

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 la técnica de clasificación de minería de datos
para mejorar los procesos de atención de citas en los
consultorios del Hospital Regional Daniel Alcides Carrión
Pasco, 2023**

Para optar el título profesional de:

Ingeniero de Sistemas y Computación

Autor:

Bach. Luis Alberto MINAYA BLAS

Asesor:

Mg. Pit Frank ALANIA RICALDI

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 la técnica de clasificación de minería de datos
para mejorar los procesos de atención de citas en los
consultorios del Hospital Regional Daniel Alcides Carrión
Pasco, 2023**

Sustentada y aprobada ante los miembros del jurado:

Mg. Lisbeth Gisela NEGRETE CARHUARICRA
PRESIDENTE

Mg. José Luis SOSA SANCHEZ
MIEMBRO

Ing. Melquiades Arturo TRINIDAD MALPARTIDA
MIEMBRO



Universidad Nacional Daniel Alcides

Carrión Facultad de Ingeniería

Unidad de Investigación

INFORME DE ORIGINALIDAD N° 130-2024-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 la técnica de clasificación de minería de datos para mejorar los procesos de atención de citas en los consultorios del Hospital Regional Daniel Alcides Carrión Pasco, 2023

Apellidos y nombres de los tesistas:

Bach. MINAYA BLAS, Luis Alberto

Apellidos y nombres del Asesor:

Mg. ALANIA RICALDI, Pit Frank

Escuela de Formación Profesional

Ingeniería Sistemas y Computación

Índice de Similitud

29%

APROBADO

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

Cerro de Pasco, 26 de setiembre del 2023


UNDA- UNIVERSIDAD NACIONAL DANIEL ALCIDES CARRIÓN
FACULTAD DE INGENIERÍA
UNIDAD DE INVESTIGACIÓN
Luis Villar Requís Carbajal
DOCTOR EN CIENCIAS - DIRECTOR

DEDICATORIA

A mi madre que desde el cielo me ilumina para seguir adelante con mis proyectos, pues sin ella no lo había logrado, por eso te dedico este trabajo en ofrenda por su paciencia y amor madre mía, te amo.

El autor

AGRADECIMIENTO

El principal agradecimiento a Dios quién me ha guiado y me ha dado la fortaleza para seguir adelante.

A mi familia por su comprensión y estímulo constante, además su apoyo incondicional a lo largo de mis estudios.

Y a todas las personas que de una y otra forma me apoyaron en la realización de este trabajo.

RESUMEN

El trabajo de investigación que realice se titula: “Aplicación de la técnica de clasificación de minería de datos para mejorar los procesos de atención de citas en los consultorios del Hospital Regional Daniel Alcides Carrión Pasco, 2023”. El objetivo principal es Determinar de qué manera la Aplicación de la técnica de clasificación de minería de datos mejora los procesos de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023. El diseño que se usó fue no experimental. La población y muestra fueron Citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023. Para la evaluación usamos el instrumento ficha de observación. Con los resultados anteriores e identificando las posibles citas a desaprovecharse, se podrá disminuir el diferimiento de cita, que actualmente hay casos que son de hasta 2 meses; esto repercutirá favorablemente al paciente cuando solicite una cita. Asimismo, esta identificación de deserción permite a los pacientes reprogramar o anular su cita, según sea el caso.

Palabras Clave: Aplicación de la técnica de clasificación de minería de datos, Procesos de atención de citas en los consultorios.

ABSTRACT

The research work that is carried out is entitled: "Application of the data mining classification technique to improve appointment care processes in the offices of the Daniel Alcides Carrión Pasco Hospital, 2023". The main objective is to determine how the application of the data mining classification technique improves appointment care processes in the offices of the Daniel Alcides Carrión Pasco Hospital, 2023. The design used was non-experimental. The population and sample were Appointments in the offices of the Daniel Alcides Carrión Pasco Hospital, 2023. For the evaluation we used the observation sheet instrument. With the previous results and identifying the possible appointments to be wasted, it will be possible to reduce the appointment deferral, which currently there are cases that are up to 2 months; this will have a favorable impact on the patient when requesting an appointment. Likewise, this desertion identification allows patients to reschedule or cancel their appointment, as the case may be.

Keywords: Application of the data mining classification technique, Appointment care processes in medical offices.

INTRODUCCIÓN

El presente trabajo de tesis se estudiará a las citas que se realiza en los consultorios Hospital Daniel Alcides Carrión – Pasco. De acuerdo con los lineamientos del Decreto Supremo, el Ministerio de Salud también aprobó formalmente, hace la entrega a la Empresa Minera del Centro del Perú (CENTROMIN PERU SAC) del antiguo "Hospital Daniel Alcides Carrión" para los fines que hubiera conveniente.

El derecho a la salud incluye el acceso a una atención sanitaria oportuna, aceptable, asequible y de calidad satisfactoria (Organización Mundial de la Salud,2013).

En definitiva, el derecho a la salud significa que los gobiernos deben crear las condiciones que permitan a todas las personas vivir lo más saludablemente posible y esas condiciones incluyen la disponibilidad garantizada de servicios de salud (OMS, 2013).

“La presente está compuesta por los siguientes capítulos y apartados mencionados a continuación”:

“Capítulo I el cual incluye los siguientes apartados: Identificación y determinación del problema, delimitación de investigación, formulación del problema, formulación de objetivos, justificaciones de la investigación y limitaciones de la investigación”.

“Capitulo II el cual incluye los siguientes apartados: Bases teóricas – científicas, definición de términos, identificación de hipótesis y variables y finalmente la definición operacional”.

“Capitulo III el cual abarca: Tipo de investigación, métodos de investigación, diseño de investigación, población y muestra, técnicas e instrumentos de recolección de datos, técnicas de procesamiento y análisis de datos, tratamiento estadístico de datos, selección, validación y confiabilidad de los instrumentos de investigación y orientación ética”.

“Capitulo IV el cual consta de: Resultados y discusión y abarca los siguientes apartados: descripción del trabajo, análisis e interpretación de resultados, prueba de hipótesis y discusión de resultados”.

“Finalmente, se muestran las conclusiones, recomendaciones, referencias bibliográficas y anexos”

El autor.

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.....	3
1.3. Formulación del problema	3
1.3.1. Problema general	3
1.3.2. Problemas específicos	3
1.4. Formulación de objetivos.....	4
1.4.1. Objetivo general.....	4
1.4.2. Objetivos específicos.....	4
1.5. Justificación de la investigación.....	4
1.6. Limitaciones de la investigación	5

CAPITULO II

MARCO TEÓRICO

2.1. Antecedentes de estudio.	6
2.2. Bases teóricas – científicas.	12
2.3. Definición de términos básicos.	26
2.4. Formulación de hipótesis.....	27
2.4.1. Hipótesis general	27
2.4.2. Hipótesis específicas	27

2.5. Identificación de variables	27
2.6. Definición operacional de variables e indicadores	28

CAPITULO III

METODOLOGÍA Y TECNICAS DE INVESTIGACIÓN

3.1. Tipo de investigación	29
3.2. Nivel de investigación	29
3.3. Métodos de investigación	29
3.4. Diseño de investigación.....	29
3.5. Población y muestra	30
3.6. Técnicas e instrumentos de recolección de datos.....	30
3.7. Selección, validación y confiabilidad de los instrumentos de investigación.....	30
3.8. Técnicas de procesamiento y análisis de datos.....	31
3.9. Tratamiento Estadístico.....	31
3.10. Orientación ética filosófica y epistémica.....	31

CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1. Descripción del trabajo de campo.....	32
4.2. Presentación, análisis e interpretación de resultados	68
4.3. Prueba de hipótesis.....	69
4.4. Discusión de resultados.....	70

CONCLUSIONES

RECOMENDACIONES

REFERENCIAS BIBLIOGRÁFICAS

ANEXOS

Instrumento de Recolección de datos

Matriz de Consistencia

ÍNDICE DE TABLAS

Tabla 1. Tareas de Data Mining.....	24
Tabla 2. Operacionalización de variables.....	28
Tabla 3. Descripción de la tabla DBF r_plano\$	33
Tabla 4. Descripción de la tabla DBF Tabla dinámicas	37
Tabla 5. Conversión de tipos de datos de campos de la tabla Admisión	43
Tabla 6. Índice de ingreso.....	72

ÍNDICE DE FIGURAS

Figura 1. Herramientas de BI	13
Figura 2. El proceso de Data Mining	16
Figura 3. Comparación de fases de las metodologías SEMMA y CRISP-DM.....	17
Figura 4. Comparación de fases de las metodologías SEMMA y CRISP-DM.....	18
Figura 5. Técnicas de Minería de Datos.....	24
Figura 6. Herramientas para Inteligencia de Negocios.....	25
Figura 7. Ubicación	33
Figura 8. Estableciendo la conexión EXCEL.....	38
Figura 9. Diseño de Datamart	38
Figura 10. Flujo de datos de la tabla tiempo.	39
Figura 11. Sentencias para la limpieza del Datamart.	44
Figura 12. Flujo de control del ETL – Poblando Hechos	44
Figura 13. Gráfico de Precisión de Minería de Datos - Estructura_Con	45
Figura 14. Dashboard de Ingresados.....	45
Figura 15. Vista de origen de datos.	49
Figura 16. Estructura de Minería de Datos – Estructura_Con	50
Figura 17. Estructura de Minería de Datos – Estructura_Proc	50
Figura 18. Ramificaciones del árbol de decisión – Motivo Enfermedad Súbita.....	51
Figura 19. Variable más influyente – Motivo Enfermedad Súbita	51
Figura 20. Detalle del nodo más cercano o influyente – Motivo Enfermedad Súbita .	52
Figura 21. Ramificaciones del árbol de decisión – Motivo Enfermedad Común.	52
Figura 22. Variable más influyente – Motivo Enfermedad Común.....	53
Figura 23. Detalle del nodo más cercano o influyente – Motivo Enfermedad Común.	53
Figura 24. Elaboración de la consulta DMX – Motivo Enfermedad Súbita.....	54
Figura 25. Relación de pacientes con probabilidad de deserción – Motivo Enfermedad Súbita	55
Figura 26. Elaboración de la consulta DMX – Motivo Enfermedad Común	56
Figura 27. Relación de pacientes con probabilidad de deserción – Motivo Enfermedad Común.....	56
Figura 28. Flujo de Control - Tareas de Consulta de minería de datos.....	56
Figura 29. Selección del modelo de minería de datos – Motivo Enfermedad Común.	57
Figura 30. Consulta DMX – Motivo Enfermedad Común.....	60
Figura 31. Salida del resultado de la consulta DMX – Actividad Procedimientos	60

Figura 32. Selección del modelo de minería de datos – Motivo Enfermedad Súbita ..	61
Figura 33. Consulta DMX – Motivo Enfermedad Súbita	63
Figura 34. Salida del resultado de la consulta DMX – Motivo Enfermedad Súbita	64
Figura 35. Flujo de Control Final	67
Figura 36. Dashboard Porcentual	68
Figura 37. Hoja de cálculo – Monitoreo enfermedad súbita.....	71
Figura 38. Hoja de cálculo – Monitoreo enfermedad Común.....	71

CAPITULO I

PROBLEMA DE INVESTIGACIÓN

1.1. Identificación y determinación del problema

“El Hospital Daniel Alcides Carrión - Pasco, cuenta con 50 años de funcionamiento en la nueva Ciudad de San Juan Pampa - Distrito de Yanacancha, Provincia y Región Pasco, desde que el 25 de Mayo de 1971 a través del Decreto de Ley N° 18863 promulgado por el presidente de la República, JUAN VELASCO ALVARADO”

“Que mediante el cual se declaró de necesidad pública e interés social la remodelación de la Ciudad de Cerro de Pasco y la construcción de la zona de expansión que albergara a la población afectada por el avance de la actividad minera mediante el sistema de tajo abierto. Es así que de conformidad con lo dispuesto por el Decreto Supremo N° 006-79-SA del 27 de setiembre de 1979, corresponde al MINISTERIO DE SALUD para que en representación de la Beneficencia Pública de Cerro de Pasco, deprecione para aquel entonces el Nuevo Local del Centro Médico Daniel Alcides Carrión debidamente equipado e inventariado , el cual a partir del 1ro de Octubre de 1979 pasa a depender del Ministerio de Salud, para cuyo efecto la sociedad de Beneficencia Pública de Cerro de Pasco, cede en uso el inmueble a las áreas que ocupa el referido

nosocomio, manteniendo su propiedad conforme al régimen establecido por Ley N° 8128”.

Así mismo el ministerio de Salud, debidamente autorizado por lo acotado al Decreto Supremo, hace la entrega a la Empresa Minera del Centro del Perú (CENTROMIN PERU SAC) del antiguo "Hospital Daniel Alcides Carrión" para los fines que hubiera conveniente.

El derecho a la salud incluye el acceso a una atención sanitaria oportuna, aceptable, asequible y de calidad satisfactoria (Organización Mundial de la Salud, 2013).

En definitiva, el derecho a la salud significa que los gobiernos deben crear las condiciones que permitan a todas las personas vivir lo más saludablemente posible y esas condiciones incluyen la disponibilidad garantizada de servicios de salud (OMS, 2013).

Según la teoría, las personas con iguales necesidades tienen igual acceso a la atención médica. La accesibilidad influye en cosas como el costo de los servicios de salud como alternativas, los tiempos de espera y las listas de espera, los modos de transporte y la distancia al centro más cercano. Esta noción de equidad respeta las diferencias interpersonales de partida y dispone un marco idóneo para la igualdad de oportunidades. (Jabalera, 2015)

Cuando hay problemas de acceso a servicios de salud, las listas de espera para las citas se hacen más largas y esto a su vez conduce a una mayor falta de asistencia (George, 2003)

La inasistencia de los usuarios a las consultas médicas programadas sin realizar la cancelación de las mismas es un problema que afecta de forma directa la calidad en la prestación de los servicios de salud, aumentando los costos sociales y económicos de la enfermedad. La no asistencia a las consultas influye negativamente en la adherencia de los pacientes a los tratamientos médicos indicados, contribuyendo al empeoramiento de su estado de salud y a la

cronificación de sus enfermedades de base, esta situación conlleva al colapso de los servicios de urgencias, de los centros de referencia de segundo y tercer nivel, redundando en el deterioro de la salud del paciente y en el incremento de los costos operacionales. (Pérez y Rendón, 2016)

Viendo esto se requiere mejorar los procesos de atención en los consultorios de Hospital Daniel Alcides Carrión Pasco por ello analizare si la aplicación de la técnica de clasificación de minería de datos no ayudara a medir la mejora.

1.2. Delimitación de la investigación.

1.2.1. Espacial

La relación entre la aplicación de la técnica de clasificación de minería de datos y el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco.

1.2.2. Temporal

Durante el proceso de investigación de recolección de datos en el año 2023, se realizó el estudio de interpretación de la información.

1.2.3. Conceptual

Se desarrolla la búsqueda teórica la técnica de clasificación de minería de datos y el proceso de atención de citas en los consultorios.

1.3. Formulación del problema

1.3.1. Problema general

¿De qué manera la Aplicación de la técnica de clasificación de minería de datos mejora los procesos de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023?

1.3.2. Problemas específicos

¿Se podrá realizar un análisis mediante los arboles de decisión para medir la mejora en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023?

¿Se podrá realizar un análisis mediante el clustering para medir la mejora en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023?

1.4. Formulación de objetivos

1.4.1. Objetivo general

Determinar de qué manera la Aplicación de la técnica de clasificación de minería de datos mejora los procesos de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023.

1.4.2. Objetivos específicos

Realizar el análisis mediante los árboles de decisión para medir la mejora en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023.

Realizar el análisis mediante el clustering para medir la mejora en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023.

1.5. Justificación de la investigación

1.5.1. A nivel tecnológico.

Se hace uso de la técnica de clasificación de minería de datos para determinar los probables pacientes que no están satisfechos con la atención en los consultorios; “desde el punto económico, permitirá disminuir el desaprovechamiento de dinero presupuestado para pacientes citados por consultorios externos y evitar también el gasto que se genera al atender pacientes por otros servicios, como Emergencia y Hospitalización; socialmente, recae en el interés continuo que prestaría la institución al estar pendiente del perfil de los pacientes que no acuden a sus citas y tomar las medidas correspondientes para mejorar la calidad de atención, con la finalidad de detectar falencias en el personal de la institución de ser el caso, y así evaluar sus procedimientos y calidad de atención, aumentando así su satisfacción por contribuir a que los pacientes

mejoren y lleven una mejor calidad de vida y de salud; científicamente, pues el desarrollo del presente trabajo servirá de guía o base para futuros proyectos de investigación con la finalidad de seguir mejorando la atención de los pacientes”.

1.5.2. Justificación Práctica

La investigación propuesta se realizó, porque se quiere tener pacientes satisfechos con la atención cuando tengan una cita.

1.6. Limitaciones de la investigación

Recursos financieros limitados para cumplir con los requisitos de preparación, presentación de informes y validación de investigaciones.

CAPITULO II

MARCO TEÓRICO

2.1. Antecedentes de estudio

2.1.1. A nivel Internacional

- Martínez (2012) en su trabajo de investigación titulado “Aplicación De Técnicas De Minería De Datos Para Mejorar El Proceso De Control De Gestión En ENTEL”. “El objetivo principal de esta tesis consiste en reducir el tiempo de cálculo de los indicadores de servicios privados de ENTEL, para lo cual se aplicó modelamiento multidimensional, técnicas de minería de datos y automatización de procesos, y de este modo poder entregar información más oportunamente. La metodología de este trabajo se basa principalmente en las etapas del proceso conocido como Knowledge Discovery in Databases (KDD), implementadas de acuerdo a la metodología CRISP-DM, la cual es usada para el desarrollo de proyectos de minería de datos. Para comenzar, se hizo un levantamiento de las métricas existentes para la gestión de la provisión de servicios. Luego, se seleccionaron y procesaron las fuentes de datos para el estudio de forma automática, almacenando las variables más relevantes en un repositorio multidimensional (Data Mart), reduciendo drásticamente el tiempo de cálculo de indicadores y liberando recursos humanos altamente calificados.

A partir de lo anterior, se aplicaron técnicas de clustering para obtener grupos de elementos con datos de clientes y servicios cuyas características fueran similares, asociándoles un valor de precio según información histórica de consumo. Por último, se generó un modelo de clasificación que asignara, de acuerdo a una medida de similitud, elementos que no habían sido facturados a los grupos previamente definidos, y de esta manera estimar los ingresos no percibidos. Con ayuda de minería de datos se logró diseñar nuevas métricas para el proceso e identificar a los clientes y servicios más críticos, lo que permite llegar a valores más exactos de los ingresos perdidos en cada segmento, y aplicar estrategias diferenciadas para hacer el cobro a sus clientes. El trabajo realizado permitió una reducción del tiempo de obtención de los indicadores en un 78%, pasando de un total de 14 horas inicialmente a tan sólo 3 horas, logrando además estimar los ingresos perdidos mensualmente por servicios no facturados en un monto de MM \$ 210, con un error de la estimación menor al 5%. Se espera que, con ayuda de este estudio, la empresa pueda tomar decisiones informadas y mejorar su capacidad de control del proceso de provisión de servicios privados, con el fin de regularizar su flujo de ingreso mensual”.

- Chamba (2015) en su trabajo de investigación titulado “Minería de Datos para segmentación de clientes en la empresa tecnológica Master PC”, la Minería de Datos aplicada en el ámbito de la comercialización permite entre otros aspectos descubrir patrones de comportamiento de clientes, que las empresas pueden utilizar para elaborar estrategias de marketing dirigidas hacia los distintos tipos de clientes que poseen. “El agrupamiento o clustering representa una de las técnicas de Minería de Datos más utilizada para este tipo de análisis, esta técnica se basa en la división de un conjunto de datos en pequeños segmentos o grupos, en donde cada segmento contiene datos similares dentro de sí y mantiene una marcada diferencia con respecto a los

otros segmentos. El presente Trabajo de Titulación tiene por objetivo obtener la segmentación de clientes en la empresa tecnológica Master PC mediante la aplicación de técnicas de Minería de Datos”, para ello se tomó en cuenta el comportamiento de compra, que permitió identificar la lealtad de los clientes de la empresa tecnológica Master PC. Se aplicó la metodología CRISP-DM para el proceso de Minería de Datos. El análisis se realizó en base al modelo RFM (Recencia, Frecuencia, Valor Monetario), y sobre este modelo se aplicaron los algoritmos de agrupamiento: k-means, k-medoids, y Self-Organizing Maps (SOM). “Para validar el resultado de los algoritmos de agrupamiento y seleccionar el que proporcione grupos de mejor calidad, se ha aplicado la técnica de evaluación en cascada aplicando un algoritmo de clasificación. Finalmente se utilizó el algoritmo Apriori para encontrar asociaciones entre productos, para cada grupo de clientes. La herramienta utilizada para el proceso de Minería de Datos fue el entorno RStudio”.

- Sánchez (2017) en su trabajo de investigación titulado “Minería de Datos de la salud: Análisis de los factores que influyen en la realización de cirugías estéticas” A nivel mundial, las cirugías estéticas se han convertido en procedimientos de constante realización y año tras año han venido en aumento. Según la Sociedad Internacional de Cirugía Plástica Estética - ISAPS, Colombia “se ha mantenido a nivel mundial entre los ocho países con mayor número de prácticas de cirugías estéticas anuales. La cirugía estética requiere, como paso indispensable, la asistencia a una cita de valoración con el especialista, y sólo el 40% de éstas, culminan en la realización de algún procedimiento quirúrgico. Por tal razón, los especialistas en cirugía plástica requieren identificar los factores que influyen en la toma de decisión de los pacientes respecto a practicarse un procedimiento estético. Se pretende entonces, mediante minería de datos, descubrir el perfil de los pacientes con mayor probabilidad de realizarse una cirugía plástica y así, definir los factores

que influyen en la decisión de someterse a cirugías estéticas como estrategia para la minimización de tiempo en espera en la asignación de consultas de valoración”.

- Amézquita (2019) en su trabajo de investigación titulado “Minería de datos para la identificación de factores de riesgo en pacientes con hipertensión arterial” La minería de datos es una técnica que ha permitido el análisis de grandes volúmenes de información dando como resultado patrones o reglas que ayudan a entender el comportamiento de un sistema. “Este trabajo se plantea como el uso de esta técnica puede beneficiarse la medicina partiendo del análisis de información de pacientes que padecieron hipertensión del área de Bogotá, determinando los factores de riesgos presentes en esta población basados en las causas descritas en el Catálogo Internacional de Enfermedades. Mediante la aplicación de diferentes algoritmos se pudo determinar variables importantes que pueden determinar la prevalencia y consecuente muerte de estos pacientes”.

2.1.2. A nivel Nacional

- Medrano (2016) en su trabajo de investigación titulado “Modelo de minería de datos usando machine learning con reconocimiento de patrones de síntomas y enfermedades respiratorias en las historias clínicas para mejorar el diagnóstico de pacientes en la ciudad de Trujillo 2016”, Esta investigación se desarrolló en la ciudad de Trujillo desde abril hasta diciembre del presente, centrándose en el problema de diagnósticos médicos y el uso de machine learning para contribuir en la mejora. “La metodología que se usó para desarrollar la investigación fue CRISP. El tipo de investigación de acuerdo al fin es aplicado, de acuerdo a la técnica de contrastación es Pre Experimental, la población es la ciudad de Trujillo, los instrumentos usados en la investigación son el cuestionario y el análisis documental. En conclusión, se logró la mejora

de los diagnósticos médicos a partir de un modelo de machine learning usando el algoritmo de árbol de decisión”.

- 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” El presente trabajo de investigación, propone aplicar un marco estándar de actividades de minería 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”
- Polo (2018) en su trabajo de investigación titulado “Aplicación de la técnica de clasificación de minería de datos para mejorar los procesos de atención de citas en el área de consultorios externos de un Hospital del Departamento de Lambayeque”, diferimiento de citas sea más prolongado, lo que dificulta cumplir con uno de los objetivos institucionales, el cual es brindar una atención oportuna a los pacientes. “A su vez y en consecuencia, no se aprovecha el presupuesto asignado; esto deriva en que la tasa de deserción a la consulta supere el estándar que es 5%. Ante esta situación, se realiza el presente

trabajo para demostrar que, aplicando técnicas de minería de datos se puede determinar la probabilidad de deserción a la consulta médica, lo que repercutirá en una mejor gestión de asignación y/o aprovechamiento en el otorgamiento de citas; siendo justificado tecnológicamente, pues se haría uso de algoritmos de minería de datos para demostrar la probabilidad de deserción; económicamente, se demostraría que existe una mejora en el aprovechamiento del presupuesto asignado; socialmente, pues estas mejoras repercutirían positivamente en los pacientes; científico, pues esta investigación quedará como base para futuros estudios. Para lograr el objetivo propuesto, se debe disminuir el índice de citas no atendidas, disminuir el presupuesto desaprovechado destinado para las atenciones en consultorios externos, y aumentar el índice de pacientes citados producto de la deserción. La presente investigación es de tipo Tecnológica Aplicada, optando como diseño de contrastación de hipótesis, la de sucesión o llamada también pre test / post test o en Línea, en la que se consideró, como población, la data histórica con la que cuenta la institución desde el año 2010 al año 2016, de la cual se tomará como muestra los pacientes atendidos durante el año 2017; y se hizo uso de encuestas y entrevistas para la recolección de datos. Como resultado del trabajo se pudo concluir que durante el periodo de prueba, se pudo disminuir 1.00% el índice de deserción en consulta ambulatoria, y un 29.79% en procedimientos. En lo que respecta al presupuesto, se aprovecharía S/.110,670.00 en consulta ambulatoria y S/.1,247,400.00 en procedimientos; asimismo, se determinó que, mediante el uso de la minería de datos, se hubiera aprovechado 2,635 citas en consulta ambulatoria y 29,700 en procedimientos, destinados a pacientes en lista de espera”

2.1.3. A nivel Local

- No se encontró evidencia o literatura de contenido similar en la biblioteca de la Universidad Nacional Daniel Alcides Carrión.

2.2. Bases teóricas – científicas

2.2.1. Aplicación de la técnica de clasificación de minería de datos.

Dado que este es un puesto de inteligencia empresarial (BI), es crucial comprender BI y todo lo que implica porque están directamente relacionados con este puesto.

2.2.1.1. Business Intelligence (Inteligencia de Negocios)

Vieira (2009), “nos dice que BI es un conjunto de conceptos y metodologías que, haciendo uso de acontecimientos (hechos) y sistemas sustentados en los mismos, apoya la toma de decisiones en los negocios. Para esto ser posible es necesario adquirir los datos, por ejemplo, por medio de un sistema de procesamiento on-line de transacciones (OLTP), almacenarlos en un sistema de base de datos, como un Data Warehouse del cual se puede generar aún un subconjunto más específico de datos, Datamart y, finalmente procesar estos datos con una herramienta de análisis que puede ser: una herramienta de procesamiento analítico online (OLAP), un sistema de informaciones para ejecutivos (EIS); un sistema de apoyo a la decisión (DSS); o aun, un sistema de descubierta y predicción (Data mining)”.

Para Curto y Conesa (2010) “Business Intelligence es el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización”.

Stair y Reynolds (2000) manifiestan “que la inteligencia de negocios es un proceso que implica recopilar información correcta,

suficiente, oportuna y utilizable, y analizarla para que tenga efecto positivo en la estrategia, tácticas y operaciones del negocio”.

Considerando las definiciones descritas anteriormente, se puede concluir que Business Intelligence permite transformar:

- Convertir datos en información.
- Tomar mejores decisiones rápidamente que ayuden a mejorar el rendimiento de una compañía o institución y que además que mejore su ventaja competitiva (objetivo primario del BI)

En la siguiente figura se puede observar, según Viera et al (2009), las herramientas de Business Intelligence:

Figura 1. Herramientas de BI

Tipo de Herramienta	Cuestión básica	Ejemplo de respuesta
Data Mining	¿Qué es interesante? ¿Qué puede suceder?	Tipos de clientes Predicción de ventas
OLAP	¿Qué sucedió y por qué?	Ventas mensuales versus variaciones de precios de los competidores
EIS/DSS	¿Qué necesito saber ahora ?	Cotizaciones diversas
Estudios y informes	¿Qué sucedió ?	Ventas del último mes

2.2.1.2. Datamart.

Según Date (2001) “un Datamart es un almacén de datos especializado, orientado a un tema, integrado, volátil y variante en el tiempo para apoyar un subconjunto específico de decisiones de administración, asimismo, el Datamart es especializado y volátil. Por especializado queremos decir que contiene datos para dar apoyo (solamente) a un área específica de análisis de negocios; por volátil queremos decir que los usuarios pueden actualizar los datos e incluso, posiblemente, crear nuevos datos (es decir, nuevas tablas) para algún propósito”.

Stair y Reynolds (2000) “nos dicen que los Datamarts contienen un subgrupo de datos relacionados con un solo aspecto de los negocios de una compañía, por ejemplo finanzas, inventario o personal son de mucha utilidad para grupos pequeños que deseen acceder a datos detallados”.

Se creó un data mart para poder utilizar técnicas de minería de datos para perfilar a los pacientes que abandonan porque se resolverían problemas con el procesamiento de las visitas médicas.

2.2.1.3. Extract, Transform and Load – ETL

“ETL es una tecnología que permite extraer datos del entorno de origen, transformarlos según nuestras necesidades de negocio para integración de datos y cargar estos datos en los entornos destino. Los entornos de origen y destino son usualmente bases de datos, ficheros, u otras fuentes estructuradas, semi estructurada o no estructurada” (Conesa y Curto 2010)

Este procedimiento fue necesario para combinar los datos preexistentes de diversas fuentes en los hospitales estudiados. Cabe recalcar que estos datos deben pasar un proceso de transformación que garantice su calidad y con ello obtener un buen perfilamiento de los pacientes desertores.

2.2.1.4. Minería de Datos (Data Mining)

Según Rodríguez et al (2003), “los orígenes de Data Mining se pueden establecer a finales de los años 80, cuando la administración de hacienda estadounidense, desarrolló un programa de investigación para detectar fraudes en las declaraciones de impuestos. Sin embargo, la gran expansión de Data Mining no se produce hasta la década de 1990 originada principalmente por tres factores”:

- Mayor potencia informática.

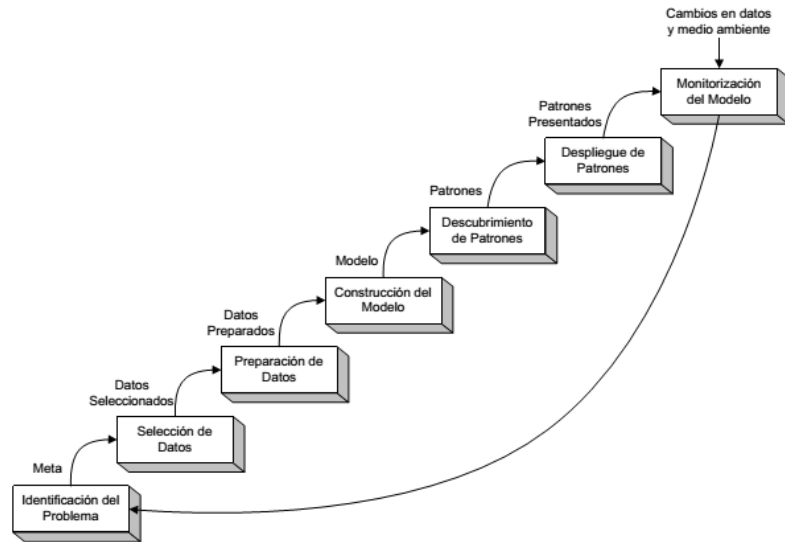
- Mayor velocidad de recopilación de datos debido al hardware más barato y la automatización del proceso de recopilación de datos.
- Aparición de nuevas tecnologías de aprendizaje y almacenamiento de datos.

Ferrer et al (2000) hace mención a la definición de minería de datos que hace (Piatetsky-Shapiro et al, 1992); donde la define como un conjunto de técnicas y herramientas aplicadas al proceso no trivial de extraer y presentar conocimiento implícito (previamente desconocido), potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con objeto de predecir de forma automatizada tendencias y comportamientos y/o describir de forma automatizada modelos previamente desconocidos.

Las disciplinas de estadística, tecnología de bases de datos, reconocimiento de patrones y visualización de datos incluyen la minería de datos.

Para determinar el perfil de los pacientes que salen de las consultas médicas se recogieron datos para este estudio

Figura 2. El proceso de Data Mining



Mediante la Figura 2, “podemos observar el proceso de minería de datos, como un proceso independiente de extracción y exploración de información oculta”.

“Sin embargo, existen diversas metodologías para los proyectos de minería de datos”.

“En el siguiente apartado se mencionará algunas metodologías existentes para proyectos de minería de datos”

2.2.1.5. Metodologías para proyectos de Data Mining

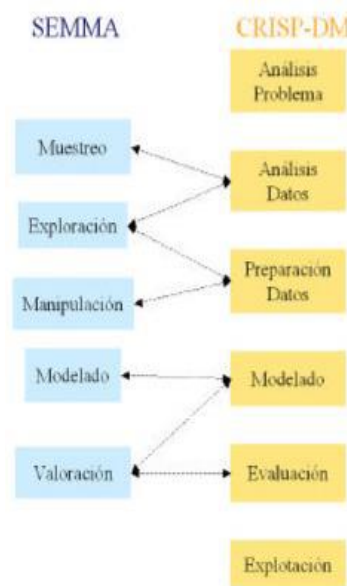
“Son diversos los modelos de proceso que han sido propuestos para el desarrollo de proyectos de Data Mining tales como”:

- SEMMA (Sample, Explore, Modify, Model, Assess) (SAS, 2003)
- DMAMC (Definir, Medir, Analizar, Mejorar, Controlar) (Isixsigma, 2005)
- CRISP-DM (Cross Industry Standard Process for Data Mining) (Chapman et al., 2000)

Básicamente las metodologías SEMMA y CRISP-DM comparten la misma naturaleza y el mismo objetivo ya que se busca solucionar un

problema que permanece en la empresa y que los resultados están ocultos dentro toda la información; sin embargo, se decidió por utilizar la metodología CRISP-DM, pues adicionalmente incorpora un análisis previo del problema existe en el negocio, tal como se muestra en la siguiente figura.

Figura 3. Comparación de fases de las metodologías SEMMA y CRISP-DM



- **Metodología CRISP-DM**

Los orígenes de CRISP-DM se remontan hacia el año 1999, cuando un importante consorcio conformado por empresas europeas tales como NCR (Dinamarca), AG (Alemania), SPSS (Inglaterra), OHRA (Holanda), Teradata, SPSS, y Daimler-Chrysler, proponen a partir de diferentes versiones de KDD (Knowledge Discovery in Databases) [Reinartz, 1995], [Adraans, 1996], [Brachman, 1996], [Fayyad, 1996], el desarrollo de una guía de referencia de libre distribución denominada CRISP-DM (Cross Industry Standard Process for Data Mining). (Gallardo, 2009)

“Por ser una metodología de libre distribución puede trabajar con cualquier herramienta para desarrollar el proyecto que esté disponible en

el mercado aplicando así una característica adicional que es el de ser una metodología equitativa”.

“Así se distingue esta metodología como un modelo que está basado en situaciones reales que ocurren en la empresa, organizando sus fases principales y estableciendo tareas secundarias que especifican con más detalle las obligaciones de cada una de las seis tareas”.

Figura 4. Comparación de fases de las metodologías SEMMA y CRISP-DM



“A continuación, se detalla las fases de la metodología CRISP-DM”, según Gallardo (2009):

Análisis del Problema o Comprensión del negocio

Esta etapa se enfoca en comprender los objetivos y especificaciones del proyecto desde un punto de vista empresarial. Esta información se convierte en conocimiento para la definición del problema y el diseño del plan para alcanzar los objetivos.

Está compuesta por las siguientes tareas o actividades:

- “Determinar los objetivos del negocio: tiene como metas, determinar cuál es el problema que se desea resolver, por qué la necesidad de utilizar Data Mining y definir los criterios de éxito”

- “Evaluación o valoración de la situación: se definen los requisitos del problema, tanto en términos de negocio como en términos de Data Mining”.
- “Determinar los objetivos de minería de datos: tiene como objetivo representar los objetivos del negocio en términos de las metas del proyecto de DM, como, por ejemplo, si el objetivo del negocio es el desarrollo de una campaña publicitaria para incrementar la asignación de créditos hipotecarios, la meta de DM será, por ejemplo, determinar el perfil de los clientes respecto de su capacidad de endeudamiento”.
- “Realizar el plan del proyecto: tiene como meta desarrollar un plan para el proyecto, que describa los pasos a seguir y las técnicas a emplear en cada paso”.

Análisis de los Datos o Comprensión de los datos

La recopilación de datos es el primer paso de esta fase, después del cual se toman medidas para desconfidencializar los datos, detectar problemas de calidad de los datos y encontrar subgrupos interesantes para plantear teorías sobre la información oculta.

Está compuesta por las siguientes actividades:

- Recopilar los datos originales: el objetivo es crear un informe que detalle los datos recopilados, dónde se ubicaron, los métodos utilizados para recopilarlos y los problemas específicos que surgieron durante el proceso.
- Descripción de los datos: Una vez recogidos los datos iniciales se deben describir.
- Exploración de datos: el objetivo es encontrar la estructura general de los datos. Para determinar las características de los datos recién

adquiridos, se aplican pruebas estadísticas básicas. En este proceso también se producen tablas de frecuencia y gráficos de distribución.

- Verificar la calidad de los datos: se verifica la coherencia de los datos de los valores de campo individuales, el número y la distribución de los valores cero y los valores fuera de rango que pueden introducir ruido en el proceso. La idea en este punto es garantizar la integridad y corrección de los datos.

Preparación de los datos

“Esta fase involucra todas las actividades para construir el conjunto final de datos a partir del conjunto inicial. Estas actividades se llevan a cabo varias veces y no en un orden predefinido. Entre estas tareas se encuentran la selección de atributos, así como la transformación y limpieza de los datos para las herramientas del modelado”.

Para cumplir con esta fase, se realiza las siguientes actividades:

- Seleccionar datos: utilizando los criterios establecidos en el paso anterior, elija un subconjunto de los datos obtenidos en el paso anterior.
- Limpieza de datos: Esta tarea complementa la anterior y es una de las que requiere más tiempo y mano de obra porque requiere una variedad de técnicas para mejorar la calidad de los datos y dejarlos listos para el modelado.
- Crear datos: esta tarea implica procedimientos de preparación de datos como construir nuevos atributos a partir de atributos preexistentes, incorporar nueva información o alterar los valores de atributos preexistentes.
- “Integrar los datos: involucra la creación de nuevas estructuras, a partir de los datos seleccionados, por ejemplo, generación de nuevos

campos a partir de otros existentes, creación de nuevos registros, fusión de tablas que contengan atributos diferentes para un mismo objeto, agregación de nuevos campos o nuevas tablas donde se resumen características de múltiples registros o de otros campos en nuevas tablas de resumen”.

- “Formateo de los datos: consiste principalmente, en la realización de transformaciones sintácticas de los datos sin modificar su significado, esto, con la idea de permitir o facilitar el empleo de alguna técnica de DM en particular, como por ejemplo la reordenación de los campos, etc”.

Modelado

“En esta fase se seleccionan y aplican varias técnicas de modelado o algoritmos y sus parámetros son calibrados para obtener los mejores resultados. Algunos algoritmos tienen requerimientos específicos para el formato de los datos, en cuyo caso se debe regresar a la fase de preparación de datos y realizar las tareas necesarias para obtener dicho formato, las veces que sean necesarias”

Las tareas que se deben realizar para cumplir con esta fase son:

- “Seleccionar la técnica de modelado: consiste en la selección de la técnica de DM más apropiada al tipo de problema a resolver. Para esta selección, se debe considerar el objetivo principal del proyecto y la relación con las herramientas de DM existentes. Por ejemplo, si el problema es de clasificación, se podrá elegir de entre árboles de decisión, k-nearest neighbour o razonamiento basado en casos (CBR); si el problema es de predicción, análisis de regresión, redes neuronales; o si el problema es de segmentación, redes neuronales, técnicas de visualización, etc”.

- “Generar el plan de prueba: se debe generar un procedimiento destinado a probar la calidad y validez del mismo. Por ejemplo, en una tarea supervisada de DM como la clasificación, es común usar la razón de error como medida de la calidad. Entonces, típicamente se separan los datos en dos conjuntos, uno de entrenamiento y otro de prueba, para luego construir el modelo basado en el conjunto de entrenamiento y medir la calidad del modelo generado con el conjunto de prueba”.
- “Construir el modelo: se ejecuta sobre los datos previamente preparados para generar uno o más modelos”.
- “Evaluar el modelo: se interpretan los modelos de acuerdo al conocimiento preexistente del dominio y los criterios de éxito preestablecidos”.

Evaluación

Esta fase examina el proceso de construcción y confirma que se han logrado los objetivos iniciales.

Las tareas a realizar son:

- “Evaluar los resultados: involucra la evaluación del modelo en relación a los objetivos del negocio y busca determinar si hay alguna razón de negocio para la cual, el modelo sea deficiente, o si es aconsejable probar el modelo, en un problema real si el tiempo y restricciones lo permiten”.
- “Proceso de revisión: se refiere a calificar al proceso entero de DM, a objeto de identificar elementos que pudieran ser mejorados”.
- “Determinación de futuras fases: si se ha determinado que las fases hasta este momento han generado resultados satisfactorios, podría pasarse a la fase siguiente, en caso contrario podría decidirse por otra

iteración desde la fase de preparación de datos o de modelación con otros parámetros”.

Explotación o Implementación

“En esta no se acaba el proyecto, puesto que, se deben hacer pruebas de su implementación, así como un seguimiento periódico”.

“Los resultados que presentan deben ser claros de tal manera que se puedan comprender fácilmente por personas que no tengan mucha intervención en el área de informática, sino también de profesionales o no que pertenezcan a otras áreas, presentando resultados de una manera visual usando cuadros estadísticos que representen las variables usadas y su aporte para la empresa”.

- Plan de implementación: Este ejercicio crea una estrategia de implementación teniendo en cuenta los hallazgos de la evaluación.
- “Plan de monitoreo y mantención: es aconsejable preparar estrategias de monitorización y mantenimiento para ser aplicadas sobre los modelos”.
- Informe final: “es la conclusión del proyecto de DM realizado”.
- Revisión del proyecto: “se evalúa qué fue lo correcto y qué lo incorrecto, qué es lo que se hizo bien y qué es lo que se requiere mejorar”

2.2.1.6. Tareas de Minería de Datos

“Una tarea es un tipo de problema de data mining. Por ejemplo: clasificar piezas en defectuosas, no defectuosas, defectuosas reparables y defectuosas no reparables es una tarea. Esta tarea se podría resolver mediante árboles de decisión o redes neuronales, entre otros métodos, éstos, son métodos o técnicas que permiten resolver las tareas” (Hernández, 2004).

“Existen dos grandes grupos en los que se pueden dividir las tareas, tal como podemos observar en la siguiente figura”:

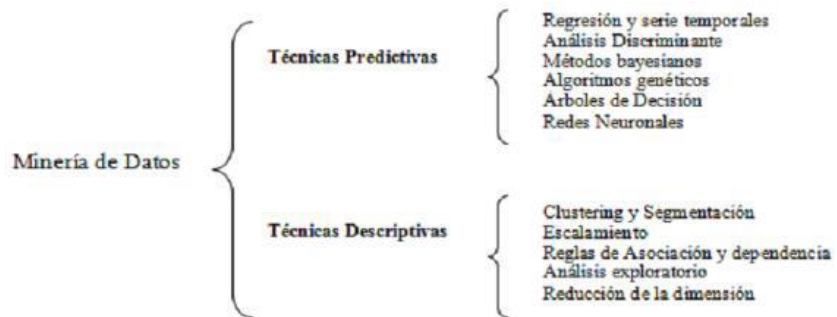
Tabla 1. Tareas de Data Mining

Tipo de Tareas	Descripción	Ejemplos
Predictivas	Se trata de problemas en los que hay que predecir uno o más valores para uno o más ejemplos	Clasificación.
		Clasificación suave.
		Estimación de probabilidad de clasificación.
		Categorización.
		Preferencias o priorización.
		Regresión
Descriptivas	Los ejemplos se presentan como un conjunto de datos sin ordenar ni etiquetar de ninguna manera. Por lo tanto, el objetivo, no es predecir nuevos datos sino describir los existentes	Clustering.
		Correlación y factorizaciones
		Reglas de asociación.
		Dependencias funcionales.
		Detección de valores e instancias anómalas.

2.2.1.7. Técnicas de Minería de Datos

“Las técnicas de la minería de datos provienen de la Inteligencia artificial y de la estadística, dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener resultados” (Belinchón, 2010).

Figura 5. Técnicas de Minería de Datos



2.2.1.8. Herramientas de Inteligencia de Negocios

“Para el desarrollo de aplicación de inteligencia de negocios, en las que incluyen el proceso de minería de datos, se cuenta con una gran diversidad de herramientas, tal como se puede visualizar en la siguiente figura”:

Figura 6. Herramientas para Inteligencia de Negocios



“Debido a que la institución en estudio cuenta con licencia de Microsoft SQL Sever y a que se cuenta con profesionales capacitados en el uso de ese gestor, se decidió utilizarla, la cual brinda una plataforma completa para el análisis de datos e inteligencia empresarial, tales como Microsoft SSIS (SQL Server Integration Services), que permite integrar datos de diferentes orígenes, construye administra y despliega soluciones de integración escalables y construye o alimenta un datawarehouse”

“Asimismo, se hace uso del servicio Microsoft SSAS (SQL Server Analysis Services), que provee una vista unificada e integrada para todos los datos de la empresa, reportes, análisis OLAP, marcadores de indicadores de desempeño y data mining”

2.2.2. Procesos de atención de citas en los consultorios.

“En el presente apartado, se considera las bases teóricas científicas que sustentan el presente trabajo; por un lado, de la institución y por el otro de la tecnología o herramienta utilizada”.

“Para entender el contexto de la problemática abordada es necesario saber que la consulta médica es la actividad prestacional que incluye evaluación integral general, diagnóstico, prescripción, tratamiento, monitoreo, seguimiento, orientación de pacientes con condiciones clínicas y/o exposición a riesgos, que

se brinda en el centro asistencial, la cual tiene un tiempo aproximada de duración de 15 minutos”.

“Asimismo, existen procesos de atención que permiten gestionar la consulta médica, los cuales son un conjunto de actividades relacionadas entre sí, cuya finalidad es brindar una atención adecuada al paciente para solicitar, reprogramar o anular una consulta médica, en adelante cita”.

- Solicitar cita: “proceso en el que el paciente solicita una consulta médica, a través de los diferentes medios que la institución brinda”
- Reprogramación de cita: “proceso en el cual, el paciente, solicita que se re programe su consulta médica ya programadas, por distintos motivos”.
- Anulación de citas: “cuando el paciente determina que le será imposible asistir a su consulta médica, debe solicitar la anulación de su cita oportunamente, para así ofrecer su cupo a un paciente que se encuentra en espera”.

2.3. Definición de términos básicos

- **Árboles de Decisión:**

Un árbol de decisión es un mapa de los posibles resultados de una serie de decisiones relacionadas. Permite que un individuo o una organización comparen posibles acciones entre sí según sus costos, probabilidades y beneficios.

- **Clustering:**

El clustering es una de las técnicas de machine learning basadas en análisis estadístico que se utiliza para analizar los datos en entornos Big Data.

- **Eliminación de datos:**

La eliminación de datos es el borrado de los datos de un ordenador o dispositivo de red de modo que sea imposible acceder a los datos o verlos.

- **Registro de datos:**

Un registro es toda la información de un elemento que se almacena en un archivo o tabla de la base de datos.

- **Reprogramación de datos:**

La reprogramación es una facilidad de pago que te permitirá extender el plazo para pagar la deuda a un número determinado de cuotas. Se aplica una tasa de interés preferencial y se coloca en cuotas fijas que te permitirán ordenar tus pagos.

2.4. Formulación de hipótesis

2.4.1. Hipótesis general

La aplicación de la técnica de clasificación de minería de datos mejorará el Procesos de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023.

2.4.2. Hipótesis específicas

El análisis mediante los árboles de decisión medirá la mejora en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023.

El análisis mediante el clustering medirá la mejora en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023.

2.5. Identificación de variables

2.5.1. Variables independientes

Aplicación de la técnica de clasificación de minería de datos.

2.5.2. Variables dependientes

Procesos de atención de citas en los consultorios.

2.6. Definición Operacional de variables e indicadores

Tabla 2. Operacionalización de variables

VARIABLE	DIMENSIONES
INDEPENDIENTE	Árboles de Decisión
	Clustering
Aplicación de la técnica de clasificación de minería de datos	Interfaz
DEPENDIENTE	Registro
	Reprogramación
Procesos de atención de citas en los consultorios	Eliminación

CAPITULO III

METODOLOGÍA Y TECNICAS DE INVESTIGACIÓN

3.1. Tipo de investigación

El tipo de investigación es aplicada puesto que, se ha generado conocimiento relevante a raíz de la aplicación de técnicas de minería de datos, cuyo diseño de contrastación es diseño de sucesión o en línea, donde se considerará los reportes y encuestas emitidas (Hernández Sampieri et al., 2014)

3.2. Nivel de investigación

El nivel será Descriptivo ya que se analizará una muestra de la población seleccionada del objeto de estudio.(Hernández Sampieri, 2014)

3.3. Métodos de investigación

La investigación que realizare es de método analítico-inductivo utiliza el razonamiento para obtener conclusiones que parten de hechos aceptados como válidos, para llegar a conclusiones.(Hernández Sampieri & Mendoza Torres, 2018)

3.4. Diseño de investigación

El diseño de contrastación de hipótesis es de sucesión o llamada también pre test / post test o en Línea, donde para determinar el perfil de los pacientes desertores, se consideró aquellas citas comprendidas entre los años 2021 y 2022.

Una vez establecido el perfil de deserción, éste se contrastará con las características de las citas del año 2023 que será el grupo experimental. (Hernández Sampieri & Mendoza Torres, 2018)

3.5. Población y muestra

3.5.1. Población

Según Hernández Sampieri (2014) menciona que: “es un conjunto de individuos que se hallan en un determinado sector y que nos apoya para adquirir la muestra y los resultados”.

La población de 1148 citas, considerada entre el año 2021 y 2022.

3.5.2. Muestra

Para Arias, Villasis y Miranda (2016) comento que “la muestra es una parte de la población que es característico de un todo y se opta por adquirir información acerca de las variables”.

Tomaremos en cuenta toda la población.

3.6. Técnicas e instrumentos de recolección de datos

“Para el presente trabajo de tesis se utiliza la técnica de búsqueda de información. Estos instrumentos se utilizan para recopilar información para el desarrollo de la investigación”.

Técnicas:

- Observación.

Instrumentos:

- Ficha de observación.

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

Selección de datos.

“La confiabilidad de los instrumentos de recolección de datos aplicados en la presente investigación fue revisada mediante la validez por juicio de expertos. Los mismos que forman parte de los anexos de la presente investigación”.

3.8. Técnicas de procesamiento y análisis de datos

“Se hizo uso para el apoyo de procesamiento y análisis de datos el software estadístico SPSS, porque permitió realizar diversos análisis de las variables (Estadística descriptiva e Inferencial)” como:

- Elaboración del reporte final de la investigación.
- Presentación del reporte final de la investigación.

3.9. Tratamiento estadístico

También se utilizó estadística inferencial para probar las hipótesis durante el procesamiento estadístico de los datos. Se utilizaron estudios descriptivos para describir, visualizar y resumir los datos.

3.10. Orientación ética filosófica y epistémica.

En la presente investigación se tuvo en cuenta lo siguiente:

Su información personal será omitida del formulario de registro porque mantendremos su confidencialidad y garantizaremos el anonimato.

Transparencia en la información de los resultados, en el sentido de que, no se añadirá información adicional ni distorsionada acerca de las fichas de registro.

CAPITULO IV

RESULTADOS Y DISCUSIÓN

4.1. Descripción del trabajo de campo

4.1.1. Organización

4.1.1.1. Descripción

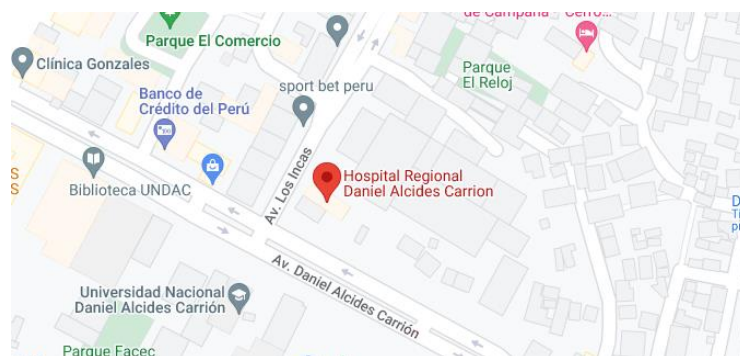
El Hospital Daniel Alcides Carrión “como objetivo institucional es mejorar el recurso humano a través de capacitaciones, estímulos y mejores condiciones laborales para que su satisfacción profesional se vea plasmada y reflejada en la atención de los usuarios”.

“El Hospital Daniel Alcides Carrión es sui generis en su desarrollo puesto que se ha ido acomodando a las necesidades de la población y en este tiempo ha cambiado para responder a nuevos retos hasta ser un hospital de gran complejidad sin perder su vínculo con la comunidad”.

4.1.1.2. Ubicación y Servicios

Se encuentra ubicado Av. Daniel Alcides Carrión 520, Cerro De Pasco.

Figura 7. Ubicación



4.1.1.3. Identificación de requerimientos

Recolectar Datos Iniciales.

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

DESCRIPCIÓN DE LOS DATOS

- **r_plano\$**

Tabla 3. Descripción de la tabla DBF r_plano\$

r_plano\$	Lista todas las actividades					
Total, de registros	1148 – 1.898 MB					
Nombre del campo	Tipo	Longitud	Decimal	Permite Nulls	Es Autonumérico	Descripción campo
cod_eess	Varchar	255	0	SI	NO	Código hospital
eess	Varchar	255	0	SI	NO	Nombre de hospital
e_ubigeo	Varchar	255	0	SI	NO	Ubigeo de hospital
E_dpto	Varchar	255	0	SI	NO	Departamento
E_prov	Varchar	255	0	SI	NO	Provincia
E_dist	Varchar	255	0	SI	NO	Distrito
E_cod_disa	Varchar	255	0	SI	NO	
E_disa	Varchar	255	0	SI	NO	
E_cod_red	Varchar	255	0	SI	NO	Código de red del hospital

E_red	Varchar	255	0	SI	NO	Nombre de red
E_cod_mred	Varchar	255	0	SI	NO	Código de micro red del hospital
E_mred	Varchar	255	0	SI	NO	Nombre de micro red
Tip_eess	Varchar	255	0	SI	NO	Tipo de centro de salud
Cat_eess	Varchar	255	0	SI	NO	Categoría de centro de salud
Nro_hc	Varchar	255	0	SI	NO	Numero de paciente
Nombre	Varchar	255	0	SI	NO	Nombre de paciente
Apellido	Varchar	255	0	SI	NO	Apellido de paciente
Dirección	Varchar	255	0	SI	NO	Dirección de paciente
Tip_doc	Varchar	255	0	SI	NO	Tipo de documento de paciente
Documento	Varchar	255	0	SI	NO	Numero de documento de paciente
Cod_etnia	Varchar	255	0	SI	NO	Código de etnia
etnia	Varchar	255	0	SI	NO	Etnia
Cod_financia	Varchar	255	0	SI	NO	Código de financiación
Financia	Varchar	255	0	SI	NO	Financiación
Codsexo	Varchar	255	0	SI	NO	Código de sexo
Sexo	Varchar	255	0	SI	NO	sexo
Edad	Float	8	0	SI	NO	Edad
Cod_t_edad	Varchar	255	0	SI	NO	Código tipo de edad
T_edad	Varchar	255	0	SI	NO	Tipo de edad
Pac_resha	Varchar	255	0	SI	NO	Código de residencia
Pac_rhdpt	Varchar	255	0	SI	NO	Residencia departamento
Pac_rhprov	Varchar	255	0	SI	NO	Residencia provincia
Pac_rhdist	Varchar	255	0	SI	NO	Residencia distrito

Pac_proce	Varchar	255	0	SI	NO	Procedencia
Pac_dpto	Varchar	255	0	SI	NO	Procedencia departamento
Pac_prov	Varchar	255	0	SI	NO	Procedencia provincia
Pac_dist	Varchar	255	0	SI	NO	Procedencia distrito
Acomp	Varchar	255	0	SI	NO	Acompañante
Acomp_doc	Varchar	255	0	SI	NO	Acompañante documento
Cod_mot_aten	Varchar	255	0	SI	NO	Código motivo de atención
Mot_aten	Varchar	255	0	SI	NO	Motivo de atención
Ocurren	Varchar	255	0	SI	NO	Ocurrencia
Ocu_dpto	Varchar	255	0	SI	NO	Ocurrencia departamento
Ocu_prov	Varchar	255	0	SI	NO	Ocurrencia provincia
Ocu_dist	Varchar	255	0	SI	NO	Ocurrencia distrito
Ocu_ups	Varchar	255	0	SI	NO	Ocurrencia ups
Ups	Varchar	255	0	SI	NO	Ups
F_aten	Datetime	8	0	SI	NO	Fecha de atención
H_aten	Varchar	255	0	SI	NO	Hora de atención
F_egreso	Datetime	8	0	SI	NO	Fecha de egreso
H_egreso	Varchar	255	0	SI	NO	Hora de egreso
estancia	Float	8	0	SI	NO	Estancia
Cod_cond_egre	Varchar	255	0	SI	NO	Código de condición de egreso
Condición_egre	Varchar	255	0	SI	NO	Condición de egreso
Observ	Varchar	255	0	SI	NO	Observación
Obs_fing	Datetime	8	0	SI	NO	Observación fecha de ingreso
Obs_hing	Varchar	255	0	SI	NO	Observación hora de ingreso
Obs_fegr	Datetime	8	0	SI	NO	Observación fecha de egreso

Obs_hegr	Varchar	255	0	SI	NO	Observación hora de egreso
Obs_cama	Varchar	255	0	SI	NO	Observación cama
Obs_prof	Varchar	255	0	SI	NO	Código observación profesional
Obs_prof_nombre	Varchar	255	0	SI	NO	Observación profesional nombre
Destino	Varchar	255	0	SI	NO	Destino
Des_estab	Varchar	255	0	SI	NO	Destino
Des_serv	Varchar	255	0	SI	NO	Destino servicio
Cod_prio	Varchar	255	0	SI	NO	Código prioridad
Prioridad	Varchar	255	0	SI	NO	Prioridad
Cod_f_ing	Varchar	255	0	SI	NO	Código fecha de ingreso
Form_ing	Varchar	255	0	SI	NO	Forma de ingreso
Cod_medico	Varchar	255	0	SI	NO	Código medico
Medico	Varchar	255	0	SI	NO	Medico
Colegio_gprof	Varchar	255	0	SI	NO	Colegio grupo de profesionales
Coddiag1	Varchar	255	0	SI	NO	Código de diagnóstico 1
Tipdiag1	Varchar	255	0	SI	NO	Tipo diagnostico 1
Coddiag2	Varchar	255	0	SI	NO	Código diagnostico 2
Tipdiag2	Varchar	255	0	SI	NO	Tipo diagnostico 2
Coddiag3	Varchar	255	0	SI	NO	Código diagnostico 3
Tipdiag3	Varchar	255	0	SI	NO	Tipo diagnostico 3
Coddiag4	Varchar	255	0	SI	NO	Código diagnostico 4
Tipdiag4	Varchar	255	0	SI	NO	Tipo diagnostico 4
Codcpt1	Varchar	255	0	SI	NO	
Codcpt2	Varchar	255	0	SI	NO	
Codcpt3	Varchar	255	0	SI	NO	
Codcpt4	Varchar	255	0	SI	NO	

Obs_diag1	Varchar	255	0	SI	NO	Observación diagnóstico 1
Obs_diag2	Varchar	255	0	SI	NO	Observación diagnóstico 2
Diag1_des	Varchar	255	0	SI	NO	Diagnóstico 1 descripción
Diag2_des	Varchar	255	0	SI	NO	Diagnóstico 2 descripción
Diag3_des	Varchar	255	0	SI	NO	Diagnóstico 3 descripción
Diag4_des	Varchar	255	0	SI	NO	Diagnóstico 4 descripción

• **Tabla dinámica\$**

Tabla 4. Descripción de la tabla DBF Tabla dinámicas

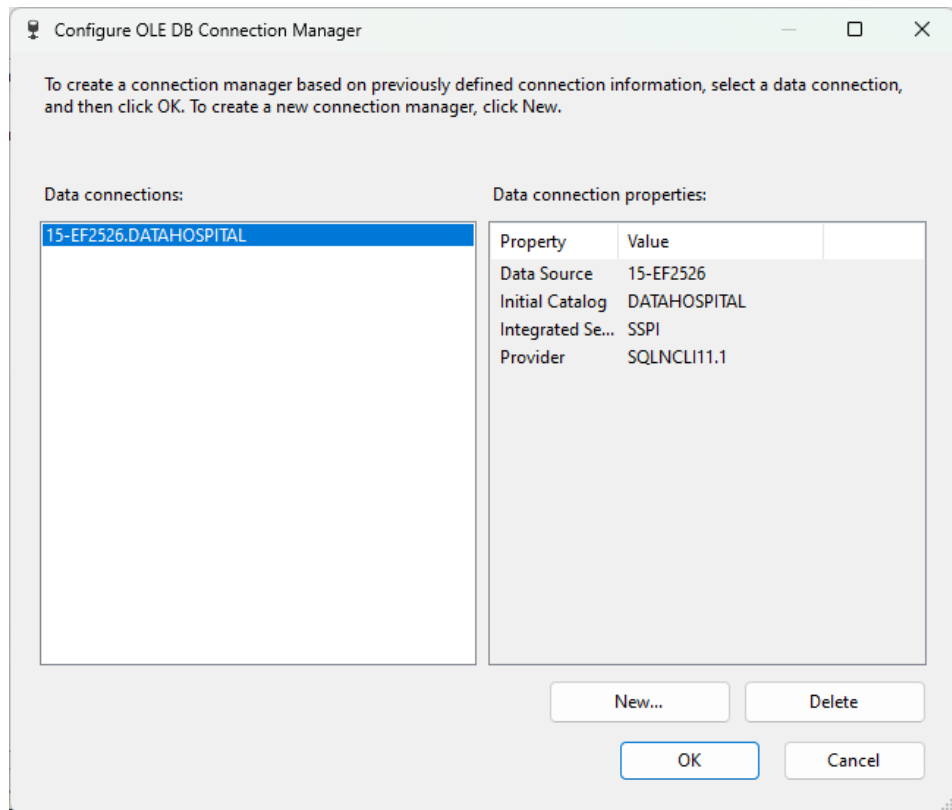
Tabla dinámica\$	Lista todas las actividades					
Total, de registros	2 – 0.008 MB					
Nombre del campo	Tipo	Longitud	Decimal	Permite Nulls	Es Autonumérico	Descripción campo
Cuenta de cod_eess	Varchar	255	0	SI	NO	Cuenta de código eess
Total	Float	8	0	SI	NO	total

Preparación de los datos

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

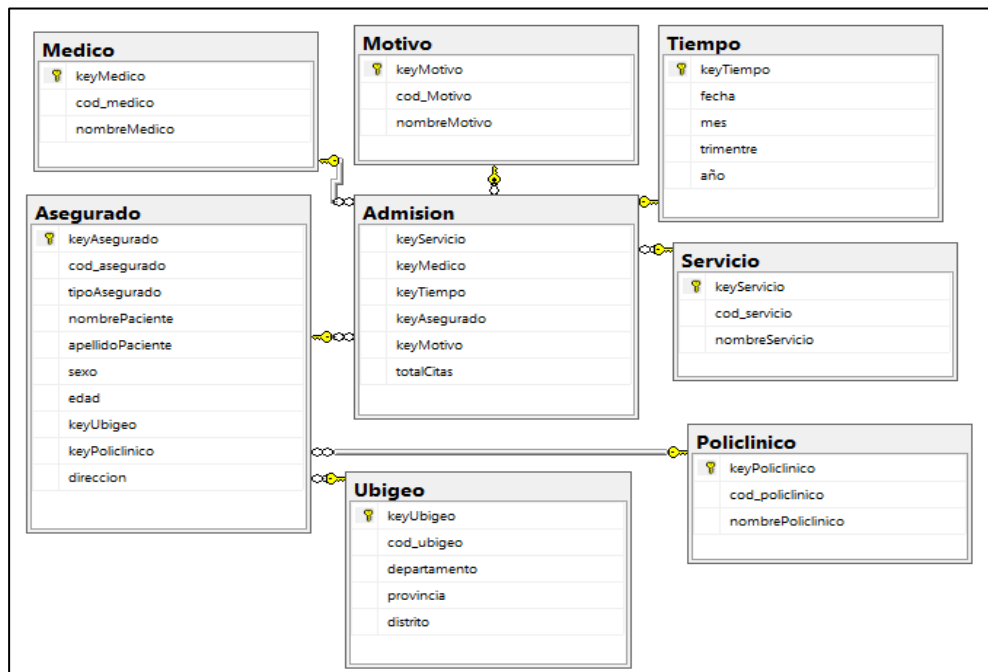
“Limpieza, estructurado, integración y formateo de los datos. Para realizar la limpieza de los datos se creó una conexión OLE DB, con la finalidad de conectarnos con las tablas DBF a utilizar; dicha conexión tiene por nombre: ConexionHospital, y ha servido para crear nuestro Datamart”

Figura 8. Estableciendo la conexión EXCEL



4.1.1.4. Base de datos

Figura 9. 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”.

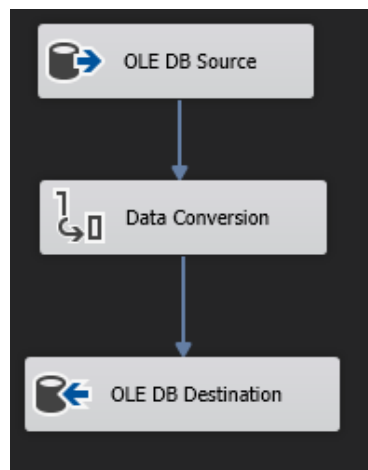
“Luego se ha creado un Paquete de SQL Server Integration Services (SSIS), llamado ETL, donde se realizará el proceso, valga la redundancia, de ETL (Extracción, transformación y carga). Es aquí donde se realiza el poblado de las Dimensiones y los Hechos, cumpliendo con las tareas de esta fase. Las dimensiones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

Dimensiones

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

```
SELECT DISTINCT cod_eess AS cod_policlinico, eess AS  
nombrePoliclinico FROM dbo.r_plano$
```

Figura 10. 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.

4.1.1.5. Dimensión Ubigeo

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

```
SELECT DISTINCT pac_resha AS cod_ubigeo, pac_dpto AS
departramento, pac_prov AS provincia, pac_dist AS distrito
FROM dbo.r_plano$
WHERE pac_resha 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.

4.1.1.6. Dimensión Servicio

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

```
SELECT DISTINCT cod_ups AS cod_servicio, ups AS
nombreServicio
FROM dbo.r_plano$
```

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

4.1.1.7. Dimensión Motivo

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

```
SELECT DISTINCT cod_mot_aten AS cod_Motivo, mot_aten AS
nombreMotivo
from dbo.r_plano$
where mot_aten is not null
```

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

4.1.1.8. Dimensión Médico

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

```
SELECT DISTINCT cod_medico, medico AS nombreMedico  
FROM dbo.r_plano$
```

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

4.1.1.9. Dimensión Tiempo

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

```
SELECT DISTINCT f_aten AS fecha, MONTH(f_aten) AS mes,  
YEAR(f_aten) AS año, DATEPART(quarter, f_aten) AS trimestre  
FROM dbo.r_plano$  
WHERE f_aten IS NOT NULL
```

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, cabe añadir aquí que se eliminaron los datos de tipo null de fecha ya que este lo requería.

4.1.1.10. Dimensión Asegurado

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

```
select polD.keyPoliclinico AS keyPoliclinico, ubiD.keyUbigeo AS  
keyUbigeo, nro_hc AS codAsegurado, financia AS tipoAsegurado, nombre  
AS nombrePaciente, apellido As apellidoPaciente, sexo, edad, direccion
```

```
from DATAHOSPITAL.dbo.r_plano$ hosData inner join
BD_HOSPITALETL.dbo.Policlinico polD on (polD.nombrePoliclinico =
hosData.eess)
```

```
inner join BD_HOSPITALETL.dbo.Ubigeo ubiD on (ubiD.districto =
hosData.pac_dist)
```

Este proceso se realizó de la misma manera que las otras dimensiones, nada más que aquí se crearon unas 2 conexiones que se requerían para esta dimensión las cuales son Ubigeo y Policlínico.

4.1.1.11. Hecho Admisión

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

```
select serD.keyServicio As keyServicio, meD.keyMedico AS
keyMedico, tiD.keyTiempo AS keyTiempo, asD.keyAsegurado AS
keyAsegurado, acD.keyMotivo As keyMotivo, ROW_NUMBER()
OVER(ORDER BY nro_hc) AS totalCitas
```

```
from DATAHOSPITAL.dbo.r_plano$ hosData inner join
BD_HOSPITALETL.dbo.Servicio serD on (serD.cod_servicio =
hosData.cod_ups)
```

```
inner join
BD_HOSPITALETL.dbo.Medico meD on (meD.cod_medico =
hosData.cod_medico)
```

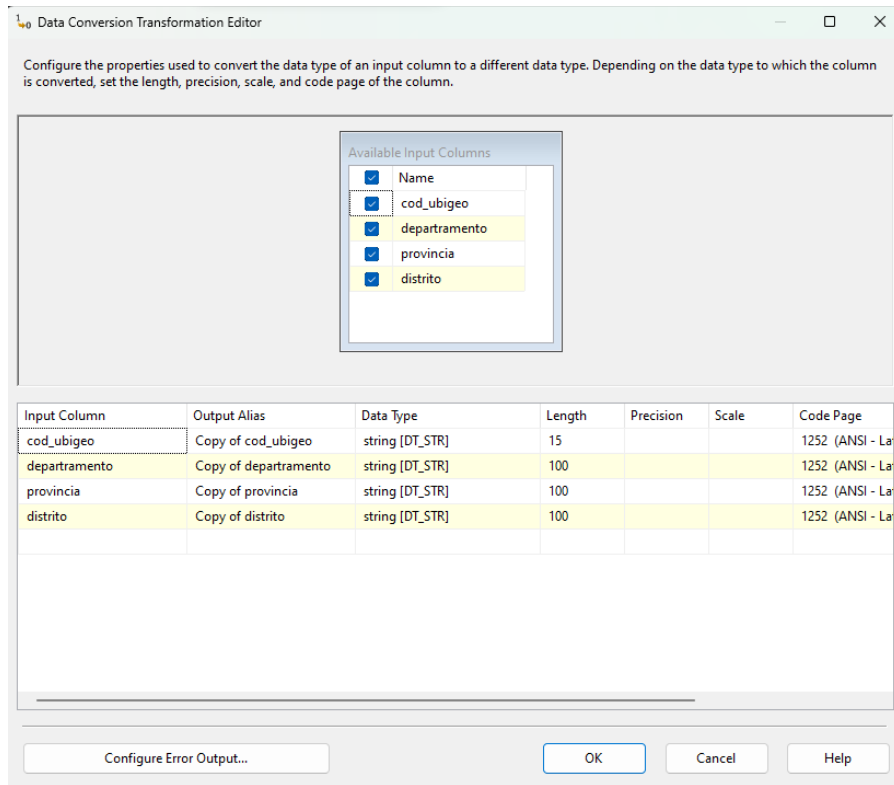
```
inner join
BD_HOSPITALETL.dbo.Tiempo tiD on (tiD.fecha = hosData.f_aten)
```

```
inner join
BD_HOSPITALETL.dbo.Asegurado asD on (asD.cod_asegurado =
hosData.nro_hc)
```


inner join

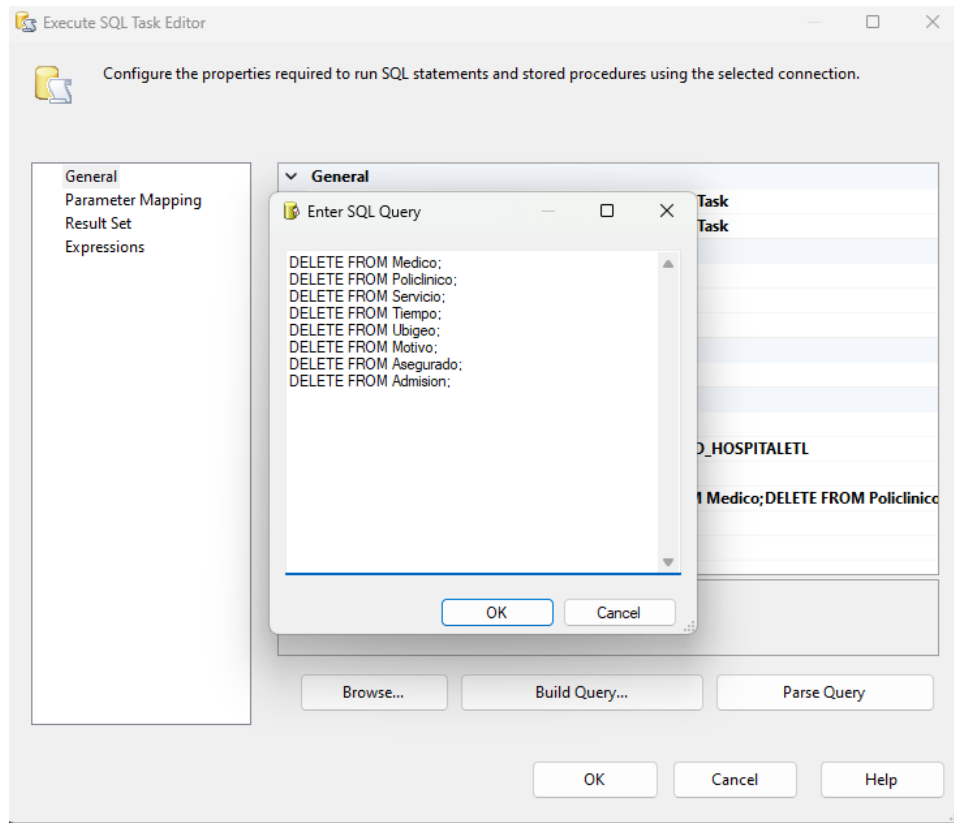
```
BD_HOSPITALETL.dbo.Motivo acD on (acD.cod_motivo =  
hosData.cod_mot_aten)
```

Tabla 5. Conversión de tipos de datos de campos de la tabla Admisión



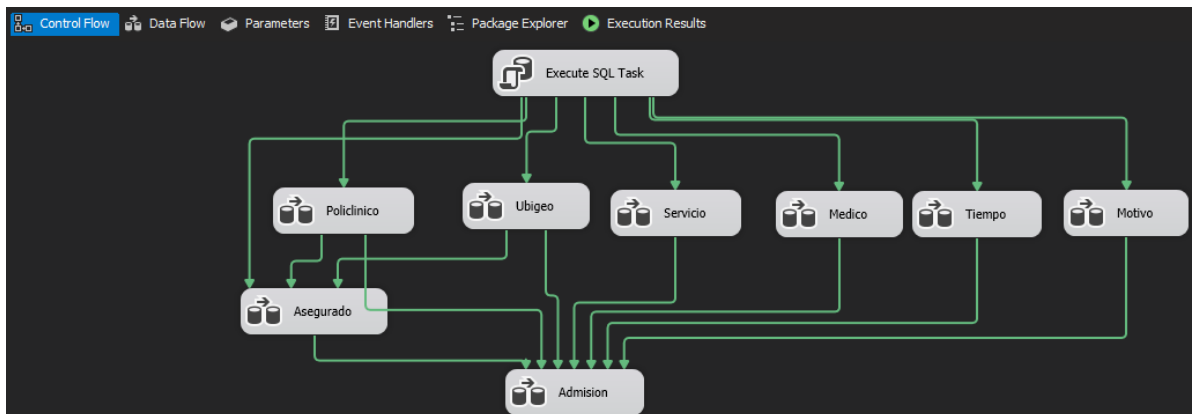
Conversión de datos, incluida la adición de columnas a conjuntos de datos y el cambio del tipo de datos de las columnas. Como se ve en la imagen, la herramienta se utiliza para cambiar el tipo de datos de la columna, como es habitual en todas las tablas.

Figura 11. 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 12. Flujo de control del ETL – Poblando Hechos



4.1.1.12. MODELADO

Seleccionar técnica de modelado

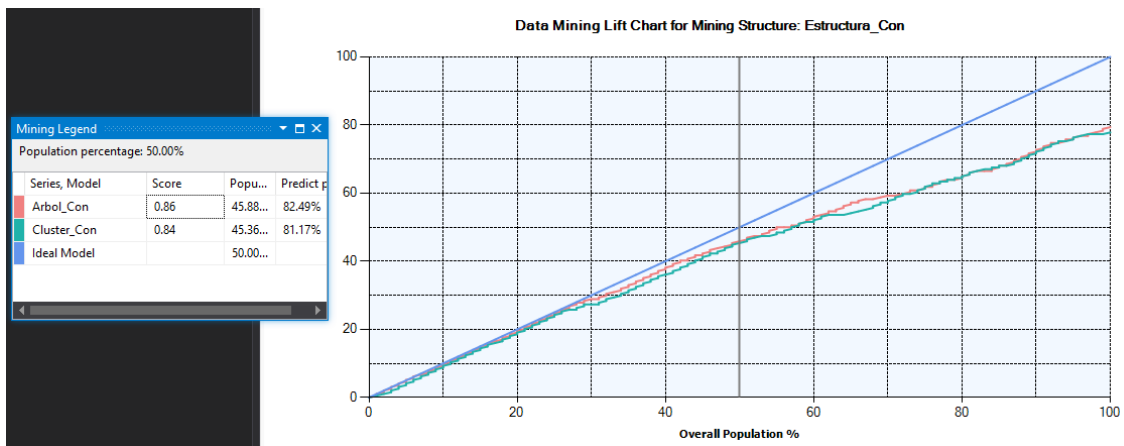
Aquí, se consideró 2 técnicas de modelado:

- 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 82.49% de probabilidad de predicción frente a un 81.17% del modelo de clustering en la estructura de motivo Enfermedad Súbita (Estructura_Con)”.

Figura 13. Gráfico de Precisión de Minería de Datos - Estructura_Con



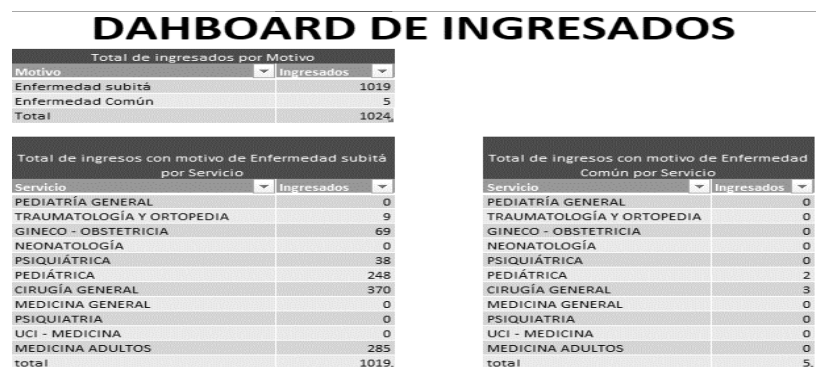
4.1.1.13. Generar el plan de prueba

La data histórica se remonta a los meses de febrero y marzo del año 2023, donde para realiza el plan de prueba.

El resultado de este plan de prueba fue el siguiente:

- Motivo Enfermedad Súbita De un total de 1019 (100%) citas registradas en el periodo del grupo de prueba.
- Actividad Procedimientos De un total de 5 (100%) citas registradas en el periodo del grupo de prueba.

Figura 14. Dashboard de Ingresados



4.1.1.14. 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, 4 vistas llamadas:

- Vista_Mineria_Con

Almacena el histórico de citas de los ingresados por enfermedad súbita; se generó la siguiente consulta SQL:

```
SELECT A.tipoAsegurado, A.sexo, A.edad,
       M.nombreMedico, P.nombrePoliclinico, U.departamento,
       U.provincia, U.districto, H.actoMedicoCitas, H.totalCitas,
       S.nombreServicio
FROM Policlinico P INNER JOIN Asegurado
A ON P.keyPoliclinico = A.keyPoliclinico INNER JOIN
Ubigeo U ON A.keyUbigeo = U.keyUbigeo INNER
JOIN Admision H ON A.keyAsegurado =
H.keyAsegurado INNER JOIN Motivo AC ON
H.keyMotivo = AC.keyMotivo 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 (AC.cod_Motivo = '10') AND (S.cod_servicio <>
'230101') AND (YEAR(T.fecha) = 2023)
```

- Vista_Mineria_Proc

Almacena el histórico de citas de los ingresados por enfermedad común; se generó la siguiente consulta SQL:

```
SELECT A.TipoAsegurado, A.Sexo, A.Edad,
       M.NombreMedico, P.NombrePoliclinico, U.Departamento,
```

```

        U.Provincia, U.Distrito, H.ActoMedicoCitas, H.totalCitas,
        S.NombreServicio
FROM Policlínico P INNER JOIN Asegurado
        A ON P.keyPoliclinico = A.keyPoliclinico INNER JOIN
        Ubigeo U ON A.keyUbigeo = U.keyUbigeo INNER
        JOIN Admision H ON A.keyAsegurado =
        H.keyAsegurado INNER JOIN Motivo AC ON
        H.keyMotivo = AC.keyMotivo 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 (AC.cod_Motivo = '11') AND (S.cod_Servicio <>
'230102') AND (YEAR(T.fecha) = 2023)

```

- Vista_Datos_Citas_Con

Almacena los registros, de motivos de ingreso por enfermedad súbita, que serán evaluados.

```

CREATE VIEW Vista_Datos_Citas_Con
AS
SELECT P.NombrePoliclinico, A.TipoAsegurado,
        A.NombrePaciente, A.Sexo, A.Edad,
        U.Departamento,          U.Provincia,          U.Distrito,
        H.ActoMedicoCitas,
        H.totalCitas, T.fecha, AC.cod_Motivo,
        S.NombreServicio,  M.NombreMedico,  M.cod_medico,
        S.cod_Servicio
FROM Policlínico P INNER JOIN Asegurado
        A ON P.keyPoliclinico = A.keyPoliclinico INNER JOIN

```

```

Ubigeo U ON A.keyUbigeo = U.keyUbigeo INNER
JOIN Admision H ON A.keyAsegurado =
H.keyAsegurado INNER JOIN Motivo AC ON
H.keyMotivo = AC.keyMotivo 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 (YEAR(T.fecha) = 2023) AND (AC.cod_Motivo <> '10')
AND (S.cod_Servicio <> '230101')

```

- Vista_Datos_Citas_Proc

Almacena los registros, de ingreso por enfermedad común, que serán evaluados.

```

CREATE VIEW Vista_Datos_Citas_Proc
AS
SELECT P.NombrePoliclinico, A.TipoAsegurado,
       A.NombrePaciente, A.Sexo, A.Edad,
       U.Departamento, U.Provincia, U.Distrito,
       H.totalCitas, T.fecha, AC.cod_Motivo,
       S.NombreServicio, M.NombreMedico, S.cod_Servicio
FROM Policlinico P INNER JOIN Asegurado
A ON P.keyPoliclinico = A.keyPoliclinico INNER JOIN
Ubigeo U ON A.keyUbigeo = U.keyUbigeo INNER
JOIN Admision H ON A.keyAsegurado =
H.keyAsegurado INNER JOIN Motivo AC ON
H.keyMotivo = AC.keyMotivo INNER JOIN
Servicio S ON H.keyServicio = S.keyServicio INNER
JOIN Tiempo T ON H.keyTiempo = T.keyTiempo

```

INNER JOIN Medico M ON

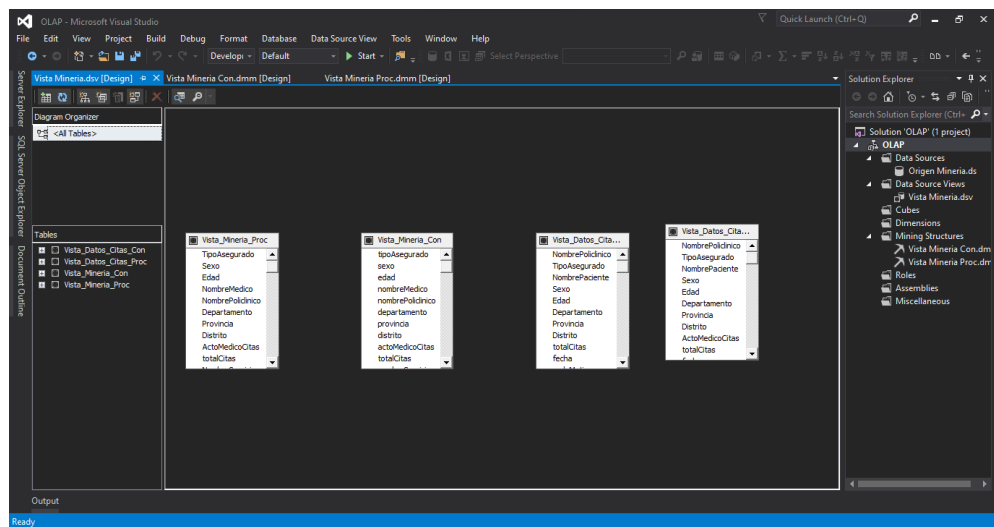
H.keyMedico = M.keyMedico

WHERE (YEAR(T.fecha) = 2023) AND (AC.cod_Motivo = '11')
AND (S.cod_Servicio <> '230102')

“Mediante el uso de la plataforma de SQL Server Analysis Services; se agregó un proyecto de analysis services llamado OLAP, 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 las cuatro vistas creadas en el SQL Server Management Studio

Figura 15. Vista de origen de datos.



“Luego de crear la vista, se creó las estructuras de minería de datos, una para citas de motivo enfermedad súbita y la otra para enfermedad común”.

Figura 16. Estructura de Minería de Datos – Estructura_Con

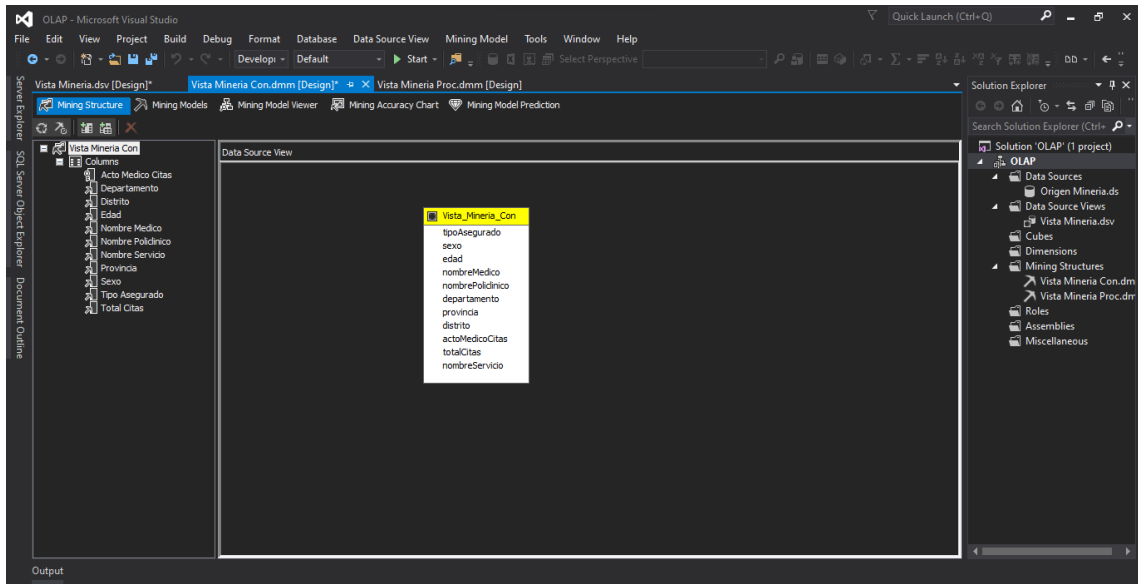
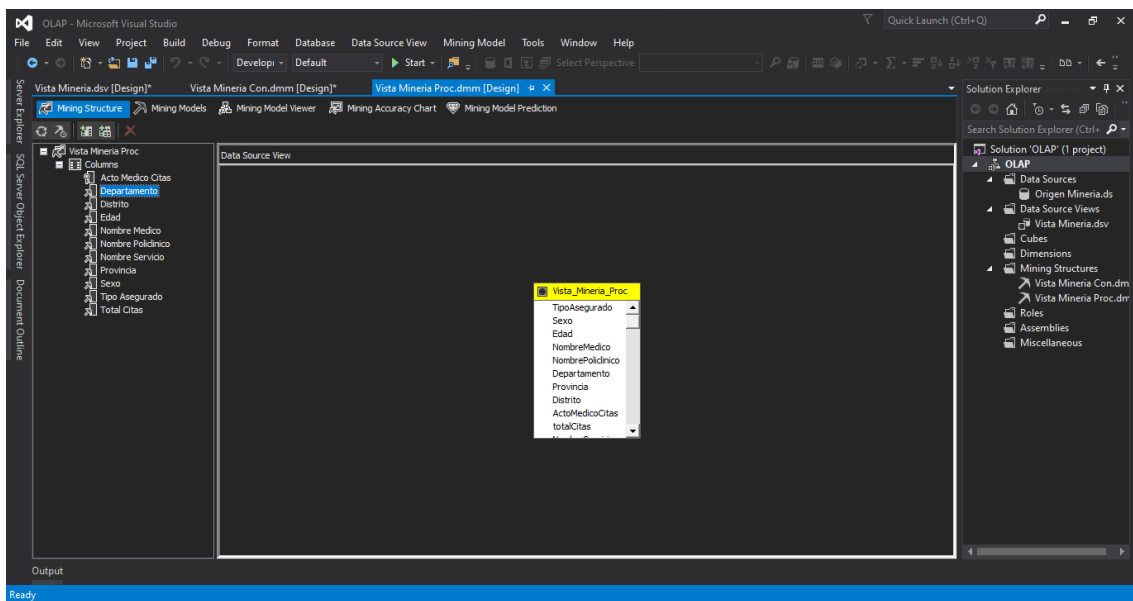


Figura 17. Estructura de Minería de Datos – Estructura_Proc



“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”.

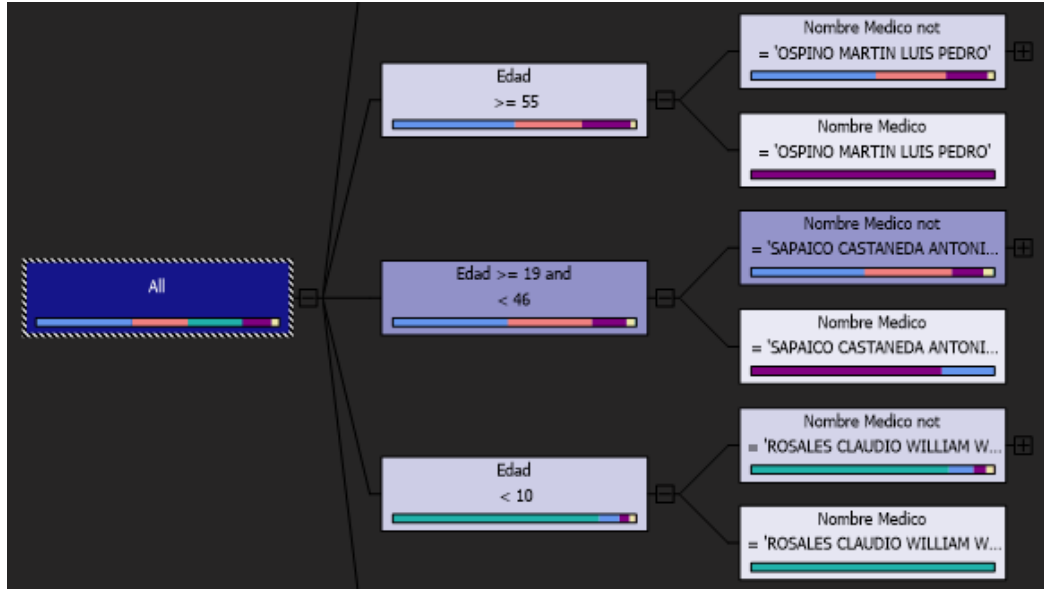
- Estructura de Actividad Consulta Ambulatoria
 - Estructura_Con – Visor de árboles de Microsoft

En este visor podemos encontrar 2 pestañas:

- Árbol de Decisión

“En este tipo de vista, podemos visualizar la estructura del árbol, la cual tiene 10 niveles, nodos o ramificaciones”.

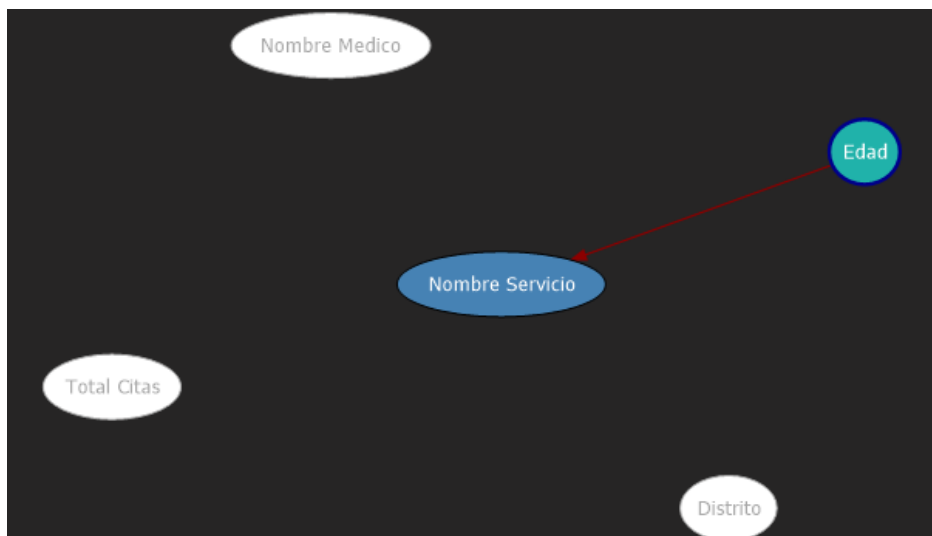
Figura 18. Ramificaciones del árbol de decisión – Motivo Enfermedad Súbita



- 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 la Edad”.

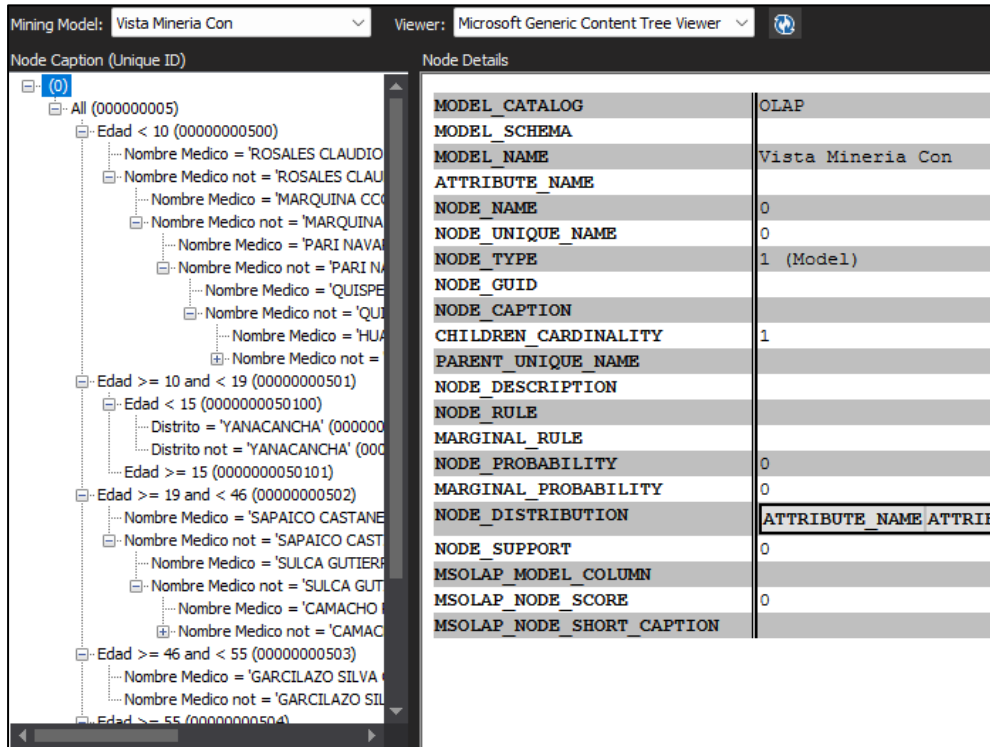
Figura 19. Variable más influyente – Motivo Enfermedad Súbita



- 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 Edad más influyente en el modelo de minería es Edad < 10”.

Figura 20. Detalle del nodo más cercano o influyente – Motivo Enfermedad Súbita



- Estructura de Actividad Procedimientos
 - Estructura_Proc – “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 4 niveles, nodos o ramificaciones”.

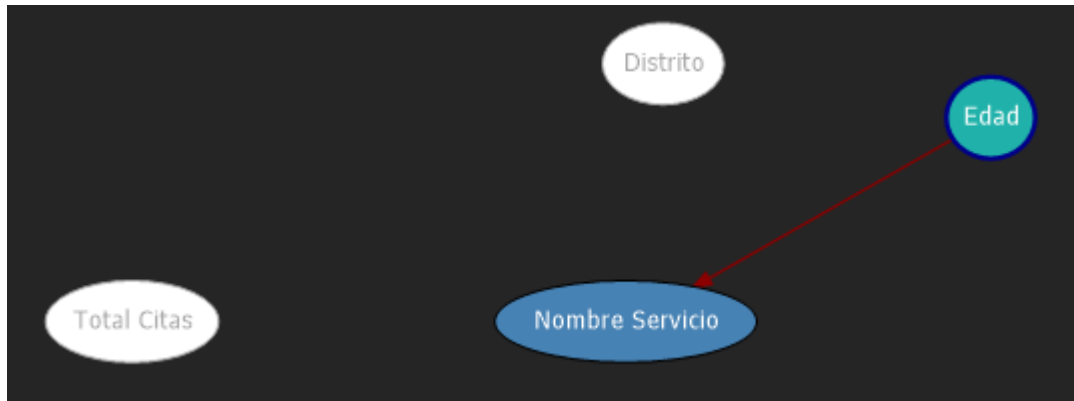
Figura 21. Ramificaciones del árbol de decisión – Motivo Enfermedad Común.



- 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 la Edad.

Figura 22. Variable más influyente – Motivo Enfermedad Común.



- Visor de árbol de contenido genérico de Microsoft

En este tipo de vista podemos ver de manera diferente lo mostrado en las figuras anteriores.

Figura 23. Detalle del nodo más cercano o influyente – Motivo Enfermedad Común.

Mining Model: Vista Minería Proc Viewer: Microsoft Generic Content Tree Viewer

Node Caption (Unique ID)	Node Details
(0)	
[-] All (000000005)	MODEL_CATALOG: OLAP
[-] Edad < 19 (00000000500)	MODEL_SCHEMA: Vista Minería Proc
[-] Edad >= 19 (00000000501)	MODEL_NAME: Nombre Servicio
[-] Total Citas < 950 (0000000050100)	ATTRIBUTE_NAME: Nombre Servicio
[-] Distrito = 'YANACANCHA' (000000005010000)	NODE_NAME: 000000005
[-] Distrito not = 'YANACANCHA' (000000005010001)	NODE_UNIQUE_NAME: 000000005
[-] Total Citas >= 950 (0000000050101)	NODE_TYPE: 2 (Tree)
	NODE_GUID:
	NODE_CAPTION: All
	CHILDREN_CARDINALITY: 2
	PARENT_UNIQUE_NAME: 0
	NODE_DESCRIPTION: All
	NODE_RULE: <compound-predicate op>
	MARGINAL_RULE:
	NODE_PROBABILITY: 1
	MARGINAL_PROBABILITY: 1
	NODE_DISTRIBUTION:
	ATTRIBUTE_NAME AT
	Nombre Servicio Missing
	Nombre Servicio CIRUGÍA
	Nombre Servicio MEDICIN
	Nombre Servicio PEDIÁTR
	Nombre Servicio PSIQUIÁ

“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”.

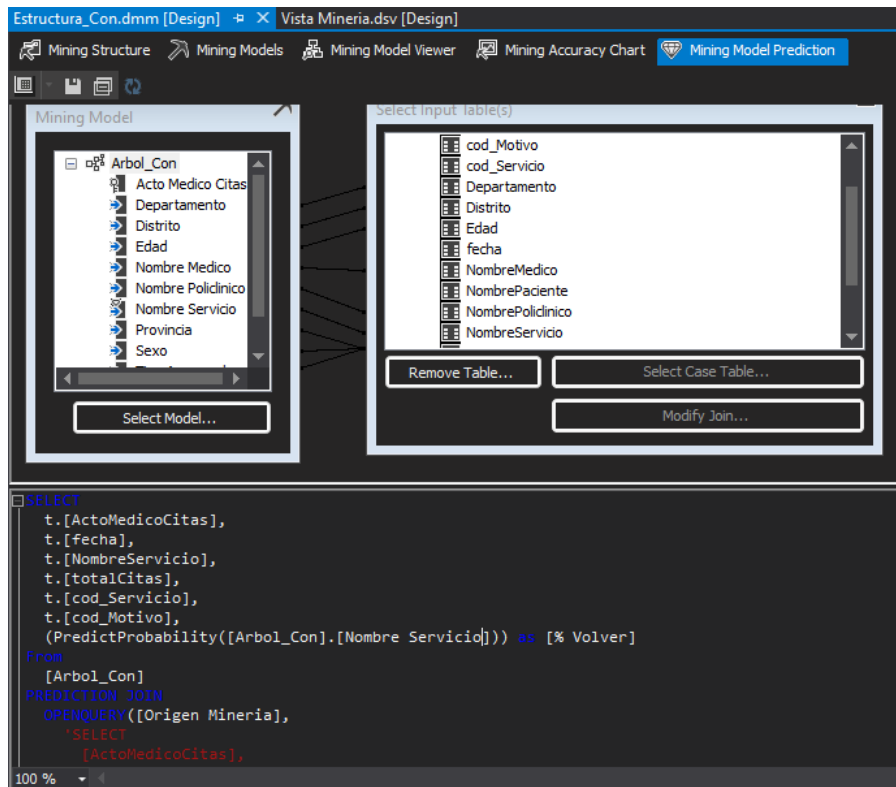
- Atención de Enfermedad súbita

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 24. Elaboración de la consulta DMX – Motivo Enfermedad Súbita



Creada la consulta, se visualiza el resultado.

Figura 25. Relación de pacientes con probabilidad de deserción – Motivo Enfermedad Súbita

ActoMedicoCitas	fecha	NombreServicio	totalCitas	cod_Servicio	cod_Motivo	% Volver
2939	4/02/2023 00:...	GINECO - OBS...	938	230102	07	0.9903563941...
2993	6/02/2023 00:...	GINECO - OBS...	992	230102	07	0.9903563941...
3010	4/02/2023 00:...	GINECO - OBS...	1009	230102	07	0.9903563941...
2991	6/02/2023 00:...	GINECO - OBS...	990	230102	07	0.9903563941...
2903	4/02/2023 00:...	GINECO - OBS...	902	230102	07	0.9903563941...
2418	16/02/2023 00:...	GINECO - OBS...	417	230102	07	0.9903563941...
2904	4/02/2023 00:...	GINECO - OBS...	903	230102	07	0.9903563941...
2551	19/02/2023 00:...	GINECO - OBS...	550	230102	07	0.9903563941...
3012	4/02/2023 00:...	GINECO - OBS...	1011	230102	07	0.9903563941...
3183	4/02/2023 00:...	GINECO - OBS...	1182	230102	07	0.9903563941...
3084	14/02/2023 00:...	GINECO - OBS...	1083	230102	07	0.9903563941...
3232	1/02/2023 00:...	GINECO - OBS...	1231	230102	07	0.9903563941...
2889	6/02/2023 00:...	GINECO - OBS...	888	230102	07	0.9903563941...
3271	3/02/2023 00:...	GINECO - OBS...	1270	230102	07	0.9903563941...
2891	6/02/2023 00:...	GINECO - OBS...	890	230102	07	0.9903563941...
3177	15/02/2023 00:...	GINECO - OBS...	1176	230102	07	0.9903563941...
3085	14/02/2023 00:...	GINECO - OBS...	1084	230102	07	0.9903563941...
3272	3/02/2023 00:...	GINECO - OBS...	1271	230102	07	0.9903563941...
2989	6/02/2023 00:...	GINECO - OBS...	988	230102	07	0.9903563941...
2940	4/02/2023 00:...	GINECO - OBS...	939	230102	07	0.9903563941...
2419	16/02/2023 00:...	GINECO - OBS...	418	230102	07	0.9903563941...
3181	4/02/2023 00:...	GINECO - OBS...	1180	230102	07	0.9903563941...
3314	5/02/2023 00:...	GINECO - OBS...	1313	230102	07	0.9903563941...
3231	1/02/2023 00:...	GINECO - OBS...	1230	230102	07	0.9903563941...

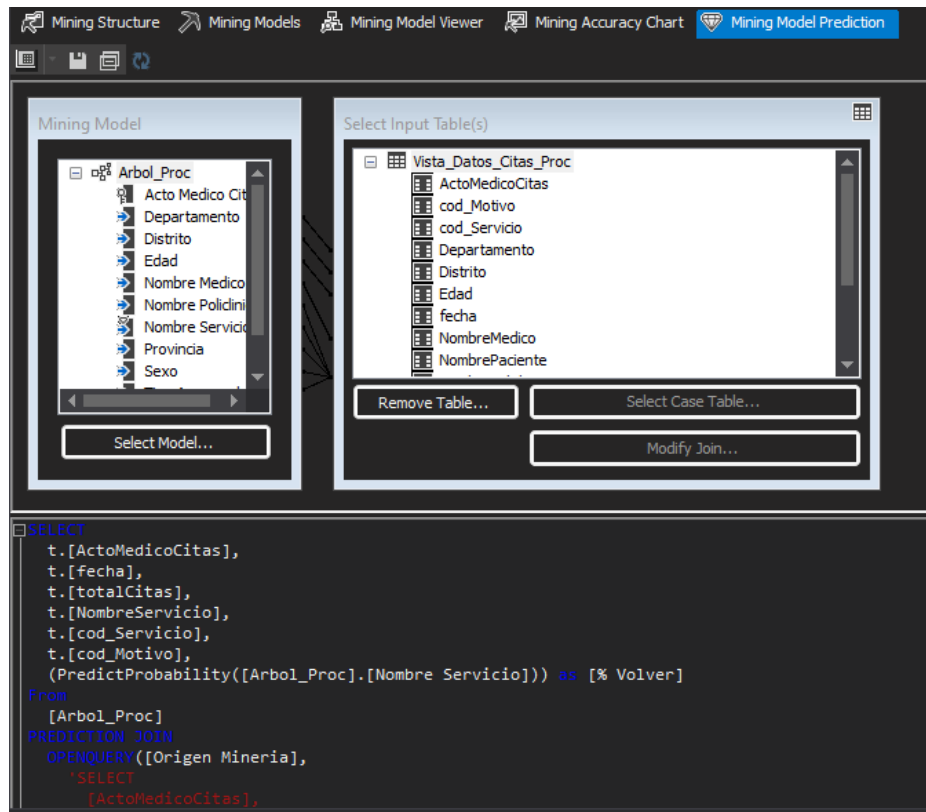
- Atención de Procedimientos

En el cuadro “Seleccionar tabla(s) de entrada”, seleccionamos la vista

Vista_Datos_Citas_Proc.

“En este tipo de predicción se filtra o selecciona a aquellos pacientes que tenga una probabilidad de volver mayor al 70%”.

Figura 26. Elaboración de la consulta DMX – Motivo Enfermedad Común



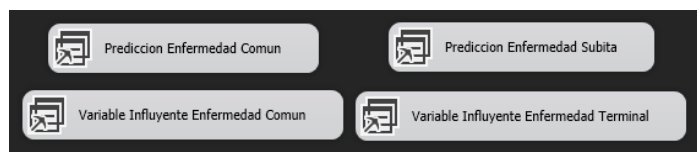
Crea la consulta, se visualiza el resultado.

Figura 27. Relación de pacientes con probabilidad de deserción – Motivo Enfermedad Común

ActoMedicoCitas	fecha	totalCitas	NombreServicio	cod_Servicio	cod_Motivo	% Volver
3508	19/02/2023 00:...	1507	PEDIÁTRICA	230106	11	0.5
2150	8/02/2023 00:...	149	CIRUGÍA GENE...	230101	11	0.5
2151	8/02/2023 00:...	150	CIRUGÍA GENE...	230101	11	0.5
2152	8/02/2023 00:...	151	CIRUGÍA GENE...	230101	11	0.5
3507	19/02/2023 00:...	1506	PEDIÁTRICA	230106	11	0.5

“Una vez definida la consulta DMX a utilizar, en el entorno del ETL, se agrega 4 controles de Tarea de Consulta de minería de datos”.

Figura 28. Flujo de Control - Tareas de Consulta de minería de datos

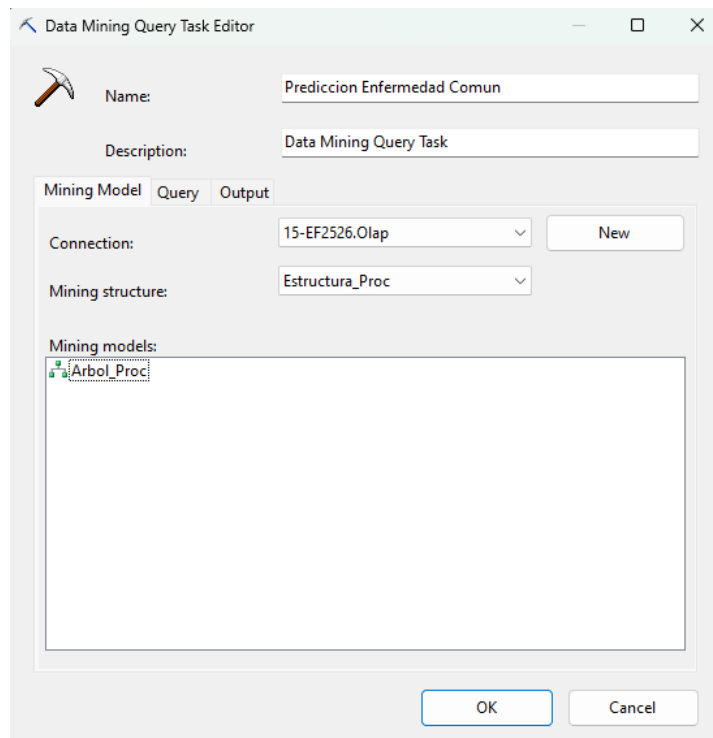


- Predicción Enfermedad común

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

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

Figura 29. Selección del modelo de minería de datos – Motivo Enfermedad Común



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

```
SELECT  
  
    t.[ActoMedicoCitas],  
  
    t.[fecha],  
  
    t.[NombreServicio],  
  
    t.[totalCitas],
```

```

t.[cod_Servicio],
t.[cod_Motivo],
(PredictProbability([Arbol_Proc].[Nombre Servicio])) as
[% Volver]
From
[Arbol_Proc]
PREDICTION JOIN
OPENQUERY([Origen Minería],
'SELECT
[ActoMedicoCitas],
[fecha],
[NombreServicio],
[totalCitas],
[cod_Servicio],
[cod_Motivo],
[NombrePoliclinico],
[TipoAsegurado],
[Sexo],
[Edad],
[Departamento],
[Provincia],
[Distrito],
[NombreMedico]
FROM
[dbo].[Vista_Datos_Citas_Proc]
') AS t
ON

```



```

[Arbol_Proc].[Nombre Policlinico] = t.[NombrePoliclinico]

AND

[Arbol_Proc].[Tipo Asegurado] = t.[TipoAsegurado] AND
[Arbol_Proc].[Sexo] = t.[Sexo] AND
[Arbol_Proc].[Edad] = t.[Edad] AND
[Arbol_Proc].[Departamento] = t.[Departamento] AND
[Arbol_Proc].[Provincia] = t.[Provincia] AND
[Arbol_Proc].[Distrito] = t.[Distrito] AND
[Arbol_Proc].[Total Citas] = t.[totalCitas] AND
[Arbol_Proc].[Nombre Servicio] = t.[NombreServicio]

AND

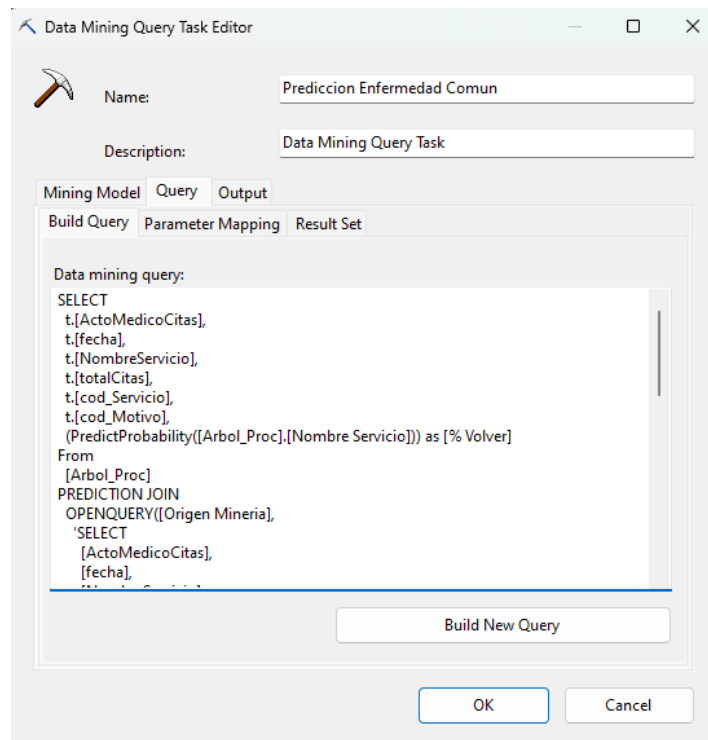
[Arbol_Proc].[Nombre Medico] = t.[NombreMedico]
where (PredictProbability([Arbol_Proc].[Nombre
Servicio])*100)>=50

ORDER BY PredictProbability([Arbol_Proc].[Nombre
Servicio])

desc

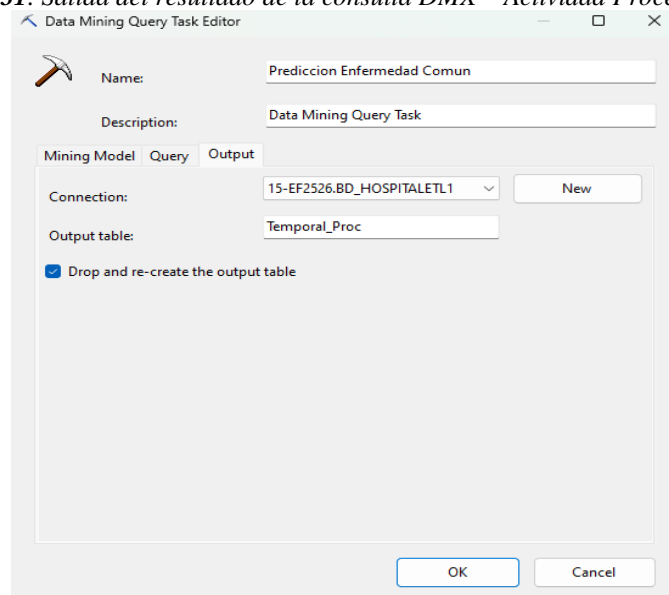
```

Figura 30. Consulta DMX – Motivo Enfermedad Común



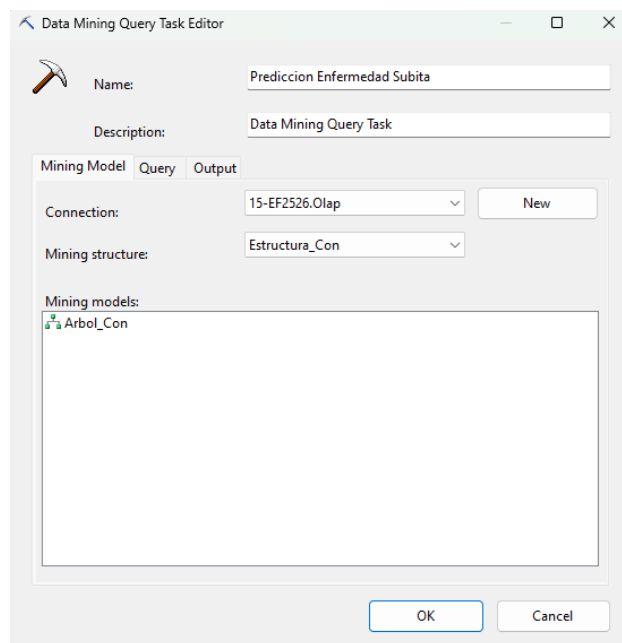
- “En la pestaña Salida, establecemos nuestra conexión de nuestro datamart DatamartDesercion, ingresamos el nombre de la tabla de salida (si no existiese la tabla, en este paso la crea), y por último, marcamos la casilla que dice Quitar y volver a crear la tabla de salida, pues es una tabla temporal;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

Figura 31. Salida del resultado de la consulta DMX – Actividad Procedimientos



- Predicción Motivo Enfermedad Súbita
- “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 32. Selección del modelo de minería de datos – Motivo Enfermedad Súbita



- “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_Motivo],
```

```

(PredictProbability([Arbol_Con].[Nombre Servicio])) as
[% Volver]

From
[Arbol_Con]
PREDICTION JOIN
OPENQUERY([Origen Minería],
'SELECT
[ActoMedicoCitas],
[fecha],
[NombreServicio],
[totalCitas],
[cod_Servicio],
[cod_Motivo],
[NombrePoliclinico],
[TipoAsegurado],
[Sexo],
[Edad],
[Departamento],
[Provincia],
[Distrito],
[NombreMedico]
FROM
[dbo].[Vista_Datos_Citas_Con]
') AS t
ON
[Arbol_Con].[Nombre Policlinico] = t.[NombrePoliclinico]
AND
[Arbol_Con].[Tipo Asegurado] = t.[TipoAsegurado] AND

```

```

[Arbol_Con].[Sexo] = t.[Sexo] AND
[Arbol_Con].[Edad] = t.[Edad] AND
[Arbol_Con].[Departamento] = t.[Departamento] AND
[Arbol_Con].[Provincia] = t.[Provincia] AND
[Arbol_Con].[Distrito] = t.[Distrito] AND
[Arbol_Con].[Total Citas] = t.[totalCitas] AND
[Arbol_Con].[Nombre Servicio] = t.[NombreServicio]

AND

[Arbol_Con].[Nombre Medico] = t.[NombreMedico]

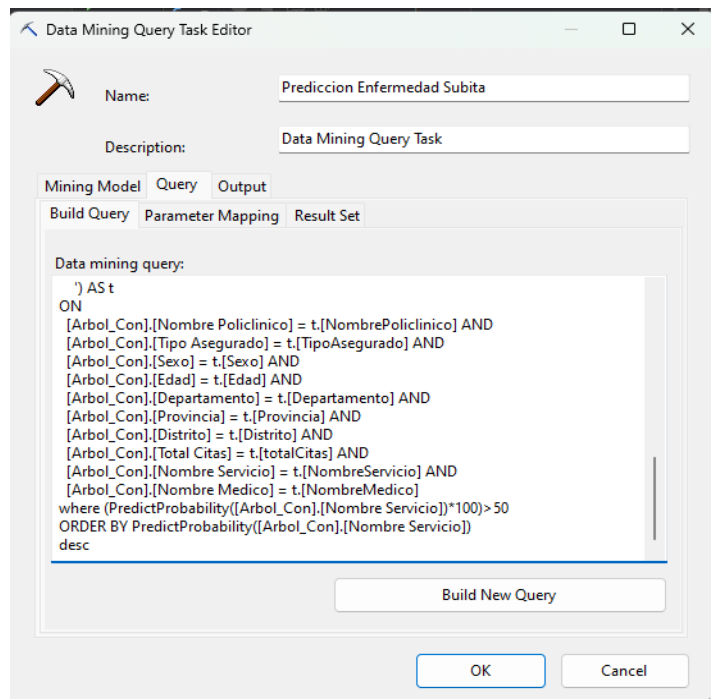
where (PredictProbability([Arbol_Con].[Nombre
Servicio])*100)>50

ORDER BY PredictProbability([Arbol_Con].[Nombre
Servicio])

Desc

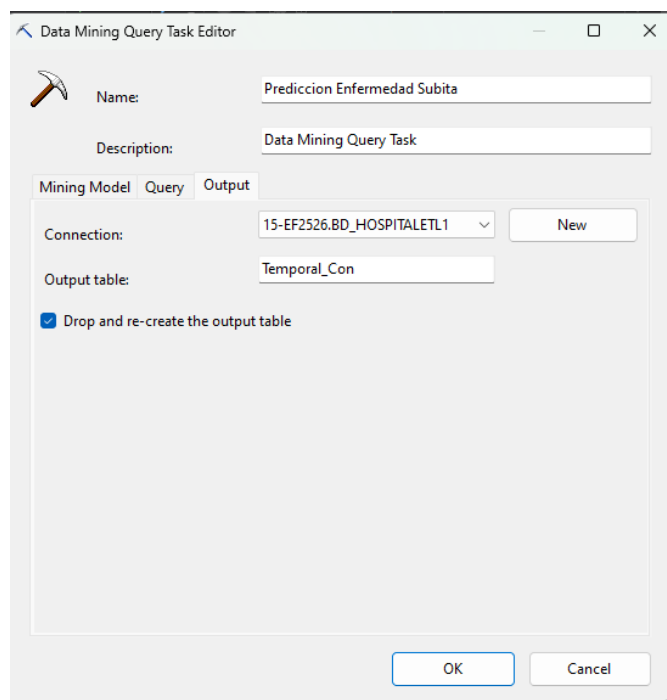
```

Figura 33. Consulta DMX – Motivo Enfermedad Súbita



- En la pestaña Salida, establecemos nuestra conexión de nuestro datamart “DB_HOSPITALETL”, 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 34. Salida del resultado de la consulta DMX – Motivo Enfermedad Súbita



- Variable Influyente Enfermedad Común

Mediante este flujo de control procedemos a poblar una tabla temporal llamada Influyente_Proc; 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_Proc, y nuestro modelo de minería de datos: Arbol_Proc.

- 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_Proc].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 “DB_HOSPITALETL”, 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, pues 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_Proc.
 - Variable Motivo Enfermedad Súbita
 - “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
```

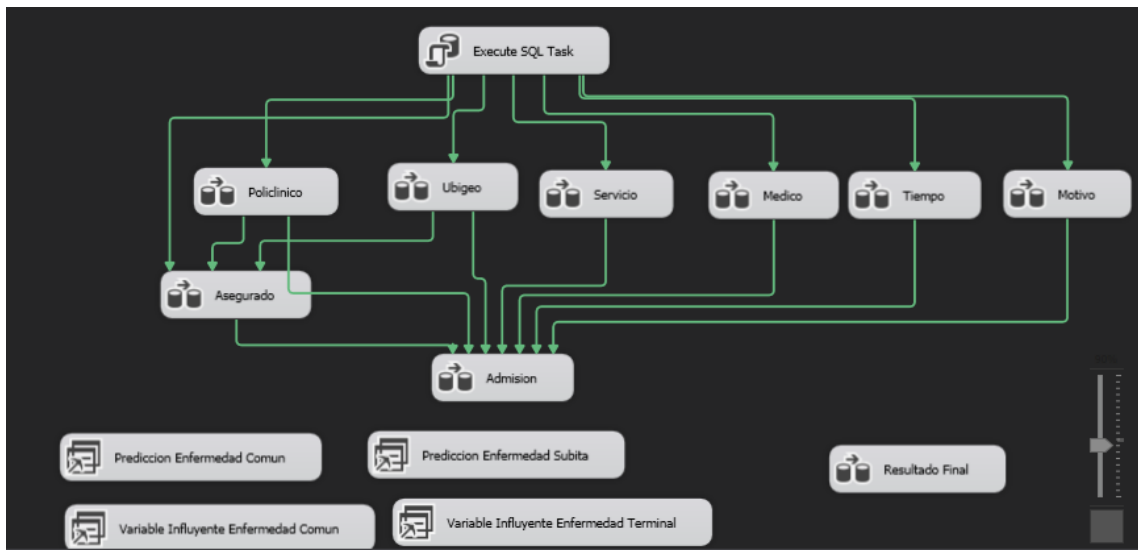
Donde:

- 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, pues 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 35. Flujo de Control Final



Evaluar el modelo

“Los modelos de árboles de decisión, de ambas estructuras (Estructura_Con y Estructura_Proc) se han acercado mucho al modelo ideal, superando al modelo de clustering con el cual se comparó su probabilidad de predicción”.

“Sin embargo, en la Estructura_Con, la diferencia entre el modelo de árbol de decisión con el de clustering es muy pequeña, con un 82.49% frente a un 81.17%, respectivamente; por lo que sería recomendable verificar esta diferencia cuando se utilice la presente aplicación”.

EVALUACIÓN

“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:

Motivo Enfermedad Súbita

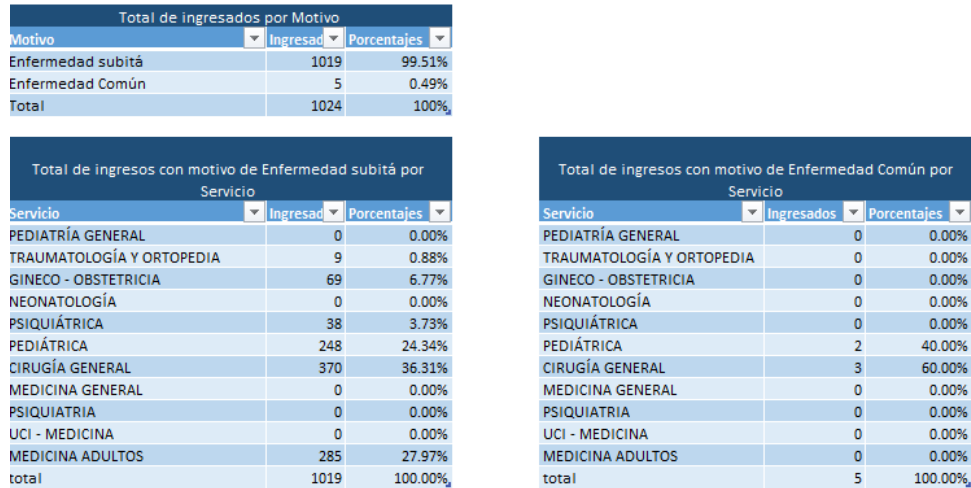
En el periodo del grupo de prueba, se obtuvo que de 1024 ingresados el 99.51% del total ingresan por enfermedad súbita.

Motivo Enfermedad común

En el periodo del grupo de prueba, se obtuvo que de 1024 ingresados el 0.48% del total ingresan por enfermedad común.

Figura 36. Dashboard Porcentual

DAHBOARD PORCENTUAL



4.2. Presentación, análisis e interpretación de resultados

4.2.1. Análisis

1. ¿Cuánto tiempo aproximadamente se demoran en la realización de un registro de citas?

Citas registradas en Excel	Citas registradas aplicando la técnica de clasificación de minería de datos
180 minutos	10 minutos

2. ¿Cuánto tiempo aproximadamente se demoran en la realización de una reprogramación de citas?

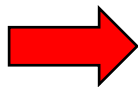
Citas registradas en Excel	Citas registradas aplicando la técnica de clasificación de minería de datos
24 horas	1 hora

3. ¿Cuánto tiempo aproximadamente se demoran en la realización de una eliminación de citas?

Citas registradas en Excel	Citas registradas aplicando la técnica de clasificación de minería de datos
10 minutos	3 minutos

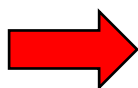
4.3. Prueba de Hipótesis

La aplicación de la técnica de clasificación de minería de datos mejorará el Procesos de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023



Una vez analizado se puede mencionar que al aplicar la técnica de minería de datos se mejoró el Procesos de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023 porque hay una reducción de tiempo al realizar los procesos de citas. En el periodo del grupo de prueba, se obtuvo que de 1024 ingresados el 99.51% del total ingresan por enfermedad súbita.

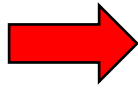
El análisis mediante los árboles de decisión medirá la mejora en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023



Una vez analizado se puede mencionar mediante los árboles de decisión se midió por lo que mejoro en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023 porque hay una reducción de tiempo al realizar la atención de citas. Sin embargo, en la Estructura_Con, la diferencia entre el modelo de árbol de decisión con el de clustering es muy pequeña, con un 82.49% frente a un 81.17%,

respectivamente; por lo que sería recomendable verificar esta diferencia cuando se utilice la presente aplicación.

El análisis mediante el clustering medirá la mejora en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023.



Una vez analizado se puede mencionar mediante el clustering se midió mejorando en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023 porque hay una reducción de tiempo al realizar la atención de citas en los consultorios.

Por lo tanto, aceptamos las hipótesis presentadas.

4.4. Discusión 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 37. Hoja de cálculo – Monitoreo enfermedad súbita

MONITOREO

ENFERMEDAD SÚBITA

Servicio ☰ 🔍

GINECO - OBSTETRICIA MEDICINA ADULTOS PEDIÁTRICA PSIQUIÁTRICA TRAUMATOLOGÍA Y OR...

Nombre Policlínico	Fecha	Servicio	codigo de Servicio	% Volver	Total Citas
DR.DANIEL ALCIDES CARRION	sábado, 4 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	938
DR.DANIEL ALCIDES CARRION	lunes, 6 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	992
DR.DANIEL ALCIDES CARRION	sábado, 4 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	1009
DR.DANIEL ALCIDES CARRION	lunes, 6 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	990
DR.DANIEL ALCIDES CARRION	sábado, 4 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	902
DR.DANIEL ALCIDES CARRION	jueves, 16 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	417
DR.DANIEL ALCIDES CARRION	sábado, 4 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	903
DR.DANIEL ALCIDES CARRION	domingo, 19 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	550
DR.DANIEL ALCIDES CARRION	sábado, 4 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	1011
DR.DANIEL ALCIDES CARRION	sábado, 4 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	1182
DR.DANIEL ALCIDES CARRION	martes, 14 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	1083
DR.DANIEL ALCIDES CARRION	miércoles, 1 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	1231
DR.DANIEL ALCIDES CARRION	lunes, 6 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	888
DR.DANIEL ALCIDES CARRION	viernes, 3 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	1270
DR.DANIEL ALCIDES CARRION	lunes, 6 de Febrero de 2023	GINECO - OBSTETRICIA	230102	0.990356394	890

Figura 38. Hoja de cálculo – Monitoreo enfermedad Común

MONITOREO

ENFERMEDAD COMÚN

NombreServicio ☰ 🔍

CIRUGÍA GENERAL PEDIÁTRICA

NombrePoliclínico	fecha	NombreServicio	cod_Servicio	% Volver	totalCitas
DR.DANIEL ALCIDES CARRION	domingo, 19 de Febrero de 2023	PEDIÁTRICA	230106	0.5	1507
DR.DANIEL ALCIDES CARRION	miércoles, 8 de Febrero de 2023	CIRUGÍA GENERAL	230101	0.5	149
DR.DANIEL ALCIDES CARRION	miércoles, 8 de Febrero de 2023	CIRUGÍA GENERAL	230101	0.5	150
DR.DANIEL ALCIDES CARRION	miércoles, 8 de Febrero de 2023	CIRUGÍA GENERAL	230101	0.5	151
DR.DANIEL ALCIDES CARRION	domingo, 19 de Febrero de 2023	PEDIÁTRICA	230106	0.5	1506

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 1148 citas, considerada entre el año 2023.

Índice de ingreso

El total de citas en la muestra, como ya vimos líneas superiores, es de 1148; el índice de ingreso en nuestra muestra es de 45.82% que hace un total de 526 ingresos a causa de enfermedad súbita y un 0.35% que hace un total de 5 ingresos a causa de enfermedad común.

A continuación, se detalla el comportamiento de este indicador en los dos motivos seleccionados: Enfermedad súbita y común.

Tabla 6. Índice de ingreso

	Enfermedad Súbita	Enfermedad Común
Total, de citas	1148	1148
Ingreso	526	5
Porcentaje de ingreso	45.82%	100%

CONCLUSIONES

Las conclusiones de la aplicación de la técnica de clasificación de minería de datos para mejorar los procesos de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco son las siguientes:

- Se demostró que todos los objetivos utilizan métodos de clasificación de extracción de datos para mejorar el proceso de atención de citas en los consultorios.
- “En la Estructura_Con, la diferencia entre el modelo de árbol de decisión con el de clustering es muy pequeña, con un 82.49% frente a un 81.17%, respectivamente; por lo que sería recomendable verificar esta diferencia cuando se utilice la presente aplicación”.
- Al hacer este tipo de predicción, incluya únicamente a los pacientes cuyas posibilidades de recurrencia sean superiores al 70%.
- “Con los resultados anteriores e identificando las posibles citas a desaprovecharse, se podrá disminuir el diferimiento de cita, que actualmente hay casos que son de hasta 2 meses; esto repercutirá favorablemente al paciente cuando solicite una cita. Asimismo, esta identificación de deserción permite a los pacientes reprogramar o anular su cita, según sea el caso”.

RECOMENDACIONES

“Las recomendaciones, en base a lo logrado con este proyecto y teniendo en cuenta posibles escenarios futuros”, son las siguientes:

- Para garantizar la calidad de los datos, se recomienda garantizar el registro adecuado de los datos en el sistema comercial soportado por la aplicación.
- Esté siempre atento a posibles cambios en el sistema comercial, manténgalo actualizado y asegúrese de que sea compatible.
- Debido a las ventajas mostradas en los puntos anteriores, se recomienda implementar el sistema propuesto.

REFERENCIAS BIBLIOGRÁFICAS

Amézquita Moreno, G. I., Patiño Cañas, N., & Salamanca Morales, C. C. (2019).

Minería de datos para la identificación de factores de riesgo en pacientes con hipertensión arterial [Fundación Universitaria Unipanamericana – Compensar].

<https://repositoriocrai.ucompensar.edu.co/handle/compensar/2293>

Chamba Jimenez, S. (2015). *Minería de Datos para segmentación de clientes en la empresa tecnológica Master PC* [Universidad Nacional de Loja - Ecuador].

<http://dspace.unl.edu.ec/jspui/bitstream/123456789/10462/1/Chamba>

Jiménez%2C Sairy Fernanda.pdf

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., Fernández Collado, C., & Baptista Lucio, P. (2014).

Metodología de la Investigación (S. A. D. C. . McGRAW-HILL /

INTERAMERICANA EDITORES (ed.); Sexta).

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

Investigación.

Martínez Álvarez, C. A. (2012). Aplicación De Técnicas De Minería De Datos Para

Mejorar El Proceso De Control De Gestión En ENTEL [UNIVERSIDAD DE

CHILE]. In *Universidad De Chile Facultad De Ciencias Físicas Y Matemáticas Departamento De Ingeniería Industrial*.

<http://repositorio.uchile.cl/bitstream/handle/2250/112065/cf->

[martinez_ca.pdf?sequence=1](http://repositorio.uchile.cl/bitstream/handle/2250/112065/cf-martinez_ca.pdf?sequence=1)

Medrano Parado, S. Z. (2016). Modelo de minería de datos usando machine learning

con reconocimiento de patrones de síntomas y enfermedades respiratorias en las

historias clínicas para mejorar el diagnóstico de pacientes en la ciudad de Trujillo
2016. In *Universidad César Vallejo*. Universidad César Vallejo.

Polo Capuñay, J. (2018). *Aplicación de la técnica de clasificación de minería de datos para mejorar los procesos de atención de citas en el área de consultorios externos de un Hospital del Departamento de Lambayeque* [UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO].

http://tesis.usat.edu.pe/bitstream/usat/1396/1/TL_PoloCapuñayJose.pdf

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.

Sánchez Sánchez, S. (2017). *Minería de Datos de la salud: Analisis de los factores que influyen en la realización de cirugias estéticas*. Universidad Pontificia Bolivariana.

ANEXOS

Instrumento de Recolección de datos

Ficha de observación

Observado	Observador

DETALLE	OBSERVACIÓN
¿Cuánto tiempo aproximadamente se demoran en la realización de un registro de citas?	_____ Minutos
¿Cuánto tiempo aproximadamente se demoran en la realización de una reprogramación de citas?	_____ Minutos
¿Cuánto tiempo aproximadamente se demoran en la realización de una eliminación de citas?	_____ Minutos

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Id_Cita	Anio	Mes	Dia	Fecha_Atencion	Lote	Num_Pag	Num_Reg	Id_Ups	UPS	Id_Estableci	Id_Paciente	Paciente
2	793569818	2022	12	1	44896	IDD	10	1	303707	NEONATOLOGÍA	979	51148490979	LOYA ESPINOZA RN CUNA 12-A
3	793569818	2022	12	1	44896	IDD	10	1	303707	NEONATOLOGÍA	979	51148490979	LOYA ESPINOZA RN CUNA 12-A
4	793569818	2022	12	1	44896	IDD	10	1	303707	NEONATOLOGÍA	979	51148490979	LOYA ESPINOZA RN CUNA 12-A
5	793569818	2022	12	1	44896	IDD	10	1	303707	NEONATOLOGÍA	979	51148490979	LOYA ESPINOZA RN CUNA 12-A
6	793569818	2022	12	1	44896	IDD	10	1	303707	NEONATOLOGÍA	979	51148490979	LOYA ESPINOZA RN CUNA 12-A
7	793569818	2022	12	1	44896	IDD	10	1	303707	NEONATOLOGÍA	979	51148490979	LOYA ESPINOZA RN CUNA 12-A
8	793569818	2022	12	1	44896	IDD	10	1	303707	NEONATOLOGÍA	979	51148490979	LOYA ESPINOZA RN CUNA 12-A
9	793569818	2022	12	1	44896	IDD	10	1	303707	NEONATOLOGÍA	979	51148490979	LOYA ESPINOZA RN CUNA 12-A
10	793570006	2022	12	1	44896	IDD	10	2	303707	NEONATOLOGÍA	979	51148492979	CASTRO LOPEZ RN CUNA 10-A
11	793570006	2022	12	1	44896	IDD	10	2	303707	NEONATOLOGÍA	979	51148492979	CASTRO LOPEZ RN CUNA 10-A
12	793570006	2022	12	1	44896	IDD	10	2	303707	NEONATOLOGÍA	979	51148492979	CASTRO LOPEZ RN CUNA 10-A
13	793570006	2022	12	1	44896	IDD	10	2	303707	NEONATOLOGÍA	979	51148492979	CASTRO LOPEZ RN CUNA 10-A
14	793570006	2022	12	1	44896	IDD	10	2	303707	NEONATOLOGÍA	979	51148492979	CASTRO LOPEZ RN CUNA 10-A
15	793570006	2022	12	1	44896	IDD	10	2	303707	NEONATOLOGÍA	979	51148492979	CASTRO LOPEZ RN CUNA 10-A
16	793570006	2022	12	1	44896	IDD	10	2	303707	NEONATOLOGÍA	979	51148492979	CASTRO LOPEZ RN CUNA 10-A
17	793570006	2022	12	1	44896	IDD	10	2	303707	NEONATOLOGÍA	979	51148492979	CASTRO LOPEZ RN CUNA 10-A
18	793584049	2022	12	1	44896	EA1	1	1	301611	MONITOREO FETAL	979	10133732979	FERNANDEZ PAREDES JHOVANA ZULM
19	793584049	2022	12	1	44896	EA1	1	1	301611	MONITOREO FETAL	979	10133732979	FERNANDEZ PAREDES JHOVANA ZULM
20	793584448	2022	12	1	44896	EA1	1	2	301611	MONITOREO FETAL	979	40156573979	VASQUEZ CALLUPE ROXANA NOEMI
21	793584448	2022	12	1	44896	EA1	1	2	301611	MONITOREO FETAL	979	40156573979	VASQUEZ CALLUPE ROXANA NOEMI
22	793584642	2022	12	1	44896	EA1	1	3	301611	MONITOREO FETAL	979	31035432979	LOYA ESPINOZA ADRIANA PILAR
23	793584642	2022	12	1	44896	EA1	1	3	301611	MONITOREO FETAL	979	31035432979	LOYA ESPINOZA ADRIANA PILAR

Matriz de Consistencia

Tema: “Aplicación de la técnica de clasificación de minería de datos para mejorar los procesos de atención de citas en los consultorios del Hospital Regional Daniel Alcides Carrión Pasco, 2023”

PROBLEMA GENERAL	OBJETIVO GENERAL	HIPÓTESIS GENERAL	VARIABLE INDEPENDIENTE	DIMENSIÓN	DISEÑO	POBLACIÓN Y MUESTRA
¿De qué manera la Aplicación de la técnica de clasificación de minería de datos mejora los procesos de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023?	Determinar de qué manera la Aplicación de la técnica de clasificación de minería de datos mejora los procesos de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023.	La aplicación de la técnica de clasificación de minería de datos mejorará el Procesos de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023.	Aplicación de la técnica de clasificación de minería de datos.	Árboles de Decisión	Diseño:	POBLACIÓN Citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023.
				Clustering	No experimental	
					Tipo de Investigación	MUESTRA Citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023
					Aplicada	
					Alcance	
					Explicativo	

PROBLEMA ESPECÍFICO	OBJETIVO ESPECÍFICO	HIPÓTESIS ESPECÍFICA	VARIABLE DEPENDIENTE	DIMENSIÓN	MÉTODO DE INVESTIGACIÓN	TÉCNICAS - INSTRUMENTOS
¿Se podrá realizar un análisis mediante los arboles de decisión para medir la mejora en el proceso de atención de citas en los consultorios	Realizar el análisis mediante los árboles de decisión para medir la mejora en el proceso de atención de citas en los consultorios del Hospital	El análisis mediante los árboles de decisión medirá la mejora en el proceso de atención de citas en los consultorios del Hospital	Procesos de atención de citas en los consultorios.	Registro	Método	Técnicas: -Ficha de observación.
				Reprogramación	Análítica, inductiva	
				Eliminación		

<p>del Hospital Daniel Alcides Carrión Pasco, 2023? ¿Se podrá realizar un análisis mediante el clustering para medir la mejora en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023?</p>	<p>Daniel Alcides Carrión Pasco, 2023. Realizar el análisis mediante el clustering para medir la mejora en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023.</p>	<p>Daniel Alcides Carrión Pasco, 2023. El análisis mediante el clustering medirá la mejora en el proceso de atención de citas en los consultorios del Hospital Daniel Alcides Carrión Pasco, 2023.</p>	<p>Enfoque Cuantitativo</p>
--	---	--	--
