; con los datos de la variable más influyente.

- En la pestaña Modelo de minería de datos, elegimos nuestra conexión de Analysis Services, la estructura de minería de datos, Estructura\_Con, y nuestro modelo de minería de datos: Arbol\_Con.
- En la pestaña Consulta, pegamos la consulta DMX que elaboramos en el paso anterior, para este caso será la siguiente:

```
SELECT TOP 3 NODE_CAPTION
FROM [Arbol_Con].content
WHERE [PARENT_UNIQUE_NAME]>'0' and
[PARENT_UNIQUE_NAME]<='00000000001'
ORDER BY MSOLAP_NODE_SCORE DESC
```

Donde:

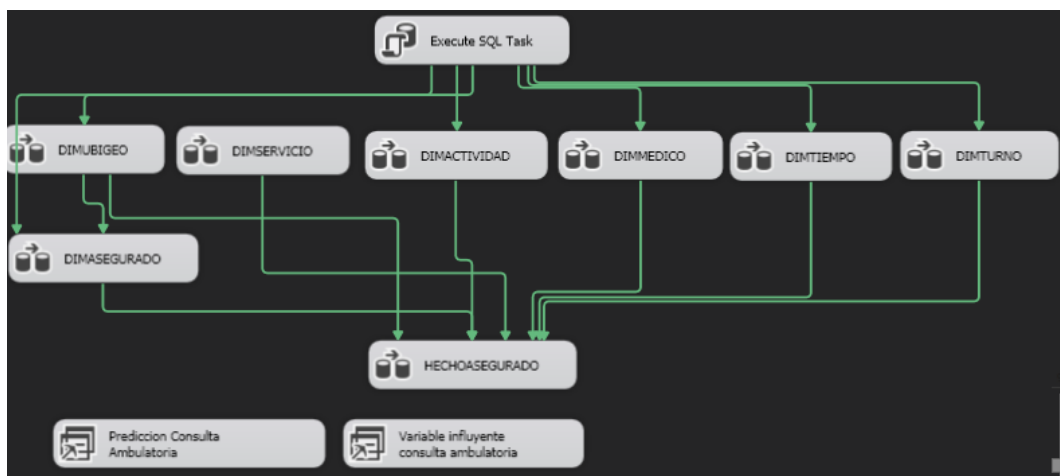
- NODE\_CAPTION = etiqueta del nodo

- PARENT\_UNIQUE\_NAME = Nombre único del nodo primario del nodo.
- MSOLAP\_NODE\_SCORE = importancia del nodo en el modelo.
- En la pestaña Salida, establecemos nuestra conexión de nuestro datamart “DatamartDesercion”, ingresamos el nombre de la tabla de salida (si no existiese la tabla, en este paso la crea), y, por último, marcamos la casilla que dice “Quitar y volver a crear la tabla de salida”, pues es una tabla temporal, cada vez que se ejecute el modelo de minería de datos determinará la variable más influyente, según sea el caso. A la tabla la llamamos Influyente\_Con.

Al ejecutarse las tareas se pueblan las tablas antes mencionadas.

Con todo el proceso seguido hasta el momento, el Flujo de Control, resultó como lo muestra la siguiente imagen:

**Figura 24.**  
*Flujo de Control Final.*



### **Fase V: Evaluación.**

- Evaluar el modelo

Los modelos de árboles de decisión, de la estructura (Estructura\_Con) se han acercado mucho al modelo ideal, superando al modelo de clustering con el cual se comparó su probabilidad de predicción.

Evaluar los resultados realizar el proceso de revisión y determinar futuras fases, son las tareas para detallar en esta fase.

Evaluar los resultados

Teniendo en cuenta los resultados obtenidos en el plan de prueba, se puede determinar que:

Del total de citas a Atención Medica Ambulatoria, ingresados por su tipo de seguro llegamos a la conclusión que.

- Obligatorio: En su mayoría los pacientes son de este tipo de seguro, lo que sugiere que es el tipo de seguro más común.
- Pensionista: Además una proporción significativa de pacientes son de programas de pensiones.
- S. Regulará Á D.LEG.1057 (CAS): Representa un porcentaje relativamente bajo de pacientes, pero aún es significativo. Esto podría indicar que los trabajadores bajo este régimen tienen acceso a servicios de salud.
- Potestativo: Una proporción mínima de pacientes llega a través de seguros potestativos, lo que sugiere que es una categoría poco común.
- Agrario Dependiente: Ambos grupos tienen una presencia muy baja en términos de pacientes, lo que indica que la agricultura puede no ser una industria dominante en términos de atención médica.
- Trabajador del Hogar: Representa una pequeña fracción de pacientes, indicando que el trabajo doméstico no es una fuente principal de demanda de servicios de salud.
- Seg. Salud Trabajad. Independientes (PEAS): Una proporción baja de pacientes llega a través de trabajadores independientes bajo el régimen PEAS.
- Beneficiario Ley 30425: Un porcentaje significativo de pacientes son beneficiarios de la Ley 30425, lo que podría indicar la importancia de esta ley en términos de acceso a la atención médica.

- Construcción Civil: Aunque baja, la presencia de pacientes de la construcción civil sugiere que esta industria contribuye a la demanda de servicios de salud.
- Trabajador Pesquero-Ley 28320: Una proporción mínima de pacientes pertenece a la categoría de trabajadores pesqueros bajo la Ley 28320.
- Salud Seguro Potestativo: Aunque relativamente bajo, algunos pacientes llegan a través de seguros potestativos.

## DASHBOARD PORCENTUAL

Total de ingresados por Actividad		
Actividad	Ingresados	Porcentajes
ATENCION MEDICA AMBULATORIA	97323	100.00%
Total	97323	100%

Total de ingresados por tipo asegurado		
Tipo seguro	Ingresados	Porcentajes
OBLIGATORIO	55101	56.617%
PENSIONISTA	30472	31.310%
S.REGULARÁ Á D.LEG.1057(CAS)	6643	6.826%
POTESTATIVO	29	0.030%
AGRARIO DEPENDIENTE	16	0.016%
TRABAJADOR DEL HOGAR	146	0.150%
SEG.SALUD TRABAJAD.INDEPENDIENTES (PEAS)	33	0.034%
AGRARIO INDEPENDIENTE	8	0.008%
BENEFICIARIO LEY 30425	3894	4.001%
CONSTRUCCION CIVIL	715	0.735%
TRABAJADOR PESQUERO-LEY 28320	1	0.001%
SALUD SEGURO POTESTATIVO	265	0.272%
TOTAL	97323	100%

### 4.2. Presentación, análisis e interpretación de resultados

- **Proceso de Revisión**

Se ha desarrollado el proceso de minería de datos, siguiendo la metodología CRISP-DM, respetando cada una de sus fases, las cuales, según mi apreciación, son muy detalladas, pues permite alinear los objetivos del negocio con los objetivos de minería de datos, incluyendo como tareas la evaluación de los resultados del modelo y de la consecución de los objetivos del negocio.

- **Determinación de futuras fases**

Según lo desarrollado, todas las fases han contribuido a desarrollar satisfactoriamente la alternativa de solución que se había planteado; con la

implementación y la puesta en marcha en la institución, se determinará si es necesario determinar nuevas fases.

- **IMPLEMENTACIÓN**

Plan de implementación

El resultado de la presente aplicación de minería de datos se ha desplegado a través de hojas de cálculo de la aplicación Microsoft Excel 2013:

*Figura 25.*  
*Hoja de cálculo – Monitoreo Consulta Ambulatoria*

## MONITOREO

Tipo Asegurado							
BENEFICIARIO LEY 30425		CONSTRUCCION CIVIL		OBLIGATORIO		PENSIONISTA	
S.REGULARÁ Á D.LEG.1057(CAS)							
Acto Medico Citas	Fecha	Nombre Servicio	Codigo Servicio	Codigo Actividad	Tipo Asegurado	% volver	Total Citas
3673635	martes, 22 de Febrero de 2022	MEDICINA GENERAL	AB1	91	OBLIGATORIO	0.99035639	77
3673635	martes, 22 de Febrero de 2022	MEDICINA GENERAL	AB1	91	OBLIGATORIO	0.99035639	81
3673635	martes, 22 de Febrero de 2022	MEDICINA GENERAL	AB1	91	OBLIGATORIO	0.99035639	85
3673635	martes, 22 de Febrero de 2022	MEDICINA GENERAL	AB1	91	OBLIGATORIO	0.99035639	89
3673635	martes, 22 de Febrero de 2022	MEDICINA GENERAL	AB1	91	OBLIGATORIO	0.99035639	93
3673635	martes, 22 de Febrero de 2022	MEDICINA GENERAL	AB1	91	OBLIGATORIO	0.99035639	97
3673635	martes, 22 de Febrero de 2022	MEDICINA GENERAL	AB1	91	OBLIGATORIO	0.99035639	101
3673635	martes, 22 de Febrero de 2022	MEDICINA GENERAL	AB1	91	OBLIGATORIO	0.99035639	105
3673673	martes, 22 de Febrero de 2022	MEDICINA GENERAL	AB1	91	PENSIONISTA	0.99035639	215
3673673	martes, 22 de Febrero de 2022	MEDICINA GENERAL	AB1	91	PENSIONISTA	0.99035639	227
3673673	martes, 22 de Febrero de 2022	MEDICINA GENERAL	AB1	91	PENSIONISTA	0.99035639	239

### Plan de monitoreo y mantención

El monitoreo del funcionamiento del modelo será realizado por personal de la Oficina de Informática, juntamente con el personal de la Oficina de Planeamiento y Control, quienes serán los usuarios responsables de verificar el buen desempeño de la aplicación.

### Discusión

En este capítulo se hará un análisis de los resultados de la aplicación del sistema desarrollado, en el hospital en estudio.

La población de 97323 citas, considerada entre el año 2022.

Índice de ingreso

El total de citas en la muestra, como ya vimos líneas superiores, es de 97323; el índice de ingreso en nuestra muestra es como se muestra en la siguiente tabla.

A continuación, se detalla el comportamiento de este indicador en los tipos de seguro de los pacientes que tuvieron citas.

*Tabla 4.*

*Índice de ingreso*

<b>Tipo seguro</b>	<b>Ingreso</b>	<b>Porcentajes</b>
<b>OBLIGATORIO</b>	55101	56.617%
<b>PENSIONISTA</b>	30472	31.310%
<b>S.REGULARÁ Á D.LEG.1057(CAS)</b>	6643	6.826%
<b>POTESTATIVO</b>	29	0.030%
<b>AGRARIO DEPENDIENTE</b>	16	0.016%
<b>TRABAJADOR DEL HOGAR</b>	146	0.150%
<b>SEG.SALUD TRABAJAD.INDEPENDIENTES (PEAS)</b>	33	0.034%
<b>AGRARIO INDEPENDIENTE</b>	8	0.008%
<b>BENEFICIARIO LEY 30425</b>	3894	4.001%
<b>CONSTRUCCION CIVIL</b>	715	0.735%
<b>TRABAJADOR PESQUERO-LEY 28320</b>	1	0.001%
<b>SALUD SEGURO POTESTATIVO</b>	265	0.272%
<b>TOTAL DE CITAS</b>	<b>97323</b>	<b>100%</b>

#### **4.3. Prueba de hipótesis**

Para la prueba de hipótesis se utilizan las fórmulas descritas en el diseño de la prueba de hipótesis:

##### **Hipótesis General**

La aplicación de un Modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo.

Según la Tabla 4 se tiene que la cantidad de datos con precisión de la predicción es 100%.

Aplicando la fórmula de Datos de la Precisión de la Predicción se tiene que con el modelo se mejora en un 100%.

#### Hipótesis Específicos

El análisis del modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo

De la Tabla 4 se tiene que la cantidad de datos históricos que se tiene actualmente sin el modelo son de 6 indicadores hospitalarios ya que la información que se almacena es en formatos preoperatorios como se explicó y con el modelo se llega a obtener información de los 12 indicadores hospitalarios ya que la información se almacena en un sistema de información automatizado.

Aplicando la fórmula de análisis predictivo se tiene que con el modelo se mejora en un 100%.

#### **4.4. Discusión de resultados**

Según (Ccopa Mamani & Chavez Viza, 2015) En esta investigación se determinó que el criterio de éxito es el control de los indicadores hospitalarios, los cuales son base para una adecuada toma de decisiones, para ello se aplicó la técnica de regresión lineal del método de regresión del modelo predictivo de minería de datos aplicando la metodología CRISP-DM, utilizando el software Weka y para la recolección de datos se desarrolló un sistema de información. Finalmente, se realizó la prueba de la hipótesis que se demuestra empíricamente, el desarrollo de un “Modelo predictivo basado en minería de datos para la mejora en la toma de decisiones es aceptable ya que mejora en un 91.97%.

En nuestro estudio se obtuvo de la Tabla 4 se tiene que la cantidad de datos históricos que se tiene actualmente sin el modelo son de 6 indicadores hospitalarios ya que la información que se almacena es en formatos preoperatorios como se explicó y con el modelo se llega a obtener información de



los 12 indicadores hospitalarios ya que la información se almacena en un sistema de información automatizado.

Aplicando la fórmula de análisis predictivo se tiene que con el modelo se mejora en un 100%.

## CONCLUSIONES

- Se desarrolló la aplicación de un Modelo predictivo basado en minería de datos para mejorar los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo, gracias al promedio del Coeficiente de Correlación de los indicadores de citas se obtuvo en la evaluación del modelo es correcta mejorando los procesos porcentaje de 100%.
- Se procesaron los datos para el modelo logrando identificar 12 indicadores de la atención médica Policlínico Metropolitano Huancayo, de esta manera se seleccionaron los datos necesarios de acuerdo a la fórmula de cada indicador, en el que se realizó el filtrado de datos en SQL finalmente se procesaron 97323 citas registros obtenidas en el 2022.
- De la Tabla 4 se tiene que la cantidad de datos históricos que se tiene actualmente sin el modelo son de 6 indicadores hospitalarios ya que la información que se almacena es en formatos preoperatorios como se explicó y con el modelo se llega a obtener información de los 12 indicadores hospitalarios ya que la información se almacena en un sistema de información automatizado
- Se logró analizar el modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo, con los datos del hospital y los datos del modelo para lo cual se aplicaron las ecuaciones obtenidas con la técnica de regresión lineal, obteniéndose un 100% de precisión, con lo cual se concluye que el modelo es válido.

## **RECOMENDACIONES**

- Adaptar el modelo predictivo de minería de datos para proyectos futuros de gestión hospitalaria, considerando el desarrollo de un software para la recolección de información y la automatización de reportes para el apoyo en la toma de decisiones.
- sugiere utilizar técnicas de minería de datos para la mejora en la toma de decisión de los diferentes centros de atención en otras instituciones hospitalarias para hacer un comparativo y adoptar mejores procesos.
- Realizar comparaciones de técnicas de minería de datos y evaluarlas con otras medidas de evaluación, de esta manera se lograrán obtener mejores resultados.

## REFERENCIAS BIBLIOGRÁFICAS

Arias, F. G. (2006). *El proyecto de investigación: Introducción a la Metodología* (6th ed.).

[https://issuu.com/fidiasgerardoarias/docs/fidias\\_g.\\_arias.\\_el\\_proyecto\\_de\\_inv](https://issuu.com/fidiasgerardoarias/docs/fidias_g._arias._el_proyecto_de_inv)

Ccopa Mamani, M., & Chavez Viza, S. V. (2015). Modelo Predictivo Basado en Minería de Datos Para la Mejora en la Toma de decisiones del departamento de Cirugía del Hospital Regional Manuel Núñez Butrón [UNIVERSIDAD NACIONAL DEL ALTIPLANO]. In *Universidad Nacional del Altiplano*.

<http://repositorio.unap.edu.pe/handle/UNAP/1911>

Edgar, M. (2023). Modelo matemático para predecir el grado de deserción de los estudiantes en el Instituto Superior Tecnológico Bolívar. In *Repositorio Institucional de la Universidad Técnica de Ambato* (Vol. 593, Issue 03). UNIVERSIDAD TÉCNICA DE AMBATO.

García, E. (2019). *Aplicación de técnicas de minería de datos para la predicción de la deserción en estudiantes de pregrado de la Universidad Pontificia Bolivariana, Sede Central Medellín*. UNIVERSIDAD PONTIFICIA BOLIVARIANA.

Hernández Sampieri, R. (2014). Metodología de la investigación. In S. A. D. C. . McGRAW-HILL / INTERAMERICANA EDITORES (Ed.), *Journal of Chemical Information and Modeling* (Sexta, Vol. 53, Issue 9).

Hernández Sampieri, R., & Mendoza Torres, C. (2018). *Metodología de la Investigación*.

MINSA. (2019, October). *Minsa reúne a expertos para fortalecer las acciones contra la enfermedad de equinocosis | Gobierno del Perú*.

Montes, J., & Elmenthaler, M. (2023). *Sistema asistido por voz para la reserva predictiva de citas médicas en un hospital de categoría III-1 utilizando redes*

*bayesianas*. UPC.

Morone, G. (2013). *Métodos y Técnicas de la Investigación*.

[http://biblioteca.ucv.cl/site/servicios/documentos/metodologias\\_investigacion.pdf](http://biblioteca.ucv.cl/site/servicios/documentos/metodologias_investigacion.pdf)

Reinoso, A. (2023). *DESARROLLO DE UN MODELO PARA PREDECIR EL RENDIMIENTO ACADÉMICO DE ESTUDIANTES DE LA EPN EN BASE A SU NIVEL DE ACCESO A TICS Y FACTORES SOCIOECONÓMICOS*. ESCUELA POLITÉCNICA NACIONAL-QUITO.

Riestra González, M. (2022). *Efectos de los patrones de aprendizaje en línea sobre el rendimiento académico desde una perspectiva de la minería de datos*.

Universidad de Oviedo.

Saire, E. (2023). *PREDICCIÓN DE LA RUTA DE RENDIMIENTO ACADÉMICO CON ALGORITMOS DE CLASIFICACIÓN [UNIVERSIDAD NACIONAL DE SAN AGUSTÍN DE AREQUIPA]*. In *Aleph* (Vol. 87, Issue 1,2).

[https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/167638/341506.pdf?](https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/167638/341506.pdf?sequence=1&isAllowed=y%0Ahttps://repositorio.ufsm.br/bitstream/handle/1/8314/LOEBLEIN%2C%20LUCINEIA)

[sequence=1&isAllowed=y%0Ahttps://repositorio.ufsm.br/bitstream/handle/1/8314/](https://repositorio.ufsm.br/bitstream/handle/1/8314/LOEBLEIN%2C%20LUCINEIA)

[LOEBLEIN%2C LUCINEIA](https://repositorio.ufsm.br/bitstream/handle/1/8314/LOEBLEIN%2C%20LUCINEIA)

[CARLA.pdf?sequence=1&isAllowed=y%0Ahttps://antigo.mdr.gov.br/saneamento/](https://repositorio.ufsm.br/bitstream/handle/1/8314/LOEBLEIN%2C%20LUCINEIA)

[proees](https://repositorio.ufsm.br/bitstream/handle/1/8314/LOEBLEIN%2C%20LUCINEIA)

Saldaña, E. (2015). *Modelo Predictivo de minería de datos de apoyo a la gestión hospitalaria sobre la morbilidad de pacientes Hospitalizados*. UNIVERSIDAD PRIVADA ANTENOR ORREGO.

Vitt, E. (2002). *Business Intelligence: Making Better Decisions Faster* (M. Pr (ed.)).

Yankovic, B. (2011). *Procesos científicos: predecir, interpretar datos, controlar*

*variables*. [https://docplayer.es/13636926-Procesos-cientificos-predecir-](https://docplayer.es/13636926-Procesos-cientificos-predecir-interpretar-datos-controlar-variables-como-trabajar-en-la-sala-de-clases-por-prof-b-yankovic-setiembre-2011.html)

[interpretar-datos-controlar-variables-como-trabajar-en-la-sala-de-clases-por-prof-](https://docplayer.es/13636926-Procesos-cientificos-predecir-interpretar-datos-controlar-variables-como-trabajar-en-la-sala-de-clases-por-prof-b-yankovic-setiembre-2011.html)

[b-yankovic-setiembre-2011.html](https://docplayer.es/13636926-Procesos-cientificos-predecir-interpretar-datos-controlar-variables-como-trabajar-en-la-sala-de-clases-por-prof-b-yankovic-setiembre-2011.html)

**ANEXOS**

# Instrumento de Investigación

## Base de Datos en Excel de Citas Médicas

AtenMed2022 - Excel (Error de activación de productos)

Herramientas de tabla | Herramientas de consultas | Inic.ses.

Archivo | Inicio | Insertar | Disposición de página | Fórmulas | Datos | Revisar | Vista | Programador | Ayuda | Acrobat | Diseño | Consulta | ¿Qué desea hacer? | Cor

Calibri 11 | Fuente | Alineación | Número | Formato condicional | Dar formato como tabla | Estilos de celda | Insertar | Eliminar | Formato | Ordenar y filtrar | Buscar y seleccionar | Edición

	CENTRO	PERIODO	COD_SERVICIO	SERVICIO	COD_ACTIVIDAD	ACTIVIDAD	COD_SUBACTIVIDAD	SUBACTIVIDAD
18695	431	202203	AB1	MEDICINA GENERAL	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18696	431	202203	D11	PEDIATRIA	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18697	431	202203	AC1	MEDICINA INTERNA	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18698	431	202203	AB1	MEDICINA GENERAL	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18699	431	202203	D11	PEDIATRIA	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18700	431	202203	AB1	MEDICINA GENERAL	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18701	431	202203	C11	GINECOLOGIA Y OBSTETRICIA	91	ATENCION MEDICA AMBULATORIA	1	CONSULTA MEDICA
18702	431	202203	A71	GERIATRIA	91	ATENCION MEDICA AMBULATORIA	459	CONSULTA EXTERNA ADULTO
18703	431	202203	AB1	MEDICINA GENERAL	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18704	431	202203	AB1	MEDICINA GENERAL	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18705	431	202203	AB1	MEDICINA GENERAL	91	ATENCION MEDICA AMBULATORIA	1	CONSULTA MEDICA
18706	431	202203	D11	PEDIATRIA	91	ATENCION MEDICA AMBULATORIA	1	CONSULTA MEDICA
18707	431	202203	D11	PEDIATRIA	91	ATENCION MEDICA AMBULATORIA	1	CONSULTA MEDICA
18708	431	202203	AB1	MEDICINA GENERAL	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18709	431	202203	AC1	MEDICINA INTERNA	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18710	431	202203	D11	PEDIATRIA	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18711	431	202203	A71	GERIATRIA	91	ATENCION MEDICA AMBULATORIA	459	CONSULTA EXTERNA ADULTO
18712	431	202203	C11	GINECOLOGIA Y OBSTETRICIA	91	ATENCION MEDICA AMBULATORIA	1	CONSULTA MEDICA
18713	431	202203	AB1	MEDICINA GENERAL	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18714	431	202203	AB1	MEDICINA GENERAL	91	ATENCION MEDICA AMBULATORIA	1	CONSULTA MEDICA
18715	431	202203	AB1	MEDICINA GENERAL	91	ATENCION MEDICA AMBULATORIA	333	TELECONSULTA
18716	431	202203	C11	GINECOLOGIA Y OBSTETRICIA	91	ATENCION MEDICA AMBULATORIA	1	CONSULTA MEDICA

## Matriz de Consistencia

“Aplicación de un Modelo predictivo basado en minería de datos para mejorar los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo”

PROBLEMA GENERAL	OBJETIVO GENERAL	HIPÓTESIS GENERAL	VARIABLE INDEPENDIENTE	DIMENSIÓN	DISEÑO	POBLACIÓN Y MUESTRA
¿De qué manera la aplicación de un Modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo?	Aplicar un Modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo.	La aplicación de un Modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo.	Aplicación de un Modelo predictivo	- Análisis	<p><b>Diseño:</b></p> <p>Experimental</p> <p><b>Tipo de Investigación</b></p> <p>Básica</p>	<p><b>POBLACIÓN</b></p> <p>Todos los datos de la atención médica ambulatoria del Policlínico Metropolitano Huancayo.</p> <p><b>MUESTRA</b></p> <p>Se trabajará con toda la población</p>
PROBLEMA ESPECÍFICO	OBJETIVO ESPECÍFICO	HIPÓTESIS ESPECIFICA	VARIABLE DEPENDIENTE	DIMENSIÓN	MÉTODO DE INVESTIGACIÓN	TÉCNICAS - INSTRUMENTOS
¿Cómo se analizará el modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo?	Analizar el modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo.	El análisis del modelo predictivo basado en minería de datos mejora los procesos de atención médica ambulatoria del Policlínico Metropolitano Huancayo	Procesos de atención médica.	- Atención médica	<p><b>Método</b></p> <p>Deductivo</p> <p><b>Enfoque</b></p> <p>Cuantitativo</p>	<p><b>Técnicas:</b></p> <p>- Encuesta.</p>