



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

1

Optimización del proceso de asignación de turnos en el servicio de consulta externa del Hospital de Especialidades de las Fuerzas Armadas N°1, utilizando técnicas de inteligencia artificial.

Huancas Arrobo, Rudel Iván y Navarrete Vera, Mario Fernando

Departamento de Ciencias de la Computación

Carrera de Ingeniería de Sistemas e Informática

Trabajo de titulación, previo a la obtención del título de Ingeniero en Sistemas e Informática

Ing. Díaz Zúñiga, Magi Paúl

11 de septiembre del 2020

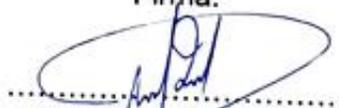
Análisis de Urkund

Document Information

Analyzed document Tesis_RudelHuancas_MarioNavarrete.pdf
(D78505600)
Submitted 9/2/2020 4:50:00 PM
Submitted by DIAZ ZUÑIGA PAUL
Submitter email mpdiaz@espe.edu.ec
Similarity 3%
Analysis address mpdiaz.espe@analysis.arkund.com

- SA** Universidad de las Fuerzas Armadas ESPE / TESIS Pillo Guano.pdf
Document TESIS Pillo Guano.pdf (D57066801) 88 1
Submitted by: mpdiaz@espe.edu.ec
Receiver: mpdiaz.espe@analysis.arkund.com
- SA** Universidad de las Fuerzas Armadas ESPE / Tesis Final Maria Ortuño.pdf
Document Tesis Final Maria Ortuño.pdf (D45669679) 88 2
Submitted by: qninahualpa@espe.edu.ec
Receiver: qninahualpa.espe@analysis.arkund.com
- SA** Universidad de las Fuerzas Armadas ESPE / TesisTatianaPillasaqua(1).docx
Document TesisTatianaPillasaqua(1).docx (D54367214) 88 2
Submitted by: rolandopreyesch@yahoo.com
Receiver: wqaquilar.espe@analysis.arkund.com
- W** URL: [http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S1815-02762004000100012 ...](http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S1815-02762004000100012...)
Fetched: 9/2/2020 4:51:00 PM 88
- W** URL: <http://bibdigital.epn.edu.ec/handle/15000/20592>
Fetched: 9/2/2020 4:51:00 PM 88 1

Firma:



Díaz Zuñiga, Magi Paúl

DIRECTOR



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

CARRERA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

Certificación

Certifico que el trabajo de titulación, **“Optimización del proceso de asignación de turnos en el servicio de consulta externa del Hospital de Especialidades de las Fuerzas Armadas N°1, utilizando técnicas de inteligencia artificial”** fue realizado por los señores **Huancas Arrobo, Rudel Iván y Navarrete Vera, Mario Fernando**; el cual ha sido revisado y analizado en su totalidad por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 11 de Septiembre del 2020.

Firma:

Díaz Zúñiga, Magi Paúl

CC: 1707249072



DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

CARRERA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

Responsabilidad de autoría

Nosotros, **Huancas Arrobo, Rudel Iván y Navarrete Vera, Mario Fernando**, con cédulas de ciudadanía N° 1713715835 y N° 1716317084 declaramos que el contenido, ideas y criterios del trabajo de titulación: **“Optimización del proceso de asignación de turnos en el servicio de consulta externa del Hospital de Especialidades de las Fuerzas Armadas N°1, utilizando técnicas de inteligencia artificial”** es de nuestra autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 11 de Septiembre del 2020.

Firmas:

Huancas Arrobo, Rudel Iván

CC: 1713715835

Navarrete Vera, Mario Fernando

CC: 1716317084



DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

CARRERA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

Autorización de publicación

Nosotros, **Huancas Arrobo, Rudel Iván** y **Navarrete Vera, Mario Fernando**, con cédulas de ciudadanía N° 1713715835 y N° 1716317084, autorizamos a la Universidad de las Fuerzas Armadas ESPE, publicar el trabajo de titulación: **“Optimización del proceso de asignación de turnos en el servicio de consulta externa del Hospital de Especialidades de las Fuerzas Armadas N°1, utilizando técnicas de inteligencia artificial”** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de nuestra responsabilidad.

Sangolquí, 11 de Septiembre del 2020.

Firmas:

Huancas Arrobo, Rudel Iván

CC: 1713715835

Navarrete Vera, Mario Fernando

CC: 1716317084

Dedicatoria

Este trabajo de titulación está dedicado de forma muy especial a mi esposa Anita María quien con su amor, paciencia y apoyo incondicional supo guiar y acompañar firmemente durante cada paso hasta conseguir este gran logro, porque a pesar de la adversidad nunca soltaste mi mano y ahora juntos podemos celebrar este triunfo; a mis hijos Romina Isabella e Iván Mateo quienes con su cariño, alegría y gran sentimiento de admiración supieron motivarme y dar ese gran impulso en los momentos más difíciles de todo este trayecto.

A mis padres Rudel y Micaela quienes desde pequeño inculcaron en mí valores inquebrantables de respeto, honor y constancia para conseguir cualquier objetivo que sea propuesto en los distintos ámbitos de mi vida.

A mis hermanos, suegros y familia en general quienes de una u otra forma a través de su apoyo constante son partícipes de la culminación exitosa de esta etapa.

Finalmente a todos quienes de corazón apoyaron y guiaron para lograr este tan anhelado triunfo.

Rudel Huancas

Dedicatoria

Este trabajo de titulación está dedicado a:

A mis padres Guillermo y María del Carmen quienes con su paciencia, cariño, amor y esfuerzo han sido los principales artífices de este gran paso en mi vida para llegar a cumplir un sueño más, gracias por dedicar su vida entera a que sea la mejor versión de mi propio ser, por acompañarme incluso cuando estuve a punto de dejarlo todo ustedes son el pilar sobre el que toda mi vida y mi ser han sido contruidos.

A mis hermanos Victor y Paúl por su cariño, apoyo y ocurrencias, durante todo este proceso de noches interminables y no dejarme solo en ningún momento, por ser un apoyo durante mi carrera universitaria y a lo largo de la vida misma.

A mi tío Leonardo por ser el modelo de lo que yo aspiro a ser, una gran persona, un gran amigo, el mejor tío y el hombre más inteligente y confiable que he tenido el placer de conocer, por enseñarme que lo único necesario para lograr algo es la dedicación y el querer aprender.

Mario Navarrete

Agradecimiento

Inicialmente mi agradecimiento a Dios por permitirme gozar de salud, fortaleza y sabiduría para que este sea un momento de regocijo con mi familia y amigos. Gracias por tus infinitas bendiciones.

De forma muy especial quiero agradecer y reconocer el apoyo de mi esposa e hijos, quienes realmente son testigos de toda la experiencia que involucró la consecución de este objetivo. Quizá muchas veces tuve que sacrificar tiempo valioso junto a ustedes, sin embargo, aquel esfuerzo ahora se ve reflejado en felicidad. Sepan que cada acción la hago pensando en el bienestar de ustedes y en la mejor enseñanza que pueda sembrar y dejar en mis pequeños hijos. Simplemente gracias por ser mi motor y fuerza cada día de mi vida, los amo mucho.

A mis suegros, hermanos, familiares y amigos quienes me apoyaron incondicionalmente en todo momento.

A mi gran amigo Mario con quien ahora festejamos este triunfo, realmente has mostrado una amistad incondicional desde el primer día de esta etapa universitaria. Que este sea solo el inicio de una vida llena de éxitos para ambos.

Un efusivo agradecimiento al Ing. Paúl Díaz y al Ing. Cristian Ruíz quienes nos guiaron con su dirección, conocimiento y amistad; y que ahora son parte de la consecución de este trabajo de investigación. Gracias estimados amigos.

Finalmente al Ejército Ecuatoriano, institución a la cual me debo, por la oportunidad de poder superarme y cursar estudios en la Universidad de Fuerzas Armadas ESPE. Todo el conocimiento adquirido permitirá engrandecer y trabajar en beneficio de la institución armada.

Rudel Huancas

Agradecimiento

Necesitaría muchas más páginas para brindar mi agradecimiento a quienes se han involucrado a lo largo de este proceso, sin embargo, mis padres y mis hermanos se merecen un reconocimiento especial por estar siempre presentes, por llenar mi vida de bendiciones, por ser las personas más importantes en mi vida.

De igual manera quiero expresar mi gratitud a mi amigo Rudel Iván, quien ha estado en los mejores y los peores momentos brindándome su apoyo y amistad a lo largo de todos estos años de carrera universitaria.

Quiero expresar mi más grande y sincero agradecimiento a Joaquín Acurio, Fernando Medina, Néstor Uvillus y Alejandro Arteaga, por brindarme su apoyo y amistad desde el primer momento, por sus enseñanzas y sus palabras de aliento, por brindarme un gran ejemplo de trabajo y honradez.

Finalmente quiero expresar mi más grande y sincero agradecimiento al Ing. Paúl Díaz y al Ing. Christian Ruiz, principales colaboradores durante todo este proceso, quienes, con su dirección, conocimiento, enseñanza y colaboración permitieron el desarrollo de este trabajo.

Mario Navarrete

Contenido

Análisis de Urkund	2
Certificación	3
Responsabilidad de autoría	4
Autorización de publicación	5
Dedicatoria	6
Agradecimiento.....	8
Índice de tablas.....	15
Índice de figuras	16
Resumen	21
Abstract	22
Capítulo I	17
Introducción.....	17
Antecedentes.....	17
Planteamiento del Problema.....	19
Justificación	21
Objetivos	21
Objetivo General	21
Objetivos Específicos	22
Alcance.....	22
Estado del Arte	24
Planteamiento de la Revisión Sistemática.....	24
Definición del grupo de control y extracción de términos	24
Construcción de la Cadena de Búsqueda	25
Selección de los Estudios Primarios	26
Elaboración del Estado de Arte	27
Capítulo II	30

	11
Marco Teórico	30
Inteligencia artificial.....	30
Algoritmos no supervisados	31
Algoritmos supervisados.....	31
Random forest.....	32
Árboles de decisión	34
Técnicas y campos de la inteligencia artificial.....	35
Minería de datos.....	35
Machine learning	35
Análítica de Datos	36
Herramientas para el análisis de datos	37
Herramienta KNIME.....	40
Componentes de KNIME.....	40
Metanodo	41
File Reader.....	42
Column Filter	43
Column Appender	44
Column Resorter	44
Row Splitter	45
Rule Engine.....	46
Concatenate	47
Math Formula	47
Table Creator	48
Group By	48
Chunk Loop Start	49
Table Row To Variable.....	50
Rule-based Row Filter.....	50

	12
Extract Table Dimension	51
Row Filter	52
Sorter.....	52
Table Column to Variable	53
Loop End.....	53
Row ID.....	54
Nominal Value Row Filter	54
Extract Date&Time Fields.....	55
Variable to Table Column	55
Date&Time Shift	55
Create Date&Time Range	56
String Manipulation.....	56
Cross Joiner	57
String to Date&Time	57
Number to String	58
Auto Binner.....	58
One to Many.....	59
PCA Compute	59
PCA Apply	60
Normalizer	60
Partitioning	61
Random Forest Learner	61
Random Forest Predictor	62
Missing Value	62
Scorer.....	63
Model Writer	63
Model Reader.....	64

	13
SQLite Connector.....	64
DB Writer.....	65
Metodología de Desarrollo.....	65
Fase 1. Comprensión del negocio.....	66
Fase 2. Estudio y comprensión de los datos.....	66
Fase 3. Análisis de los datos y selección de las características.....	66
Fase 4. Modelado.....	66
Fase 5. Evaluación.....	67
Fase 6. Despliegue o Implementación.....	67
Capítulo III.....	69
Desarrollo del modelo en base a la metodología CRISP-DM.....	69
Arquitectura del modelo.....	69
Adquisición de datos.....	70
Comprensión del negocio y de los Datos.....	71
Análisis de los datos.....	75
Toma de la muestra.....	76
Clasificación de enfermedades catastróficas.....	78
Selección de variables.....	82
Limpieza de la data.....	83
Análisis clúster.....	84
Asignación de pesos a variables de análisis.....	85
Modelado.....	87
Análisis de componentes principales (PCA).....	87
Clasificación de pacientes.....	89
Entrenamiento de modelos con inteligencia artificial.....	90
Modelo utilizando Random Forest.....	91
Modelo utilizando Árboles de Decisión.....	92

	14
Modelo utilizando Perceptrón	93
Análisis y elección del modelo óptimo	93
Predicción para asignación de turnos	95
Capítulo IV	96
Simulación y análisis de resultados.....	96
Simulación	97
Análisis de resultados	102
Capítulo V.....	104
Conclusiones y recomendaciones	104
Conclusiones.....	105
Recomendaciones	106
Bibliografía.....	108

Índice de tablas

Tabla 1. <i>Artículos que conforman el grupo de control</i>	25
Tabla 2. <i>Estudios primarios</i>	27
Tabla 3. <i>Comparación entre KNIME y Rapidminer</i>	38
Tabla 4. <i>Comparación entre Tableau y Microsoft Power BI</i>	39

Índice de figuras

Figura 1. <i>Aprendizaje no supervisado</i>	31
Figura 2. <i>Aprendizaje supervisado</i>	32
Figura 3. <i>Algoritmo random forest</i>	34
Figura 4. <i>Enfermedades analizadas</i>	37
Figura 5. <i>Componente</i>	41
Figura 6. <i>Metanodo</i>	41
Figura 7. <i>Componentes internos de un metanodo</i>	42
Figura 8. <i>File reader</i>	43
Figura 9. <i>Column filter</i>	43
Figura 10. <i>Column appender</i>	44
Figura 11. <i>Column resorter</i>	45
Figura 12. <i>Row splitter</i>	46
Figura 13. <i>Reglas definidas en el componente rule engine</i>	46
Figura 14. <i>Rule engine</i>	47
Figura 15. <i>Concatenate</i>	47
Figura 16. <i>Math formula</i>	48
Figura 17. <i>Table creator</i>	48
Figura 18. <i>Group by</i>	49
Figura 19. <i>Chunk loop start</i>	50
Figura 20. <i>Table row to variable</i>	50
Figura 21. <i>Rule-based row filter</i>	51
Figura 22. <i>Extract table dimension</i>	51
Figura 23. <i>Row filter</i>	52

	17
Figura 24. <i>Sorter</i>	52
Figura 25. <i>Table column to variable</i>	53
Figura 26. <i>Loop end</i>	53
Figura 27. <i>Row id</i>	54
Figura 28. <i>Nominal value row filter</i>	54
Figura 29. <i>Extract date&time fields</i>	55
Figura 30. <i>Variable to table column</i>	55
Figura 31. <i>Date&time shift</i>	56
Figura 32. <i>Create date&time range</i>	56
Figura 33. <i>String manipulation</i>	57
Figura 34. <i>Cross joiner</i>	57
Figura 35. <i>String to date&time</i>	58
Figura 36. <i>Number to string</i>	58
Figura 37. <i>Auto binner</i>	59
Figura 38. <i>One to many</i>	59
Figura 39. <i>PCA compute</i>	60
Figura 40. <i>PCA apply</i>	60
Figura 41. <i>Normalizer</i>	61
Figura 42. <i>Partitioning</i>	61
Figura 43. <i>Random forest learner</i>	62
Figura 44. <i>Random forest predictor</i>	62
Figura 45. <i>Missing value</i>	63
Figura 46. <i>Missing value</i>	63
Figura 47. <i>Model writer</i>	64
Figura 48. <i>Model reader</i>	64

	18
Figura 49. <i>SQLite connector</i>	65
Figura 50. <i>DB writer</i>	65
Figura 51. <i>Fases de la metodología CRISP-DM</i>	68
Figura 52. <i>Diseño de la arquitectura</i>	70
Figura 53. <i>Obtención de datos para análisis</i>	70
Figura 54. <i>Análisis de los días de espera en base a consultorio</i>	72
Figura 55. <i>Análisis de los días de espera en base a especialidad</i>	72
Figura 56. <i>Análisis del número de pacientes de acuerdo al tipo de afiliación</i>	73
Figura 57. <i>Análisis de los días de espera en base al tipo de afiliación</i>	74
Figura 58. <i>Análisis de los días de espera en base al tipo de turno/atención</i>	74
Figura 59. <i>Análisis del tiempo de atención en turnos de primera vez</i>	75
Figura 60. <i>Tamaño de la base de datos inicial</i>	76
Figura 61. <i>Estratificación de la muestra</i>	77
Figura 62. <i>Resultado de la muestra</i>	78
Figura 63. <i>Catálogo de enfermedades a comparar</i>	79
Figura 64. <i>Valor de la distancia de Levenshtein de las enfermedades</i>	80
Figura 65. <i>Motor de reglas crea columna de enfermedades catastróficas</i>	80
Figura 66. <i>Clasificación de enfermedades catastróficas y no catastróficas</i>	81
Figura 67. <i>Proceso de clasificación de enfermedades catastróficas</i>	82
Figura 68. <i>Filtrado de variables</i>	82
Figura 69. <i>Proceso de análisis para limpieza en tiempos de consulta</i>	83
Figura 70. <i>Proceso de estandarización de fechas</i>	84
Figura 71. <i>Proceso de agrupamiento de datos</i>	84
Figura 72. <i>Muestra de clúster clasificado por colores</i>	85
Figura 73. <i>Proceso de asignación de pesos a variables</i>	86

	19
Figura 74. <i>Conversión de filas en columnas</i>	87
Figura 75. <i>Flujo de análisis de componentes principales</i>	88
Figura 76. <i>Selección de variables para análisis de componentes principales</i>	88
Figura 77. <i>Fraccionamiento y representación de la información</i>	89
Figura 78. <i>Dimensiones generadas por el análisis de componentes principales</i>	89
Figura 79. <i>Priorización de pacientes en base a su clasificación</i>	90
Figura 80. <i>Técnicas de inteligencia artificial utilizadas para entrenamiento</i>	91
Figura 81. <i>Flujo de random forest utilizado para entrenamiento</i>	92
Figura 82. <i>Flujo de árboles de decisión utilizado para entrenamiento</i>	93
Figura 83. <i>Flujo de perceptrón utilizado para entrenamiento</i>	93
Figura 84. <i>Matriz de confusión del modelo de random forest</i>	94
Figura 85. <i>Matriz de confusión del modelo de árboles de decisión</i>	94
Figura 86. <i>Matriz de confusión del modelo de perceptrón</i>	94
Figura 87. <i>Datos de predicción para asignación de turnos</i>	95
Figura 88. <i>Flujo de trabajo para simular asignación de turnos</i>	96
Figura 89. <i>Selección de especialidad desde la base de datos sqlite</i>	97
Figura 90. <i>Selección de especialidad para simulación</i>	97
Figura 91. <i>Especialidad seleccionada</i>	98
Figura 92. <i>Horarios de médicos especialistas</i>	98
Figura 93. <i>Horarios generados por el modelo, sin turnos asignados</i>	99
Figura 94. <i>Turnos de 1ra vez y subsecuentes generados aleatoriamente</i>	100
Figura 95. <i>Flujo de trabajo para generación aleatoria del tipo de turno</i>	100
Figura 96. <i>Días de espera para exámenes en turnos subsecuentes</i>	101
Figura 97. <i>Chunk óptimo y estrategia de búsqueda bayesiana</i>	101
Figura 98. <i>Tiempo necesario para asignación de turnos antes de optimizar</i>	102

Figura 99. *Tiempo necesario para asignación de turnos después de optimizar.....* 103

Figura 100. *Pasos para crear el modelo en base a CRISP-DM.....* 104

Resumen

El Hospital de Especialidades de Fuerzas Armadas N°1 requiere mejorar la forma en la que se agendan los turnos en el área de consulta externa, para de esta manera ofrecer una atención rápida y eficiente a sus pacientes quienes acuden a esta casa de salud por un servicio hospitalario de calidad. Como parte de una solución se propone un modelo que permita optimizar el proceso de agendamiento de turnos en el área de consulta externa mediante la aplicación de técnicas de inteligencia artificial, a través de las cuales se realizará la clasificación y priorización de pacientes de acuerdo al análisis de factores que inciden en el proceso de asignación actual.

Dentro de las herramientas tecnológicas utilizadas está TABLEAU para el análisis y representación estadístico de los datos, y KNIME para realizar el modelo que incluye el análisis inicial de la data, tratamiento, limpieza y entrenamiento a través de distintas técnicas de inteligencia artificial que permitan definir la óptima para el actual proceso. Además se realiza la simulación utilizando el mismo software para demostrar la validez y los resultados de la investigación que se llevó a cabo. La metodología utilizada es la CRISP-DM, que a través de sus etapas permiten al investigador comprender el giro del negocio y la naturaleza de sus datos, tratar esta información para descartar datos erróneos o que no son representativos para su posterior análisis, simulación y representación.

Palabras Clave:

- **INTELIGENCIA ARTIFICIAL**
- **AGENDAMIENTO DE TURNOS**
- **MINERÍA DE DATOS**

Abstract

The Armed Forces Specialty Hospital No. 1 needs to improve the way in which shifts are scheduled in the outpatient department, in order to offer fast and efficient care to its patients who come to this health center for quality hospital service. As part of a solution to the problem found, a model is proposed that allows for the optimization of the shift scheduling process in the outpatient department through the application of artificial intelligence techniques, through which patients will be classified and prioritized according to the analysis of factors that affect the current assignment process.

Among the technological tools used is TABLEAU for the analysis and statistical representation of the data, and KNIME to carry out the model that includes the initial analysis of the data, treatment, cleaning and training through different artificial intelligence techniques that allow defining the optimal for the current process. In addition, the simulation is carried out using the same software to demonstrate the validity and results of the research. The methodology that allows executing all the steps sequentially to achieve the proposed objectives is the CRISP-DM, which through its stages allows the researcher to understand the business and the nature of their data, to treat this information to discard data that are erroneous or not representative for further analysis.

Keywords:

- **ARTIFICIAL INTELLIGENCE**
- **SHIFT SCHEDULING**
- **DATA MINING**

Capítulo I

Introducción

En este capítulo se realiza la descripción de antecedentes, problema y alcance del proyecto; temática que permitirá obtener una comprensión general de la presente investigación. Una vez que se tenga claro el concepto se marcan objetivos que serán cumplidos durante la elaboración del proyecto: “Optimización del proceso de asignación de turnos en el servicio de consulta externa del hospital de especialidades de las fuerzas armadas #1, utilizando técnicas de inteligencia artificial”.

Antecedentes

El Hospital de Especialidades de las Fuerzas Armadas N°1, pertenece al Sistema de Salud de las Fuerzas Armadas y está dentro del tercer nivel de atención, de acuerdo a lo que establece el Ministerio de Salud Pública. Brinda atención de salud integral tanto a la comunidad militar como a la civil dentro de la Red Pública Integral de Salud.

En el Hospital de Especialidades de Fuerzas Armadas N° 1 laboran un total de 1666 personas, de los cuales 1378 corresponde a personal civil y 238 a personal militar.

Para la atención ambulatoria el hospital cuenta con 131 ambientes entre consultorios y áreas para procedimientos menores, brindando atención en 37 especialidades entre clínicas, quirúrgicas y materno infantiles.

Además cuenta con cinco servicios complementarios para el diagnóstico y tratamiento que son los siguientes:

- Emergencia

- Hospitalización
- Consulta externa
- Auxiliares de diagnóstico
- Odontología.

En la actualidad se dispone de un Hospital Information System (HIS), que a través de sus 18 módulos se encarga del manejo y administración de la información de todos los servicios hospitalarios que brinda la casa de salud, los cuales se detalla a continuación:

- Admisión y estadística
- Gestión de consulta externa
- Cuidados intensivos
- Agendamiento y horarios médicos
- Odonto Web
- Emergencia
- Bodega
- Planillaje
- Financiero
- B.I
- Permanencia y atenciones
- Labores de enfermería

- Cuidado y manejo diario
- Seguridades
- Gestión de usuarios
- Banco de sangre
- RIS / PACS
- Integración con laboratorio clínico

En los últimos años la demanda de atención en el hospital se ha incrementado, sin embargo la capacidad de atención, así como el proceso de agendamiento de citas médicas se ha mantenido igual, motivo por el cual un paciente debe esperar entre dos a tres meses para obtener un turno a través del área de consulta externa, esto sin considerar que para llegar al Hospital de Especialidades debió pasar tiempo esperando la atención en un policlínico hasta que el mismo realice la transferencia del paciente a la casa de salud de tercer nivel.

Planteamiento del Problema

Actualmente el Hospital de Especialidades de las Fuerzas Armadas N°1, atiende aproximadamente 1200 personas diariamente distribuidas en 37 especialidades y 5 servicios. La asignación de los turnos se realiza en función a tres tipos: turnos de primera vez, subsecuentes y extras.

Los turnos de primera vez son otorgados en las primeras horas del día hasta aproximadamente las 10:00 a.m, mientras que los turnos subsecuentes son a partir de esta hora hasta finalizar las citas planificadas para el día, y los turnos extras son otorgados por el médico de acuerdo a su criterio al final del día. Para otorgar el turno en

caso de ser primera vez se lo realiza a través de caja, cuando es subsecuente o extra se realiza mediante el médico tratante.

El tiempo de atención por cada turno varía de acuerdo a la especialidad, por ejemplo psicología con duración de 1 hora y dermatología 15 minutos, esto se considera para agendar en forma diaria la cantidad de turnos a cada médico tratante.

El problema se identifica en que para otorgar los turnos dentro de sus tres tipos no se prioriza o clasifica sus pacientes (Ej. Embarazadas, personas con discapacidad, enfermedades catastróficas, edad del paciente), sino que simplemente se asigna de acuerdo a la disponibilidad de la agenda. Con la generación de turnos de acuerdo al método actual se presentan varios inconvenientes en el hospital, por ejemplo la saturación de turnos en el área de emergencia debido a que la atención es más rápida que cuando se realiza a través de consulta externa, esto incluso causa pérdidas económicas a la casa de salud ya que cuando un paciente ingresa por emergencias no se cobra inicialmente ningún valor, es decir en el proceso del triage o toma de signos.

Otro problema se presenta en la falta de atención oportuna a pacientes considerados como críticos, quienes por su condición patológica pueden tener afectados uno o más sistemas, lo cual influye directamente en su salud poniendo en riesgo potencial su vida (Grupo de Estudios de Etica Clínica, 2003).

El sistema actual de agendamiento de turnos no permite realizar una planificación correcta entre los horarios médicos y la calendarización de consulta externa. Esto causa que se sature la agenda de ciertos médicos mientras que otros tienen una agenda holgada. Al no existir una distribución equilibrada de pacientes para los médicos disponibles del hospital también genera que la disponibilidad de turnos no

sea para fechas próximas y que los usuarios deban esperar más tiempo para obtener una cita médica. De igual manera no se dispone de un método de re agendamiento óptimo, el cual es útil en caso de no existir atención médica por motivos de fuerza mayor, por ejemplo, en el caso de estado de excepción la casa de salud tuvo que dejar de atender pacientes especialmente en el área de consulta externa.

Justificación

Debido a la investigación realizada se puede evidenciar que la atención en el área de consulta externa del Hospital de Especialidades de Fuerzas Armadas N°1 no es la adecuada ya que los usuarios deben esperar periodos de tiempo mayores a un mes para recibir la atención médica en sus distintas especialidades. Dado el vertiginoso incremento en la demanda de pacientes y la importancia de esta casa de salud dentro de las Fuerzas Armadas conviene desarrollar soluciones que mejoren la atención médica a sus usuarios.

La utilización de técnicas de inteligencia artificial plantea un reto para optimizar un sistema de agendamiento que no ha tenido cambios en los últimos años. Si bien se llega a atender a todos los pacientes que acuden al hospital, la mayoría de ellos no se muestran conformes por los problemas referentes al tiempo de espera para la obtención de turnos y a las limitadas opciones en médicos que se ofrece en el área de consulta externa.

Objetivos

Objetivo General

Optimizar el proceso de asignación de turnos del servicio de consulta externa del Hospital de Especialidades de las Fuerzas Armadas N°1 mediante el uso de técnicas de inteligencia artificial.

Objetivos Específicos

- Identificar la situación actual referente a la demanda diaria y disponibilidad de consultorios y médicos en el área de consulta externa del Hospital Militar.
- Analizar las técnicas adecuadas para la priorización y clasificación de pacientes que acuden por citas médicas en el área de consulta externa.
- Diseñar la arquitectura y desarrollar un modelo que permita la optimización del proceso de asignación de turnos en consulta externa, mediante la utilización de inteligencia artificial.
- Realizar la evaluación, validación y socialización de los resultados obtenidos.

Alcance

Para delimitar el alcance del proyecto de investigación a desarrollarse se ha dividido en las siguientes fases:

1. Estudio y obtención de datos sobre la situación actual en el área de consulta externa del Hospital de Especialidades de Fuerzas Armadas # 1.
 - a. Identificar la demanda diaria de pacientes por cada una de las especialidades.
 - b. Identificar la cantidad de consultorios disponibles por especialidad para efectuar la atención médica.
 - c. Determinar la capacidad diaria de atención de acuerdo a los tiempos establecidos por consulta en cada especialidad.
2. Identificación de la técnica adecuada de inteligencia artificial para la clasificación y priorización en el agendamiento de pacientes en el área de consulta externa.

- a. Investigar sobre las técnicas de algoritmos supervisados para aplicarla en el modelo de generación de turnos en consulta externa.
 - b. Seleccionar las herramientas software adecuadas que a través de la utilización de técnicas de inteligencia artificial adecuadas, permita la clasificación y priorización de pacientes que acuden por consulta externa.
3. Estructuración del diseño y arquitectura para la construcción del modelo, mediante inteligencia artificial:
- a. Análisis de los estudios relacionados sobre la construcción de modelos para la priorización y clasificación de pacientes.
 - b. Identificación de la metodología a utilizar.
 - c. Definición de las características que debe poseer el modelo de inteligencia artificial.
 - d. Construcción del modelo de inteligencia artificial para la clasificación y priorización de pacientes en el agendamiento de turnos del área de consulta externa.
4. Simulación y evaluación del modelo de optimización para la clasificación y priorización de pacientes en el agendamiento de turnos del área de consulta externa.
- a. Identificar el método adecuado para la evaluación.
 - b. Simular el modelo de clasificación y priorización de pacientes.
 - c. Exponer los resultados obtenidos del modelo desarrollado.

Estado del Arte

Para la revisión de la literatura sobre la presente investigación, se realizó el mapeo sistemático de literatura haciendo referencia a las guías propuestas por Kitchenham, mismas que se mencionan a continuación: (1) Planteamiento del estudio sistemático de literatura, (2) Definición del grupo de control y extracción de términos, (3) Construcción de la cadena de búsqueda, (4) Selección de los estudios primarios y (5) Elaboración del estado del arte (Kitchenham & Charters, 2007) . Las fases mencionadas se describen posteriormente:

Planteamiento de la Revisión Sistemática

Como fase inicial para el desarrollo del mapeo sistemático de literatura se realizó la descripción del problema central del proyecto de investigación, la definición de las preguntas de investigación y la especificación de los criterios de inclusión y exclusión para la extracción de los estudios primarios.

Definición del grupo de control y extracción de términos

La segunda fase permite la identificación de los estudios que conforman el grupo de control de manera que deben estar estrechamente relacionados con los criterios de inclusión y exclusión, además de pertenecer a revistas o bases digitales de impacto.

Luego de realizar el análisis de varios estudios científicos se seleccionó artículos que conforman el grupo de control, los que se describen en la tabla 1:

Tabla 1.

Artículos que conforman el grupo de control

Código	Título	Palabras Clave
EC1	A Web-based Application for Innovative Hospital Appointment Scheduling using Neural Network	Neural networks, intelligence, hospital appointment
EC2	Heuristics To Solve Appointment Scheduling In Chemotherapy	Predictive models, Hospitals, Logistics, Data models, Analytical models, Real-time systems
EC3	Integrated Innovative Solutions to Improve Healthcare Scheduling	Fuzzy Petri Nets, EN13606, openEHR, Healthcare, Chronic Disease, Scheduling
EC4	Improving Outpatient Waiting Time Using Simulation Approach	Outpatient scheduling; Average Waiting Time; Simulation Modeling and Analysis; simulation-based decision support system

El grupo de control (GC), permitió la selección de las palabras claves y con mayor frecuencia en los artículos científicos, considerando que deben encontrarse estrechamente alineadas al objetivo de la investigación, las cuales son: Hospital appointment, scheduling, modeling and analysis, data models, data mining, simulation.

Construcción de la Cadena de Búsqueda

Identificados los términos, en esta fase se procede a crear y probar las posibles cadenas de búsqueda en la base digital escogida, para este caso es IEEE XPLORE. Se construyó inicialmente como se muestra a continuación:

((("All Metadata":data mining) AND "All Metadata":hospital scheduling) AND "All Metadata":Hospital appointment)

El retorno de la cadena de búsqueda es de 2 artículos científicos, los cuales tienen una estrecha relación a la temática y la combinación de los términos clave utilizan los conectores específicos de la base digital IEEE Xplore acorde al contexto de la búsqueda.

Selección de los Estudios Primarios

La búsqueda en la base digital IEEE Xplore, empleando la cadena final retornó 2 artículos científicos, en los cuales se encontró una relación directa con la mayor cantidad de artículos pertenecientes al grupo de control, considerando este aspecto se determinó como la cadena idónea.

Se aplicaron los siguientes filtros adicionales de manera que los estudios candidatos sean válidos al ser considerados como parte del estado de la temática:

Se seleccionaron cinco artículos científicos de acuerdo a la búsqueda del grupo de control y a la cadena de búsqueda como estudios primarios y se detallan en la tabla 2.

Tabla 2.*Estudios primarios*

Código	Título	Cita
EP1	A machine learning framework for auto classification of imaging system exams in hospital setting for utilization optimization	(Meru, Ravindra, Krishnamoorthy, & Jacob, 2016)
EP2	Appointment scheduling of oncology outpatients	(Sadki, Xie, & Chauvin, 2011)
EP3	Integrated Innovative Solutions to Improve Healthcare Scheduling	(Ovidiu Stan, & Iulia Stefan, 2016)
EP4	A Web-based Application for Innovative Hospital Appointment Scheduling using Neural Network	(Rammos, Papadimitriou, Virvou, & Alepis, 2018)

Elaboración del Estado de Arte

EP1 (Meru, Ravindra, Krishnamoorthy, & Jacob, 2016): A machine-learning framework for auto classification of imaging system exams in hospital setting for utilization optimization.

Los autores desarrollan un método para clasificar exámenes de tipo anatómico aplicando técnicas de aprendizaje automático (SVM, KNN y árboles de decisión) para registrar información de los sistemas. El resultado de la clasificación muestra una precisión superior al 90%. Los autores crean una metodología completa que abarca desde la extracción de datos de EnWiki hasta la capacitación y pruebas utilizando varios algoritmos de aprendizaje automático. En esta metodología, proponen un conjunto de características para utilizar en el análisis automático de datos con el fin de detectar las

cuentas de Sockpuppets creadas en EnWiki. Los aplican en una base de datos de 10.000 cuentas de usuario. Los resultados comparan varios algoritmos de aprendizaje automático para mostrar que las nuevas características y datos de entrenamiento permiten detectar el 99% de las cuentas falsas, mejorando los resultados previos.

EP2 (Sadki, Xie, & Chauvin, 2011): Appointment scheduling of oncology outpatients

Los autores abordan el problema de agendamiento de citas de pacientes de una unidad ambulatoria de oncología. Cada paciente es visto por primera vez por su oncólogo, y luego el medicamento se prepara e inyecta al paciente. Se consideran dos recursos principales, los oncólogos para consulta y las camas para inyección. Los tiempos de inactividad de los oncólogos no están permitidos. El problema consiste en agendar el tiempo de cita de cada paciente y el tiempo de inicio de la inyección para minimizar el costo ponderado total incurrido por los periodos de espera del paciente. Se propone una heurística Lagrangiana y una heurística de optimización local. Los resultados numéricos muestran que la heurística basada en la relajación Lagrangiana siempre supera a la heurística de optimización local.

EP3 (Ovidiu Stan, & Iulia Stefan, 2016): Integrated Innovative Solutions to Improve Healthcare Scheduling

Los autores presentan un nuevo método de programación de citas médicas para pacientes crónicos. Teniendo en cuenta la situación actual, en la que debido al creciente número de pacientes crónicos, tanto los hospitales como las unidades de atención primaria tienen que hacer frente a un número cada vez mayor de citas, uno de los objetivos principales del método propuesto es equilibrar la carga de trabajo del personal

médico. Se eligió un enfoque difuso y se consideraron muchos factores, como el contexto epidemiológico, la carga de trabajo actual del personal médico, las preferencias individuales, los períodos de vacaciones y las variaciones estacionales del número de pacientes en el proceso de planificación. Estos factores se dividieron en grupos y las reglas de inferencia difusa se aplicaron en dos etapas. Primero, diferentes cargas del agendamiento de citas según el contexto se determinaron en intervalos de tiempo predefinidos. En la segunda etapa, considerando los rasgos característicos de la enfermedad crónica y las evoluciones individuales actuales, los pacientes se distribuyeron en los plazos establecidos previamente. Se utilizaron redes de Petri difusas para modelar la aplicación. El método propuesto es flexible y ofrece la oportunidad de utilizar algunas características específicas de un monitoreo correspondiente de pacientes crónicos para mejorar y equilibrar la carga de trabajo del personal médico.

EP4 (Rammos, Papadimitriou, Virvou, & Alepis, 2018): A Web-based Application for Innovative Hospital Appointment Scheduling using Neural Network

En un intento por mejorar el sistema de citas en el hospital y reducir el tiempo de espera de los pacientes, los autores desarrollaron una aplicación inteligente basada en una aplicación web para programar citas. Se utiliza una red neuronal para definir dinámicamente los intervalos de tiempo de la programación de cada médico. Se presenta la codificación de las variables de la red neuronal y describen las tres etapas que siguieron hasta concluir en su configuración más adecuada. Finalmente, se compara sus resultados con los esperados demostrando su fiabilidad.

Capítulo II

Marco Teórico

En este capítulo se aborda aspectos teóricos y conceptuales sobre inteligencia artificial, sus técnicas, tipos de algoritmos, además de un breve estudio sobre las herramientas que permiten desarrollar sistemas expertos y de soporte a la decisión (DSS). Los DSS permiten realizar el análisis sobre los datos de una organización para resolver un problema en concreto (Sinnexus, 2020).

Además, se incluye información de la metodología utilizada para el desarrollo del modelo basado en inteligencia artificial.

Inteligencia artificial

La Inteligencia Artificial data su existencia desde el apareamiento de la vida humana cuando los pueblos primitivos iniciaron su deseo de mejorar sus cualidades tanto físicas como intelectuales. Es así que el concepto de Inteligencia Artificial está presente en varios momentos y actividades de la historia, su principal objetivo ha sido imitar tanto al pensamiento como a la capacidad de aprender del ser humano. (Munera, 1991)

En años recientes se ha evidenciado un crecimiento acelerado en el uso y la aplicación de sistemas inteligentes en distintas áreas como por ejemplo: seguridad, comercio, educación, medicina, entre otros. En el caso de la medicina existe una alta demanda que ha dado paso al desarrollo de sistemas utilizando inteligencia artificial se vaya perfeccionando y mejorando hasta convertirse actualmente en una técnica potente y flexible a la vez, permitiendo de esta manera manipular e interpretar grandes cantidades de datos. (Shadabi & Sharma, 2008)

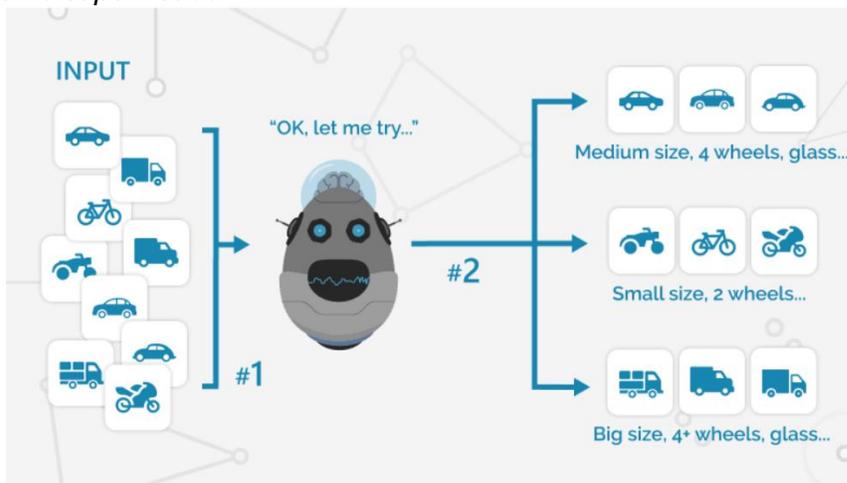
La inteligencia artificial es aplicada a través de los algoritmos que utiliza: aprendizaje supervisado, no supervisado y por refuerzo; los cuales son conocidos en forma específica como algoritmos de Machine Learning considerando que la inteligencia artificial es un “Súper conjunto de tecnologías” (Aura Portal, 2019).

Algoritmos no supervisados

Los algoritmos de aprendizaje no supervisados se diferencian de los supervisados en que únicamente consideran los datos de entrada dentro del sistema, de esta manera los resultados se dan en función de un auto entrenamiento en el cual no intervienen reglas externas.

Figura 1.

Aprendizaje no supervisado



Nota: La figura representa las entradas y salidas que se realiza en el proceso de un aprendizaje no supervisado. Tomado de (Aura Portal, 2019).

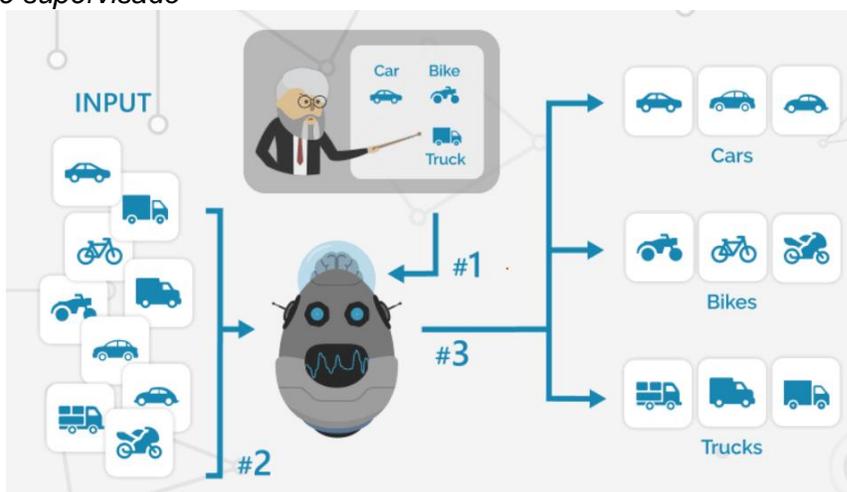
Algoritmos supervisados

La aplicación de algoritmos supervisados se realiza cuando se dispone de datos de entrenamiento como parte de las distintas entradas que tienen los modelos predictivos,

para que mediante el manejo de reglas consideradas como indicaciones externas se pueda obtener los resultados esperados (Aura Portal, 2019).

Figura 2.

Aprendizaje supervisado



Nota: La figura representa las entradas, reglas y salidas que se realiza en el proceso de un aprendizaje supervisado. Tomado de (Aura Portal, 2019).

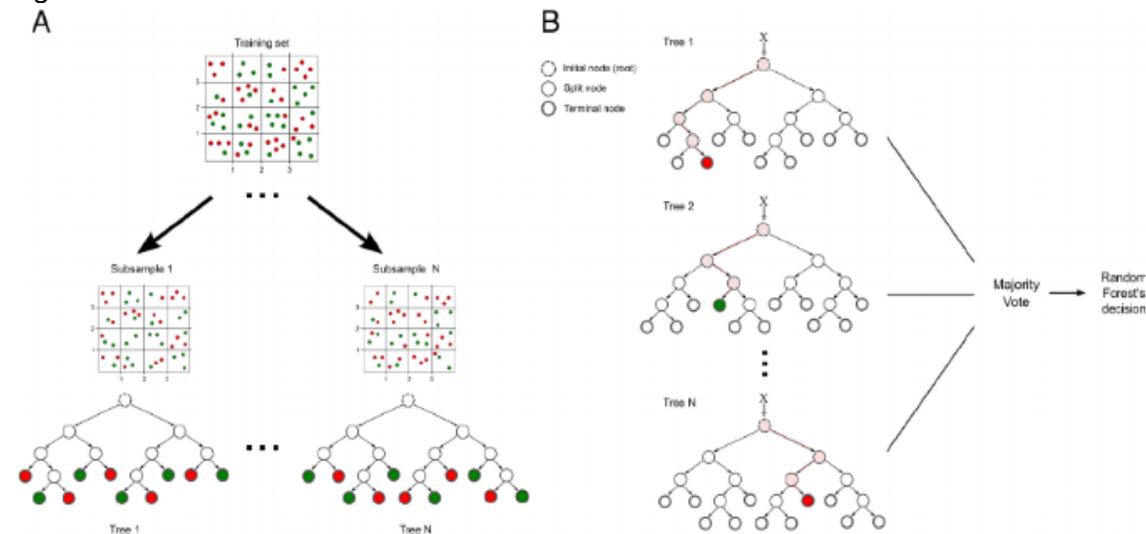
El aprendizaje supervisado cuenta con distintos algoritmos que permiten realizar la categorización de los datos, los cuales deben ser seleccionados de manera adecuada puesto que no todos permitirán ejecutar la clasificación dentro de un sistema. La selección de los algoritmos dependerá del comportamiento de estos ante factores como disponibilidad, características y clases que presenten los datos (Román, 2019).

Random forest

El algoritmo random forest se basa en la combinación de árboles de decisión para la clasificación que están fundamentados bajo ciertas condiciones para optimizar el resultado. Además se considera un método preciso, estable y dinámico que busca el sesgo y la varianza del bosque (Pérez Tatamués, 2019).

En este algoritmo son importantes los factores de fuerza y correlación. Estos realizan la medición de la heterogeneidad existente entre los árboles predictores y su precisión; quiere decir que mientras mayor sea esta relación se obtendrá una calidad superior en cuanto a la productividad del conjunto.

Se considera este algoritmo para la solución del problema planteado ya que cumple ciertas condiciones, pues al pertenecer a la inteligencia artificial y a su subconjunto como es el aprendizaje automático permite manejar gran cantidad de datos, que en este estudio son basados en el historial de atenciones médicas realizadas durante el año 2018. Dentro del proceso del algoritmo se realiza el entrenamiento y la clasificación, para la primera etapa cada árbol de decisión del conjunto se basa en una muestra de arranque aleatoria de los datos originales, que contiene ejemplos positivos y negativos; mientras que la predicción de clases para nuevas instancias se basa en un procedimiento de votación por mayoría entre todos los árboles individuales. El procedimiento que se lleva a cabo para cada árbol es el siguiente: para cada nuevo punto de datos, el algoritmo comienza en el nodo raíz de un árbol de decisión y atraviesa el árbol probando los valores de las variables en cada una de las divisiones visitadas. Este proceso se repite hasta que se alcanza un nodo hoja, que asigna una clase a esta instancia: los nodos verdes predicen para la clase positiva, los nodos rojos predicen para la clase negativa. Al final del proceso, cada árbol emite un voto por la etiqueta de clase preferida y el modo de los resultados se elige a manera de predicción final, como se observa en la siguiente figura.

Figura 3.*Algoritmo random forest*

Nota: La figura representa el proceso de entrenamiento y clasificación que se usa en random forest. A) Se representa el proceso de entrenamiento en la que los ejemplos positivos están representados por etiquetas verdes y negativos por rojas. B) Muestra la clasificación y predicción de clases en el que cada nuevo punto de datos se representa por X, el camino que atraviesa se expresa mediante las ramas resaltadas. Los nodos que predicen la clase negativa se representan por el color rojo y las clases positivas por el color verde. Tomado de (Machado, 2015).

Árboles de decisión

Es un algoritmo que se considera dentro del aprendizaje supervisado ya que toma inicialmente o requiere de datos que hayan sido etiquetados. Dentro de su estructura está compuesto de reglas subsecuentes las cuales finalizan cuando cada uno de los elementos tiene asignada una determinada etiqueta. Este algoritmo permite realizar una interpretación sencilla de sus datos ya que no funciona como una caja negra en la cual las reglas o condiciones no son visibles en forma notable para cualquier usuario, sin

embargo en el árbol de decisión se pueden observar explícitamente durante todo el proceso. Finalmente, para realizar la distribución de los datos no parte de un modelo estadístico como lo requieren algoritmos como la regresión logística múltiple, mínimos cuadrados generalizados y la regresión lineal múltiple.

Técnicas y campos de la inteligencia artificial

Minería de datos

La minería de datos surge como una herramienta utilizada para la interpretación y el conocimiento de toda la información que está distribuida mundialmente. El avance tecnológico para gestionar bases de datos permite integrar variables de distinta índole como video, texto, imagen, entre otros, dentro de una sola base, permitiendo de esta manera la administración sencilla y un mejor procesamiento de la información.

Dentro de las tareas propias de la minería de datos están el descubrimiento de patrones relevantes o características que permitan describir los datos, y la predicción de las salidas en función de la clasificación de nuevos datos para revelar sus relaciones; para el efecto se utilizan algoritmos sofisticados que permiten descubrir e interpretar la estructura de una gran cantidad de datos que están almacenados dentro de los data warehouses (Riquelme Santos, Ruiz, & Gilbert, 2006).

Machine learning

Esta rama perteneciente a la inteligencia artificial se encarga de desarrollar algoritmos que permitan a un determinado sistema aprender distintos datos en función de la información que se proporcione inicialmente, para lo cual analiza la estructura y los patrones que se encuentren inmersos en dicha información, los consolida y crea un

nuevo conocimiento que permita la toma de decisiones y la ejecución de acciones sobre los nuevos datos que sean proporcionados como entrada (Pérez Tatamués, 2019).

El machine learning permite resolver problemas referentes a la clasificación de datos usando algoritmos. Para efecto de esta investigación se utiliza algoritmos de aprendizaje supervisados por el tipo de datos que son tomados como entrada del sistema.

Analítica de Datos

La analítica de datos hace posible que cualquier tipo de información pueda ser sometida a técnicas de análisis para obtener información relevante, llegando incluso a señalar lo que se conoce como 'cuellos de botella' dentro de los ambientes de producción (Mario Bochicchio, 2016).

La analítica de datos es uno de los enfoques más conocidos en cuanto al análisis de los datos, refiriéndonos al Big Data en particular, que nos permite llegar a obtener conclusiones, para llegar a tomar decisiones estratégicas que ayuden a aumentar el volumen de negocios en determinada empresa. En este caso la asignación eficiente de los turnos en el Hospital de Especialidades de Fuerzas Armadas N°1 en base a los datos históricos recopilados por la Unidad de Tecnologías de la Información, utilizando la analítica de datos es posible mejorar la operatividad del hospital.

En este caso se utiliza el Análisis de Datos Exploratorios ya que permite realizar una búsqueda supervisada de patrones y relaciones dentro de los datos que consideramos de entrada, de esta forma se efectúa un análisis tanto cuantitativo como cualitativo, ya que las variables encargadas de manejar el tiempo de atención de los pacientes son numéricas y los datos de las enfermedades son cadenas de caracteres.

Estas son comparadas con el documento que detalla las más de catorce mil enfermedades conocidas actualmente, conocido como CIE-10 (Organización Panamericana de la Salud, 1995), lo cual hace que los datos obtenidos sean información que pueda llegar a ser procesada.

Figura 4.

Enfermedades analizadas

	I	EDAD	S	GRUPO...	S	ENFERMEDAD	S	CODIG...	S	DIAGN...
44		20-49 ANIOS				CONVULSIONES NO CLASIFICADAS EN OTRA PARTE		R56		PRIMARIO
3		1-4 ANIOS				ADENOMEGALIA		R59		PRESUNTIVO
35		20-49 ANIOS				ADENOMEGALIA LOCALIZADA		R59.0		PRIMARIO
1		1-4 ANIOS				FALTA DEL DESARROLLO FISIOLÓGICO NORMAL E...		R62		PRIMARIO
1		1-4 ANIOS				RETARDO EN DESARROLLO		R62.0		PRESUNTIVO
58		50-64 ANIOS				NIVEL ELEVADO DE GLUCOSA EN SANGRE		R73		PRESUNTIVO
32		20-49 ANIOS				ABDOMEN AGUDO		R10.0		PRESUNTIVO
18		15-19 ANIOS				DOLOR ABDOMINAL LOCALIZADO EN PARTE SUPE...		R10.1		PRESUNTIVO
42		20-49 ANIOS				DOLOR ABDOMINAL LOCALIZADO EN PARTE SUPE...		R10.1		PRESUNTIVO
54		50-64 ANIOS				DOLOR LOCALIZADO EN OTRAS PARTES INFERIO...		R10.3		PRESUNTIVO
1		1-4 ANIOS				HEPATOMEGALIA, NO CLASIFICADA EN OTRA PA...		R16.0		PRESUNTIVO
51		50-64 ANIOS				TUMEFACCION, MASA O PROMINENCIA INTRAAB...		R19.0		PRESUNTIVO
14		10-14 ANIOS				TRAUMATISMO SUPERFICIAL DE LA CABEZA		S00		PRIMARIO
59		50-64 ANIOS				OTROS TRAUMATISMOS DE MIEMBRO INFERIOR, ...		T13		PRESUNTIVO
24		20-49 ANIOS				HERIDA DEL TOBILLO Y DEL PIE		S91		PRESUNTIVO
25		20-49 ANIOS				FRACTURA DE HUESO DEL METATARSO		S92.3		PRIMARIO
42		20-49 ANIOS				FRACTURA DE LA EPIFISIS INFERIOR DEL RADIO		S52.5		PRIMARIO
43		20-49 ANIOS				FRACTURA DE LA EPIFISIS INFERIOR DEL RADIO		S52.5		PRESUNTIVO
34		20-49 ANIOS				FRACTURA DE LA EPIFISIS INFERIOR DEL CUBITO...		S52.6		PRIMARIO
20		20-49 ANIOS				CONTUSION DE DEDO(S) DE LA MANO, SIN DAÑO ...		S60.0		PRESUNTIVO
6		5-9 ANIOS				FRACTURA DE OTRAS PARTES Y DE LAS NO ESPE...		S62.8		PRESUNTIVO
92		65 ANIOS Y ...				FRACTURA DE OTRAS PARTES Y DE LAS NO ESPE...		S62.8		PRESUNTIVO
73		65 ANIOS Y ...				FRACTURA DE OTRAS PARTES Y DE LAS NO ESPE...		S62.8		PRESUNTIVO
16		15-19 ANIOS				LUXACION DE DEDOS DE LA MANO		S63.1		PRESUNTIVO
12		10-14 ANIOS				TRAUMATISMO DEL OJO Y DE LA ORBITA, NO ESP...		S05.9		PRESUNTIVO

Nota: La información del sistema hospitalario fue previamente tratada haciendo uso de algoritmos capaces de reconocer la distancia de Levenshtein entre estos datos y los obtenidos del documento CIE-10.

Herramientas para el análisis de datos

Para el análisis de los datos del Hospital de Especialidades de Fuerzas Armadas N°1 se consideró varias de las herramientas más populares actualmente, pero se procedió a estudiar 2 de ellas en cada una de las etapas del análisis (Minería y Visualización).

Para el proceso de Minería de Datos se analizó KNIME y RapidMiner, siendo más eficiente KNIME. Esta herramienta se recomienda para personas que están iniciando en el manejo de software para el análisis de datos, como para aquellas que tienen experiencia en este tema, además el software presenta componentes muy robustos que incluso permiten incorporar adicionales mediante el uso de otras librerías (Dwivedi, Kasliwal, & Soni, 2016).

Tabla 3.

Comparación entre KNIME y Rapidminer

Métrica	KNIME	RapidMiner
Modelo de Precios	Gratis, Basado en cotización	Gratis, Suscripción Anual
Limitaciones en versión gratuita	Los componentes necesarios para analítica de datos deben ser instalados manualmente.	Los componentes para analítica de datos restringen el número de datos que pueden analizarse (10000 filas).
Ejecución de procesos en segundo plano	Disponible	Disponible a partir de paquete 'Mediano' (\$5000 / usuario / año)
Integración	Tableau, R, Python, Java	Tableau
Tamaño de empresas	Pequeñas, Medianas, Grandes	Pequeñas, Medianas
Sistema Operativo	Windows, Linux, MacOS	Windows, Linux, MacOS
Soporte por Formación	Disponible	No Disponible
Algoritmos de Inteligencia Artificial	Disponible	Disponible a partir de paquete 'Mediano' (\$5000 / usuario / año)
Calificación de Algoritmos de Predicción según Gartner	4.7 / 5	4.3 / 5
Satisfacción de la Comunidad	100%	96%
Calificación SmartScore	8.7 / 10	8.9 / 10

En base a estos parámetros se toma la decisión de utilizar la plataforma de KNIME, ya que esta permite acceder a todos los algoritmos de inteligencia artificial sin realizar ningún pago, adicionalmente Gartner le otorga una calificación superior en cuanto a sus algoritmos de predicción (Shraddha Dwivedi, Comprehensive Study of Data Analytics Tools, 2016).

Para representar los datos y variables de la investigación se ha tomado en cuenta a Tableau y a Microsoft Power BI, ambas plataformas de análisis y generación de tableros de visualización a partir de entradas dadas como ejes de investigación, siendo el elegido Tableau, esto en base a los parámetros de la tabla 4.

Tabla 4.

Comparación entre Tableau y Microsoft Power BI

Métrica	Tableau	Microsoft Power BI
Volumen de Datos	Teóricamente ilimitado	Limitado
Funcionalidad	Mejor con datos en la nube	No es buena con datos en la nube
Usuarios	Analistas experimentados y novatos	Analistas experimentados
Aprendizaje	Es poco difícil de aprender	Es fácil de aprender
Almacenamiento	Propio en servidores de Tableau	Se concentra en el modelado analítico, no en almacenar datos
Tamaño de Empresas	Medianas, Grandes	Pequeñas, Medianas, Grandes
Medición	Por dimensiones	Por columnas
Fuentes de Datos	Teóricamente ilimitado	Limitada a las que Microsoft provee
Creación de Informes	Desafiante	Fácil

Considerando estos parámetros se toma la decisión de utilizar la plataforma de Tableau, ya que esta brinda facilidades superiores, pero principalmente por sus métodos de conexión teóricamente ilimitados.

Herramienta KNIME

“Konstanz Information Miner” conocido comercialmente como “KNIME”, es un software de código abierto para aplicaciones de ciencia de datos enfocada a la minería de datos, aunque su comunidad ha logrado desarrollar varias configuraciones, extensiones y herramientas que pueden ser utilizadas incluso para el análisis de imágenes. Esto ha hecho posible la consecuente creación y desarrollo de modelos amigables con el usuario final porque se maneja un entorno completamente visual, evitando el uso innecesario de líneas de código y programación, permitiendo acceder a cualquiera de las operaciones realizadas por los componentes utilizados (Network Analysis, 2019).

KNIME posee componentes capaces de aprovechar por completo los algoritmos más comunes de aprendizaje automático e inteligencia artificial, lo único que se necesita es arrastrar y soltar el componente específico para la tarea que deseamos desarrollar. Es un software multiplataforma (Linux, Windows, Mac OSX), ya que es construido con el IDE Eclipse, se ejecutará en cualquier máquina que soporte la máquina virtual de Java (Network Analysis, 2019).

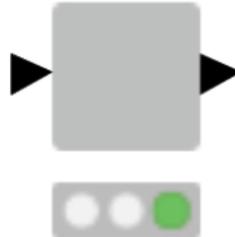
Componentes de KNIME

Los componentes son nodos que contienen un sub-flujo de trabajo, el cual permite agrupar una funcionalidad en concreto que depende del componente a utilizarse, agrupando la funcionalidad para compartir y reutilizar el flujo creado (Network Analysis, 2018). Los componentes encapsulan y abstraen la funcionalidad para la cual están destinados, cada uno de ellos tiene su propio diálogo de configuración lo que permite de cierta manera ocultar la complejidad detrás del nodo en cuestión. De esta

manera se logra un flujo de trabajo sumamente fácil de leer y por consecuencia entender, permitiendo que posteriormente se reutilicen los componentes o nodos.

Figura 5.

Componente



Nota: La figura representa un componente en la plataforma KNIME. Tomado de (Network Analysis, 2019)

Metanodo

Los metanodos, en contraste a los componentes, son utilizados para organizar de mejor manera los flujos de trabajo, se puede tomar una gran parte o partes del flujo de trabajo en el que se está trabajando y colapsarlo todo dentro de un metanodo, permitiendo esconder de cierta manera la funcionalidad, haciendo más comfortable la lectura y compresión del modelo en cuestión (Network Analysis, 2015).

Figura 6.

Metanodo

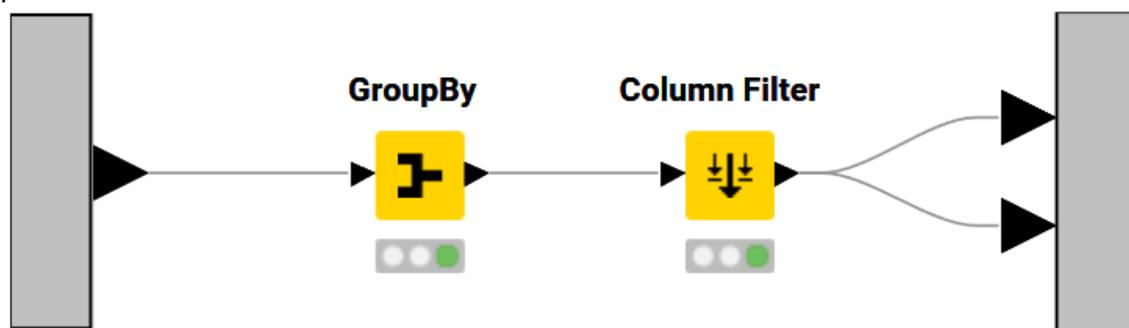


Nota: La figura representa un metanodo en la plataforma KNIME. Tomado de (Network Analysis, 2015)

A diferencia de los componentes, estos no pueden compartirse por las plataformas que proporciona KNIME, tampoco tienen diálogos de configuración, esto debido a que únicamente son componentes visuales que ayudan a entender de una manera general los pasos que se han generado dentro del flujo de trabajo.

Figura 7.

Componentes internos de un metanodo



Nota: La figura representa los componentes dentro de un metanodo en la plataforma KNIME. Tomado de (Network Analysis, 2015)

File Reader

La forma más común de almacenar cantidades relativamente pequeñas de datos sigue siendo un archivo de texto. Entre los archivos de texto, el formato más común ha sido hasta ahora el formato CSV (versión separada por comas). La "coma" en el acrónimo CSV es solo uno de los posibles caracteres para separar los datos dentro del archivo. Los signos de punto y coma, dos puntos, tabulación y muchos otros son igualmente aceptables (Network Analysis, 2018).

El nodo más versátil para acceder a un archivo de texto, cualquiera que sea su formato, es el Lector de archivos.

Figura 8.*File reader*

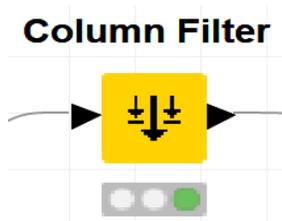
Nota: El componente File Reader no es el único capaz de realizar esta función, ya que existen lectores específicos para cada tipo de archivo, por lo que al utilizar un archivo de texto sin formato y con codificación UTF-8 este componente es el más eficiente.

Tomado de (Network Analysis, 2018)

Column Filter

En un conjunto de datos, no todas las columnas de datos contienen información o están llenas de valores útiles. En caso de que queramos deshacernos de las columnas no deseadas, la operación ETL a ejecutar es el filtrado de columnas.

Este componente permite eliminar columnas no deseadas, ya sea de forma manual o automática en función de algunos conocimientos previos, haciendo uso de tablas de selección (Network Analysis, 2018).

Figura 9.*Column filter*

Nota: La figura representa el componente Column Filter en la plataforma KNIME.

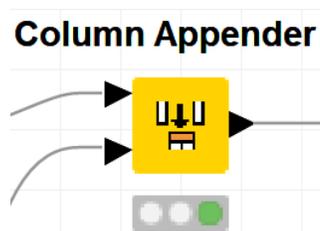
Tomado de (Network Analysis, 2018).

Column Appender

El componente Column Appender toma dos tablas y las combina de manera que agrega las columnas de la segunda tabla a la primera tabla. Realiza una función similar a la de Joiner, pero resulta ser más rápida para tablas grandes (Hub Knime, 2017). Sin embargo, es posible utilizar solo si se han cumplido ciertas condiciones previas, la más importante es evitar por completo las claves de fila idénticas. Una vez finalizado el proceso dentro del componente, este creará una nueva tabla con los datos en conjuntos de las tablas de entrada.

Figura 10.

Column appender



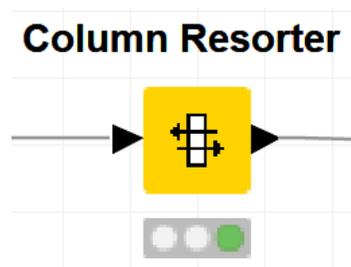
Column Resorter

El componente Column Resorter se encarga de cambiar el orden de las columnas de entrada, mediante un flujo de trabajo y configuraciones definidos por el usuario, las columnas en cuestión pueden ser cambiadas mediante un simple paso, pueden ser movidas hacia la izquierda o la derecha, al igual pueden moverse al inicio o al final de la tabla con la que estamos trabajando, incluso existe la posibilidad de ordenar las columnas de manera alfabética y la tabla ordenada será el resultado de este componente, de esta manera se puede usar su puerto de salida para realizar operaciones con la nueva tabla (Hub Knime, 2020).

Una vez que se ha configurado el componente con reglas específicas, es posible ingresar una nueva tabla con estructura diferente y ejecutar el componente sin necesidad de configurarlo nuevamente. Las columnas nuevas y antiguas que resulten desconocidas para el nodo serán insertadas en la marca de posición que ha sido configurada previamente y este puede ser posicionado en cualquier lugar de la tabla en que el usuario lo encuentre necesario.

Figura 11.

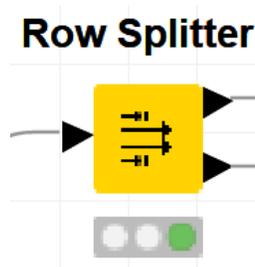
Column resorter



Row Splitter

El componente Row Splitter tiene una funcionalidad similar al componente Row Filter, excepto que tiene una salida adicional, la cual proporciona los datos de las filas que han sido filtradas (Node Pit, 2016).

El nodo permite el filtrado de filas de acuerdo con los criterios que han sido generados por el usuario, esto permite que se puedan incluir o excluir ciertos rangos, filas con cierto ID de fila, incluso filas con un cierto valor en una columna seleccionable.

Figura 12.*Row splitter***Rule Engine**

El componente Rule Engine toma una regla o lista de reglas definidas por el usuario y verifica que cada fila en la tabla de entrada coincida con estas. Si una regla coincide, el resultado se agrega a una nueva columna que será añadida a la tabla de salida (Hub Knime, 2018).

La primera regla que se cumpla, en orden de definición, determina el resultado. Cada regla está representada por una línea de código, que se detalla a continuación.

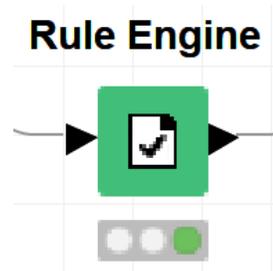
Figura 13.*Reglas definidas en el componente rule engine*

```

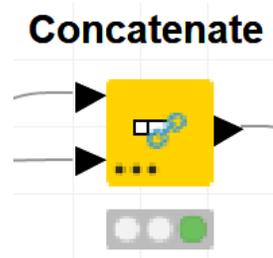
Expression
? 1 // enter ordered set of rules, e.g.:
? 2 // $double column name$ > 5.0 => "Large"
? 3 // $string column name$ LIKE "*blue*" => "small and blue"
? 4 // TRUE => "default outcome"
S 5 TRUE => $ESPECIALIDAD CONSULTORIO$

```

El resultado de una regla puede ser una cadena de caracteres, un número, una constante booleana, una referencia a otra columna o el valor de una variable del flujo de trabajo. El tipo de la columna resultado se conoce como 'súper tipo común' que puede almacenar todos los resultados posibles.

Figura 14.*Rule engine****Concatenate***

El componente Concatenate nos permite concatenar por completo dos tablas, las cuales serán ingresadas en los puertos de entrada, esto hace que las columnas que poseen nombres iguales en las dos tablas sean concatenadas (Network Analysis, 2019).

Figura 15.*Concatenate****Math Formula***

El componente Math Formula permite evaluar una expresión matemática basada en operaciones sobre las filas de la tabla de entrada. Los resultados calculados pueden agregarse como una nueva columna o usarse para reemplazar una columna de entrada.

Figura 16.

Math formula

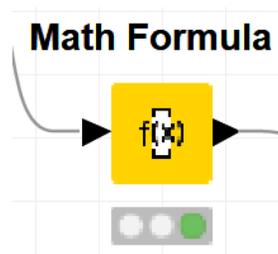


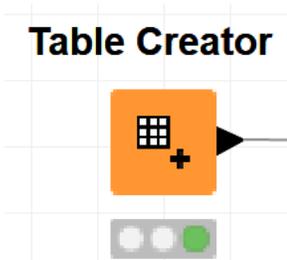
Table Creator

El componente Table Creator es un método rápido, simplificado y eficiente, que permite generar datos de entrada, los cuales son considerados ad-hoc. Este componente resulta ser sumamente poderoso porque permite generar datos con un ciclo de vida interno y completamente dependiente del flujo de trabajo en KNIME.

Una micro tabla o conjunto de tablas puede ser utilizado como fuente de información, lo que permite crear un almacén de datos modular al cual se le podrán realizar operaciones internas sin problemas o necesidad de instalar plugins externos.

Figura 17.

Table creator



Group By

El componente Group By agrupa las filas de una tabla de entrada por sus identificadores o valores únicos dentro del grupo seleccionado, para lograr esto crea una

fila para cada conjunto único de valores de la columna, las columnas restantes se agregarán según la configuración que ha sido previamente especificada por el usuario, esto permite que la tabla de salida posea una fila para cada combinación en base al valor único o identificador de las columnas seleccionadas (Node Pit, 2017).

El componente permite agregar las columnas seleccionadas por el usuario directamente, pero en un estudio esto podría afectar la calidad de la muestra a analizar, por lo que se recomienda incluso en la documentación de KNIME que se utilice la opción 'Agregación basada en patrones' o 'Agregación basada en tipos', para que la muestra tomada sea completamente fidedigna y pueda ser utilizada sin problema alguno dentro del flujo de trabajo y su posterior análisis.

Figura 18.

Group by



Chunk Loop Start

El componente Chunk Loop Start es el inicio de un bucle de fragmentación, esto implica que cada iteración procesa una nueva porción consecutiva de filas. La fragmentación nos permite establecer un número fijo de fragmentos, lo cual se conoce como el número de iteraciones, y será el número que rige el bucle dentro del flujo de trabajo (Hub Knime, 2020).

Figura 19.

Chunk loop start

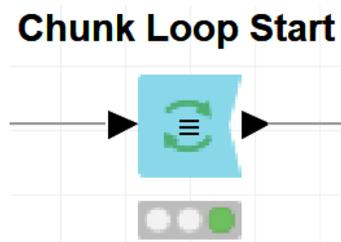
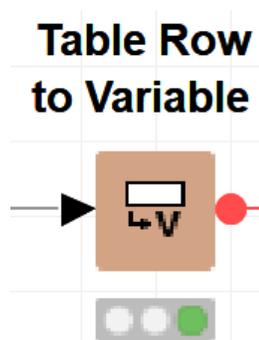


Table Row To Variable

El componente Table Row to Variable usa la primera fila de los datos de entrada para definir un nuevo flujo de variables. Los nombres de las variables son definidos por la columna de nombres en conjunto con la asignación de variables. Esto simplifica y otorga mayor eficiencia al momento de analizar individualmente los datos de una fila que forma parte de los datos de entrada.

Figura 20.

Table row to variable



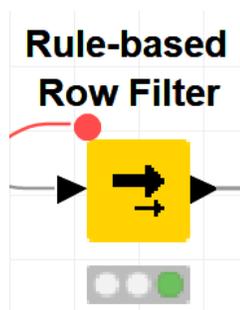
Rule-based Row Filter

El componente Rule-based Row Filter toma como entrada una lista de reglas definidas por el usuario, e intenta hacerlas coincidir con cada una de las filas que se encuentran definidas en la tabla de entrada (Hub Knime, 2016). Si la primera regla posee coincidencia tiene un resultado de 'Verdadero' o 'True', la fila será seleccionada

para su posterior inclusión, de lo contrario esta será excluida por completo de los datos de salida del propio componente, el cual es capaz de invertir la selección con una simple acción por parte del usuario dentro de su diálogo de configuración.

Figura 21.

Rule-based row filter



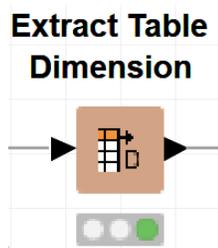
Extract Table Dimension

El componente Extract Table Dimension permite obtener tanto el número de filas, como el número de columnas de la tabla de entrada, posteriormente estos datos son escritos dentro de la tabla de salida (Hub Knime, 2017).

La salida de este componente consiste en dos filas y 1 columna, la primera fila contiene el número de filas y la segunda fila contiene el número de columnas de la tabla de entrada.

Figura 22.

Extract table dimension

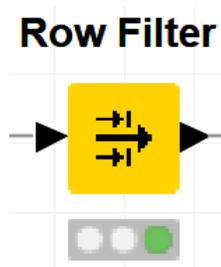


Row Filter

El componente Row Filter permite filtrar filas basado en reglas definidas por el usuario, lo que permite incluir o excluir por igual ciertos rangos de valores, ciertas filas con un determinado identificador (Node Pit, 2016).

Figura 23.

Row filter



Sorter

El componente Sorter permite ordenar toda la tabla, parte de esta, o simplemente filas o columnas en base a reglas definidas por el usuario (Node Pit, 2015). En el diálogo de configuración de este componente se puede incluso seleccionar que datos deben ser ordenados, esto puede escalar incluso a un orden alfabético, sea este ascendente o descendente.

Figura 24.

Sorter

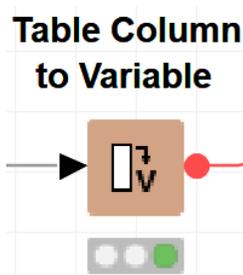


Table Column to Variable

El componente Table Column To Variable convierte los valores de una tabla de entrada en variables de flujo, en base a sus propios identificadores, todo esto analizando cada una de las columnas seleccionadas como valores de las variables independientes (Hub Knime, 2019).

Figura 25.

Table column to variable

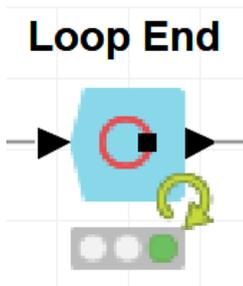


Loop End

El componente Loop End es utilizado para marcar el final de un bucle dentro del flujo de trabajo, de esta manera en su puerto de salida tendremos la colección de resultados intermedios y su completa concatenación con la(s) tabla(s) de entrada (Node Pit, 2018).

Figura 26.

Loop end

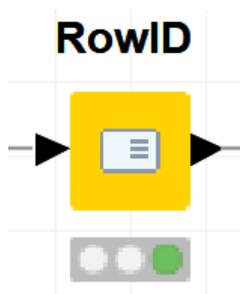


Row ID

El componente Row ID es utilizado para reemplazar el identificador de las filas de la tabla de entrada con valores pertenecientes a otra columna, convirtiendo los valores a una cadena de caracteres implícitamente o generando un identificador propio con el formato: 'Row0, Row1, Row2, ...'.

Figura 27.

Row id

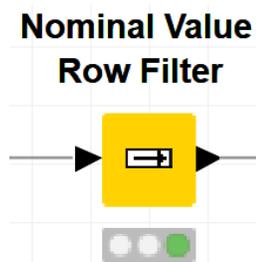


Nominal Value Row Filter

El componente Nominal Value Row Filter filtra las filas según el valor seleccionado de un atributo nominal perteneciente a la tabla de entrada, además se puede seleccionar una columna nominal en conjunto con uno o más valores nominales pertenecientes al atributo seleccionado (Node Pit, 2017).

Figura 28.

Nominal value row filter

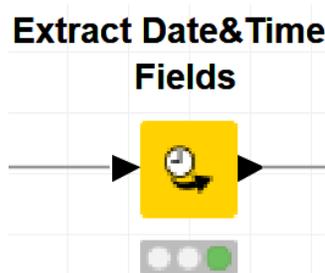


Extract Date&Time Fields

El componente Extract Date&Time Fields permite obtener y convertir a variable una cadena de caracteres que posea el formato de LocalDate, Local Date Time o Zoned Date Time, posteriormente añadirá los valores en la tabla de salida como variables numéricas (Node Pit, 2019).

Figura 29.

Extract date&time fields

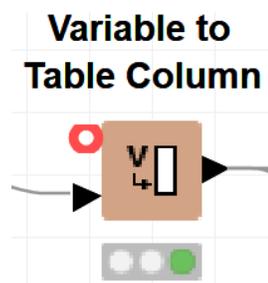


Variable to Table Column

El componente Variable to Table Column extrae las variables que han sido manejadas durante el flujo de trabajo y las añade a la tabla de entrada (Node Pit, 2020).

Figura 30.

Variable to table column



Date&Time Shift

El componente Date&Time Shift cambia una fecha u hora en base a las reglas previamente definidas por el usuario previamente, el componente permite seleccionar

las columnas de fecha y hora para cambiar, junto con el valor de cambio (Node Pit, 2019).

Figura 31.

Date&time shift

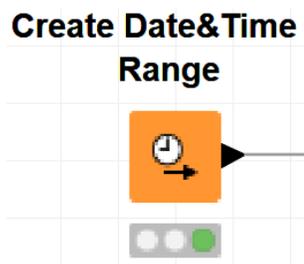


Create Date&Time Range

El componente Create Date&Time Range genera valores de fecha y hora, en cualquiera de los formatos soportados por KNIME (Hub Knime, 2019). Existen tres modelos de creación: seleccionar inicio y final de las filas, ingresar el número de filas o seleccionar un intervalo que se repetirá posteriormente en la tabla de entrada.

Figura 32.

Create date&time range



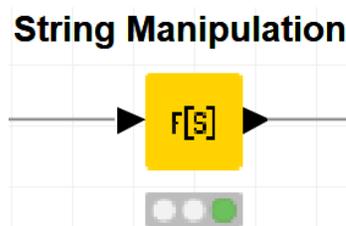
String Manipulation

El componente String Manipulation manipula cadenas de caracteres y puede ejecutar las operaciones: reemplazar, capitalizar, eliminar, eliminar espacios. De esta manera se puede utilizar cadenas de caracteres y buscar 'similares' para ejecutar

operaciones internas y obtener el dato necesario para ejecutar determinado proceso dentro del flujo de trabajo (Hub Knime, 2018).

Figura 33.

String manipulation

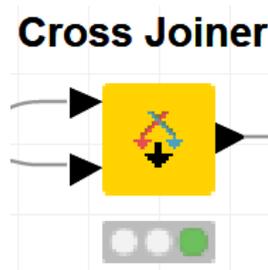


Cross Joiner

El componente Cross Joiner realiza una unión cruzada de dos tablas. Cada fila de la tabla superior se une con cada fila de la tabla inferior. Esta es una operación extremadamente costosa en términos de procesamiento computacional, ya que el número de filas en la salida es el producto de los dos recuentos de filas de la tabla de entrada.

Figura 34.

Cross joiner

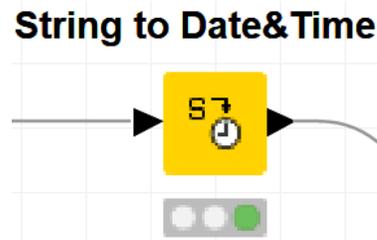


String to Date&Time

El componente String to Date&Time analiza las cadenas en las columnas seleccionadas y las convierte en celdas de fecha y hora utilizando el patrón de formato proporcionado.

Figura 35.

String to date&time

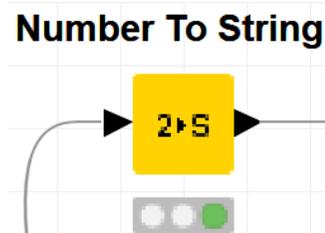


Number to String

El componente Number to String transforma números en una columna o en un conjunto de columnas a cadenas, el nodo permite incluso redondear variables de tipo doble dentro de su diálogo de configuración (Node Pit, 2018).

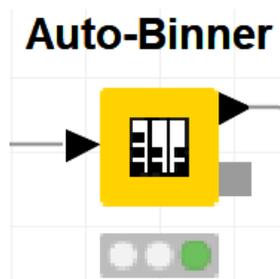
Figura 36.

Number to string

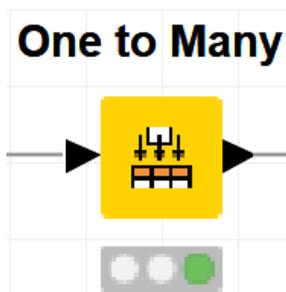


Auto Binner

El componente Auto Binner permite agrupar datos numéricos en intervalos, llamados contenedores. Hay dos opciones de nomenclatura para los contenedores y dos métodos que definen el número y el rango de valores que se encuentran en un contenedor.

Figura 37.*Auto binner****One to Many***

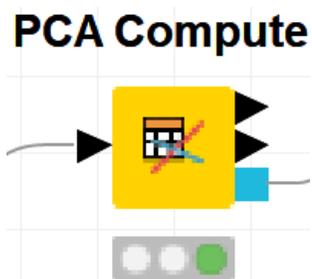
El componente One to Many transforma todos los valores posibles en una columna seleccionada. El valor se establece como el nombre de la nueva columna, los valores de celda en esa columna son 1 si esa fila contiene este posible valor, o 0 si no (Hub Knime, 2018).

Figura 38.*One to many****PCA Compute***

El componente PCA Compute realiza un análisis de componentes principales (PCA) en los datos de entrada. Las direcciones de varianza máxima o componentes principales se extraen y se pueden utilizar posteriormente para proyectar la entrada en un espacio de dimensión inferior mientras se preserva la máxima información (Node Pit, 2017).

Figura 39.

PCA compute

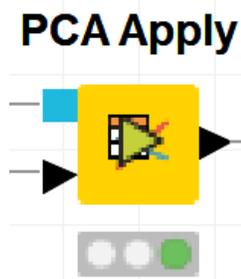


PCA Apply

El componente PCA Apply aplica una proyección a los componentes principales en los datos de entrada. El modelo de datos del componente 'PCA Compute' se puede aplicar a datos arbitrarios para reducirlo a un número dado de dimensiones (Hub Knime, 2017).

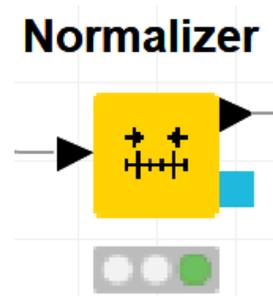
Figura 40.

PCA apply

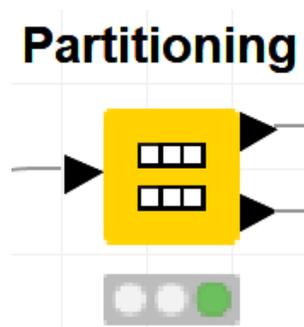


Normalizer

El componente Normalizer normaliza los valores de todas las columnas numéricas. En el diálogo de configuración, se puede elegir las columnas en las que desea trabajar.

Figura 41.*Normalizer****Partitioning***

El componente Partitioning permite que la tabla de entrada se divida en dos particiones, entrenamiento y prueba de datos. Las dos particiones están disponibles en los dos puertos de salida (Hub Knime, 2019).

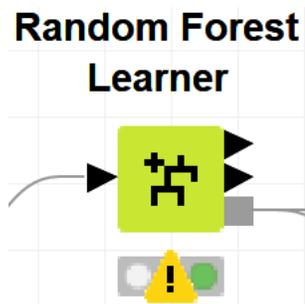
Figura 42.*Partitioning****Random Forest Learner***

El componente Random Forest Learner consiste en un número elegido de árboles de decisión. Cada uno de los modelos de árbol de decisión se aprende en un conjunto diferente de filas o registros y un conjunto diferente de columnas que describe los atributos, por lo que el último también puede ser un vector de bits o un descriptor de

vector de bytes (Node Pit, 2018). Los conjuntos de filas para cada árbol de decisión se crean mediante bootstrapping y tienen el mismo tamaño que la tabla de entrada original.

Figura 43.

Random forest learner

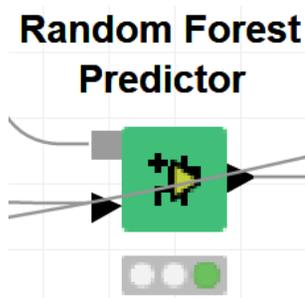


Random Forest Predictor

El componente Random Forest Predictor permite predecir patrones de acuerdo con el conjunto de datos generados por cada uno de los árboles individuales en un modelo de random forest (Node Pit, 2018).

Figura 44.

Random forest predictor



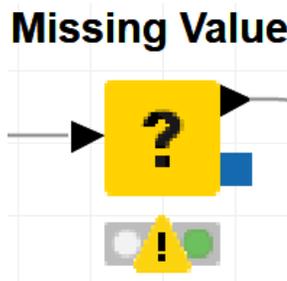
Missing Value

El componente Missing Value ayuda a manejar los valores faltantes encontrados en las celdas de la tabla de entrada. La primera pestaña del diálogo de configuración

que posee la etiquetada 'Predeterminado' proporciona opciones de manejo predeterminadas para todas las columnas de un tipo dado (Node Pit, 2016).

Figura 45.

Missing value

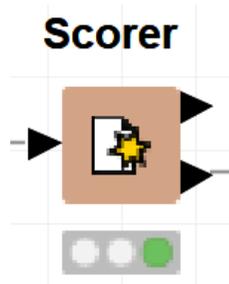


Scorer

El componente Scorer compara dos columnas por sus pares de valores de atributos y muestra la matriz de confusión, es decir, cuántas filas de cada atributo y su clasificación coinciden. Además, es posible resaltar las celdas de esta matriz para determinar las filas subyacentes.

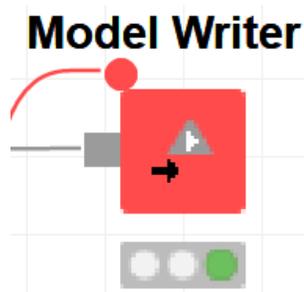
Figura 46.

Missing value

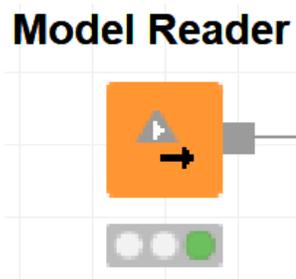


Model Writer

El componente Model Writer escribe objetos como parte del entrenamiento de los algoritmos de aprendizaje no supervisado de KNIME en un archivo, que se puede leer con el componente Model Reader (Node Pit, 2018).

Figura 47.*Model writer****Model Reader***

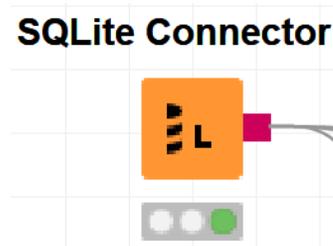
El componente Model Reader lee los objetos del puerto de entrada y los convierte en un modelo que puedan operar los componentes de KNIME (Node Pit, 2018).

Figura 48.*Model reader****SQLite Connector***

El componente SQLite Connector crea una conexión a un archivo de base de datos SQLite a través de su controlador JDBC. El usuario debe proporcionar la ruta al archivo de la base de datos o el nombre cuando utilice el modo de memoria, adicionalmente no se requiere nombre de usuario ni contraseña para las bases de datos SQLite (Node Pit, 2018).

Figura 49.

SQLite connector

***DB Writer***

El componente DB Writer inserta filas de datos en la base de datos en función de las columnas seleccionadas de la tabla de entrada. Si la tabla de la base de datos no existe, será creada por el componente antes de insertar los valores. La nueva tabla tendrá una columna para cada columna de entrada. El nombre de la columna de la base de datos será el mismo que el de la columna KNIME de entrada. Todas las columnas de la base de datos permitirán valores faltantes, los cuales serán reconocidos como 'NULL' (Hub Knime, 2019).

Figura 50.

DB writer



Metodología de Desarrollo

Los proyectos que usan la metodología Cross Industry Standard Process for Data Mining, conocido por sus siglas como CRISP-DM, se encarga de abarcar todas las

etapas y las tareas por realizar, de forma comparable a la ingeniería de software y las metodologías que se usan en esta. Como analogía se puede decir que el ciclo de vida para un proyecto de minería de datos consta de 6 etapas bien definidas, que son las siguientes:

Fase 1. Comprensión del negocio

Esta fase está completamente centrada en la comprensión total de los objetivos detallados dentro del proyecto y la creación de un plan preliminar que se encuentre específicamente diseñado para cumplir con las metas propuestas.

Fase 2. Estudio y comprensión de los datos

Esta fase inicia con la recolección de data que posteriormente será analizada, continuando con las tareas y actividades que admite la familiarización por parte del investigador. Este paso permite llegar a obtener información que puede ser considerada como conocimiento y de la misma manera llegar a revelar conjuntos interesantes que se encuentren ocultos dentro de la información.

Fase 3. Análisis de los datos y selección de las características

Esta fase está completamente centrada en el tratamiento de la data de entrada, para que de esta manera esta se convierta en útil y necesaria, capaz de construir el conjunto definitivo que ingrese al análisis, en concreto datos que sirvan para tomar decisiones en base al aprendizaje y el modelo que posteriormente puedan desarrollarse.

Fase 4. Modelado

En esta fase, deben seleccionarse y aplicarse métodos necesarios para la mejor resolución del problema y la consecuente toma de decisiones por parte del Hospital de Especialidades de las Fuerzas Armadas N°1, ya que existen varias técnicas para la

minería de datos. Se debe tener en cuenta que para algunas de estas hay que volver a la fase 3 de la metodología, puesto que será necesario recurrir a datos que en otros análisis podrían ser considerados como irrelevantes.

Fase 5. Evaluación

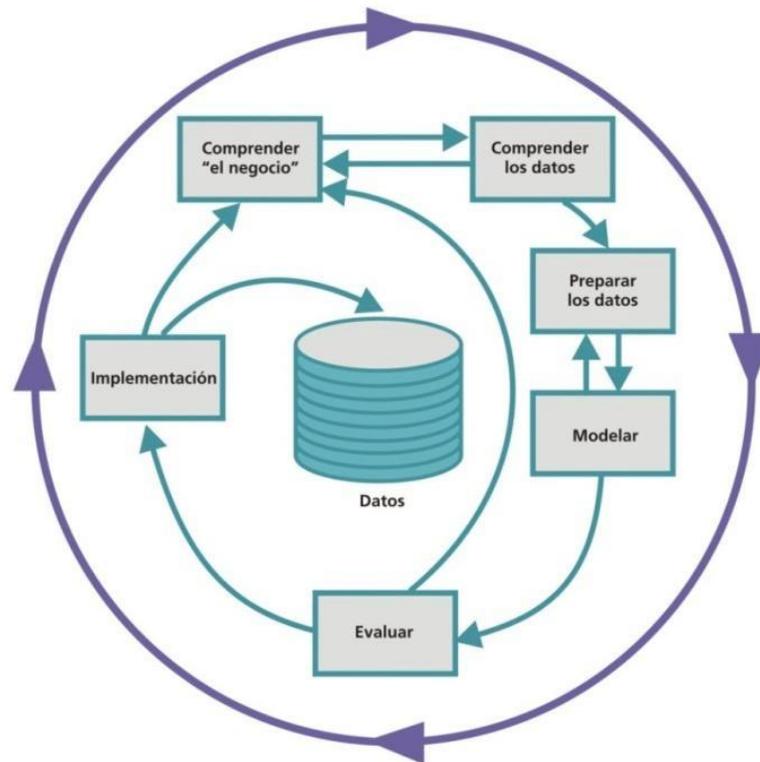
Al llegar a esta fase del proyecto ya se han construido varios modelos capaces de resolver el problema de asignación de turnos en el Hospital de Especialidades de las Fuerzas Armadas N°1, por lo que es necesario evaluar minuciosamente y verificar los pasos seguidos para llegar a cada uno de los modelos obtenidos.

Fase 6. Despliegue o Implementación

Esta fase intenta crear un nuevo conocimiento acerca de los datos de entrada que han sido analizados y sometidos a varios tipos de modelado, por lo que toda esta información tendrá que posteriormente ser evaluada por completo y durante este proceso verificar los pasos y tareas ejecutadas antes de la culminación del proyecto. De esta manera se puede generar un informe o incluso automatizar un proceso, como en el caso de este estudio.

Figura 51.

Fases de la metodología CRISP-DM



Nota: La figura representa las fases que se siguieron para la realización del proyecto en base a la metodología CRISP-DM, tomando en cuenta que es un proceso cíclico, lo que permite volver a ciertas fases de ser necesario, logrando una metodología "ágil" para minería de datos. Tomado de (Bellini Saibene, 2014)

Capítulo III

Desarrollo del modelo en base a la metodología CRISP-DM

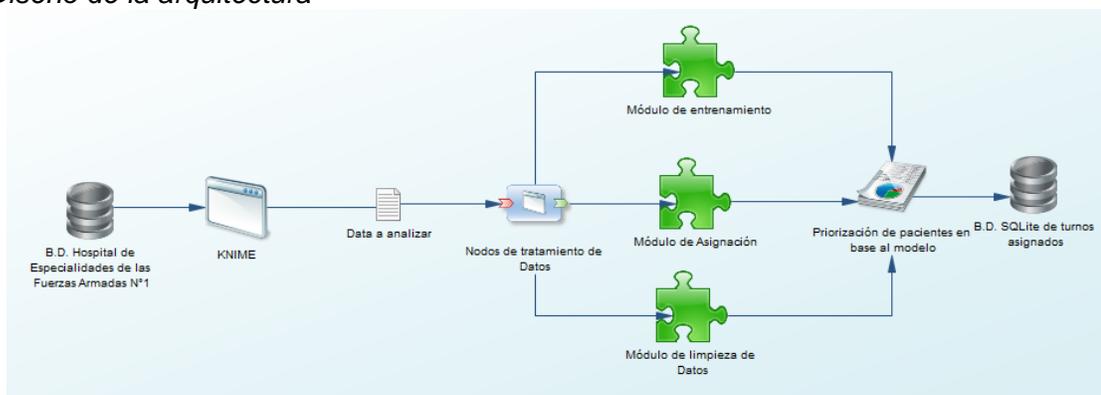
Arquitectura del modelo

El modelo propuesto para realizar la priorización de los pacientes y su calendarización se basa en una arquitectura que toma inicialmente los datos obtenidos de la base de datos del Hospital de Especialidades de las Fuerzas Armadas N°1, posteriormente ingresan a la herramienta de análisis Knime en la cual mediante un archivo plano generamos la data que será tratada aplicando la metodología CRISP-DM. Mediante la integración de los componentes y los nodos utilizados en el flujo de trabajo se desarrollan módulos previos a la obtención de un modelo funcional que permita la priorización de pacientes, estos módulos son: limpieza de datos, entrenamiento y asignación.

Para la consiguiente priorización de los pacientes y su calendarización en el Sistema de Gestión Hospitalaria, se realiza desde una base de datos SQLite que almacena los resultados obtenidos del modelo y permite la comunicación entre el sistema de optimización y el Sistema de Gestión Hospitalaria de mencionada casa de salud, como se muestra en la siguiente figura.

Figura 52.

Diseño de la arquitectura



Adquisición de datos

El primer paso para el desarrollo del modelo es netamente la adquisición de los datos que serán posteriormente analizados. Estos provienen del Sistema de Gestión Hospitalaria utilizado por el Hospital de Especialidades de las Fuerzas Armadas N°1, se almacenan en su propio Mainframe para mantener la más estricta confidencialidad de los pacientes.

Figura 53.

Obtención de datos para análisis

Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
DE ID DAT	DE ID DAT	DE ID DAT	DE ID DAT	DE ID DAT	DE ID DAT	DE ID DAT	DE ID DAT	DE ID DAT	DE ID DAT	DE ID DAT	DE ID DAT
AFILIADO_A	ESPECIALIDAD C...	NOMBRE CONSUL...	CODIGO CONSUL...	FECHA_ATENCION	ESTADO	ID_TURNO	NUMERO_HC	FECHA_NACIMIE...	DISCAPACIDAD	SEXO	PERTENENCIA_ET...
ISSFA	ODONTOLOGIA	C8-DR. UNDA PATRICIO			ATENDIDO			4/11/1974		FEMENINO	MESTIZO
ISSFA	PEDIATRIA	C1-DR. GUTIERREZ M.			ATENDIDO			8/28/2014		MASCULINO	MESTIZO
ISSFA	MEDICINA INTERNA	C5-DR. CASTELO M.			ATENDIDO			10/17/1982		FEMENINO	MESTIZO
ISSFA	PEDIATRIA	C1-DR. GUTIERREZ M.			ATENDIDO			5/6/2016		MASCULINO	MESTIZO
ISSFA	FISIATRIA	C3-DR. BONIFAZ C.			ATENDIDO			6/22/2016		MASCULINO	MESTIZO
ISSFA	MEDICINA INTERNA	C5-DR. CASTELO M.			ATENDIDO			2/23/1960		MASCULINO	MESTIZO
ISSFA	CIRUGIA GENERAL	CONSULTORIO 1			ATENDIDO			11/23/1985		MASCULINO	MESTIZO
ISSFA	GASTROENTEROLOGIA	C5-DR. CANGO L.			ATENDIDO			10/19/1999		FEMENINO	MESTIZO
ISSFA	GASTROENTEROLOGIA	C5-DR. CANGO L.			ATENDIDO			6/10/1975		MASCULINO	MESTIZO
IESS - ISSFA	GINECOLOGIA	C3C-ORA. PANCHO R.			ATENDIDO			6/24/1963		FEMENINO	MESTIZO
ISSFA	PEDIATRIA	C3-ORA. MONZAVI M.			ATENDIDO			3/16/2017		FEMENINO	MESTIZO
ISSFA	ONCOLOGIA	C3-CIRU-PAJULA CARLOS			ATENDIDO			5/15/1966		MASCULINO	MESTIZO
ISSFA	NEUROPEDIATRIA	C1-DR. GOMEZ A.			ATENDIDO			11/19/2003		MASCULINO	MESTIZO
ISSFA	TRAUMATOLOGIA	C1B-ERAZO-M-INFERI...			ATENDIDO			4/20/1959		MASCULINO	MESTIZO
ISSFA	TRAUMATOLOGIA	C1B-ERAZO-M-INFERI...			ATENDIDO			4/20/1994		MASCULINO	MESTIZO

Nota: La figura muestra los datos que han sido obtenidos para el análisis, los datos que no se muestran han sido calificados como privados por el hospital, por lo que no estarán disponibles en la figura.

Comprensión del negocio y de los Datos

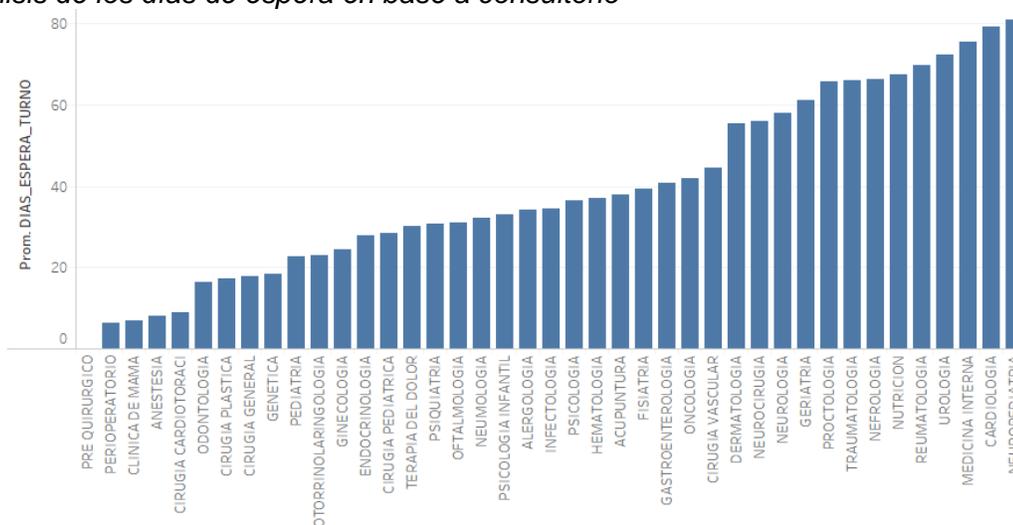
Una vez finalizada la etapa de adquisición de manera correcta, se procede a la etapa de Comprensión del Negocio, esto permite que las personas involucradas en el análisis de los datos obtenidos previamente se encuentren completamente capacitadas para la obtención de resultados útiles. También se abarca la Comprensión de los Datos, ya que estos representan por completo el giro de negocio del hospital, es decir los dos pasos son considerados como individuales, pero incluyentes dentro de esta etapa del desarrollo, logrando que el análisis previo resulte mucho más sencillo en conjunto con el uso de la herramienta Tableau.

Dentro del análisis realizado como parte de la comprensión de los datos se tomó en cuenta la relación de variables que puedan incidir directamente en el desarrollo del modelo, considerando disponibilidad, características y clases de las mismas. Se consideró la relación días de espera – consultorio, días de espera – especialidad, días de espera – tipo de afiliación, días de espera – tipo de turno / atención, tiempo de atención en turnos de primera vez y número de pacientes atendidos por afiliación.

En las siguientes figuras se muestra la representación gráfica del análisis realizado utilizando la herramienta Tableau.

Figura 54.

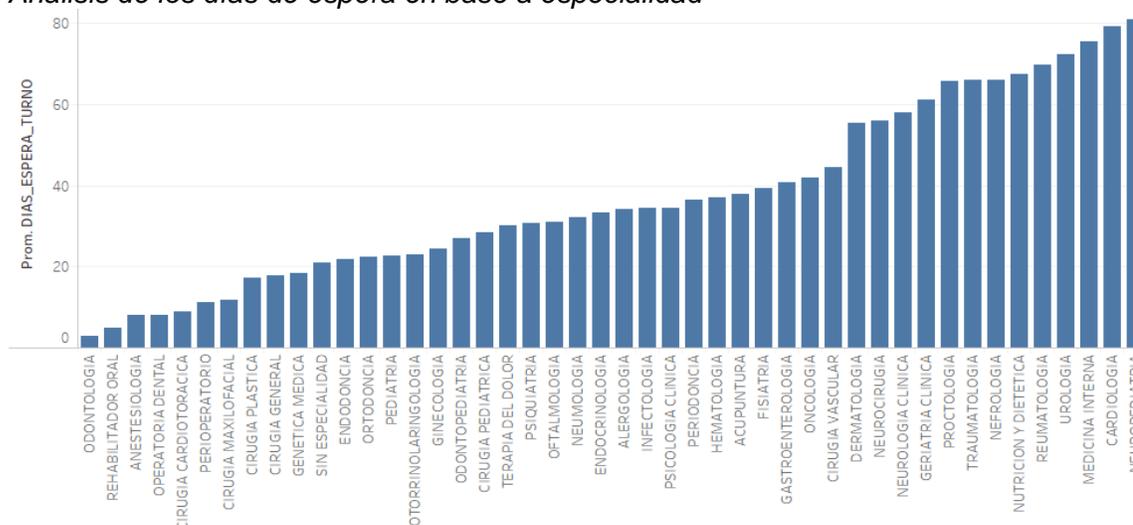
Análisis de los días de espera en base a consultorio



Nota: La figura representa el análisis de los días de espera en relación a cada uno de los consultorios. Se deduce que el menor tiempo de espera sin considerar el Pre Quirúrgico es el consultorio de Perioperatorio con 6 días, mientras que el mayor tiempo se establece en el consultorio de Neuropediatria con 81 días.

Figura 55.

Análisis de los días de espera en base a especialidad



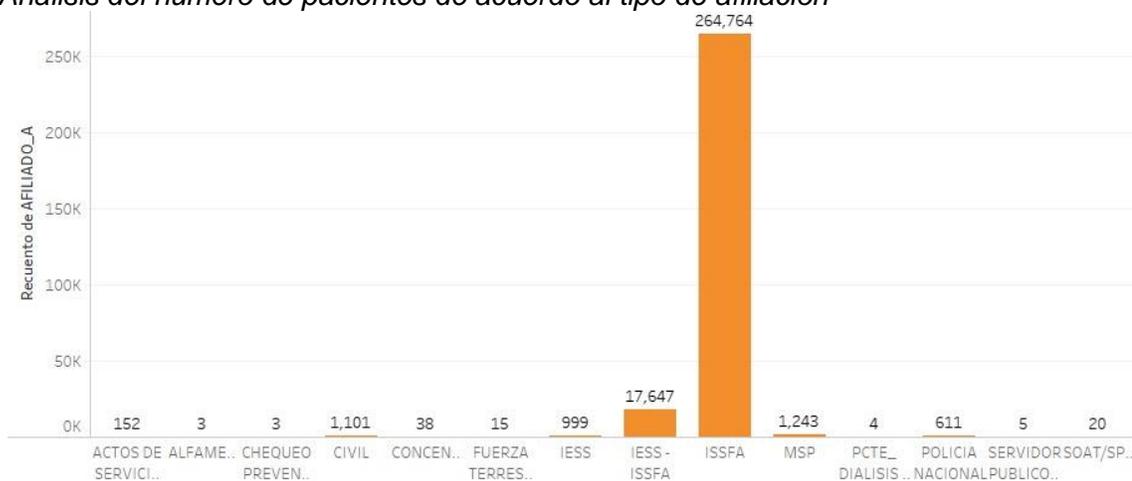
Nota: La figura representa el análisis de los días de espera en relación a la especialidad. Se deduce que el menor tiempo de espera es en la especialidad de Odontología con 3

días, mientras que el mayor tiempo se establece en el consultorio de Neuropediatría con 81 días.

Es importante analizar la cantidad de pacientes que acuden al hospital de acuerdo a su tipo de afiliación ya que permite interpretar que la mayor cantidad de pacientes que reciben atención pertenecen al Instituto de Seguridad Social de Fuerzas Armadas (ISSFA), lo cual influye posteriormente cuando se realiza el análisis de los días de espera en relación al tipo de afiliación ya que se observa que el mayor tiempo es para pacientes ISSFA, sin embargo esto se da por la demanda de este tipo de afiliados mas no por el tiempo real que deben esperar para obtener su turno de primera vez o subsecuente. Este análisis se observa en la figura 56 y figura 57 respectivamente.

Figura 56.

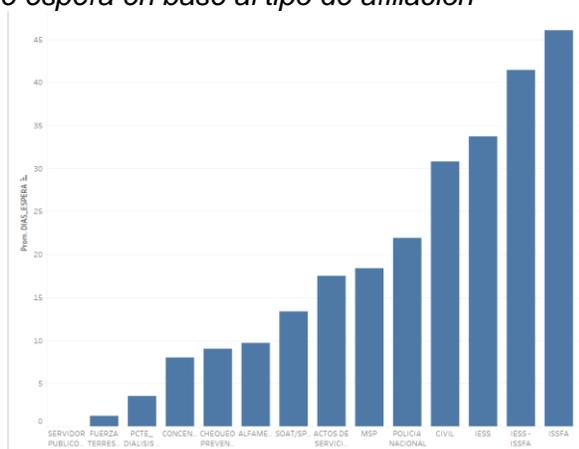
Análisis del número de pacientes de acuerdo al tipo de afiliación



Nota: La figura representa el número de pacientes de acuerdo al tipo de afiliación, en el cual la mayor cantidad pertenece al ISSFA.

Figura 57.

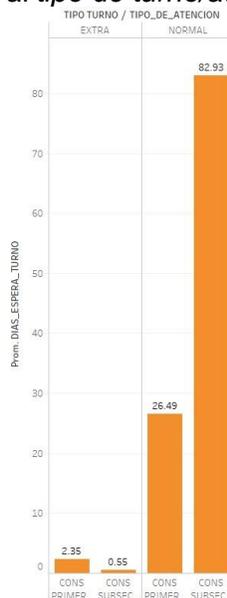
Análisis de los días de espera en base al tipo de afiliación



Nota: La figura representa el análisis de los días de espera en relación al tipo de afiliación. Se observa que la mayor cantidad de días de espera es para afiliación tipo ISSFA, lo cual se da por la mayor demanda de pacientes.

Figura 58.

Análisis de los días de espera en base al tipo de turno/atención

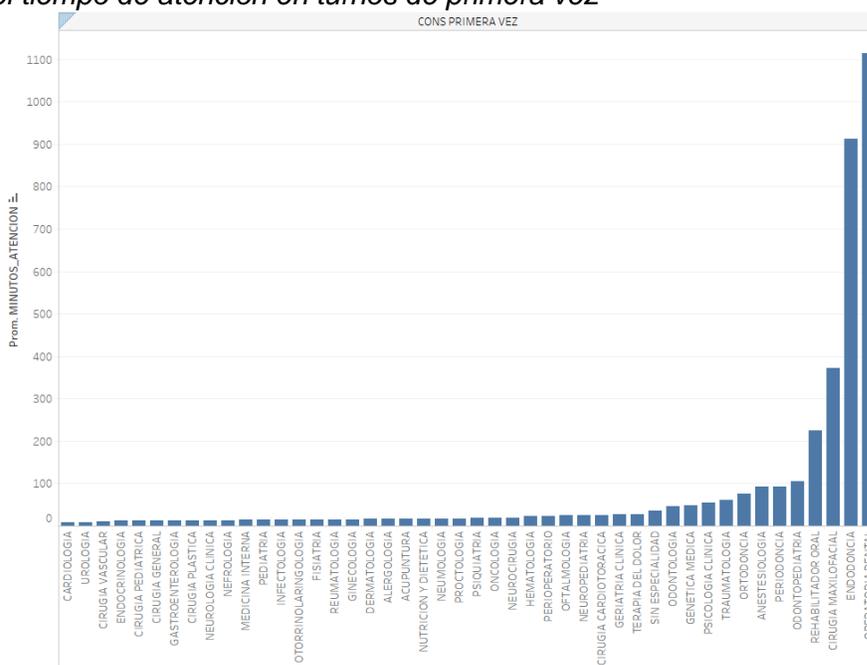


Nota: La figura representa el análisis de los días de espera en relación al tipo de turno o atención. Se maneja dos tipos de turnos: extra y normal, y en cada uno atención de

primera vez y subsecuente. Se observa que en turnos extras el tiempo de espera es mínimo por el limitante en la asignación de este tipo de turnos que lo realiza cada médica, mientras que el mayor tiempo de espera se da en los turnos normales con un promedio de 26 días para atención primera vez y 83 días para subsecuentes.

Figura 59.

Análisis del tiempo de atención en turnos de primera vez



Nota: La figura representa el análisis del tiempo de atención (minutos) en turnos de primera vez. Es importante observar que en ciertas especialidades reflejan tiempos excesivos de atención, lo cual se da porque cada médico debe registrar el momento en que inicia y termina su atención, sin embargo existen consultorios que no registran la finalización. Este un parámetro que se debe tomar en cuenta al momento de realizar la limpieza de la data.

Análisis de los datos

Como parte de las actividades de análisis se establece la muestra de datos con la que se va a trabajar, además se realiza el tratamiento y limpieza de los mismos.

Los insumos considerados para el análisis de los datos son: los turnos agendados en consulta externa durante el año 2018 (286.605 registros), la calendarización que refleja las fechas en la que se otorgó cada turno, los análisis realizados en Tableau, las excepciones en el horario de cada médico, y las fechas libres en cada consultorio (vacaciones, salidas a capacitación, feriados).

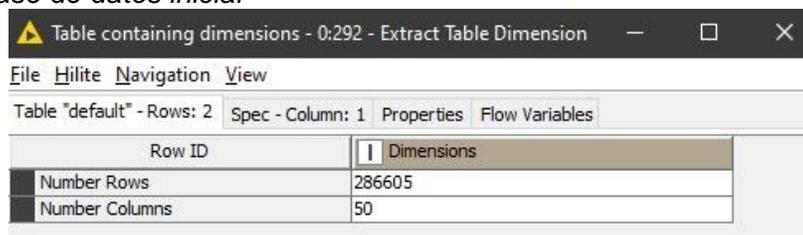
Toma de la muestra

Se considera como muestra a la parte del universo en la cual se realizará la investigación, para el efecto se aplican métodos que permiten seleccionar esta muestra en base a reglas, criterios y procedimientos, a éste procedimiento se conoce como muestreo.

La base de datos inicial en la que constan todas las atenciones realizadas en consulta externa durante el año 2018 cuenta con 286.605 registros, es decir el universo, que representa una cantidad elevada para realizar el análisis, incurriendo en mayor tiempo de realización del estudio, mayor gasto de recursos, menor profundidad y control al momento del análisis de las variables.

Figura 60.

Tamaño de la base de datos inicial



The image shows a screenshot of a Tableau window titled "Table containing dimensions - 0:292 - Extract Table Dimension". The window displays a table with the following data:

Row ID	Dimensions
Number Rows	286605
Number Columns	50

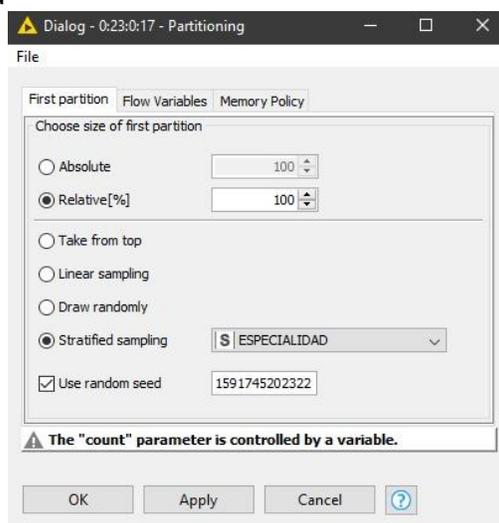
Nota: La figura representa la cantidad de registros y variables que se tiene en el universo inicial.

De acuerdo a las características de la población inicial que presentan variabilidad o distribución conocida que es necesaria al momento de extraer la muestra se opta por un tipo de muestreo que permita dividir el universo en distintos estratos para permitir su estudio. Este procedimiento se conoce como muestreo estratificado el cual permite respetar en la muestra la distribución inicial de la población, es decir se mantiene las características del universo con menos cantidad de registros, reduciendo de esta manera posibles desbalances (Lopez, 2004).

Una vez identificada y definida la población se realiza el cálculo de la muestra en Knime utilizando el nodo Partitioning que permite ingresar el porcentaje relativo de la data que queremos representar, en este caso se trabaja con el 100%, además se escoge la opción “stratified sampling” que representa el muestreo estratificado como se identifica en la siguiente figura.

Figura 61.

Estratificación de la muestra

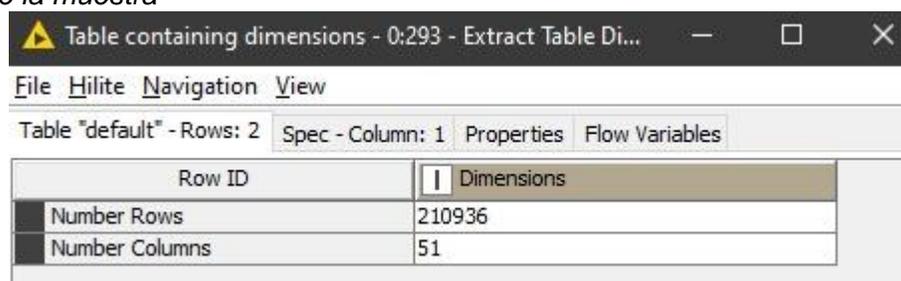


Nota: La figura representa el porcentaje relativo de la data que se utiliza a partir del muestreo estratificado.

Después de obtener la nueva muestra se cuenta con 210.936 registros, los cuales ingresarán al tratamiento de los datos. La base de datos actual cuenta con las mismas características y variables de la base inicial, manteniendo de esta manera la integridad de los datos.

Figura 62.

Resultado de la muestra



Row ID	Dimensions
Number Rows	210936
Number Columns	51

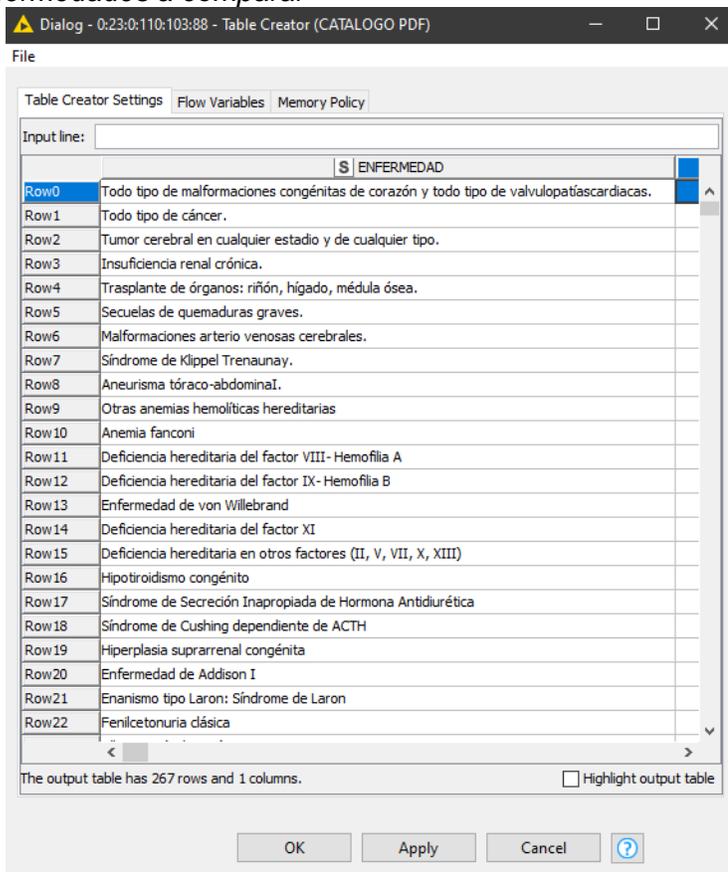
Nota: La figura muestra la cantidad de registros y variables obtenidas después de aplicar el muestreo estratificado. Referente al número de columnas aumenta en uno ya que se considera como variable adicional las enfermedades catastróficas que no estaban en la base de datos inicial.

Clasificación de enfermedades catastróficas

Después de haber obtenido la muestra estratificada, se procede al análisis de las enfermedades presentes en la base del documento CIE-10 (Organización Panamericana de la Salud, 1995). Utilizando el nodo Table Creator se crea un catálogo de las enfermedades catastróficas.

Figura 63.

Catálogo de enfermedades a comparar



Se manipula las cadenas de caracteres obtenidas de la tabla creada anteriormente y se transforma cada fila en una variable dentro de un bucle que permita abarcar todas las filas, permitiendo realizar futuras operaciones. Obtenida la información se procede al cálculo de la Distancia de Levenshtein entre la variable recuperada del catálogo y los datos del hospital utilizando el propio algoritmo recursivamente. De esta manera se automatiza la comparación entre la tabla que se tenía inicialmente con los registros del hospital y la tabla generada a partir del catálogo CIE-10.

Figura 64.

Valor de la distancia de Levenshtein de las enfermedades

Row ID	S ENFERMEDAD	S CODIG...	I nearest...	S nearest...	D distance
Row0_1	(OSTEO)ARTROSIS PRIMARIA GENERALIZADA	M15.0	0	Row266	31
Row1_1	ABDOMEN AGUDO	R10.0	0	Row266	12
Row2_1	ABORTADORA HABITUAL	N96	0	Row266	17
Row3_1	ABORTO ESPONTANEO	O03	0	Row266	14
Row4_1	ABORTO ESPONTANEO: COMPLETO O NO ES...	O03.5	0	Row266	82
Row5_1	ABORTO ESPONTANEO: COMPLETO O NO ES...	O03.9	0	Row266	55
Row6_1	ABORTO ESPONTANEO: INCOMPLETO, COM...	O03.1	0	Row266	65
Row7_1	ABORTO ESPONTANEO: INCOMPLETO, SIN ...	O03.4	0	Row266	39
Row8_1	ABORTO NO ESPECIFICADO	O06	0	Row266	18
Row9_1	ABORTO NO ESPECIFICADO: COMPLETO O ...	O06.6	0	Row266	86
Row10_1	ABORTO NO ESPECIFICADO: COMPLETO O ...	O06.9	0	Row266	59
Row11_1	ABORTO NO ESPECIFICADO: INCOMPLETO, ...	O06.0	0	Row266	71
Row12_1	ABORTO NO ESPECIFICADO: INCOMPLETO, ...	O06.4	0	Row266	44
Row13_1	ABORTO RETENIDO	O02.1	0	Row266	13
Row14_1	ABSCESO AMEBIANO DEL HIGADO	A06.4	0	Row266	22
Row15_1	ABSCESO ANAL	K61.0	0	Row266	10
Row16_1	ABSCESO ANORRECTAL	K61.2	0	Row266	16
Row17_1	ABSCESO CUTANEO, FURUNCULO Y ANTRAX	L02	0	Row266	32
Row18_1	ABSCESO CUTANEO, FURUNCULO Y ANTRAX...	L02.3	0	Row266	43
Row19_1	ABSCESO CUTANEO, FURUNCULO Y ANTRAX...	L02.0	0	Row266	42
Row20_1	ABSCESO CUTANEO, FURUNCULO Y ANTRAX...	L02.1	0	Row266	45
Row21_1	ABSCESO CUTANEO, FURUNCULO Y ANTRAX...	L02.4	0	Row266	42
Row22_1	ABSCESO CUTANEO, FURUNCULO Y ANTRAX...	L02.8	0	Row266	46
Row23_1	ABSCESO CUTANEO, FURUNCULO Y ANTRAX...	L02.9	0	Row266	54
Row24_1	ABSCESO CUTANEO, FURUNCULO Y ANTRAX...	L02.2	0	Row266	43
Row25_1	ABSCESO DE LA GLANDULA DE BARTHOLIN	N75.1	0	Row266	31
Row26_1	ABSCESO DE LAS REGIONES ANAL Y RECTAL	K61	0	Row266	32
Row27_1	ABSCESO DEL HIGADO	K75.0	0	Row266	14
Row28_1	ABSCESO DEL OIDO EXTERNO	H60.0	0	Row266	21
Row29_1	ABSCESO EXTRADURAL Y SUBDURAL, NO ES...	G06.2	0	Row266	40
Row30_1	ABSCESO PERIAMIGDALINO	J36	0	Row266	18
Row31_1	ABSCESO PERIAPICAL CON FISTULA	K04.6	0	Row266	24
Row32_1	ABSCESO PERIAPICAL SIN FISTULA	K04.7	0	Row266	24
Row33_1	ABSCESO RECTAL	K61.1	0	Row266	12
Row34_1	ABSCESO RENAL Y PERIRRENAL	N15.1	0	Row266	21
Row35_1	ABSCESO RETROFARINGEO Y PARAFARINGEO	J39.0	0	Row266	31
Row36_1	ABSCESO VULVAR	N76.4	0	Row266	12
Row37_1	ABSCESO Y GRANULOMA INTRACRANEAL	G06.0	0	Row266	28

El momento que se conoce a que enfermedades se refieren los códigos presentados en la base de datos del hospital se procede a utilizar un motor de reglas para clasificar si las enfermedades en cuestión son catalogadas como catastróficas.

Figura 65.

Motor de reglas crea columna de enfermedades catastróficas

```

Expression
? 1 // enter ordered set of rules, e.g.:
? 2 // $double column name$ > 5.0 => "large"
? 3 // $string column name$ LIKE "*blue*" => "small and blue"
? 4 // TRUE => "default outcome"
S 5 $CATASTROFICA$ = "S" => "S"
S 6 NOT $CATASTROFICA$ = "S" => "N"

```

Una vez cumplidos estos pasos se realiza la limpieza superficial de los datos que presenta la nueva tabla de valores con la columna que clasifica si las enfermedades son catastróficas.

Figura 66.

Clasificación de enfermedades catastróficas y no catastróficas

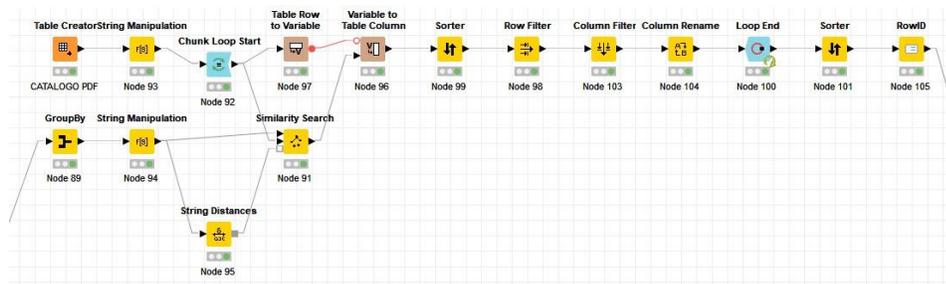
The screenshot shows a window titled 'Outport 1 - 0:23 - TAMAÑO DE LA MUESTRA'. The main area displays a table with the following columns: Row ID, NUMER..., CREADO POR, TELEFO..., FECHA_PRO..., and CATAS... The table contains 36 rows of data, each representing a patient record. The 'CATAS...' column contains either 'S' (indicating a catastrophic disease) or 'N' (indicating a non-catastrophic disease). A red box highlights the 'CATAS...' column.

Row ID	NUMER...	CREADO POR	TELEFO...	FECHA_PRO...	CATAS...
Row2_Row195		SALAZAR MALDONADO BYRON EDUARDO	?	31/12/2018 1:28:55	N
Row115_Row...		PAULA AVILES CARLOS EDUARDO	?	31/12/2018 1:29:02	S
Row116_Row...		RUIZ LLERENA LUIS ARNULFO	?	31/12/2018 1:29:02	S
Row156_Row...		LIZARZABURU BUNGER DIEGO JAVIER	?	31/12/2018 1:29:02	S
Row253_Row...		SALAZAR MALDONADO BYRON EDUARDO	?	31/12/2018 1:28:55	S
Row258_Row54		ARMAS CORRALES CARLOS XAVIER	0988438974	31/12/2018 1:28:55	S
Row258_Row62		ARMAS CORRALES CARLOS XAVIER	0988438974	31/12/2018 1:28:55	N
Row259_Row54		GOMEZ GARCIA ARIEL OSMANI	?	31/12/2018 1:28:55	S
Row259_Row62		GOMEZ GARCIA ARIEL OSMANI	?	31/12/2018 1:28:55	N
Row260_Row54		CONSTANTE COBO ANA MARIA	?	31/12/2018 1:28:55	S
Row260_Row62		CONSTANTE COBO ANA MARIA	?	31/12/2018 1:28:55	N
Row268_Row88		MELLENDEZ VISCARRA PAULINA PAMELA	?	31/12/2018 1:28:55	N
Row278_Row34		PAIS CEDEÑO PEDRO CRISTIAN	0997691949	31/12/2018 1:28:55	S
Row281_Row54		GOMEZ GARCIA ARIEL OSMANI	0986262642	31/12/2018 1:28:55	S
Row281_Row62		GOMEZ GARCIA ARIEL OSMANI	0986262642	31/12/2018 1:28:55	N
Row282_Row54		GOMEZ GARCIA ARIEL OSMANI	?	31/12/2018 1:28:55	S
Row282_Row62		GOMEZ GARCIA ARIEL OSMANI	?	31/12/2018 1:28:55	N
Row283_Row54		NARVAEZ CASTILLO BEATRIZ AUGUSTA	?	31/12/2018 1:28:55	S
Row283_Row62		NARVAEZ CASTILLO BEATRIZ AUGUSTA	?	31/12/2018 1:28:55	N
Row284_Row54		GOMEZ GARCIA ARIEL OSMANI	?	31/12/2018 1:28:55	S
Row284_Row62		GOMEZ GARCIA ARIEL OSMANI	?	31/12/2018 1:28:55	N
Row285_Row54		GARCIA CAMPANA CARLOS EDUARDO	?	31/12/2018 1:28:55	S
Row285_Row62		GARCIA CAMPANA CARLOS EDUARDO	?	31/12/2018 1:28:55	N
Row286_Row54		VILLA INFANTE FERNANDO PATRICIO	0979372829	31/12/2018 1:28:55	S
Row286_Row62		VILLA INFANTE FERNANDO PATRICIO	0979372829	31/12/2018 1:28:55	N
Row287_Row54		VELARDE LASSO AUGUSTO ENRIQUE	0983214696	31/12/2018 1:28:55	S
Row287_Row62		VELARDE LASSO AUGUSTO ENRIQUE	0983214696	31/12/2018 1:28:55	N
Row288_Row54		VELARDE LASSO AUGUSTO ENRIQUE	0959030059	31/12/2018 1:28:55	S
Row288_Row62		VELARDE LASSO AUGUSTO ENRIQUE	0959030059	31/12/2018 1:28:55	N
Row289_Row54		SALAZAR MALDONADO BYRON EDUARDO	?	31/12/2018 1:28:55	S
Row289_Row62		SALAZAR MALDONADO BYRON EDUARDO	?	31/12/2018 1:28:55	N
Row291_Row54		VELARDE LASSO AUGUSTO ENRIQUE	?	31/12/2018 1:28:55	S
Row291_Row62		VELARDE LASSO AUGUSTO ENRIQUE	?	31/12/2018 1:28:55	N
Row323_Row54		VELARDE LASSO AUGUSTO ENRIQUE	0999078713	31/12/2018 1:28:55	S
Row323_Row62		VELARDE LASSO AUGUSTO ENRIQUE	0999078713	31/12/2018 1:28:55	N
Row324_Row54		NARVAEZ CASTILLO BEATRIZ AUGUSTA	0994754394	31/12/2018 1:28:55	S
Row324_Row62		NARVAEZ CASTILLO BEATRIZ AUGUSTA	0994754394	31/12/2018 1:28:55	N

Nota: La figura representa la asignación de “S” cuando el paciente presenta enfermedades catastróficas y “N” cuando no.

Figura 67.

Proceso de clasificación de enfermedades catastróficas



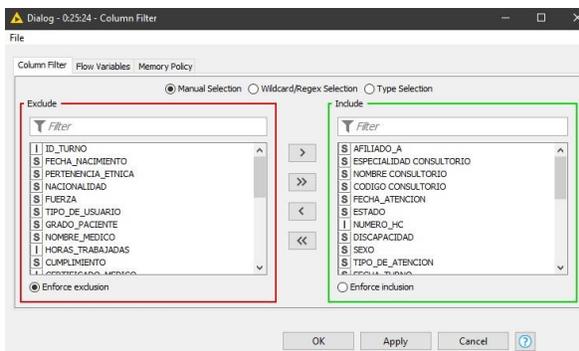
Nota: La figura muestra los nodos y el flujo de trabajo que han sido utilizados para la clasificación de enfermedades catastróficas y no catastróficas a partir del catálogo de enfermedades CIE-10.

Selección de variables

A continuación se procede a la selección de variables que resultan útiles para el análisis. En cuestión de los registros disponibles se trabaja con todos puesto que ya son parte de la muestra seleccionada, sin embargo, existen campos innecesarios que no son tomados en cuenta por ser considerados irrelevantes para el objetivo de minería de datos, de esta manera se reduce de 51 a 26 variables.

Figura 68.

Filtrado de variables



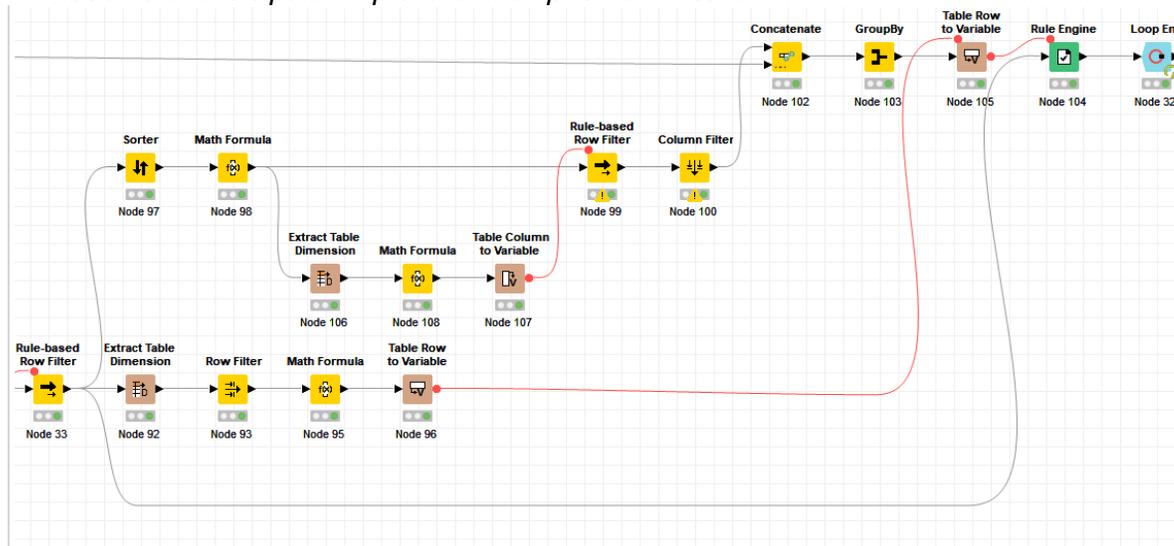
Nota: La figura muestra la exclusión de algunos campos ya que son irrelevantes para el análisis de acuerdo a lo definido en la fase de comprensión del negocio.

Limpieza de la data

Cuando todas las variables relevantes han sido seleccionadas se estandariza los datos. Este paso consiste en corregir ciertos errores que son causados en gran parte por índole humana, el más común se refiere a médicos que no realizan el cierre de sesión al terminar las consultas o tratamientos de los pacientes, ocasionando que los datos que se obtienen del tiempo de atención sean inverosímiles y nada útiles al momento su análisis.

Figura 69.

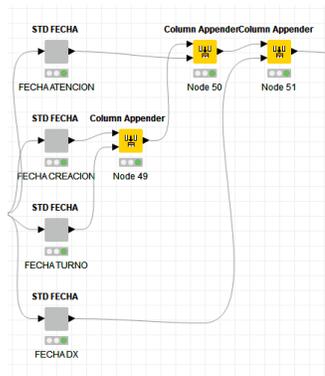
Proceso de análisis para limpieza en tiempos de consulta



Además, se realiza la estandarización de las fechas, ya que en ciertos módulos del Sistema de Gestión Hospitalaria esta se guarda en la base sin el tiempo (horas, minutos, segundos) que conlleva a la inutilización de los datos. Para corregir este error se toma cada especialidad en un bucle, determinando el número de datos con los que se va a trabajar para obtener el tiempo que tarda la atención por especialidad y así encontrar y tratar las anomalías que se generan por no cerrar sesión en el sistema del hospital. Luego se obtiene el promedio de las atenciones por especialidad, dato que se usa para evitar valores exagerados en el campo de atención.

Figura 70.

Proceso de estandarización de fechas



Nota: La figura indica las fechas que se dispone en la muestra del modelo, las cuales son tratadas para evitar datos erróneos posteriores.

Análisis clúster

Se realiza este análisis para clasificar los individuos en grupos homogéneos. El fin es determinar qué conjunto de variables que podemos obtener proveniente de los datos del hospital, por lo que todos los individuos que presenten características similares entre si serán agrupados dentro de un mismo clúster. El agrupamiento se realiza utilizando la técnica conocida como k-medias que hace posible la medición de las distancias entre los datos seleccionados, logrando que aquellos con particularidades similares sean agrupados. Finalmente se utiliza un nodo que pinta los registros de un mismo clúster permitiendo que sean mucho más entendibles por el usuario final.

Figura 71.

Proceso de agrupamiento de datos

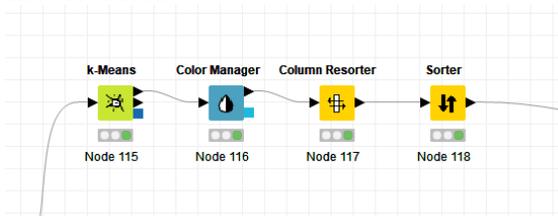


Figura 72.

Muestra de clúster clasificado por colores

Sorted Table - 0:265:118 - Sorter

File Hilite Navigation View

Table "default" - Rows: 154 Spec - Columns: 8 Properties Flow Variables

Row ID	S Cluster	S 1VEZ_E...	S CODIG...	S 1VEZ_NO...	I 1VEZ_T...	I SUBS_T...	I %1VEZ	I %SUBS
Row11_Row10	duster_0	CARDIOLOGIA	M309	C1-DR. LASSO J.	273	1330	17	83
Row72_Row73	duster_1	CARDIOLOGIA	M314	C2-DRA. SEGA...	719	3404	17	83
Row21_Row20	duster_1	NUTRICION ...	M381	C1-DRA. ALVA...	923	2935	24	76
Row23_Row22	duster_10	DERMATOLO...	M353	C1-DRA. BAYA...	1291	1150	53	47
Row107_Row...	duster_10	ENDOCRINO...	M321	C4-DR. VARGA...	1355	1156	54	46
Row157_Row...	duster_11	ENDOCRINO...	M320	CONSULTORIO 3	149	32	82	18
Row3_Row2	duster_11	PERIOPERAT...	MI79	C1-DR. PIEDR...	203	47	81	19
Row122_Row...	duster_12	ENDODONCIA	M652	C4B-DR. ORDO...	192	492	28	72
Row83_Row84	duster_13	FISIATRIA	M380	C3-DR. BONIF...	751	627	54	46
Row103_Row...	duster_14	FISIATRIA	M913	C4-DR. FABAR...	1435	1063	57	43
Row28_Row27	duster_15	HEMATOLOGIA	M376	C1-DRA. LEIVA...	233	848	22	78
Row81_Row82	duster_15	ONCOLOGIA	M372	C3-CIRU-PAUL...	276	815	25	75
Row44_Row44	duster_16	GASTROENT...	M537	C1B-DRA. SAM...	424	1039	29	71
Row78_Row79	duster_16	GASTROENT...	M538	C2B-DRA. ORD...	357	1029	26	74
Row1_Row0	duster_16	TRAUMATOL...	M520	C1-BALAREZO-...	439	1047	30	70
Row144_Row...	duster_17	CARDIOLOGIA	M836	C8-DR. IZAGUI...	567	673	46	54
Row96_Row98	duster_17	GASTROENT...	M349	C3B-DR. PAZM...	509	740	41	59
Row156_Row...	duster_17	UROLOGIA	M312	CONSULTORIO 3	502	786	39	61
Row158_Row...	duster_17	UROLOGIA	M313	CONSULTORIO 4	520	783	40	60
Row126_Row...	duster_18	GASTROENT...	MD04	C5-DR. CANGO L.	202	1297	13	87
Row104_Row...	duster_2	CARDIOLOGIA	M316	C4-DR. LOPEZ J.	336	1478	19	81
Row108_Row...	duster_20	OFTALMOLO...	M334	C4-DR. VITERI M.	551	293	65	35
Row19_Row18	duster_21	GINECOLOGIA	M403	C1-DR. YANEZ P.	551	823	40	60
Row46_Row46	duster_22	GINECOLOGIA	M523	C1B-PROCEDI...	210	564	27	73
Row60_Row61	duster_23	GINECOLOGIA	M404	C2-DR. PROAÑ...	662	1329	33	67
Row101_Row...	duster_23	MEDICINA I...	M993	C4-DR. DAVILA L.	535	1351	28	72
Row74_Row75	duster_24	GINECOLOGIA	M510	C2B-DR. ESTR...	796	1468	35	65
Row80_Row81	duster_24	GINECOLOGIA	M541	C2D-DRA NAV...	795	1377	37	63
Row31_Row30	duster_24	NEUROLOGI...	M367	C1-DRA. NARV...	872	1461	37	63
Row6_Row5	duster_24	NEUROPEDI...	M392	C1-DR. GOMEZ...	903	1555	37	63
Row95_Row97	duster_25	GINECOLOGIA	M542	C3B BRIONES ...	1	4	20	80
Row111_Row...	duster_25	ODONTOLOGIA	M645	C4-DRA. ROME...	1	30	3	97
Row5_Row4	duster_26	ODONTOPE...	M649	C1-DR. CARRI...	275	1020	21	79
Row87_Row89	duster_27	CIRUGIA PL...	M325	C3-DR. RAMOS...	143	722	17	83
Row110_Row...	duster_27	CIRUGIA PL...	M326	C4-DRA. PRAD...	141	726	16	84
Row47_Row47	duster_27	HEMATOLOGIA	M725	C1C-DRA. ROD...	182	685	21	79
Row115_Row...	duster_28	ENDODONCIA	M654	C4A ENDODON...	134	387	26	74
Row25_Row24	duster_28	INFECTOLOGIA	M377	C1-DRA. CELI A.	98	356	22	78

Asignación de pesos a variables de análisis

En este paso se utiliza principalmente fórmulas matemáticas de suma, resta y multiplicación en conjunto con nodos de motor de regla, para usar los pesos que han sido asignados previamente a cada una de las variables que se analizarán.

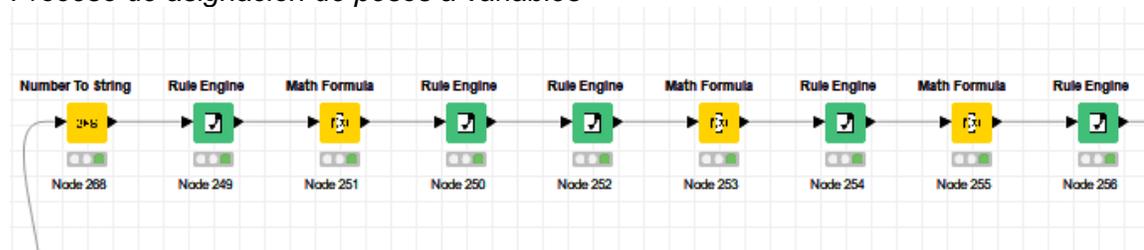
El aprendizaje de la red consiste en modificar el valor de los pesos determinados en respuesta al grupo de entrenamiento, consiguiendo que aprenda a solucionar un

problema. Los criterios que se utiliza para modificar estos parámetros son denominados reglas de aprendizaje, que consisten en algoritmos matemáticos que ajusta el supervisor para conseguir una salida que se aproxime a la deseada. Una vez que se obtiene los pesos óptimos en la fase de entrenamiento se mide la eficacia de la red, durante este proceso no se realizan más modificaciones (Palmer Pol & Montaña Moreno, 1999).

De esta manera la asignación de los pesos en las variables que se determinaron como relevantes para el proceso de clasificación de los pacientes, se realizó considerando el análisis de los datos estadísticos a través de la herramienta Tableau, y mediante una reunión con el responsable de la administración del Sistema de Gestión Hospitalaria que en este caso es el experto en la administración de la base del hospital.

Figura 73.

Proceso de asignación de pesos a variables



Nota: La figura muestra los nodos y el flujo de trabajo que han sido utilizados para el proceso de asignación de pesos a variables, mientras que los pesos en cuestión deben permanecer como privados por políticas propias del hospital, las cuales salvaguardan los datos de los pacientes.

Completada la asignación de los pesos, se procede al ordenamiento de los datos a través de un proceso conocido como “data tidying” que permite modelar, manipular y visualizar los datos de manera más sencilla. Además se reestructura la tabla con un nodo One To Many que convierte las variables en columnas y les asigna valores de 0 o 1, interpretando 0 como falso y 1 como verdadero.

Figura 74.

Conversión de filas en columnas

Row ID	ISSFA	IESS - I...	ACTOS ...	POLIC...	MSP	CIVIL	IESS	SOAT/S...	FUERZ...	CHEQU...	SERVID...	CONCE...	PCTE...	ALFAM...	NINGUNA
Row0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row7	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row8	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row11	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row13	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row17	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row18	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row20	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Row21	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Nota: La figura muestra el resultado de la transformación de filas en columnas para el posterior análisis de componentes principales.

Modelado

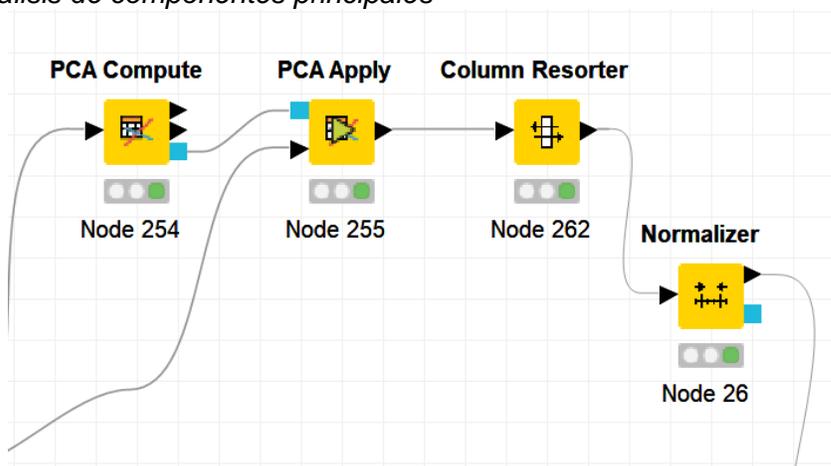
Dentro de esta fase se realiza la selección de las técnicas de inteligencia artificial más adecuadas para este proyecto. Se debe considerar criterios en base a si es apropiada para la solución del problema, si cumple con los requisitos planteados, se dispone de los datos y tiempo necesarios, y finalmente si se tiene conocimiento de las técnicas.

Análisis de componentes principales (PCA)

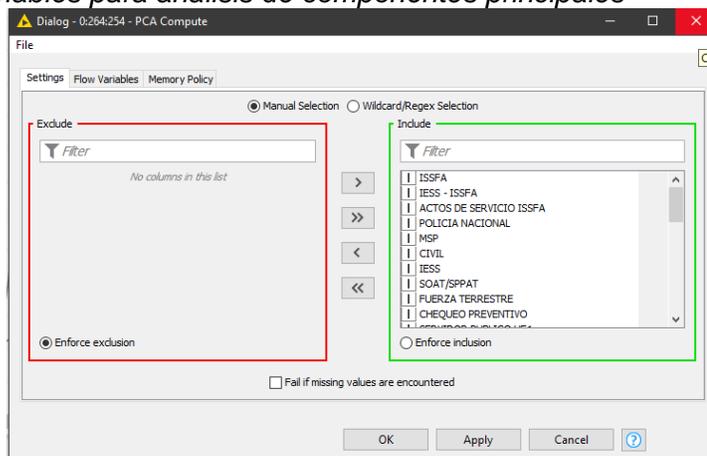
Para iniciar con la etapa de análisis de componentes principales se debe seleccionar las variables a utilizar, en este caso todas las generadas en la etapa anterior ya que el tratamiento y la limpieza fueron previamente efectuados.

Figura 75.

Flujo de análisis de componentes principales

**Figura 76.**

Selección de variables para análisis de componentes principales



Posteriormente se determina la fracción mínima de información, la cual representa la cantidad mínima de datos que serán representados dentro de las dimensiones obtenidas por el nodo, el análisis fue realizado en base al estudio llamado “A Combination of Multiple Imputation and Principal Components” (Novita Anindita, 2017). En función al porcentaje que se detalla en dicho estudio se obtienen 5

dimensiones que representan de manera correcta a todo el conjunto de datos de ésta investigación, consiguiendo que el desarrollo y entrenamiento del modelo sea mucho más eficiente.

Figura 77.

Fraccionamiento y representación de la información

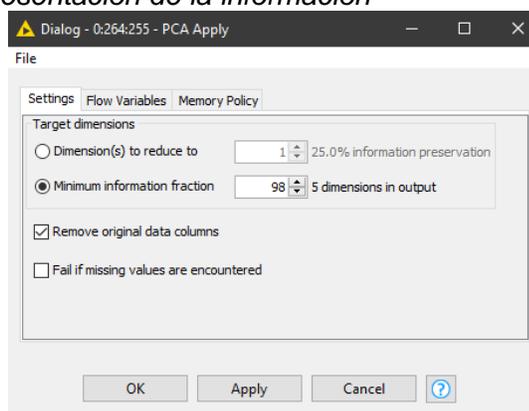


Figura 78.

Dimensiones generadas por el análisis de componentes principales

D	PCA di...	D	PCA di...	D	PCA di...	D	PCA di...	D	PCA di...
0.532	0.811	-0.925	-0.017	0.116	0.721	0.986	0.315	0.6	0.045
0.546	0.719	-0.977	-0.043	-1.292	0.532	0.811	-0.925	-0.017	0.116
0.641	-0.586	0.01	-0.022	-1.209	0.641	-0.586	0.01	-0.022	-1.209
0.627	-0.494	0.062	0.004	0.198	0.641	0.814	-0.15	-0.107	-1.294
0.627	0.906	-0.098	-0.081	0.114	0.627	0.906	-0.098	-0.081	0.114
0.68	-0.507	0.392	-0.722	0.164	0.68	-0.507	0.392	-0.722	0.164
0.735	0.895	0.263	0.573	-1.362	0.735	0.895	0.263	0.573	-1.362
0.721	0.986	0.315	0.6	0.045	0.721	0.986	0.315	0.6	0.045
0.721	0.986	0.315	0.6	0.045	0.721	0.986	0.315	0.6	0.045
0.679	0.893	0.232	-0.808	0.08	0.679	0.893	0.232	-0.808	0.08
0.679	0.893	0.232	-0.808	0.08	0.679	0.893	0.232	-0.808	0.08
0.679	0.893	0.232	-0.808	0.08	0.679	0.893	0.232	-0.808	0.08

Normalización permite transformar todas las variables que se encuentran en distintos rangos a rangos establecidos por el usuario en este caso de -2 a 2. Estandariza la información en rangos similares, hace que la tasa sea más certera.

Clasificación de pacientes

La clasificación de pacientes se realizó en base a los pesos de cada una de las variables que componen el total de sus datos, por lo que la mayor prioridad la tendrán

pacientes con enfermedades crónicas o que necesiten tratamientos urgentes, mientras que los pacientes que tengan cuadros de diagnóstico no crónicos o sean revisiones rutinarias tendrán una menor prioridad. De esta manera quienes realmente necesiten un turno serán los primeros en ser considerados sin dejar de lado a los pacientes de menor prioridad, esto permite que el entrenamiento de la inteligencia artificial sea más fiable y no exista intervención humana que pueda retrasar las citas médicas de los pacientes que más lo necesiten.

Figura 79.

Priorización de pacientes en base a su clasificación

TIPO_DE_AT...	FECHA_CREA...	FECHA_ATEN...	CLASS
CONS PRIMERA VEZ	2018-09-21T11:51...	2018-12-05T08:59...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-09-11T12:48...	2018-11-15T10:02...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-08-14T08:14...	2018-10-11T09:28...	PRIORIDAD-2
CONS PRIMERA VEZ	2018-08-14T08:14...	2018-10-11T09:28...	PRIORIDAD-3
CONS SUBSECUENTE	2018-07-31T09:12...	2018-09-17T10:13...	PRIORIDAD-3
CONS SUBSECUENTE	2018-07-31T09:12...	2018-09-17T10:13...	PRIORIDAD-4
CONS PRIMERA VEZ	2018-07-16T11:11...	2018-09-13T09:40...	PRIORIDAD-2
CONS PRIMERA VEZ	2018-07-16T11:11...	2018-09-13T09:40...	PRIORIDAD-4
CONS SUBSECUENTE	2018-08-09T12:56...	2018-08-20T09:42...	PRIORIDAD-4
CONS PRIMERA VEZ	2018-05-28T14:06...	2018-07-19T09:58...	PRIORIDAD-2
CONS PRIMERA VEZ	2018-05-28T14:06...	2018-07-19T09:58...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-05-09T07:14...	2018-07-10T08:18...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-05-17T10:51...	2018-07-10T09:52...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-04-12T10:28...	2018-06-13T09:59...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-04-05T14:28...	2018-06-05T09:36...	PRIORIDAD-3
CONS PRIMERA VEZ	2017-12-14T09:40...	2018-02-26T08:46...	PRIORIDAD-3
CONS PRIMERA VEZ	2017-10-26T10:12...	2018-01-16T08:59...	PRIORIDAD-2
CONS PRIMERA VEZ	2017-11-14T11:41...	2018-01-25T09:21...	PRIORIDAD-2
CONS PRIMERA VEZ	2017-11-14T11:41...	2018-01-25T09:21...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-03-07T11:32...	2018-05-03T09:24...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-03-07T11:33...	2018-05-03T09:43...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-03-07T11:50...	2018-05-04T10:14...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-03-07T11:56...	2018-05-04T08:51...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-03-07T10:29...	2018-05-03T09:11...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-03-07T10:16...	2018-05-03T08:47...	PRIORIDAD-3
CONS PRIMERA VEZ	2017-12-13T10:38...	2018-02-22T09:28...	PRIORIDAD-3
CONS PRIMERA VEZ	2017-12-13T10:19...	2018-02-22T08:03...	PRIORIDAD-4
CONS PRIMERA VEZ	2017-12-13T10:24...	2018-02-22T09:41...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-10-02T10:52...	2018-12-14T11:14...	PRIORIDAD-3
CONS SUBSECUENTE	2018-09-27T11:50...	2018-12-14T09:34...	PRIORIDAD-4
CONS PRIMERA VEZ	2018-10-02T10:33...	2018-12-14T09:26...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-10-01T14:25...	2018-12-14T10:12...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-09-28T11:51...	2018-12-11T10:07...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-12-13T08:54...	2018-12-14T10:04...	PRIORIDAD-3
CONS PRIMERA VEZ	2018-10-01T11:19...	2018-12-13T10:14...	PRIORIDAD-3
CONS SUBSECUENTE	2018-10-01T08:46...	2018-12-11T08:16...	PRIORIDAD-4
CONS SUBSECUENTE	2018-11-12T11:29...	2018-12-12T11:06...	PRIORIDAD-4

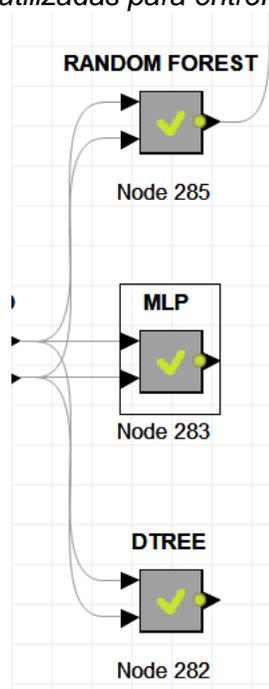
Entrenamiento de modelos con inteligencia artificial

Para la etapa de entrenamiento de modelos con inteligencia artificial se utilizan 3 tipos de algoritmos que son entrenados y posteriormente evaluados mediante los resultados obtenidos en la matriz de confusión, por lo que a cada una de las técnicas se

le realiza un análisis independiente, permitiendo obtener el mayor porcentaje de aciertos entre estos y conocer cuál es el óptimo.

Figura 80.

Técnicas de inteligencia artificial utilizadas para entrenamiento



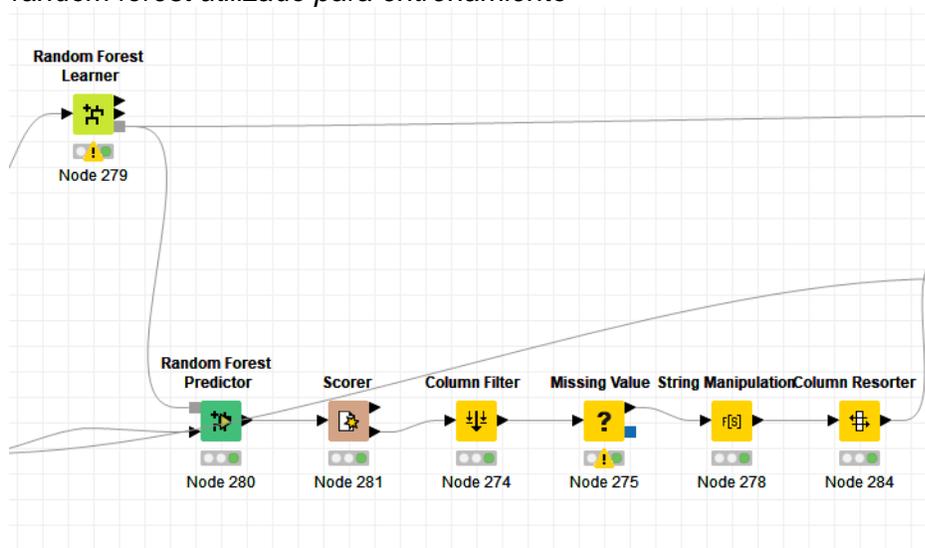
Nota: La figura muestra los metadatos correspondientes a las técnicas de inteligencia artificial que se utilizan para el modelado.

Modelo utilizando Random Forest

Este algoritmo de aprendizaje es uno de los más certeros que se encuentran actualmente disponibles, pero necesita una cantidad de datos sumamente grande para ser capaz de crear un clasificador certero, mientras que puede analizar y manejar cientos de variables sin que su rendimiento se vea disminuido, esto implica que es óptimo para utilizarlo con la base de datos del hospital.

Figura 81.

Flujo de random forest utilizado para entrenamiento

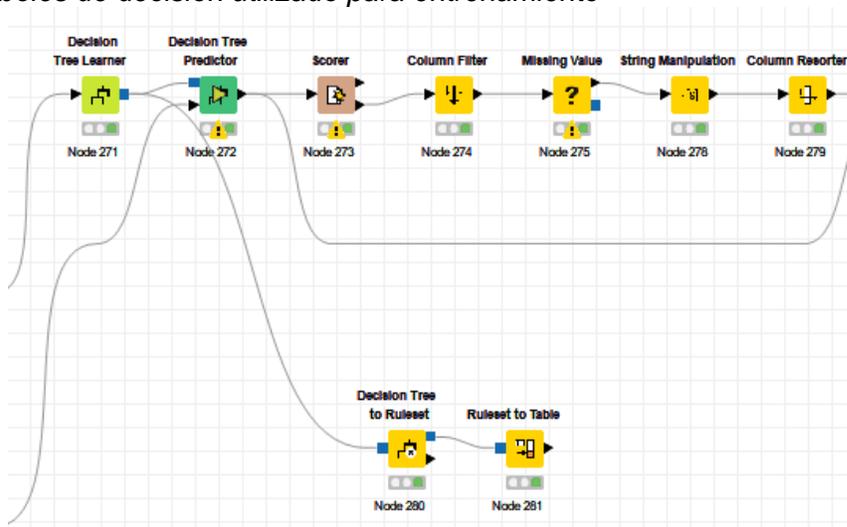


Modelo utilizando Árboles de Decisión

Este algoritmo de aprendizaje es un método analítico capaz de crear representaciones de manera esquemática para todas las alternativas existentes en un problema determinado, esto implica la creación de un mapa que contiene ramificaciones de cada una de las posibles variaciones y resultados. Resulta el modelo óptimo de análisis con resultados certeros, pero de la misma manera es el menos eficiente, ya que este analiza uno a uno todos los escenarios y decisiones posibles a tomar, haciendo que la tarea de entrenamiento resulte ser una gran carga computacional, por lo que otros sistemas pertenecientes al Hospital de Especialidades de las Fuerzas Armadas N°1 podrían verse perjudicados al momento de realizar el entrenamiento del modelo para la posterior asignación de turnos.

Figura 82.

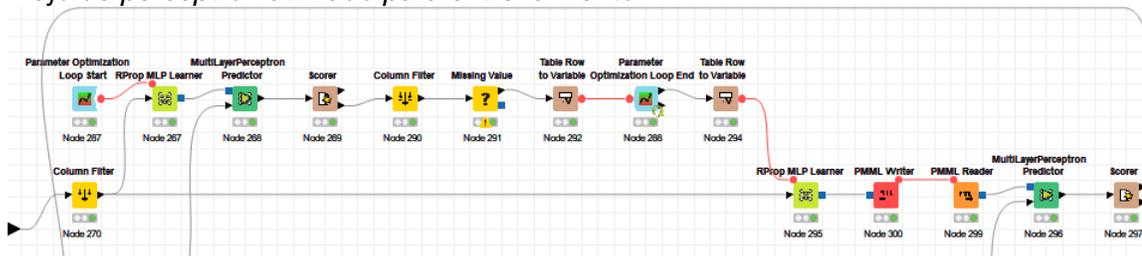
Flujo de árboles de decisión utilizado para entrenamiento



Modelo utilizando Perceptrón

Figura 83.

Flujo de perceptrón utilizado para entrenamiento



Análisis y elección del modelo óptimo

Dentro de la herramienta KNIME se utiliza el nodo Scorer, que muestra la matriz de confusión de cada uno de los modelos. Esta permite que el usuario final pueda entender de manera eficiente sin recurrir a explicaciones complejas por parte de los analistas y tomar una decisión en base a los parámetros expuestos, tal y como se observa en las tres figuras siguientes.

Figura 84.*Matriz de confusión del modelo de random forest*

Row ID	PRIORI...	PRIORI...	PRIORI...	PRIORI...	PRIORI...
PRIORIDAD-3	10407	3	1	0	0
PRIORIDAD-2	1	3188	0	0	1
PRIORIDAD-4	2	0	26955	0	0
PRIORIDAD-5	0	0	4	1600	0
PRIORIDAD-1	0	1	0	0	25

Nota: La figura muestra que existe una clasificación certera con mayor presencia de valores válidos en cada prioridad, y menos resultados fuera de rango.

Figura 85.*Matriz de confusión del modelo de árboles de decisión*

Row ID	PRIORI...	PRIORI...	PRIORI...	PRIORI...	PRIORI...
PRIORIDAD-3	10384	11	4	0	2
PRIORIDAD-2	5	3182	0	0	1
PRIORIDAD-4	1	4	26942	9	0
PRIORIDAD-5	1	2	19	1579	0
PRIORIDAD-1	0	8	0	0	17

Nota: La figura muestra que existe una clasificación acertada de acuerdo a los valores válidos en cada prioridad, sin embargo se observa mayor cantidad de resultados fuera de rango que en el modelo de random forest.

Figura 86.*Matriz de confusión del modelo de perceptrón*

Row ID	PRIORI...	PRIORI...	PRIORI...	PRIORI...	PRIORI...
PRIORIDAD-3	9752	361	296	2	0
PRIORIDAD-2	34	3140	16	0	0
PRIORIDAD-4	11	4	26879	63	0
PRIORIDAD-5	2	0	378	1224	0
PRIORIDAD-1	3	23	0	0	0

Nota: La figura muestra que no existe una clasificación acertada puesto que no ubica valores en todas las prioridades, además se observa gran cantidad de resultados fuera de rango.

Predicción para asignación de turnos

La predicción para asignación de turnos se realiza en base al modelo óptimo de Random Forest a través de un nodo propio de KNIME conocido como predictor, el cual se encarga de realizar los cálculos pertinentes para hacer uso del modelo que ha sido entrenado, de esta manera se hace uso de la prioridad que se les ha sido asignada y se procede a la predicción para comparar y contrastar los resultados de la predicción.

Figura 87.

Datos de predicción para asignación de turnos

FECHA_ATEN...	CLASS	Predicti...	Predicti...	ASIGN...
2018-12-05T08:59...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-11-15T10:02...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-10-11T09:28...	PRIORIDAD-2	PRIORIDAD-2	0.99	0
2018-09-17T10:13...	PRIORIDAD-3	PRIORIDAD-3	0.95	0
2018-09-13T09:40...	PRIORIDAD-2	PRIORIDAD-2	0.99	0
2018-08-20T09:42...	PRIORIDAD-4	PRIORIDAD-4	1	0
2018-07-19T09:58...	PRIORIDAD-2	PRIORIDAD-2	1	0
2018-07-10T08:18...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-07-10T09:52...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-06-05T09:36...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-02-26T08:46...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-01-16T08:59...	PRIORIDAD-2	PRIORIDAD-2	0.99	0
2018-01-25T09:21...	PRIORIDAD-2	PRIORIDAD-2	0.99	0
2018-05-03T09:24...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-05-03T09:43...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-05-04T10:14...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-05-04T08:51...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-05-03T09:11...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-02-22T08:03...	PRIORIDAD-4	PRIORIDAD-4	1	0
2018-12-14T11:14...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-12-14T09:34...	PRIORIDAD-4	PRIORIDAD-4	1	0
2018-12-14T09:26...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-12-14T10:12...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-12-11T10:07...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-12-14T10:04...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-12-13T10:14...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-12-11T08:16...	PRIORIDAD-4	PRIORIDAD-4	1	0
2018-12-10T09:29...	PRIORIDAD-4	PRIORIDAD-4	1	0
2018-12-05T08:35...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-12-13T10:40	PRIORIDAD-5	PRIORIDAD-5	1	0
2018-12-12T08:36...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-12-13T09:52...	PRIORIDAD-4	PRIORIDAD-4	1	0
2018-12-13T10:32...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-12-06T09:20...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-12-06T08:24...	PRIORIDAD-3	PRIORIDAD-3	1	0
2018-12-04T09:21...	PRIORIDAD-4	PRIORIDAD-4	1	0

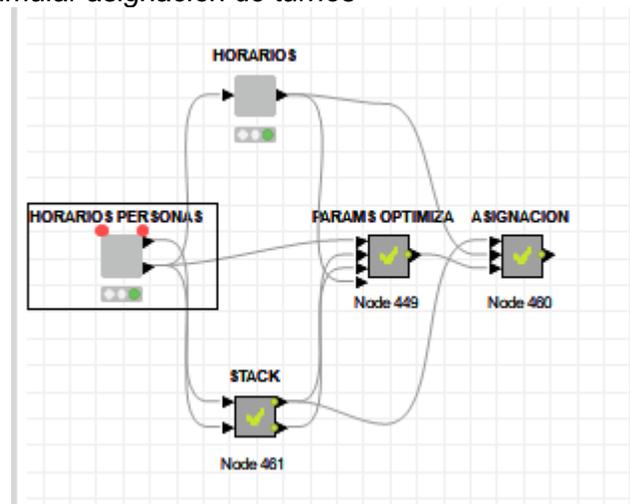
Capítulo IV

Simulación y análisis de resultados

Una vez que el entrenamiento y la predicción en cuanto a la asignación de los turnos finalizó satisfaciendo los parámetros necesarios, se procede a realizar la simulación del modelo asegurando que esta sea óptima y que no incluya valores externos que puedan influir en la capacidad y velocidad del análisis. Para realizar esto inicialmente se considerará la selección de cualquier especialidad en la que los pacientes acuden por un turno, de esta manera se manejarán todos los campos que el sistema actual utiliza para el agendamiento, asimilando el procedimiento a la realidad. Cuando se obtenga los datos finales se realizará el contraste con la información obtenida en los análisis ejecutados en Tableau.

Figura 88.

Flujo de trabajo para simular asignación de turnos

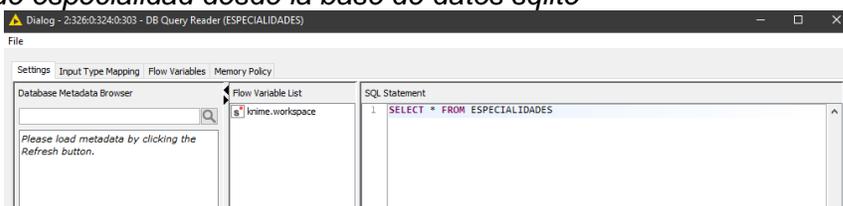


Simulación

Se usa una conexión a la base de datos previamente generada por el modelo entrenado haciendo uso del nodo SQLite Connector, pasando a la selección de especialidad.

Figura 89.

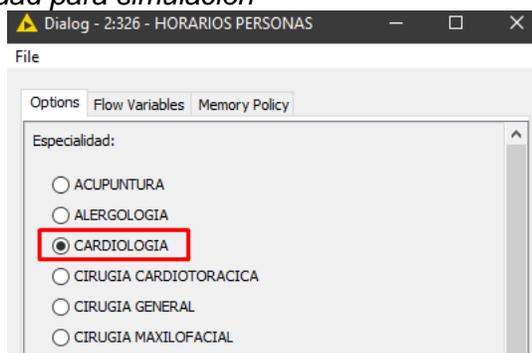
Selección de especialidad desde la base de datos sqlite



Se convierte las variables obtenidas de la base de datos SQLite y se transforma en filas de la tabla, permitiendo realizar la selección de cada especialidad para que el análisis posterior sea sumamente dinámico y se adapte a las necesidades del usuario final, en este caso se seleccionó Cardiología, como se observa en la siguiente figura.

Figura 90.

Selección de especialidad para simulación



Con la selección del usuario se analiza únicamente la especialidad requerida y no toda la base de datos logrando un análisis óptimo y poco demandante para el

hardware encargado del proceso. Como parte de las fuentes utilizadas para la asignación está el horario de los médicos como primera entrada, así el modelo asegura que los turnos asignados estarán en la agenda del especialista y no se generen fuera de esta debido a que muchos de los doctores vinculados no poseen un itinerario continuo, existiendo espacios sin asignación a las que el modelo se ajustará por completo.

Figura 91.

Especialidad seleccionada

Row ID	ESPECIALIDAD SELECCION
values	CARDIOLOGIA

Figura 92.

Horarios de médicos especialistas

Row ID	CODIG...	ESPECI...	NOMBRE ...	DIA	Day of ...	FECHA ...	HORA ...	ORDEN...	TURNO...	TIEMPO	HORA_INI...	HORA_FINAL	HC_ASI...
Row0	M309	CARDIOLOGIA	C1-DR. LASSO J.	3	miércoles	2020-01-01	07:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row1130	M314	CARDIOLOGIA	C2-DRA. SEGA...	3	miércoles	2020-01-01	07:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row5490	M317	CARDIOLOGIA	C5-DR. JUMBO L.	3	miércoles	2020-01-01	07:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row1	M309	CARDIOLOGIA	C1-DR. LASSO J.	3	miércoles	2020-01-01	07:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row1131	M314	CARDIOLOGIA	C2-DRA. SEGA...	3	miércoles	2020-01-01	07:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row5491	M317	CARDIOLOGIA	C5-DR. JUMBO L.	3	miércoles	2020-01-01	07:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row2	M309	CARDIOLOGIA	C1-DR. LASSO J.	3	miércoles	2020-01-01	08:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row1132	M314	CARDIOLOGIA	C2-DRA. SEGA...	3	miércoles	2020-01-01	08:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row5492	M317	CARDIOLOGIA	C5-DR. JUMBO L.	3	miércoles	2020-01-01	08:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row3	M309	CARDIOLOGIA	C1-DR. LASSO J.	3	miércoles	2020-01-01	08:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row1133	M314	CARDIOLOGIA	C2-DRA. SEGA...	3	miércoles	2020-01-01	08:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row5493	M317	CARDIOLOGIA	C5-DR. JUMBO L.	3	miércoles	2020-01-01	08:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row4	M309	CARDIOLOGIA	C1-DR. LASSO J.	3	miércoles	2020-01-01	09:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row1134	M314	CARDIOLOGIA	C2-DRA. SEGA...	3	miércoles	2020-01-01	09:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row5494	M317	CARDIOLOGIA	C5-DR. JUMBO L.	3	miércoles	2020-01-01	09:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row5	M309	CARDIOLOGIA	C1-DR. LASSO J.	3	miércoles	2020-01-01	09:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row1135	M314	CARDIOLOGIA	C2-DRA. SEGA...	3	miércoles	2020-01-01	09:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row5495	M317	CARDIOLOGIA	C5-DR. JUMBO L.	3	miércoles	2020-01-01	09:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row6	M309	CARDIOLOGIA	C1-DR. LASSO J.	3	miércoles	2020-01-01	10:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row1136	M314	CARDIOLOGIA	C2-DRA. SEGA...	3	miércoles	2020-01-01	10:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row5496	M317	CARDIOLOGIA	C5-DR. JUMBO L.	3	miércoles	2020-01-01	10:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row7	M309	CARDIOLOGIA	C1-DR. LASSO J.	3	miércoles	2020-01-01	10:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row1137	M314	CARDIOLOGIA	C2-DRA. SEGA...	3	miércoles	2020-01-01	10:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row5497	M317	CARDIOLOGIA	C5-DR. JUMBO L.	3	miércoles	2020-01-01	10:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row8	M309	CARDIOLOGIA	C1-DR. LASSO J.	3	miércoles	2020-01-01	11:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row1138	M314	CARDIOLOGIA	C2-DRA. SEGA...	3	miércoles	2020-01-01	11:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row5498	M317	CARDIOLOGIA	C5-DR. JUMBO L.	3	miércoles	2020-01-01	11:00	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row9	M309	CARDIOLOGIA	C1-DR. LASSO J.	3	miércoles	2020-01-01	11:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row1139	M314	CARDIOLOGIA	C2-DRA. SEGA...	3	miércoles	2020-01-01	11:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row5499	M317	CARDIOLOGIA	C5-DR. JUMBO L.	3	miércoles	2020-01-01	11:30	1	10	30	2020-06-01T07...	2020-06-01T12...	0
Row2900	M315	CARDIOLOGIA	C3-DR. SAENZ B.	3	miércoles	2020-01-01	13:00	1	8	30	2020-06-01T13...	2020-06-01T17...	0
Row6270	M836	CARDIOLOGIA	C8-DR. IZAGUI...	3	miércoles	2020-01-01	13:00	1	8	15	2020-06-01T13...	2020-06-01T17...	0
Row6271	M836	CARDIOLOGIA	C8-DR. IZAGUI...	3	miércoles	2020-01-01	13:15	1	8	15	2020-06-01T13...	2020-06-01T17...	0
Row2901	M315	CARDIOLOGIA	C3-DR. SAENZ B.	3	miércoles	2020-01-01	13:30	1	8	30	2020-06-01T13...	2020-06-01T17...	0
Row6272	M836	CARDIOLOGIA	C8-DR. IZAGUI...	3	miércoles	2020-01-01	13:30	1	8	15	2020-06-01T13...	2020-06-01T17...	0
Row6273	M836	CARDIOLOGIA	C8-DR. IZAGUI...	3	miércoles	2020-01-01	13:45	1	8	15	2020-06-01T13...	2020-06-01T17...	0
Row2902	M315	CARDIOLOGIA	C3-DR. SAENZ B.	3	miércoles	2020-01-01	14:00	1	8	30	2020-06-01T13...	2020-06-01T17...	0

Se utiliza colas que son tablas dentro de la base SQLite, estas se conectan tanto con bases transaccionales como con bases de archivo, se extrae las fechas y se coloca las horas, las mismas a las que inicialmente se asigna un valor de 0, ya que pertenecen a los pacientes y estos aún no han sido asignados, posteriormente con el análisis respectivo este campo será llenado con la historia clínica del paciente.

Figura 93.

Horarios generados por el modelo, sin turnos asignados

Row ID	S NUMER...	S ESPECI...	I CONS P...	I CONS S...	D RATIO ...	I DEM DI...	I DEM DI...	S ENFERMEDAD	S DISCAP...	S Predict...	I DELTA ...	I ACTUA...	S TIPO
Row0	10003	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row1	10005	CARDIOLOGIA	2166	5174	2.389	8	16	FIBRILACION Y ALETEO AURICULAR	NINGUNA	PRIORIDAD-4	98	0	
Row2	10007	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row3	10017	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row4	100202	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row5	100316	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row6	10034	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row7	100354	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row8	100372	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row9	10039	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row10	100477	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row11	100489	CARDIOLOGIA	2166	5174	2.389	8	16	INFARTO AGUDO DEL MIOCARDIO	NINGUNA	PRIORIDAD-4	98	0	
Row12	100528	CARDIOLOGIA	2166	5174	2.389	8	16	FIBRILACION Y ALETEO AURICULAR	NINGUNA	PRIORIDAD-4	98	0	
Row13	10056	CARDIOLOGIA	2166	5174	2.389	8	16	ARRITMIA CARDIACA, NO ESPECIFI...	NINGUNA	PRIORIDAD-4	98	0	
Row14	100571	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row15	10065	CARDIOLOGIA	2166	5174	2.389	8	16	CHEQUEO PREQUIRURGICO	NINGUNA	PRIORIDAD-3	98	0	
Row16	100665	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row17	100683	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row18	10069	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row19	1007	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-3	98	0	
Row20	10076	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-5	98	0	
Row21	100874	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-3	98	0	
Row22	100922	CARDIOLOGIA	2166	5174	2.389	8	16	CHEQUEO PREQUIRURGICO	NINGUNA	PRIORIDAD-3	98	0	
Row23	100951	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row24	100988	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row25	101018	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row26	10102	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row27	101036	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row28	101062	CARDIOLOGIA	2166	5174	2.389	8	16	ENFERMEDAD ISQUEMICA CRONICA...	NINGUNA	PRIORIDAD-4	98	0	
Row29	101079	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row30	10120	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row31	10130	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row32	101318	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	
Row33	101411	CARDIOLOGIA	2166	5174	2.389	8	16	CHEQUEO PREQUIRURGICO	NINGUNA	PRIORIDAD-3	98	0	
Row34	10144	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-3	98	0	
Row35	101456	CARDIOLOGIA	2166	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	

Para que la simulación sea aún más real se han generado aleatoriamente tanto turnos de primera vez como subsecuentes mediante los históricos estadísticos analizados y clasificados por el modelo. Además, se utiliza el algoritmo de shuffle dos veces en esta etapa de forma consecutiva, logrando que los turnos generados se combinen dentro de la tabla, de esta forma llegamos a un ambiente que asemeja mucho más el escenario que se presenta en el Hospital de Especialidades de las Fuerzas Armadas N°1.

Figura 94.

Turnos de 1ra vez y subsecuentes generados aleatoriamente

Classified values - 2461:379:372 - Rule Engine

File Hilitte Navigation View

Table "default" - Rows: 7340 Spec - Columns: 15 Propertes Flow Variables

Row ID	CONS P...	I CONS S...	D RATIO ...	I DEM DI...	DEM DI...	S ENFERMEDAD	S DISCAP...	S Predicti...	I DELTA ...	I ACTUA...	S TIPO
Row5222	56	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	1VEZ
Row2246	56	5174	2.389	8	16	CHEQUEO PREQUIRURGICO	NINGUNA	PRIORIDAD-4	98	0	SUBSECUENTE
Row5235	56	5174	2.389	8	16	OTRAS ARRITMIAS CARDIACAS	NINGUNA	PRIORIDAD-4	98	0	1VEZ
Row7325	56	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	1VEZ
Row2220	56	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	1VEZ
Row1637	56	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-5	98	0	1VEZ
Row1681	56	5174	2.389	8	16	ENFERMEDAD ISQUEMICA CRONICA...	NINGUNA	PRIORIDAD-4	98	0	SUBSECUENTE
Row421	56	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	SUBSECUENTE
Row5349	56	5174	2.389	8	16	CHEQUEO PREQUIRURGICO	NINGUNA	PRIORIDAD-4	98	0	SUBSECUENTE
Row781	56	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-5	98	0	SUBSECUENTE
Row1126	56	5174	2.389	8	16	BOCIO DIFUSO NO TOXICO	NINGUNA	PRIORIDAD-3	98	0	1VEZ
Row1212	56	5174	2.389	8	16	CHEQUEO PREQUIRURGICO	NINGUNA	PRIORIDAD-3	98	0	1VEZ
Row1215	56	5174	2.389	8	16	BRADICARDIA, NO ESPECIFICADA	NINGUNA	PRIORIDAD-3	98	0	1VEZ
Row3230	56	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	1VEZ
Row6494	56	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	1VEZ
Row4682	56	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	1VEZ
Row4750	56	5174	2.389	8	16	TAQUICARDIA SUPRAVENTRICULAR	NINGUNA	PRIORIDAD-4	98	0	SUBSECUENTE
Row6194	56	5174	2.389	8	16	DISNEA	NINGUNA	PRIORIDAD-3	98	0	1VEZ
Row2299	56	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	SUBSECUENTE
Row7291	56	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-3	98	0	SUBSECUENTE
Row4312	56	5174	2.389	8	16	HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0	SUBSECUENTE
Row7197	56	5174	2.389	8	16	OTROS TUMORES MALIGNO DE LA PIEL	NINGUNA	PRIORIDAD-3	98	0	1VEZ

Otro valor considerado para la simulación es la cantidad de días de espera que tiene un paciente para realizarse exámenes referentes a la consulta, en este parámetro se establece un rango de entre 4 y 15 días dependiendo del diagnóstico brindado. Este dato se toma en cuenta para que en los turnos subsecuentes no se asigne horario a un paciente antes de que los resultados sean entregados, por lo que el modelo se encarga de validar y verificar esto con un motor de reglas.

Figura 95.

Flujo de trabajo para generación aleatoria del tipo de turno

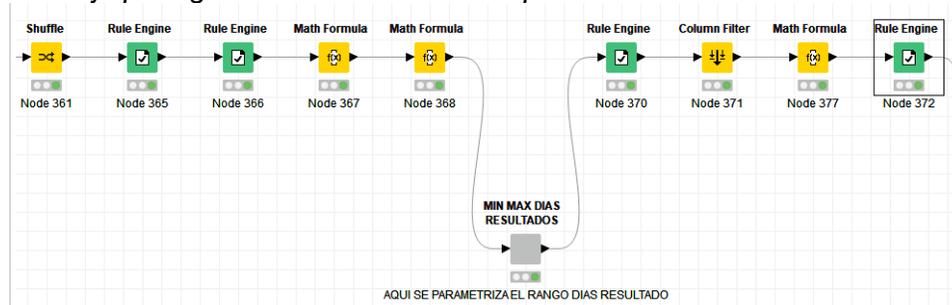


Figura 96.

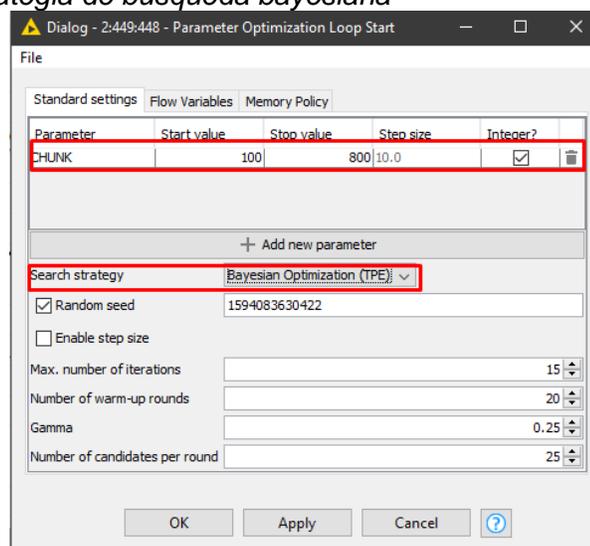
Días de espera para exámenes en turnos subsecuentes

98	0	SUBSECUENTE	S	4
98	0	SUBSECUENTE	S	16
98	0	1VEZ	N	0
98	0	1VEZ	N	0
98	0	SUBSECUENTE	S	14
98	0	SUBSECUENTE	S	9
98	0	1VEZ	N	0

Se realiza el cálculo de la media entre turnos para minimizar el número de días mediante los datos de las colas, se utiliza por tercera vez el algoritmo de shuffle para asegurar que lo generado anteriormente por el modelo no se vea condicionado por ningún tipo de agrupamiento. Utilizando chunk se agenda de manera óptima durante las horas de atención, para lograr esto se utiliza un optimizador en conjunto con nodos de bucle que permiten realizar una búsqueda bayesiana recorriendo cada uno de los datos. Al final de este proceso se obtiene el chunk óptimo y se procede a la asignación de los turnos.

Figura 97.

Chunk óptimo y estrategia de búsqueda bayesiana



Análisis de resultados

Inicialmente el tiempo de espera promedio para recibir atención en la especialidad de cardiología era de 98 días aproximadamente, además no se consideraba prioridades en sus pacientes, es decir simplemente se agendaba de acuerdo a la disponibilidad de horario.

Figura 98.

Tiempo necesario para asignación de turnos antes de optimizar

S ENFERMEDAD	S DISCAP...	S Predicti...	I DELTA ...	I ACTUA...
ENFERMEDAD ISQUEMICA CRONICA DEL CORAZON	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
FIBRILACION Y ALETEO AURICULAR	NINGUNA	PRIORIDAD-5	98	0
SOPLO CARDIACO, NO ESPECIFICADO	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
CARDIOMIOPATIA DILATADA	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
OTROS TRASTORNOS DEL DESARROLLO Y CRECI...	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-5	98	0
DEFECTO DEL TABIQUE AURICULAR	NINGUNA	PRIORIDAD-3	98	0
ENFERMEDAD ISQUEMIA CRONICA DEL CORAZO...	NINGUNA	PRIORIDAD-4	98	0
CHEQUEO PREQUIRURGICO	NINGUNA	PRIORIDAD-3	98	0
OTROS TUMORES MALIGNO DE LA PIEL	NINGUNA	PRIORIDAD-3	98	0
ENFERMEDAD CARDIACA HIPERTENSIVA	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
CHEQUEO PREQUIRURGICO	NINGUNA	PRIORIDAD-3	98	0
ESTENOSIS AORTICA REUMATICA	NINGUNA	PRIORIDAD-5	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-4	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-3	98	0
HIPERTENSION ESENCIAL (PRIMARIA)	NINGUNA	PRIORIDAD-3	98	0

Terminada la simulación se puede observar en la siguiente figura que el promedio de días de espera se reduce a 65, además se considera la clasificación de los pacientes de acuerdo a su prioridad, y se ubica atenciones de tipo primera vez y subsecuentes. Esto demuestra que el sistema de agendamiento cumple con todos los requisitos viéndose reducido en tiempo y optimizado gracias al presente modelo.

Figura 99.

Tiempo necesario para asignación de turnos después de optimizar

Processed data - 0:460:471 - RowID

File Hilite Navigation View

Table "default" - Rows: 8326 Spec - Columns: 15 Properties Flow Variables

Row ID	Iteration	CHUNK	Objecti...	ESPECI...	Day of ...	FECHA ...	HORA ...	Predicti...	RANGO...	TIPO
Row2100	3	700	65	CARDIOLOGIA	viernes	2020-03-27	08:30	PRIORIDAD-2	0	1VEZ
Row2101	3	700	65	CARDIOLOGIA	viernes	2020-03-27	08:30	PRIORIDAD-2	0	1VEZ
Row2102	3	700	65	CARDIOLOGIA	viernes	2020-03-27	09:00	PRIORIDAD-2	0	SUBSECUENTE
Row2103	3	700	65	CARDIOLOGIA	viernes	2020-03-27	09:00	PRIORIDAD-2	0	1VEZ
Row2104	3	700	65	CARDIOLOGIA	viernes	2020-03-27	09:00	PRIORIDAD-2	4	SUBSECUENTE
Row2105	3	700	65	CARDIOLOGIA	viernes	2020-03-27	09:30	PRIORIDAD-3	0	1VEZ
Row2106	3	700	65	CARDIOLOGIA	viernes	2020-03-27	09:30	PRIORIDAD-3	0	1VEZ
Row2107	3	700	65	CARDIOLOGIA	viernes	2020-03-27	09:30	PRIORIDAD-3	0	1VEZ
Row2108	3	700	65	CARDIOLOGIA	viernes	2020-03-27	10:00	PRIORIDAD-3	0	1VEZ
Row2109	3	700	65	CARDIOLOGIA	viernes	2020-03-27	10:00	PRIORIDAD-3	0	1VEZ
Row2110	3	700	65	CARDIOLOGIA	viernes	2020-03-27	10:00	PRIORIDAD-3	0	1VEZ
Row2111	3	700	65	CARDIOLOGIA	viernes	2020-03-27	10:30	PRIORIDAD-3	0	1VEZ
Row2112	3	700	65	CARDIOLOGIA	viernes	2020-03-27	10:30	PRIORIDAD-3	0	1VEZ
Row2113	3	700	65	CARDIOLOGIA	viernes	2020-03-27	10:30	PRIORIDAD-3	0	1VEZ
Row2114	3	700	65	CARDIOLOGIA	viernes	2020-03-27	11:00	PRIORIDAD-3	0	SUBSECUENTE
Row2115	3	700	65	CARDIOLOGIA	viernes	2020-03-27	11:00	PRIORIDAD-3	0	1VEZ
Row2116	3	700	65	CARDIOLOGIA	viernes	2020-03-27	11:00	PRIORIDAD-3	0	1VEZ
Row2117	3	700	65	CARDIOLOGIA	viernes	2020-03-27	11:30	PRIORIDAD-3	0	1VEZ
Row2118	3	700	65	CARDIOLOGIA	viernes	2020-03-27	11:30	PRIORIDAD-3	0	1VEZ
Row2119	3	700	65	CARDIOLOGIA	viernes	2020-03-27	11:30	PRIORIDAD-3	0	1VEZ

Nota: Los días de espera en la especialidad de cardiología se han reducido a 65 días, respetando prioridad de paciente y tipo de atención.

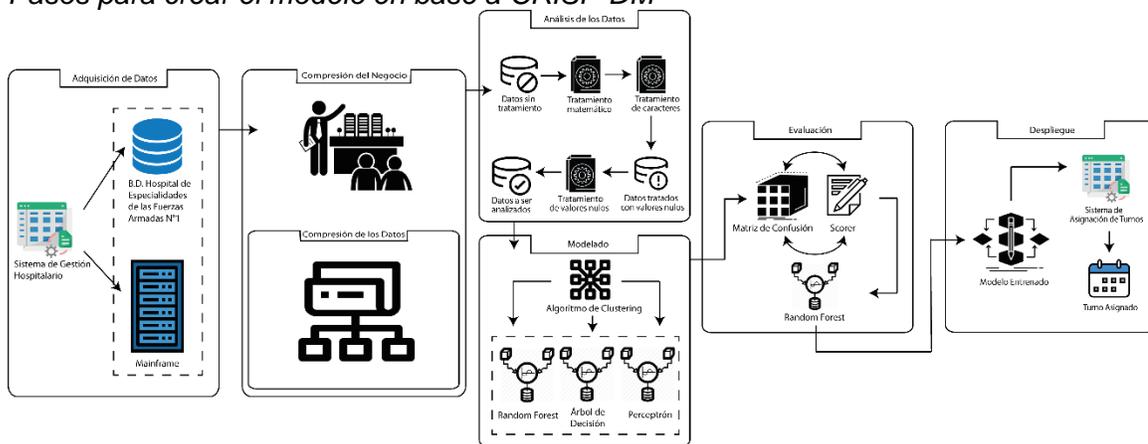
Capítulo V

Conclusiones y recomendaciones

El modelo desarrollado se realizó en base a la metodología CRISP-DM, la misma que de acuerdo a sus pasos permitió aplicar las distintas técnicas de inteligencia artificial para abordar todos los objetivos propuestos. En síntesis se ejecutaron las siguientes fases con sus respectivas actividades, las cuales se describen en la siguiente figura.

Figura 100.

Pasos para crear el modelo en base a CRISP-DM



Nota: Se inicia con la *Adquisición de Datos* trabajando con los registros del Sistema de Gestión Hospitalario que se encuentran almacenados en un mainframe; se continúa con la *Comprensión del negocio y de los Datos* para tener claramente el concepto de las reglas del negocio y variables que se puede utilizar; en el *Análisis de los Datos* en primera instancia se obtiene información sin tratamiento que mediante la utilización de herramientas y técnicas de limpieza se pudo obtener la base final que entrará a *Modelamiento*. A esta data se aplica algoritmos de clustering para que pueda ser analizada mediante técnicas de aprendizaje supervisado, que fueron evaluadas

mediante la matriz de confusión resultando óptimo el de Random Forest. Finalmente se obtiene un modelo entrenado al cual se aplica la *Simulación* para verificar su validez y eficacia; para el *Despliegue* se debe concatenar con la base de producción del Hospital para aplicarlo en el sistema de Asignación de turnos.

Conclusiones

La metodología CRISP-DM utilizada para el desarrollo se adaptó perfectamente a las características del presente proyecto, como resultado de esto se obtuvo un modelo funcional que se adapta a las exigencias del Hospital de Especialidades de Fuerzas Armadas N°1; además el uso de la herramienta KNIME proporcionó flexibilidad y portabilidad para el tratamiento y análisis de los datos permitiendo cubrir todas las necesidades en cuanto al agendamiento de turnos en el área de consulta externa.

Mediante la simulación del modelo obtenido a través de la técnica de aprendizaje supervisado Random Forest, se determinó la reducción considerable de los tiempos para el otorgamiento de turnos en las distintas especialidades de consulta externa, esto beneficia de manera significativa tanto a la casa de salud como a sus pacientes ya que podrán efectuar de manera más ágil la atención manteniendo la calidad del servicio.

Además con la ayuda de la investigación realizada se encontró problemas adicionales que se dan en los distintos consultorios, como por ejemplo cuando el médico tratante no realiza el cierre de sus consultas impidiendo determinar un promedio de tiempo real de atención a cada paciente; sin embargo se utilizaron técnicas estadísticas y de limpieza de datos para corregir estos errores y permitir que el análisis de la información sea la más adecuada.

La validación de los resultados en cuanto al proceso de clasificación y priorización de los pacientes se llevó a cabo mediante la aplicación de distintas técnicas de inteligencia artificial: Árboles de decisión, Perceptrón y Random Forest; mediante la interpretación y análisis de la matriz de confusión de cada proceso se determinó cual era la que más se acoplaba a la investigación. Esto permite concluir que el modelo se fundamentó en Minería puesto que aquí se estipula la técnica apropiada después de todo el análisis de datos, mientras que en Machine Learning antes de iniciar con el descubrimiento de la data ya se establece un modelo a utilizar.

Recomendaciones

Para realizar el despliegue del modelo se debe analizar minuciosamente las características técnicas con las que cuenta actualmente el Hospital ya que esto demanda afectación a los sistemas con los que se encuentra funcionando, considerando esto se determinan las siguientes alternativas para la puesta en producción.

Cambiar la arquitectura de los aplicativos que actualmente son cliente – servidor, deben formar parte de una estructura analítica utilizando por ejemplo Python que permita transformar la data del hospital que se encuentra en forma transaccional y que a través de sus librerías permitan realizar la conexión de los aplicativos con el modelo desarrollado.

Teniendo de por medio una estructura distribuida se puede utilizar Python como lenguaje y como framework Django y Flask, mientras que a nivel de Java mediante las librerías de Deep4j para el análisis, predicciones y asignaciones aplicados directamente a los sistemas actuales.

Sin embargo, el método más sencillo y que representaría un menor impacto es mediante la realización de todo el proceso de minería con KNIME en forma visual, posteriormente transformar toda la información para concatenar con la base de datos que se encuentra en producción. Esto significa la utilización de menos recursos, esfuerzo y personal técnico en relación a las anteriores opciones para efectuar el despliegue.

Bibliografía

- Ahmed, K. (2017). Analysis of Data Mining Tools for Disease Prediction. *Journal of Pharmaceutical Science and Research*, 1886–1888.
- Anindita, N., Nugroho, H. A., & Adji, T. B. (2017). A Combination of Multiple Imputation and Principal Components. *IEEE*, 3111–3115.
- Aura Portal. (2019). *Tipos de algoritmos de Inteligencia Artificial y Machine Learning*.
- Bochicchio, M., Cuzzocrea, A., & Vaira, L. (2016). A Big Data Analytics Framework for Supporting. *IEEE*, 508–513.
- Dwivedi, S., Kasliwal, P., & Soni, S. (2016a). Comprehensive Study of Data Analytics. *IEEE*, 978–985.
- Dwivedi, S., Kasliwal, P., & Soni, S. (2016b). Comprehensive Study of Data Analytics Tools (RapidMiner, Weka, R tool, Knime). *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*.
- Gómez, V. (n.d.). *Instinto Binario*. <https://instintobinario.com/scrum-la-metodologia-agil-mas-usada/>
- Grupo de Estudios de Etica Clínica. (2003). *Sobre las acciones médicas proporcionadas y el uso de métodos extraordinarios de tratamiento*. 689–696.
- Hub Knime. (2016). *Rule-based Row Filter (Dictionary)*.
<https://hub.knime.com/knime/extensions/org.knime.features.javasnipet/latest/org.knime.base.node.rules.engine.twoports.RuleEngineFilter2PortsNodeFactory>
- Hub Knime. (2017a). *Column Appender (deprecated) – KNIME Hub*.

<https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.preproc.columnappend.ColumnAppenderNodeFactory>

Hub Knime. (2017b). *Extract Table Dimension*.

<https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.util.extracttabledimension.ExtractTableDimensionNodeFactory>

Hub Knime. (2017c). *Normalizer*.

<https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.preproc.normalize3.Normalizer3NodeFactory>

Hub Knime. (2017d). *PCA Apply*.

<https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.mine.transformation.pca.apply.PCA2ApplyNodeFactory>

Hub Knime. (2018a). *Model Reader*.

<https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.io.predictor.PredictorReaderNodeFactory>

Hub Knime. (2018b). *One to Many*.

<https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.preproc.columntrans2.One2ManyCol2NodeFactory>

Hub Knime. (2018c). *Rule Engine*.

<https://hub.knime.com/knime/extensions/org.knime.features.javasnippet/latest/org.knime.base.node.rules.engine.RuleEngineNodeFactory>

Hub Knime. (2018d). *String Manipulation*.

<https://hub.knime.com/knime/extensions/org.knime.features.javasnippet/latest/org.knime.base.node.stringmanipulation.StringManipulationNodeFactory>

nime.base.node.preproc.stringmanipulation.StringManipulationNodeFactory

Hub Knime. (2019a). *Create Date&Time Range*.

<https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.time.node.create.createdatetime.CreateDateTimeNodeFactory>

Hub Knime. (2019b). *DB Writer*.

<https://hub.knime.com/knime/extensions/org.knime.features.database/latest/org.knime.database.node.io.write.DBWriteNodeFactory>

Hub Knime. (2019c). *Partitioning*.

<https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.preproc.partition.PartitionNodeFactory>

Hub Knime. (2019d). *Table Column to Variable*.

<https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.flowvariable.tablecoltvariable3.TableColumnToVariable3NodeFactory>

Hub Knime. (2020a). *Auto Binner Hub*. <https://www.knime.com/knime-introductory-course/chapter3/section1/advanced-row-filter>

Hub Knime. (2020b). *Chunk Loop Start Hub*.

<https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.meta.looper.chunk.LoopStartChunkNodeFactory>

Hub Knime. (2020c). *Column Resorter*.

<https://hub.knime.com/knime/extensions/org.knime.features.base/latest/org.knime.base.node.preproc.columnresorter.ColumnResorterNodeFactory>

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature*

Reviews in Software Engineering.

Koppad, S. H., & Kumar, A. (2016). Application of Big Data Analytics in Healthcare.

IEEE, 108–112.

Lopez, P. L. (2004). Población, muestra y muestreo. *Punto Cero*, 69–74.

Machado, G. (2015). What variables are important in predicting bovine viral diarrhoea virus? A random forest approach. *Veterinary Research*.

Mancini, M. (2017). Exploiting big data for improving healthcare services. *Journal of E-Learning and Knowledge Society*, 23–33.

Marquardt, J., Newman, S., Hattarki, D., Srinivasan, R., & Shanu, S. (2017).

HealthSCOPE: An Interactive Distributed Data Mining Framework for Scalable Prediction of Healthcare Costs. *IEEE*, 1227–1230.

Meru, P., Ravindra, P., Krishnamoorthy, P., & Jacob, J. (2016). A Machine Learning

Framework for Auto Classification of Imaging. *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*.

Monserrat, S. (2018). Minería de Datos en Base de Datos de Servicios de Salud. *Cidisi –*

UTN – FRSF, Ingar UTN- CONICET, 103–109.

Munera, L. E. (1991). Inteligencia artificial y sistemas expertos. *ICESI*.

Network Analysis. (2015). *Meta Nodes | KNIME*. <https://www.knime.com/metanodes>

Network Analysis. (2018a). *Advanced Row Filter | KNIME*.

<https://www.knime.com/knime-introductory-course/chapter3/section1/advanced-row-filter>

Network Analysis. (2018b). *Basic Row Filter* | KNIME. <https://www.knime.com/knime-introductory-course/chapter3/section1/basic-row-filter>

Network Analysis. (2018c). *Column Filter* | KNIME. <https://www.knime.com/knime-introductory-course/chapter3/section1/column-filter>

Network Analysis. (2018d). *File Reader Node* | KNIME. <https://www.knime.com/knime-introductory-course/chapter2/section1/file-reader-node>

Network Analysis. (2018e). *Knime Components Guide*. https://docs.knime.com/2020-07/analytics_platform_components_guide/index.html

Network Analysis. (2019a). *Concatenation* | KNIME. <https://www.knime.com/knime-introductory-course/chapter3/section3/concatenation>

Network Analysis. (2019b). *Documentation* | KNIME.
<https://www.knime.com/documentation-3>

Network Analysis. (2019c). *Knime Analytics Platform* | KNIME.
<https://www.knime.com/knime-analytics-platform>

Network Analysis. (2020). *Filtering* | KNIME. <https://www.knime.com/nodeguide/etl-data-manipulation/filtering>

Node Pit. (2015a). *RowID* — *NodePit*.
<https://nodepit.com/node/org.knime.base.node.preproc.rowkey2.RowKeyNodeFactory2>

Node Pit. (2015b). *Sorter*.
<https://nodepit.com/node/org.knime.base.node.preproc.sorter.SorterNodeFactory>

Node Pit. (2016a). *Missing Value*.

<https://nodepit.com/node/org.knime.base.node.preproc.pmml.missingval.compute.MissingValueHandlingNodeFactory>

Node Pit. (2016b). *Row Filter*.

<https://nodepit.com/node/org.knime.base.node.preproc.filter.row.RowFilterNodeFactory>

Node Pit. (2016c). *Row Splitter*.

<https://nodepit.com/node/org.knime.base.node.preproc.filter.row.RowFilter2PortNodeFactory>

Node Pit. (2016d). *Scorer*.

<https://nodepit.com/node/org.knime.base.node.mine.scorer.accuracy.AccuracyScorerNodeFactory>

Node Pit. (2017a). *GroupBy*.

<https://nodepit.com/node/org.knime.base.node.preproc.groupby.GroupByNodeFactory>

Node Pit. (2017b). *Nominal Value Row Filter*.

<https://nodepit.com/node/org.knime.base.node.preproc.filter.nominal.NominalValueRowFilterNodeFactory>

Node Pit. (2017c). *PCA Compute*.

<https://nodepit.com/node/org.knime.base.node.mine.pca.PCAComputeNodeFactory>

Node Pit. (2018a). *Loop End*.

<https://nodepit.com/node/org.knime.base.node.meta.looper.LoopEndNodeFactory>

Node Pit. (2018b). *Model Reader*.

<https://nodepit.com/node/org.knime.base.node.io.predictor.PredictorReaderNodeFactory>

Node Pit. (2018c). *Model Writer*.

<https://nodepit.com/node/org.knime.base.node.io.predictor.PredictorWriterNodeFactory>

Node Pit. (2018d). *Number To String*.

<https://nodepit.com/node/org.knime.base.node.preproc.colconvert.numbertostring.NumberToStringNodeFactory>

Node Pit. (2018e). *Random Forest Learner*.

<https://nodepit.com/node/org.knime.base.node.mine.treeensemble.node.randomforest.learner.classification.RandomForestClassificationLearnerNodeFactory>

Node Pit. (2018f). *Random Forest Predictor*.

<https://nodepit.com/node/org.knime.base.node.mine.treeensemble2.node.randomforest.predictor.classification.RandomForestClassificationPredictorNodeFactory>

Node Pit. (2018g). *SQLite Connector*.

<https://nodepit.com/node/org.knime.database.extension.sqlite.node.connector.SQLiteDBConnectorNodeFactory>

Node Pit. (2018h). *String to Date&Time*.

<https://nodepit.com/node/org.knime.time.node.convert.stringtodatetime.StringToDateTimeNodeFactory>

Node Pit. (2019a). *Auto Binner - NodePit*.

<https://nodepit.com/node/org.knime.base.node.preproc.autobinner3.AutoBinnerLearner>

nNodeFactory

Node Pit. (2019b). *Concatenate*.

<https://nodepit.com/node/org.knime.base.node.preproc.append.row.AppendedRowsNodeFactory>

Node Pit. (2019c). *Create Date&Time Range*.

<https://nodepit.com/node/org.knime.time.node.create.createdatetime.CreateDateTimeNodeFactory>

Node Pit. (2019d). *Date&Time Shift*.

<https://nodepit.com/node/org.knime.time.node.manipulate.datetimeshift.DateTimeShiftNodeFactory>

Node Pit. (2019e). *Extract Date&Time Fields - NodePit*.

<https://nodepit.com/node/org.knime.time.node.extract.datetime.ExtractDateTimeFieldsNodeFactory>

Node Pit. (2020a). *Cross Joiner*.

<https://nodepit.com/node/org.knime.base.node.preproc.crossjoin.CrossJoinerNodeFactory>

Node Pit. (2020b). *Variable to Table Column*.

<https://nodepit.com/node/org.knime.base.node.flowvariable.appendvariabletotable2.AppendVariableToTable2NodeFactory>

Organización Panamericana de la Salud. (1995). *Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud* (Programa d).

Palmer Pol, A., & Montaña Moreno, J. . (1999). ¿Qué son las redes neuronales

artificiales? Aplicaciones realizadas en el ámbito de las adicciones. *Adicciones*, 243–255.

Pérez Tatamués, E. A. (2019). *Algoritmo de random forest aplicado a la detección de fraude en el sistema bancario ecuatoriano*.

Rammos, K., Papadimitriou, S., Virvou, M., & Alepis, E. (2018). A Web-based Application for Innovative Hospital. *9th International Conference on Information, Intelligence, Systems and Applications (IISA)*.

Riquelme Santos, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial*, 11–18.

Román, V. (2019). *Aprendizaje Supervisado: Introducción a la Clasificación y Principales Algoritmos*. <https://medium.com/datos-y-ciencia/aprendizaje-supervisado-introducción-a-la-clasificación-y-principales-algoritmos-dadee99c9407>

Sadki, A., Xie, X., & Chauvin, F. (2011). Appointment Scheduling of Oncology Outpatients. *2011 IEEE International Conference*.

Saibene, B., & Yanina, N. (2014). *Research Gate*. https://www.researchgate.net/figure/Fases-del-proceso-de-CRISP-DM-Adaptado-de-10_fig2_306959832

Shadabi, F., & Sharma, D. (2008). *Artificial Intelligence and Data Mining Techniques in Medicine – Success*. 235–242.

Sinnexus. (2020). *Sistemas de Soporte a la Decisión (DSS)*. https://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx

Stan, O., Avram, C., Stefan, I., & Astilean, A. (2016). Integrated Innovative Solutions to Improve. *IEEE International Conference on Automation, Quality and Testing Robotics (AQTR)*.