



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Aplicación de minería de datos para mejorar el diagnóstico de un grupo de enfermedades respiratorias en un hospital de Trujillo

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

AUTORES:

Alva Mariños, Ronald Segundo (ORCID: 0000-0001-5816-417X)

Cruz Isla, Luis Felipe (ORCID: 0000-0001-9151-7731)

ASESOR:

Mtro. Cieza Mostacero, Segundo Edwin (ORCID: 0000-0002-3520-4383)

LÍNEA DE INVESTIGACIÓN:

Sistema de Información y Comunicaciones

TRUJILLO – PERÚ

2021

Dedicatoria

Dedicamos el presente trabajo de investigación a nuestros padres quienes con su amor, paciencia y esfuerzo nos han permitido llegar a cumplir hoy un sueño más, gracias por inculcarnos el ejemplo de esfuerzo y valentía, de no temer las adversidades porque Dios está conmigo siempre.

Agradecimiento

Agradecemos a todas las autoridades y personal que forman parte de CAP III Metropolitano Trujillo, por confiar en nosotros, abrirnos las puertas y permitirnos realizar todo el proceso investigativo dentro de su establecimiento.

Índice de contenido

Carátula.....	i
Dedicatoria	ii
Agradecimiento	iii
Índice de Contenidos	iv
Índice de tablas	v
Índice de figuras	vii
Resumen.....	viii
Abstract.....	ix
I. INTRODUCCIÓN.....	1
II. MARCO TEÓRICO	6
III. METODOLOGÍA.....	17
3.1. Tipo y diseño de investigación	17
3.2. Variables y operacionalización.....	17
3.3. Población, muestra y muestreo.....	18
3.4. Técnicas e instrumentos de recolección de datos.....	18
3.5. Procedimiento	21
3.6. Método de análisis de datos.....	23
3.7. Aspectos éticos	27
IV. RESULTADOS	28
V. DISCUSIÓN.....	42
VI. CONCLUSIONES.....	44
VII. RECOMENDACIONES.....	45
REFERENCIAS.....	46

Índice de tablas

Tabla 1. Tabla resumen de validación – indicador nivel de morbilidad	19
Tabla 2. Tabla resumen de validación – indicador tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias	20
Tabla 3. Tabla resumen de validación – indicador costo promedio de diagnóstico de un grupo de enfermedades respiratorias.....	20
Tabla 4. Hipótesis para el indicador nivel de morbilidad	23
Tabla 5. Hipótesis para el indicador tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias.....	24
Tabla 6. Hipótesis para el indicador costo promedio de diagnósticos de un grupo de enfermedades respiratorias.....	25
Tabla 7. Fechas de recolección de datos por tipo	28
Tabla 8. Medidas descriptivas del indicador nivel de morbilidad.....	28
Tabla 9. Prueba de normalidad del indicador nivel de morbilidad	29
Tabla 10. Hipótesis para el indicador nivel de morbilidad	30
Tabla 11. Prueba de rangos con signo de Wilcoxon del indicador nivel de morbilidad	30
Tabla 12. Prueba Z del indicador nivel de morbilidad.....	31
Tabla 13. Medidas descriptivas del indicador tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias	32
Tabla 14. Prueba de normalidad del indicador tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias.....	33
Tabla 15. Hipótesis para el indicador tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias.....	33
Tabla 16. Prueba de rangos con signo de Wilcoxon del indicador tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias.....	34
Tabla 17. Prueba Z del indicador tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias.....	34
Tabla 18: Medidas descriptivas del indicador costo promedio de diagnóstico de un grupo de enfermedades respiratorias.....	36
Tabla 19. Prueba de normalidad del indicador disminuir el costo promedio de diagnóstico de un grupo de enfermedades respiratorias.....	37

Tabla 20. Hipótesis para el indicador costo promedio de diagnóstico de un grupo de enfermedades respiratorias.....	37
Tabla 21. Prueba de rangos con signo de Wilcoxon del indicador costo promedio de diagnóstico de un grupo de enfermedades respiratorias.....	38
Tabla 22. Prueba Z del indicador costo promedio de diagnóstico de un grupo de enfermedades respiratorias.....	38
Tabla 23. Indicadores con resultado esperado antes de la implementación y después de la implementación	41
Tabla 24. Hipótesis general.....	41

Índice de figuras

Figura 1. Diseño de investigación	17
Figura 2. Indicador Nivel de morbilidad – Antes y después de la implementación	29
Figura 3. Aceptación de la hipótesis nula del indicador.....	31
Figura 4. Indicador tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias – Antes y después de la implementación.....	32
Figura 5. Aceptación de la hipótesis alterna del indicador	35
Figura 6. Indicador costo promedio de diagnóstico de un grupo de enfermedades respiratorias – Antes y después de la implementación	36
Figura 7. Aceptación de la hipótesis alterna del indicador	39
Figura 8. Resultado general antes de la implementación y después de la implementación	40

Resumen

El siguiente trabajo de investigación tuvo como objetivo general mejorar el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano de Trujillo con la implementación de una aplicación de minería de datos. En esta ocasión se realizó una investigación de grado pre-experimental; haciendo uso de herramientas de recolección de datos, como fichas de registro y fichas de observación, las cuales fueron validadas por juicio de expertos, y cuya confiabilidad se utilizó el coeficiente V de Aiken que dio como resultado un 95% de validez. En el proceso del proyecto se dio uso a tecnologías como Pentaho Data Integration, Pentaho Server y MySQL. Una vez con el desarrollo del software se siguieron las fases de la metodología CRISP-DM, que comprende: Comprensión de negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación y Despliegue. Se contó con una población de 56 pacientes, de la cual la muestra tomada fue de 19 pacientes; los resultados obtenidos después de la implementación de la aplicación de minería de datos fue el aumento en el nivel de morbilidad de un 4.96%, así mismo la reducción en el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias, en 20 minutos, también se redujo costo promedio de diagnóstico de dicho grupo en S/327.95. Finalmente se logró una mejora significativa en el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano de Trujillo.

Palabras clave: Diagnósticos, Morbilidad, Tiempo de diagnósticos, Costos de diagnósticos.

Abstract

The following research work had the general objective of improving the diagnosis of a group of respiratory diseases in the CAP III Metropolitan Primary Care Center of Trujillo with the implementation of a data mining application. On this occasion, a pre-experimental research was carried out; making use of data collection tools, such as registration cards and observation cards, which were validated by expert judgment, and whose reliability the Aiken V coefficient was used, which resulted in 95% validity. Technologies such as Pentaho Data Integration, Pentaho Server and MySQL were used in the project process. Once with the development of the software, the phases of the CRISP-DM methodology were followed, which includes: Business understanding, Data understanding, Data preparation, Modeling, Evaluation and Deployment. There was a population of 56 patients, of which the sample taken was 19 patients; The results obtained after the implementation of the data mining application was the increase in the level of morbidity of 4.96%, as well as the reduction in the average time to identify the existence of a group of respiratory diseases, in 20 minutes, The average cost of diagnosis for said group was also reduced by S / 327.95. Finally, a significant improvement was achieved in the diagnosis of a group of respiratory diseases at the CAP III Metropolitan Primary Care Center in Trujillo.

Keywords: Diagnoses, Morbidity, Diagnosis time, Diagnosis costs.

I. INTRODUCCIÓN

Al año 2020, el avance sobre el proceso de diagnóstico de enfermedades respiratorias ha ido en aumento, tales como el mejoramiento de sus técnicas como el estudio por imagen y exámenes biológicos, no obstante, aún sigue presente la dificultad de este, viéndose perjudicado la integridad del paciente. Asimismo, el Instituto de Medicina Americana de Texas encontró que, la causa de 400 mil muertes fue provocado debido al procesos mal elaborados, generando errores médicos prevenibles (To err IS human, 2016).

Además, al año 2019, la Organización Mundial de la Salud o también conocida como “OMS”, informó sobre cómo se vieron perjudicados más de 138 millones de pacientes por errores médicos, demostrando la existencia permanente hasta la actualidad. Las presentes estadísticas fueron referentes países de bajo y medio nivel de ingresos, abarcando el 80% de los habitantes globalmente (Gestión, 2019). De igual manera, en Estados Unidos, dicha organización llevo a cabo un estudio donde se evidenciaron errores de diagnóstico, generando alrededor de un 10% de mortalidad (Índice de personas que mueren) en los hospitalizados, lo cual represento entre el 6 % y 17 % de daños a la integridad del paciente (OMS, 2019).

En Cuba, en el Centro Provincial de Información de Ciencias Médicas de Camagüey, se realizó un estudio logrando detectar que las Infecciones Respiratorias Agudas (IRA), son causa principal de morbilidad (Índice de personas que se enferman); además fueron las más consultadas en personas de todas las edades, donde se determinó que estas infecciones afectaban principalmente a menores de 15 años (Reus y Ortiz, 2013).

Por otro lado, una investigación realizada en el Hospital Distrital de Sayaxché de Guatemala, se encontró como causante principal en las consultas médicas una fiebre elevada o en el peor de ellos caso un bronconeumonía en menores de cinco años; además mencionaron con enfoque que la causa entre el 80 y 90% de Infecciones Respiratorias Agudas (IRA), son de causa viral y se ha demostrado la ineficacia del tratamiento precoz con antibióticos, generando una alta resistencia de las enfermedades y el encarecimiento de los tratamientos (Juy et al., 2014).

En el Perú, una publicación del Diario El Peruano (2020), indicó que, el cambio climático (bajas temperaturas) generó un aumento de enfermedades respiratorias, entre ellas y la más común en ese entonces, era la gripe, el cual era confundido por el COVID-19, debido a la similitud de sus síntomas como: elevada fiebre, tos seca y problemas respiratorios, ocasionando una menor precisión a la hora de diagnosticar dicho problema.

Además, El Ministerio de Salud (MINSA) difundió que, en el año 2019, la mortalidad de la neumonía ha incrementado, alcanzado a un total de 38 mil casos nuevos en los niños menores de 5 años, siendo la población con nivel de riesgo alto (Expreso, 2019).

Por otro lado, en una publicación realizada por el Diario Peru21 (2019) mencionó alrededor de los meses de abril y agosto los meses de abril y agosto (épocas de temperaturas bajas), los niños de esas mismas edades, presentaron 42,073 casos de Infecciones Respiratorias Agudas (IRA), lo cual fue un factor determinante en el aumento de casos de dicho grupo de enfermedades.

Asimismo, la Oficina General de Estadística e Informática del Ministerio de Salud menciona que, los hospitales se incorporó variadas alternativas con base de apoyo tecnológico sobre los campos conocidos, entre ellos “TeleSalud” ofreció para las zonas rurales de las regiones de La Libertad una atención a distancia con la aplicación de citas en línea, teniendo al inicio un programador que brindaba el historial médico. Se usaron diversas Tecnologías de información y Comunicación ante la existencia de problemas y así darle pronta solución, aunque este inconveniente se mantuvo vigente (Curioso, 2014).

Una publicación de Andina (2018), indicó que la Contraloría General de la Republica llevo a cabo una auditoría al Hospital Regional Docente de Trujillo, en donde se reflejó un inadecuado seguimiento al realizar la totalidad de los exámenes auxiliares requeridos, aumentando el riesgo de definir un mal diagnóstico y tratamiento, lo que generó gastos adicionales a las personas que cuentan Seguro Integral de Salud (SIS).

En la entrevista no formal realizada en el Centro de Atención Primaria CAP III Metropolitano Trujillo realizada en el mes de junio del 2020, se dedicaba a realizar consultas externas; estuvo ubicada en la avenida Francisco Adrianzen 326 de dicha ciudad y la administradora era la Ing. Margarita Sairitupac Paredes. La entidad mencionada con anterioridad contaba con un total de 67 trabajadores, quienes se encargaban de los diferentes servicios como: Medicina General, Medicina Interna, Terapia, Obstetricia y Pediatría, donde se identificó un incremento de casos, por lo que, en la mayoría de estos fueron enfermedades respiratorias, los cuales afectaron a niños entre 1 a 5 años, viéndose reflejado la poca eficacia en el proceso de diagnóstico médico y tratamientos en el departamento de Pediatría de dicho conjunto de enfermedades, ocasionando un aumento en el nivel de morbilidad, el promedio era nivel del tiempo que se requería para conocer una presencia de enfermedades respiratorias en conjunto y costo promedio de diagnóstico del dicho grupo.

Basándonos en la realidad mencionada anteriormente, se planteó la pregunta de investigación: ¿De qué manera una aplicación de minería de datos influye en el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano de Trujillo?

Esta investigación se justificará teóricamente porque gracias a la implementación de la aplicación de minería de datos, disminuirá el nivel de morbilidad, también el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias y el costo promedio de diagnóstico de dicho grupo. De la misma forma se justificará metodológicamente donde se usará las fichas de observación y de registro para analizar los cambios del nivel de morbilidad, mejorando los tiempos promedio para identificar la existencia de un grupo de enfermedades respiratorias y el costo promedio de diagnóstico de dicho grupo. También, se justificará de forma práctica, debido que se propondrá una solución a un problema que afecta a los niños del Perú, reduciendo así el nivel de morbilidad, logrando disminuir el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias y el costo promedio de diagnóstico del grupo mencionado anteriormente.

La investigación tuvo como principal objetivo mejorar el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano de Trujillo, a través de la implementación de una aplicación de minería de datos, y como objetivos específicos: disminuir el nivel de morbilidad, expuestas bajo la inoportuna situación de riesgo en diagnósticos y tratamiento, disminuir el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias bajo la deficiencia de los exámenes médicos y la falta de precisión al realizar al definir una situación de salud, por último, disminuir el costo promedio de diagnóstico del grupo mencionado anteriormente, consecuente del costo extra en diversos exámenes complementarios que darían solución al diagnóstico respectivo.

Por lo mencionado anteriormente, la presente investigación buscó mejorar el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano de Trujillo, a través de la implementación de una aplicación de minería de datos en el año 2020.

II. MARCO TEÓRICO

Orientado en el objeto de estudio, Según Juy et al. (2014) con el estudio titulado “Morbilidad por Infecciones Respiratorias Agudas en menores de 5 años”, presentado en MEDISAN. Estudio realizado con un enfoque descriptivo y transversal en el hospital de Sayaxche, cuyo objetivo fue identificar la causa de consultas e ingresos por Infecciones Respiratorias Agudas, una población que cuenta con 253 pacientes (menor de 5 años de edad), con diagnóstico de Infecciones Respiratorias Agudas (IRA), se obtuvieron como resultados que, el primer motivo de ingreso en los hospitales, fueron manifestaciones en fiebre, representando en un 98% de los casos, el cual tuvo un impacto mayor en diagnósticos de neumonía con un 88.6% en niños varones menores de un año. Se concluyó que la primera causa de consultas fueron los diagnósticos por Infecciones Respiratorias Agudas en niños menores de 5 años.

Asimismo, Ordóñez (2019) cuya investigación titulada “Situación epidemiológica de las Infecciones Respiratorias Agudas (IRA) en el Perú”, presentado en el “Centro Nacional de Epidemiología, Prevención y Control de Enfermedades”. Se obtuvo como resultado sobre el impacto de episodios de neumonía que, el 44.3% de los casos reportados fueron en menores de 5 años, una diferencia reveladora siendo que, el 28.9% fueron mayores de 60 años (medidas realizadas en la libertad), concluyendo que la población o segmento más vulnerable son los niños menores de 5 años.

Estos estudios brindaron la identificación de la unidad de estudio, a base a la magnitud significativa reportada sobre niños menores de 5 años, como unidad a trabajar en la presente tesis.

La investigación de Cordova et al. (2020) llamada "Prevalencia de Infecciones Respiratorias Agudas en niños menores de 5 años", presentada en el Centro Materno-Infantil, ubicado en Lima (Perú). Tiene un enfoque de tipo descriptivo, con el objetivo de la prevalencia de Infecciones Respiratorias Agudas (IRA) en niños menores de 5 años, con 4050 historias clínicas como población, dichos datos se recolectaron mediante fichas de registros. Se obtuvieron como resultados que, la prevalencia de IRAS en el centro materno-infantil fue de un 30.47% casos reportados, dentro de esos porcentajes, impactó más en los niños de menores de 3 años. Se concluyó que la prevalencia de Infecciones Respiratorias Agudas afectó mayormente a niños menores de 3 años de edad.

Como aval de la presente investigación, esta información contribuyó como fuente de justificación en el origen de datos, para medir el objetivo llamado: "nivel de morbilidad".

Según Saldaña (2016) con la investigación "Modelo predictivo de minería de datos de apoyo a la gestión hospitalaria sobre la morbilidad de pacientes hospitalizados". Se concluyó que el modelo predictivo de minería de datos, ayudó notablemente en la distribución de recursos, generando una mejora en la condición de su servicio calidad de servicio y reduciendo todo nivel en morbilidad dentro de esta región de la investigación, mediante futuras planificaciones en medicamentos o insumos en el hospital VRG (Víctor Ramos Guardia).

Esta investigación ayudó a justificar que la minería de datos ayuda a reducir la morbilidad en pacientes.

También los autores Marimón y Navarro (2017) con el estudio titulado “Métodos de diagnóstico rápido de las infecciones respiratorias”, presentada por la Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica. Realizado el estudio con un enfoque descriptivo, tuvo el objetivo de realizar evaluaciones de las técnicas de diagnóstico en las Infecciones Respiratorias (IR) siendo comunes; se obtuvo como resultados, sobre las técnicas de diagnóstico rápido poseen fácil ejecución, pueden durar entre 15 – 30 minutos, teniendo la duración menos eficiente comparado con la ejecución la misma técnica en otros instantes. Se concluyó que las técnicas de diagnóstico son menos eficientes.

Con fines de aval de la presente investigación; este antecedente contribuyó como medio de justificación para el objetivo llamado "tiempo promedio de identificación de un grupo de enfermedades respiratorias".

Por ello Bernal, Carvajal y Alvis (2017)) con la investigación titulada como “Costos económicos de la infección respiratoria aguda en un Municipio de Colombia”, presentada en la Universidad Santo Tomás de Bogotá, ubicado en Colombia, con un análisis de población de 16 087 habitantes, a través del enfoque descriptivo aplicando el proceso llamada “micro-costeo”, para realizar una estimación de costos económicos enfocado en Infecciones Respiratorias Agudas (a base de los registros de facturación de cada paciente), se obtuvo como resultado que, la carga significativa de municipios enfocado desde lo económico fue alcanzando a más de 49 mil pesos en costos totales en solo consultas. Se concluyó que esta categoría de infecciones mencionada, representaron una gran carga económica para los municipios.

Con fines de aval de la presente investigación; este antecedente contribuyó como medio de justificación para el objetivo llamado "Costo promedio de diagnóstico de un grupo de enfermedades respiratorias".

También, según Dávila y Sánchez (2012) con la investigación titulada “Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas”, cuyo objetivo ha sido el estudio y análisis de las técnicas (clasificación y agrupamiento), presentada en Revista Cubana de Informática Médica, se concluyó que las técnicas de información a base modelos matemáticos, ayudó al proceso de análisis de información de los especialistas en la toma de decisiones médicas.

Este estudio determinó las posibles técnicas relacionadas más adecuadas para el tratado de la información para la presente tesis.

Además, según Calabrese et al. (2019) llevo a cabo una estudio titulado “Guía para evaluar calidad de datos basada en ISO/IEC 25012” que como objetivo era crear la guía de medición base el proceso definido en ISO/IEC 25040 a una base de datos con acceso libre como caso de estudio, concluyendo sobre su análisis que en base a la norma ISO/IEC 25012 resulta oneroso y complicado adquirirlo si se trata de que presente careciente experiencia; pero la ayuda brindada es significativa para el análisis a través de una accesible aplicación para todo el usuario que desee obtenerlo.

Esta investigación ayudó a establecer como una guía de calidad de datos para la presente tesis.

Según Racero (2018), con el estudio titulado “Epidemiología de la faringitis aguda bacteriana en un hospital general de agudos”, presentada en Argentina. Dicho estudio realizado con una orientación descriptivo y transversal, como objetivo es identificar la causa de consultas e ingresos por esta enfermedad, con una población de 3246 muestras diagnosticadas en el laboratorio de bacteriología del HIGA Evita de Lanús, se obtuvieron como resultado que el 64.8% fueron diagnosticadas con Faringitis Aguda bacteriana (FAB) en menores de 15 años. Se concluyó que la enfermedad es una de las causas más frecuentes de consultas médicas.

Este estudio mencionado, proporcionó la información ideal para esclarecer la enfermedad a trata, con fines de investigación, basándose en los numerosos casos impactados en menores de 5 años.

Por ello, según Alomía (2019), con el estudio titulado “Infecciones respiratorias agudas en infantes menores de 5 años del Centro de Salud Javier Loyola, Ecuador”, presentada en Ecuador. Dicho estudio realizado con un enfoque descriptivo y retrospectivo, cuyo objetivo fue determinar las patologías respiratorias infecciosas y las características clínicas principales en niños menores de 5 años atendidos en el Centro de salud de la investigación, con una población de 146 pacientes, se obtuvieron como resultado que la edad más frecuente en grupo fue de 2 a 5 años varones con un 52.7% con el diagnóstico de rinofaringitis. Se concluyó que la enfermedad más frecuente de tipo IRA es rinofaringitis con predominio en varones menores de 5 años.

Este estudio mencionado, proporcionó la información ideal para esclarecer la enfermedad a tratar, con fines de investigación, basándose en los numerosos casos impactados en menores de 5 años.

Además, según Bernal, Carvajal y Alvis (2017), con el estudio titulado “Costos económicos de la infección respiratoria aguda en un Municipio de Colombia”, dada a conocer en Colombia. Se realizó desde el planteamiento descriptivo y transversal por medio de una técnica micro-costeo, cuyo objetivo era valorar costos económicos sobre infecciones respiratorias agudas (IRA) en el municipio, con una población de 1576 pacientes, se obtuvieron como resultado que el 81.4% en consulta externas por IRA, llevando por un costo promedio por paciente hospitalizado a \$759.437. Se concluye que las IRAS son una elevada carga económica al municipio que se realizó la investigación.

Este estudio mencionado, ayudó a establecer una enfermedad del grupo a tratar, basándose en el impacto económico por la enfermedad.

Por consiguiente, se compartirá definiciones indispensables que guardan relación con el objeto de estudio, como el análisis de salud da conocimiento sobre el perfil de salud-enfermedad, basándose en tendencias y daños que identifican los riesgos y así diseñar una estrategia de solución ante dicho problema Águila et al. (2019).

Un diagnóstico según Ramos (2006) o como el autor lo definió como Diagnóstico de la Situación de Salud (DSS), el cual dicho medio instaurado en la población o comunidad, permitiendo tener un conocimiento donde se analizará el estado de salud, pero es importante tener en cuenta que la fase final de dicho diagnóstico es el control o solución del problema presente en el paciente. En cambio, para el Diccionario de cáncer (2011) es el procedimiento aplicado a una población o comunidad, permitiendo tener un conocimiento donde se analizará el estado de salud, pero es importante tener en cuenta que la fase final de dicho diagnóstico es el control o solución del problema presente en el paciente.

Una enfermedad según en la Real Academia Española (2020) es una variación o alteración leve o grave de salud, también se define como anomalía del mismo. De acuerdo con ello y como complemento, una enfermedad respiratoria según el Diccionario de cáncer (2011), es aquella complicación dañina en todo el aparato respiratorio, causado por el consumo excesivo del tabaco, presencia de neumonía, asma y enfermedad obstructiva crónica.

La patogenia según Real Academia Española (2020) definida como todo estudio sobre el origen y desarrollo en enfermedades. Por otro lado, un error médico según Alvarado y Flores (2009) se le conoce como falla o error durante la evaluación provocando daño en la salud del paciente.

Una definición fundamental es la morbilidad, definido según la Real Academia Española (2020) como una cantidad de individuos afectados dentro de un contexto en un tiempo específico, otra definición dada por el Ministerio de Salud (2018) señalan que son criterios que miden la prevalencia de una enfermedad, la cual se ve relacionada ciertos fallos durante su diagnóstico Alvarado y Flores (2009).

La OPS (2016) define la gravedad de la enfermedad cómo cada enfermedad presenta un cierto nivel de gravedad, dependerá mucha de la edad, estilo de vida, herencia genética y estado en el que se diagnostica al paciente dicha enfermedad. El paciente puede recaer nuevamente o desencadenar otras enfermedades ligadas a la enfermedad principal.

Las reacciones adversas a medicamentos (RAM) según COFZ (2018) es una reacción a un medicamento que se produce tras la administración de un fármaco. Por otro lado, la faringitis según Pavez et al. (2019) es una infección en la faringe y es uno de las enfermedades más consultadas en pediatría y médicos de atención primaria. También la rinofaringitis según la Biblioteca Nacional de Medicina de los Estados Unidos (2010) se conoce como infección común dada en primera infancia, es contagiosa que a su vez es causante inicial en consulta en pediatría, su tratamiento tiene un coste elevado.

La bronquitis aguda según Dubón (2016) es una inflamación de la tráquea de origen viral y sus síntomas pueden presentar como tos constante, también Duque (2008) define que una de las causas más comunes en los hospitales de consulta externa.

Los exámenes auxiliares o complementarios según Ortiz (1996) son aquellos exámenes que proporcionan datos clínicos acerca de las condiciones internas sobre el paciente. Otro término para definir es sobre el tratamiento farmacológico que según el Instituto Nacional del Cáncer (2020) consiste en consumir sustancias que ayudan a disminuir o aliviar dichos síntomas generados por alguna enfermedad u estado anormal que presente el paciente en tratamiento.

Para una correcta codificación de las enfermedades, emplearon el manual denominado Clasificación Internacional de Enfermedades (CIE) versión 10, según el Minsalud (2015) es la clasificación de enfermedades, lesiones y motivos de consulta que deben cumplir ciertos criterios que se encuentran estipulados. Comprende el rango de enfermedades con terminología médica (nomenclatura internacional de enfermedades).

Teniendo en cuenta las definiciones base previamente mencionado, según Fernández (2014) hace énfasis sobre la prevalencia por su falta de conocimiento, automedicación y con el ello incertidumbre en el diagnóstico –patogenia sobre las infecciones respiratorias. Se evidenció un tratamiento precoz con uso de los antibióticos que generó la presencia de otros fenómenos con base de resistencia bacteriana y finalmente que esto sea inaccesible, esto sobre las infecciones respiratorias.

A continuación, definiremos cada una de las enfermedades escogidas del grupo de infecciones respiratorias agudas. Según OMS (2020) define la tuberculosis daña principalmente los pulmones que es provocado por una bacteria denominada “ *Mycobacterium tuberculosis*” y se llega a transmitir al estar en contacto con la persona que padece de dicha infección.

Además, el Diccionario de cáncer (2020) define que la bronquitis es una afección pulmonar que ataca los bronquios, causando inflamación y cicatrices que como consecuencia produce en los bronquios elevadas cantidades de moco, lo que conlleva a la vez a una tos crónica con problemas respiratorios, por otro lado, se conoce que es causa principal de dicha enfermedad, el consumo de cigarrillo, exponerse al aire contaminado por vapores químicos o tabaco. El tipo común de bronquitis es la bronquitis crónica, al ser una enfermedad pulmonar obstructiva no se logra desaparecer en su totalidad del organismo de la persona. También Ependium (2020) define que neumonía adquirida en una comunidad es toda enfermedad reconocida por sus síntomas de carácter infección aguda en las vías respiratorias bajas y opacidades que se ven en una radiografía del tórax que hasta ahora se desconoce su origen.

A continuación, se presentaron varias definiciones, términos indispensables que ayuden a la comprensión en la investigación, en contexto del campo de acción, una de ellas, y muy importante es la minería de datos, definido por el autor Velarde (2003), como instrumento útil al explorar extensas bases de datos, utilizando modelos de agrupación, predictivos y segmentación, como una herramienta de mucha utilidad para la exploración de grandes bases de datos, utilizando modelos de agrupación, predictivos, y de segmentación, en el conjunto presente hacia la ayuda de tomar decisiones dentro de diversas entidades. Agregando, para insightsoftware (2019) dicha herramienta implica un proceso donde extraen información variada según las fuentes consultadas como puntos de venta, máquinas con capacidad Internet de las cosas (IoT) o herramientas de logística, convirtiéndolo en información útil; generando conocimiento a las empresas, además, aclaró que es toda herramienta integral de Inteligencia de negocios (BI). Pero basándose de la gran aplicabilidad de la herramienta, está asociada ante un medio existente, una base de datos, según Date (2001) es aquel contenedor que resguarda cualquier tipo de información; una colección organizada de datos perteneciente a un mismo contexto o tema en común y un usuario con acceso puede manipular la información; esa interacción realizada se denomina como Procesamiento de Transacciones en Línea (OLTP). Asimismo, Microsoft (2020) hoy en día lo define como, una herramienta para poder recopilar y organizar la información, pudiendo almacenar diferente información como personas, productos entre otros.

Las OLTP son sistemas diseñados con el propósito de procesar y almacenar acciones en forma de transacciones, de una manera eficiente a gracia de la normalización de los datos; su finalidad es prevalecer la integridad de la información, sin embargo, este tipo de base de datos, por su forma compleja de estructurar los datos, no ha sido diseñado con el propósito de responder las dudas de negocio mediante el análisis, en otras palabras no sería un modelo óptimo porque consumiría más requerimiento computacional por las consultas complejas (DML).

Es por ello que existe los sistemas Procesamiento Analítico en Línea (OLAP) (Microsoft, 2019), que son un diseño más adecuado al responder las incógnitas del negocio porque su forma de estructurar los datos, que está orientado a grandes almacenes de datos, proporcionando una gran aplicabilidad de análisis complejos (Microsoft, 2018).

Mediante la metodología se aplicará CRISP-DM o también dicho como Cross-Industry Standard Process for Data Mining, según (IBM, 2020) define que, es un método orientado a trabajos relacionados con la minería de datos, en el cual se compone en 6 fases. Además, se define como un modelo manejable o flexible ante cualquier cambio, e imparcial por la variedad de herramientas para el desarrollo siendo libre y gratuita ante su distribución (Leonard y Castro, 2013). También la metodología facilita a proyectos grandes o pequeños relacionados, siendo que sea fiable manejable y rápido de aplicar (Rivo et al., 2012)

Entre las 6 fases, La primera denominada "Compresión de negocio" según Botero (2019) y Colina (2017), con la finalidad llegar a las necesidades; describen los objetivos (requerimientos) desde una perspectiva de negocio; dentro del catálogo de actividades es la resolución de los objetivos comerciales, subdivididos como criterios de rendimiento comercial u objetivos comerciales, desglosado como análisis ante costo, gestión de riesgos y contingencias, etc.

La fase 2 denominada "Comprensión de los datos", mismo nombre que describe la fase; es el conocimiento y recolección de los datos, para la prevalencia de la calidad del mismo (detecta anomalías para mantener todo en orden); entre las actividades más impactantes son la recopilación, definición y exploración de los datos.

La tercera fase denominada “Preparación de datos” se conforma del en el preparado de los datos fines a realizar o utilizar, desarrollando una depuración y creación de variables clave para la transformación, entre las tareas más impactante de este punto en desarrollo son la evaluación de los datos ya verificados (realiza un análisis a la base de datos con función de evaluar la calidad de la información) y también realiza un análisis de exclusión e inclusión de la información extraída.

La cuarta fase denominada “Modelado”, en esta parte se realiza un proceso selectivo de técnicas más eficaces para el tratado de información y construcción del modelo de datos (dentro de ello se realiza una configuración y testeo de este).

La penúltima fase denominada “Evaluación”, es el proceso de contratación del producto realizado con las metas descritas (objetivos) en la primera fase del proyecto, una vez aprobado, también se realiza un catalogado de acciones posteriores u nuevos objetivos comerciales tras la entrega total del proyecto.

La sexta o última fase denominada como “Despliegue”, previamente ante una planificación uso técnico y mantenimiento general para un monitoreo constante del servicio, es el procedimiento en el cual, de toda la información obtenida, procesada y tratada se presenta a los usuarios en producción, mediante graficas o forma de contenido más fácil de comprender (esta presentación se realiza mediante algún aplicativo o reporte, etc).

La tecnología principal para la aplicación de este estudio, es sobre el conjunto de herramientas o Kit, llamado Pentaho (Kettle), que de acuerdo con Gravitar (2014) lo define como un conjunto de herramienta opensource, enfocado en el desarrollo de Inteligencia de Negocios o también conocido como BI, cuya finalidad es gestionar la información, y que compuesta con su conjunto de programas se convierte en un solución ideal para la toma de decisiones; uno de los programas es Pentaho Data Integration (PDI) que tiene la finalidad de extraer, tratar y cargar (también con el acrónimo de ETL), todo tipo de información necesaria con fines empresariales.

III. METODOLOGÍA

3.1. Tipo y diseño de investigación

La investigación es un tipo de estudio aplicada y sobre su diseño es Experimental de tipo preexperimental (Hernández,2003).

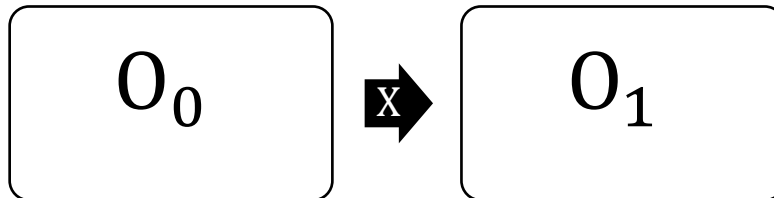


Figura 1. Diseño de investigación

Fuente: elaboración propia de los autores.

Dónde:

O_0 : diagnóstico de un grupo de enfermedades respiratorias antes de la implementación de una aplicación de minería de datos.

X : aplicación de minería de datos.

O_1 : diagnóstico de un grupo de enfermedades respiratorias después de la implementación de una aplicación de minería de datos.

3.2. Variables y operacionalización

La variable independiente consiste en la aplicación de minería de datos, mientras que la dependiente es el diagnóstico de un grupo de enfermedades respiratorias.

El cuadro de operacionalización de variables se encuentra en la sección de anexos (Anexo 1.1).

3.3. Población, muestra y muestreo

La investigación abordada presenta a 56 pacientes como población; la muestra es 19 niños entre 1 a 5 años, mediante la técnica llamada “muestreo por conveniencia” dentro del muestreo no probabilístico, como criterios de inclusión se seleccionó aquellos sujetos que tenían o que acudían por la enfermedad de Rinofaringitis aguda [resfriado común], Faringitis aguda, no especificada y Bronquitis aguda, no especificada, que tienen un historial médico en la entidad y hayan sido atendidos por consulta externa, en lo correspondiente a criterios de exclusión se descartaron aquellos niños que no cumple los criterios.

3.4. Técnicas e instrumentos de recolección de datos

De acuerdo con Huaman (2005) en su libro “Manual de técnicas de investigación Conceptos y Aplicaciones” definió que el fichaje es una técnica auxiliar que consisten en la recolección o recopilación de la información relevante de un estudio, volviéndolo un instrumento valioso, ahorra dinero y tiempo, la elaboración se realiza por ficha o formatos *produciendo* un valor propio, hoy en día es común recolectar dicha información útil para la investigación desde una base de datos, aunque las fichas de forma tradicional son cartulinas. Además, según Taffarel (2009) en su libro “La creación del conocimiento” indicó que el fichaje consiste en extraer y recopilar de manera ordenada la información que se requiere saber, permitiendo recolectar los datos para los indicadores de la investigación mencionados anteriormente.

Según López y Martel (2001) en su libro “La escritura en UOHH” definió que la ficha de registro es uno de los instrumentos de investigación que identifica toda información de las fuentes que han sido consultadas. Teniendo el aval de estos autores, se realizó una ficha de registro para los indicadores ya mencionados.

Según Carrasco (2005) en su libro “Metodología de la investigación científica” definió que la ficha de observación registra los datos que se obtienen del contacto directo entre el observador y la realidad observada. Se elaboro una ficha de observación para el indicador de tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias.

Para confirmar la validez de los instrumentos, se utilizó la técnica del juicio de expertos (Anexo 3,4 y 5), que según Ecurra (1988) en su artículo titulado “Cuantificación de la validez de contenido por criterio de jueces” consiste en requerir la aceptación o rechazo de los instrumentos por parte de varios expertos en el tema, cuyo número puede variar dependiendo de cada investigación y para la confiabilidad se utilizó el coeficiente V de Aiken (Anexo 6) el cual computa a partir de un dato conseguido sobre la suma máxima de los valores posibles, los cuales pueden ser calculados utilizando las valoraciones de un grupo de expertos con relación a los ítems, dichas valoraciones pueden ser dicotómicas donde reciben valores de (0 o 1) o politómicas con valores de (0 a 5).

Tabla 1. Tabla resumen de validación – indicador nivel de morbilidad

Expertos	Ficha de registro
	Nivel de morbilidad %
Mg. Icanaque Ordinola, Juan Damian	91.42
Qf. Moreno Zavaleta, Hilda Edith	100.00
Ing. Seguro Murga, Leopoldo	80.00
Total	90.47

Fuente: elaboración propia de los autores

En la tabla 1 visualizada se aprecia la ficha de registro que fue validada por los tres expertos y la calificación obtenida de la evaluación del indicador mencionado, obteniendo un promedio de 90.47% de validez, la cual se encuentra en la sección de anexos de la presente investigación (Anexo 7).

Tabla 2. Tabla resumen de validación – indicador tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias

Expertos	Ficha de Observación
	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias %
Mg. Icanaque Ordinola, Juan Damian	91.42
Qf. Moreno Zavaleta, Hilda Edith	100.00
Ing. Seguro Murga, Leopoldo	80.00
Total	90.47

Fuente: elaboración propia de los autores

En la tabla 2 se visualiza la ficha de observación que fue validada por los tres expertos y la calificación obtenida de la evaluación para el indicador mencionado, obteniendo un promedio de 90.47% de validez, la cual se encuentra en la sección de anexos de la presente investigación (Anexo 7).

Tabla 3. Tabla resumen de validación – indicador costo promedio de diagnóstico de un grupo de enfermedades respiratorias

Expertos	Ficha de Registro
	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias %
Mg. Icanaque Ordinola, Juan Damian	91.42
Qf. Moreno Zavaleta, Hilda Edith	100.00
Ing. Seguro Murga, Leopoldo	80.00
Total	90.47

Fuente: elaboración propia de los autores

En la tabla 3 se visualiza la ficha de registro que fue validada por los tres expertos y la calificación obtenida de la evaluación para el indicador mencionado, obteniendo un promedio de 90.47% de validez, la cual se encuentra en la sección de anexos de la presente investigación (Anexo 7).

3.5. Procedimiento

Primero se realizó un análisis base el panorama principal, se aplicó una entrevista informal no estructurado al gerente general del Centro de Atención Primaria CAP III Metropolitano Trujillo para poder conocer su punto de vista, el estado actual y/o realidad problemática. Luego se realizará un análisis descriptivo que consistirá en aplicarlo antes de la implementación del diagnóstico médico, donde se recopilará la información de dicho proceso en términos cuantitativos, tales como el nivel de morbilidad, realizadas al mes, tiempo de promedio para identificar la existencia de una enfermedad respiratoria realizadas al mes, costo promedio de diagnóstico de una enfermedad respiratoria realizadas al mes relacionados a la mejora del diagnóstico alcanzados utilizando la técnica de fichaje y como instrumentos la ficha de observación y las fichas de registro (anexos 4,5 y 6, para el proceso de validación se consultó con expertos en temas relacionadas al diagnósticos y errores médicos. Cabe señalar que la empresa brindó un documento como consentimiento para la realización de esta investigación la cual se encuentra en el Anexo 9 del presente informe. Para la implementación de la aplicación de minería de datos se utilizará la metodología CRISP-DM (Anexo 10), la cual comprende cinco fases, primero se realizará la primera fase que es la comprensión del negocio en la cual consistirá en una recopilación de información del hospital, definir los objetivos del negocio, definir los criterios de rendimiento comercial, después se aplicará la valoración de la situación en el cual es el definir requisitos, restricciones, riesgos y contingencias dentro de ello también está en analizar la terminología del negocio(esto es importante porque ayuda a mantener la relación base el contexto que ellos viven) y elaborar análisis de costes / beneficios también se aplicará la determinación de los objetivos estratégicos o de minería de datos que es definir la casuística del negocio conjuntamente relacionado con las técnicas de minería, finalizando la fase se aplicará la producción del plan de proyecto en el cual es el desarrollo del mismo realizando un hincapié en la valoración de herramientas y técnicas.

La segunda fase se aplicará la comprensión de los datos en el cual es la realización de la recopilación de datos iniciales, describir los datos obtenidos, aplicar una exploración de datos y por último una validación de calidad de datos.

En la tercera fase se efectuará la preparación de los datos en el cual es la realización de la selección que basa en la exclusión e inclusión de datos, depuración que son entregables, construcción que es derivación de atributos, integración y formateado de datos.

En la cuarta fase se realizará el modelado de datos en cual se definirán las técnicas a emplear para el modelado, se realizará un diseño de prueba y como resultado de estos pasos se elaborará un modelo de datos y una configuración de parámetros, concluyendo con la fase se evaluará el modelo de datos creado.

Como quinta fase se realizará una evaluación en la cual se constatará los criterios de rendimiento, se elaborarán resúmenes para actividades, decisiones y sugerencias de cada fase realizada en el proyecto. Como última fase se efectuará la distribución del modelo de datos, se elaborará un plan de distribución completa y precisa, un plan de control y mantenimiento para el modelo, se elaborará un informe final para la presentación del proyecto y como culminación del proyecto se efectuará una revisión final.

Posteriormente, luego de la implementación del modelo, se realizará la evaluación antes de la implementación, donde se extraerá la información necesaria (en términos cuantitativos a fin de estudio), con los instrumentos aplicados en la evaluación después de la implementación de los indicadores mencionados anteriormente.

Al culminar se determinará el impacto que tuvo la aplicación de minería en el diagnóstico de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo. Se realizará un análisis inferencial donde se comprobará la hipótesis planteada para cada indicador y para determinar la influencia, los datos obtenidos serán plasmados en gráficos y tablas para su correcto análisis y posteriormente se les aplicarán pruebas estadísticas como la prueba de normalidad y la contrastación de hipótesis.

3.6. Método de análisis de datos

Se plantearon las siguientes hipótesis específicas para cada indicador de esta investigación. A continuación, las mismas:

Tabla 4. Hipótesis para el indicador nivel de morbilidad

Indicador	Nivel de morbilidad
	<p>H_1: La aplicación de minería de datos disminuirá el nivel de morbilidad en el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p>H_0: La aplicación de minería de datos no disminuirá el nivel de morbilidad en el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p>
	<p>Donde:</p> <p>NMa: nivel de morbilidad antes de utilizar la aplicación de minería de datos.</p> <p>NMd: nivel de morbilidad después de utilizar la aplicación de minería de datos.</p>
	<p>Hipótesis Nula H_0: La aplicación de minería de datos no disminuirá el nivel de morbilidad en el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p style="text-align: center;">$H_0: NMa - NMd \leq 0$</p>
	<p>Hipótesis Alterna H_a: La aplicación de minería de datos disminuirá el nivel de morbilidad en el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p style="text-align: center;">$H_a: NMa - NMd > 0$</p>

Fuente: elaboración propia de los autores

Para el segundo indicador se plantea la hipótesis:

Tabla 5. Hipótesis para el indicador tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias

Indicador	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias
<p>H_1: La aplicación de minería de datos disminuirá el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p>H_0: La aplicación de minería de datos no disminuirá el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p>	
<p>Donde:</p> <p>TPIEGERa: tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias antes de utilizar la aplicación de minería de datos.</p> <p>TPIEGERd: tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias después de utilizar la aplicación de minería de datos.</p>	
<p>Hipótesis Nula H_0: La aplicación de minería de datos no disminuirá el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo</p> $H_0: \text{TPIEGERa} - \text{TPIEGERd} \leq 0$	
<p>Hipótesis Alterna H_a: La aplicación de minería de datos disminuirá el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> $H_a: \text{TPIEGERa} - \text{TPIEGERd} > 0$	

Fuente: elaboración propia de los autores

Para el último indicador se plantea la hipótesis:

Tabla 6. Hipótesis para el indicador costo promedio de diagnósticos de un grupo de enfermedades respiratorias

Indicador	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias
<p>H_1: La aplicación de minería de datos disminuirá el costo promedio de diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p>H_1: La aplicación de minería de datos no disminuirá el costo promedio de diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p>	
<p>Donde:</p> <p>CPDGERa: costo promedio de diagnóstico de un grupo de enfermedades respiratorias antes de utilizar la aplicación de minería de datos.</p> <p>CPDGERd: costo promedio de diagnóstico de un grupo de enfermedades respiratorias después de utilizar la aplicación de minería de datos.</p>	
<p>Hipótesis Nula H_0: La aplicación de minería de datos no disminuirá el costo promedio de diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p style="text-align: center;">$H_0: \text{CPDGERa} - \text{CPDGERd} \leq 0$</p>	
<p>Hipótesis Alterna H_a: La aplicación de minería de datos disminuye el costo promedio de diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p style="text-align: center;">$H_a: \text{CPDGERa} - \text{CPDGERd} > 0$</p>	

Fuente: elaboración propia de los autores

Análisis Descriptivo

En esta investigación se implementará una aplicación de minería de datos para evaluar el nivel de morbilidad, tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias y el costo promedio de diagnóstico de un grupo de enfermedades respiratorias en los pacientes del Centro de Atención Primaria CAP III Metropolitano Trujillo. Para realizar la medición de los indicadores propuestos se realizaron instrumentos como ficha de observación y de registro el cual se aplicará una evaluación antes de la implementación que permitirá conocer los objetivos mencionados anteriormente. Posteriormente se implementará una aplicación de minería de datos durante un mes, este tiempo permitirá que la aplicación influya en los indicadores anteriormente mencionados. Luego de la implementación se realizará la evaluación después de la implementación, para registrar la variación en los indicadores. Estos resultados se verán representados mediante gráficos de barras, gráficos lineales o tablas para cada de indicador, para su correcto análisis y evaluación.

Análisis inferencial

Los datos obtenidos en esta investigación se realizarán de la siguiente manera: en primer lugar, se optará por usar la técnica de prueba de normalidad para saber si los recopilado posee o no una distribución normal, esta misma será realizada con la prueba...Shapiro-Wilk utilizada para muestras menores a 35 como lo menciona Romero (2016) ,contamos con una población desconocida y en este caso la muestra es de 19 niños que tienen una enfermedad respiratoria y se utilizara la herramienta SPSS (versión 25), esta prueba se realizará para cada uno de los indicadores. Luego se determinará si lo recopilado posee o no una distribución normal, de acuerdo en ello se procederá a realizar la prueba de hipótesis correspondiente por cada indicador, usando la técnica T de Student o Wilconxon, dependiendo del resultado obtenido de la prueba de normalidad, se determinará que hipótesis es la que se acepta.

3.7. Aspectos éticos

La presente investigación cumple con el código de ética de la Universidad Cesar Vallejo, donde señala la participación igualitaria por partes de los pacientes del Centro de Atención Primaria CAP III Metropolitano Trujillo, así como la transparencia de la información obtenida, corroborando la autenticidad de la investigación por parte de los autores (Anexo 12), además se realizó cumpliendo estrictamente los requisitos éticos, legales y de seguridad.

IV. RESULTADOS

Análisis Descriptivo:

En la presente investigación se implementó una aplicación de minería de datos para mejorar el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano de Trujillo, para el cual se aplicó una evaluación antes de la implementación, en donde se ejecutaron los indicadores planteados que permitieron conocer el diagnóstico de un grupo de enfermedades respiratorias, se procedió con la implementación de la aplicación de minería de datos y después de la implementación se realizó nuevamente la evaluación. El resultado que se obtuvo al procesar la información recolectada se puede encontrar en la sección de anexos de este informe (anexo 14).

Tabla 7. Fechas de recolección de datos por tipo

Tipo	Fecha de Inicio	Fecha de Fin
Antes de la implementación	01/11/2020	30/11/2020
Después de la implementación	01/12/2020	31/12/2020

Fuente: Elaboración propia de los autores

A continuación, se muestra el análisis descriptivo e inferencial por cada indicador.

Indicador 1: Nivel de morbilidad

Análisis Descriptivo:

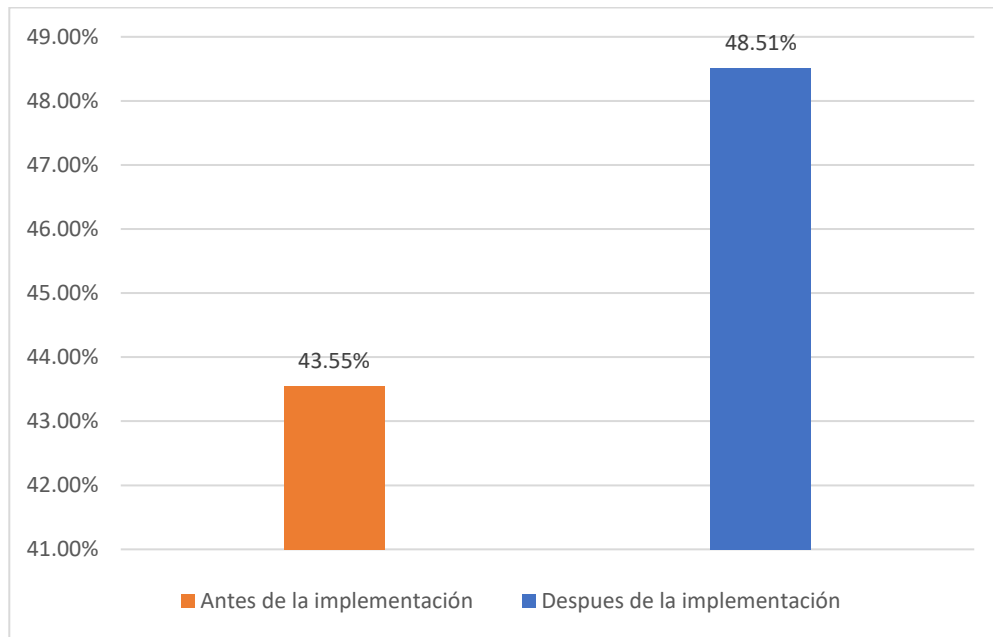
Tabla 8. Medidas descriptivas del indicador

Estadísticos descriptivos				
	N	Mínimo	Máximo	Media
Antes de la implementación	19	7.69	61.53	43.5579
Después de la implementación	19	9.09	100.00	48.5142
N válido (por lista)	19			

Fuente: Elaboración propia de los autores.

Se observa que en la tabla 8 obtuvo como nivel de morbilidad un mínimo 7.69% y como máximo 62% de morbilidad y de igual manera se observa que después de la implementación se obtuvo un mínimo de 9.09% y un máximo de 100% de morbilidad.

Figura 2. Indicador Nivel de morbilidad – Antes y después de la implementación



Fuente: Elaboración propia de los autores

Se observa en la figura 2 que el nivel de morbilidad antes de la implementación fue de 43.55% y después de la implementación fue de 48.51%. Como se puede visualizar, hay un aumento de 4.91% después de la implementación.

Análisis Inferencial:

Tabla 9. Prueba de normalidad del indicador

Pruebas de normalidad			
	Shapiro – Wilk		
	Estadístico	gl	Sig.
DIFERENCIA	.854	19	.008

a. Corrección de significación de Lilliefors

Fuente: Elaboración propia de los autores

Se observa en la tabla 9 que el "Sig." es de valor 0.008, cuyo valor es menor que 0.05, es por ello que los datos no cumplen o siguen una distribución normal, por eso se optó utilizar una prueba no paramétrica, la cual fue Wilcoxon por ser una población menor a 30.

Prueba de Hipótesis:

Tabla 10. Hipótesis del indicador

Indicador:	Nivel de morbilidad
<p>H_1: La aplicación de minería de datos disminuirá el nivel de morbilidad en el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p>H_0: La aplicación de minería de datos no disminuirá el nivel de morbilidad en el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p>	
<p>Donde:</p> <p>NMa: nivel de morbilidad antes de utilizar la aplicación de minería de datos.</p> <p>NMd: nivel de morbilidad después de utilizar la aplicación de minería de datos.</p>	
<p>Hipótesis Nula H_0: La aplicación de minería de datos no disminuirá el nivel de morbilidad en el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p style="text-align: center;">$H_0: NMa - NMd \leq 0$</p>	
<p>Hipótesis Alterna H_a: La aplicación de minería de datos disminuirá el nivel de morbilidad en el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p style="text-align: center;">$H_a: NMa - NMd > 0$</p>	

Fuente: Elaboración propia de los autores

Tabla 11. Prueba de rangos con signo de Wilcoxon del indicador

Rangos				
		N	Rango promedio	Suma de rangos
Después de la implementación - Antes de la implementación	Rangos negativos	5 ^a	12.80	64.00
	Rangos positivos	14 ^b	9.00	126.00
	Empates	0 ^c		
	Total	19		
a. Después de la implementación < Antes de la implementación				
b. Después de la implementación > Antes de la implementación				
c. Después de la implementación = Antes de la implementación				

Fuente: Elaboración propia de los autores

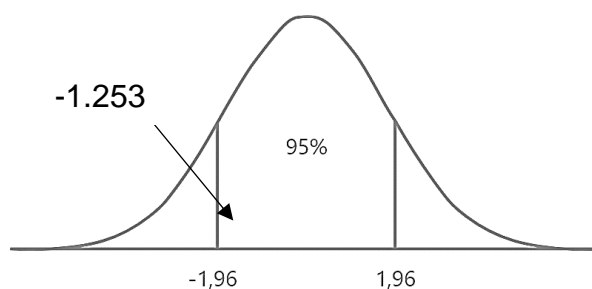
Tabla 12. Prueba Z del indicador

Estadísticos de prueba^a	
	Antes de la implementación - Después de la implementación
Z	-1.253 ^b
Sig. asintótica(bilateral)	.210
a. Prueba de rangos con signo de Wilcoxon	
b. Se basa en rangos positivos.	

Fuente: Elaboración propia de los autores

Se acepta la hipótesis nula con un 95% de confianza, donde la aplicación de minería de datos no disminuye el nivel de morbilidad en el diagnóstico de un grupo de enfermedades respiratorias, dado que $Z = -1.253$ así como p (Sig.) es mayor que 0.05 ($0.210 > 0.05$) y se rechaza la hipótesis alterna.

Figura 3. Aceptación de la hipótesis nula del indicador



Fuente: Elaboración propia de los autores

En la figura N° 3 se indica que el valor de $Z = -1.253$ se encuentra en la zona de aceptación de la campana de Gauss; por ello, se acepta la hipótesis nula de este indicador.

Indicador 2: Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias

Análisis Descriptivo:

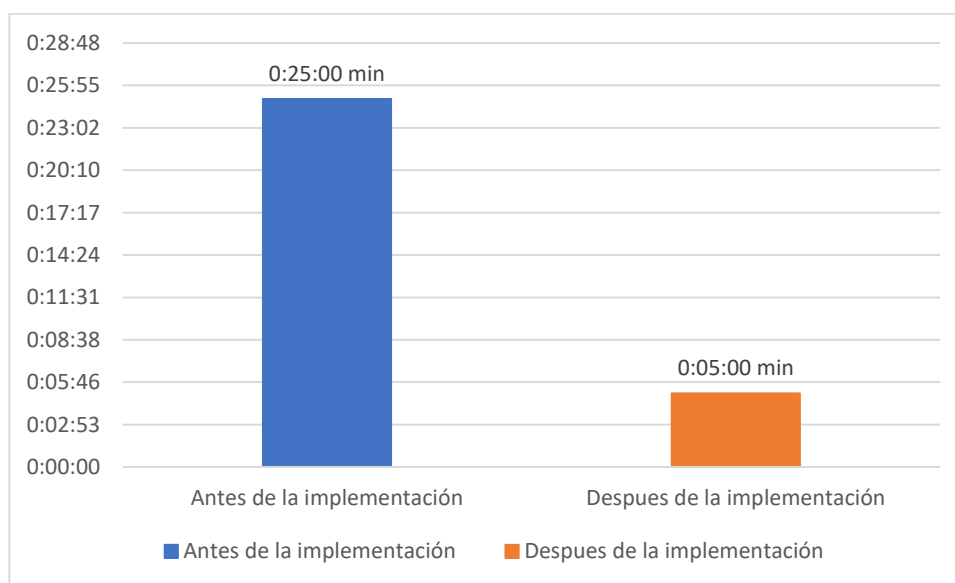
Tabla 13. Medidas descriptivas del indicador

Estadísticos descriptivos				
	N	Mínimo	Máximo	Media
Antes de la implementación	19	0.12	0.40	0.2553
Después de la implementación	19	0.00	0.07	0.0537
N válido (por lista)	19			

Fuente: Elaboración propia

Se observa que en la tabla 13 obtuvo como tiempo promedio un mínimo de 00:12 minutos y como máximo de 00:40 minutos, y de igual manera se observa que después de la implementación se obtuvo un mínimo de 0:00 minutos y un máximo de 0:07 minutos.

Figura 4. Indicador tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias – Antes y después de la implementación



Fuente: Elaboración propia

Se observa en la figura 4 que el tiempo promedio antes de la implementación fue 00:25 minutos y después de la implementación fue de 00:05 minutos. Como se puede visualizar hay una disminución de 00:20 minutos después de la implementación.

Análisis inferencial:

Tabla 14. Prueba de normalidad del indicador

Pruebas de normalidad			
	Shapiro – Wilk		
	Estadístico	gl	Sig.
DIFERENCIA	.729	19	.000

a. Corrección de significación de Lilliefors

Fuente: Elaboración propia

Se observa en la tabla 14 que el "Sig." es de valor 0.000, cuyo valor es menor que 0.05, es por ello que los datos no cumplen o siguen una distribución normal, por eso se optó utilizar una prueba no paramétrica, la cual fue Wilcoxon por ser una población menor a 30.

Prueba de Hipótesis:

Tabla 15. Hipótesis del indicador

Indicador:	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias
<p>H_2: La aplicación de minería de datos disminuirá el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p>H_0: La aplicación de minería de datos no disminuirá el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p>	
<p>Donde:</p> <p>TPIEGERa: tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias antes de utilizar la aplicación de minería de datos.</p> <p>TPIEGERd: tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias después de utilizar la aplicación de minería de datos.</p>	
<p>Hipótesis Nula H_0: La aplicación de minería de datos no disminuirá el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo</p> <p style="text-align: center;">$H_0: TPIEGERa - TPIEGERd \leq 0$</p>	

Hipótesis Alternativa H_a : La aplicación de minería de datos disminuirá el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.

$$H_a: TPIEGERa - TPIEGERd > 0$$

Fuente: Elaboración propia

Tabla 16. Prueba de rangos con signo de Wilcoxon del indicador

Rangos				
		N	Rango promedio	Suma de rangos
Después de la implementación - Antes de la implementación	Rangos negativos	19 ^a	10.00	190.00
	Rangos positivos	0 ^b	0.00	0.00
	Empates	0 ^c		
	Total	19		
a. Después de la implementación < Antes de la implementación				
b. Después de la implementación > Antes de la implementación				
c. Después de la implementación = Antes de la implementación				

Fuente: Elaboración propia de los autores

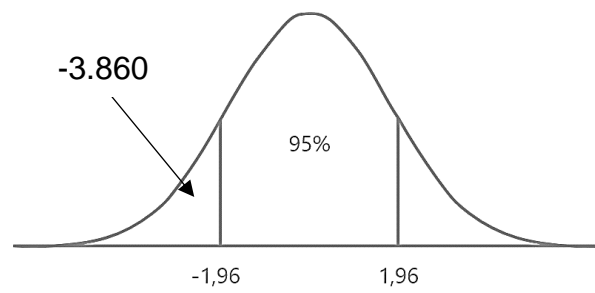
Tabla 17. Prueba Z del indicador

Estadísticos de prueba^a	
	Antes de la implementación - Después de la implementación
Z	-3.860 ^b
Sig. asintótica(bilateral)	.000
a. Prueba de rangos con signo de Wilcoxon	
b. Se basa en rangos positivos.	

Fuente: Elaboración propia de los autores

Se acepta la hipótesis alternativa con un 95% de confianza, donde la aplicación móvil multiplataforma incrementa la cantidad promedio de productos de primera necesidad en stock, dado que $Z = -3.860$ así como p (Sig.) es menor que 0.05 ($0.000 < 0.05$) y se rechaza la hipótesis nula.

Figura 5. Aceptación de la hipótesis alterna del indicador



Fuente: Elaboración propia de los autores

En la figura N° 5 se indica que el valor de $Z = -3.860$ se encuentra en la zona de aceptación de la campana de Gauss; por ello, se acepta la hipótesis alterna de este indicador.

Indicador 3: Costo promedio de diagnóstico de un grupo de enfermedades respiratorias

Análisis Descriptivo:

Tabla 18: Medidas descriptivas del indicador

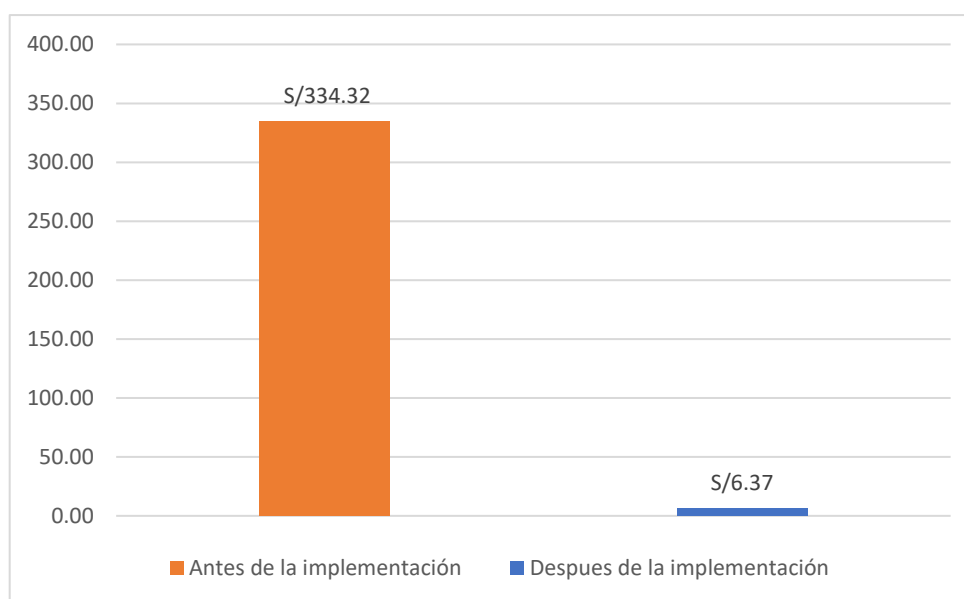
Estadísticos descriptivos				
	N	Mínimo	Máximo	Media
Antes de la implementación	19	6.00	417.00	334.3158
Después de la implementación	19	0.00	7.00	6.3684
N válido (por lista)	19			

Fuente: Elaboración propia

Se observa que en la tabla 18 obtuvo como costo promedio un mínimo de S/6.00 y como máximo de S/417.00, y de igual manera se observa que después de la implementación se obtuvo un mínimo de S/0.00 y un máximo de S/7.00.

Figura 6. Indicador costo promedio de diagnóstico de un grupo de enfermedades respiratorias – Antes y después de la implementación

Fuente: Elaboración propia



Se observa en la figura 4 que el costo promedio antes de la implementación fue de S/334.31 y después de la implementación fue de S/6.36. Como se puede visualizar hay una disminución de S/327.95 después de la implementación.

Análisis Inferencial:

Tabla 19. Prueba de normalidad del indicador

Pruebas de normalidad			
	Shapiro – Wilk		
	Estadístico	gl	Sig.
DIFERENCIA	.649	19	.000
a. Corrección de significación de Lilliefors			

Fuente: Elaboración propia

Se observa en la tabla 19 que el "Sig." es de valor 0.000, cuyo valor es menor que 0.05, es por ello que los datos no cumplen o siguen una distribución normal, por eso se optó utilizar la técnica de pruebas no paramétricas, la cual fue Wilcoxon por ser una población menor a 30.

Prueba de Hipótesis:

Tabla 20. Hipótesis del indicador

Indicador:	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias
<p>H_3: La aplicación de minería de datos disminuirá el costo promedio de diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p>H_0: La aplicación de minería de datos no disminuirá el costo promedio de diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p>	
<p>Donde:</p> <p>CPDGERa: costo promedio de diagnóstico de un grupo de enfermedades respiratorias antes de utilizar la aplicación de minería de datos.</p> <p>CPDGERd: costo promedio de diagnóstico de un grupo de enfermedades respiratorias después de utilizar la aplicación de minería de datos.</p>	
<p>Hipótesis Nula H_0: La aplicación de minería de datos no disminuirá el costo promedio de diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.</p> <p style="text-align: center;">$H_0: CPDGERa - CPDGERd \leq 0$</p>	

Hipótesis Alterna H_a : La aplicación de minería de datos disminuirá el costo promedio de diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano Trujillo.

$$H_a: \text{CPDGERa} - \text{CPDGERd} > 0$$

Fuente: Elaboración propia

Tabla 21. Prueba de rangos con signo de Wilcoxon del indicador

Rangos				
		N	Rango promedio	Suma de rangos
Después de la implementación - Antes de la implementación	Rangos negativos	19 ^a	10.00	190.00
	Rangos positivos	0 ^b	0.00	0.00
	Empates	0 ^c		
	Total	19		
a. Después de la implementación < Antes de la implementación				
b. Después de la implementación > Antes de la implementación				
c. Después de la implementación = Antes de la implementación				

Fuente: Elaboración propia

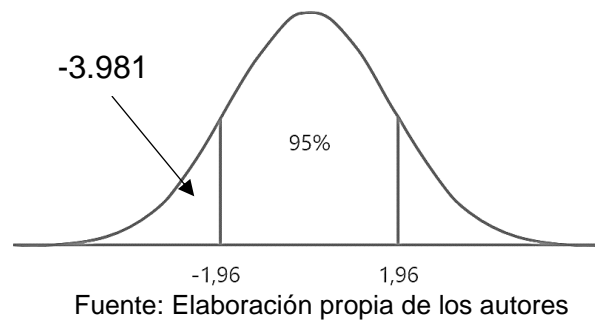
Tabla 22. Prueba Z del indicador

Estadísticos de prueba^a	
	Antes de la implementación - Después de la implementación
Z	-3.981 ^b
Sig. asintótica(bilateral)	.000
a. Prueba de rangos con signo de Wilcoxon	
b. Se basa en rangos positivos.	

Fuente: Elaboración propia

Se acepta la hipótesis alterna con un 95% de confianza, donde la aplicación móvil multiplataforma incrementa la cantidad promedio de productos de primera necesidad en stock, dado que $Z = -3.981$ así como p (Sig.) es menor que 0.05 ($0.000 < 0.05$) y se rechaza la hipótesis nula.

Figura 7. Aceptación de la hipótesis alterna del indicador

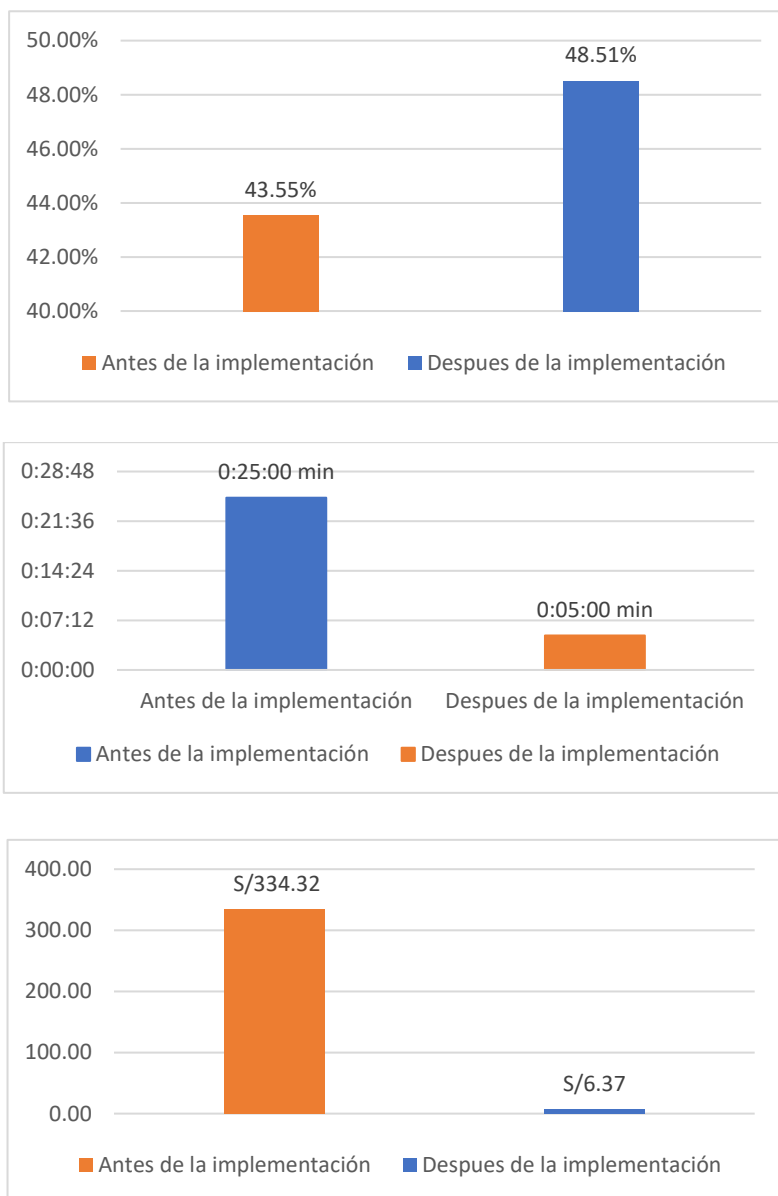


En la figura N° 7 se indica que el valor de $Z = -3.981$ se encuentra en la zona de aceptación de la campana de Gauss; por ello, se acepta la hipótesis alterna de este indicador.

Objetivo general

Mejorar el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano de Trujillo a través de la implementación de una aplicación de minería de datos.

Figura 8. Resultado general antes de la implementación y después de la implementación



Fuente: Elaboración propia

Los indicadores reflejan el resultado general antes de la implementación y después de la implementación por cada indicador según los objetivos específicos.

Tabla 23. Indicadores con resultado esperado antes de la implementación y después de la implementación

Indicadores	Antes de la implementación	Después de la implementación	Resultado Esperado
Nivel de morbilidad	43:55%	48:51%	Disminuir (-)
Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias	00:25:00 min	00:05:00 min	Disminuir (-)
Costo promedio de diagnóstico de un grupo de enfermedades respiratorias	S/334.32	S/6.37	Disminuir (-)

Fuente: Elaboración propia

En la tabla N° 23, se observa que en el indicador 1 no se logra el resultado esperado que es disminuir, en el indicador 2 se logra el resultado esperado que es disminuir y en el indicador 3 se logra el resultado esperado que es disminuir.

Tabla 24. Hipótesis general

Hipótesis general
Hipótesis Nula H_0 : Una aplicación de minería de datos no mejora significativamente el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano de Trujillo.
Hipótesis Nula H_a : Una aplicación de minería de datos mejora significativamente el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano de Trujillo.

Fuente: Elaboración propia

Frente a los resultados obtenidos, se acepta la hipótesis alterna donde se logra una mejora significativa en el diagnóstico de un grupo de enfermedades respiratorias en el Centro de Atención Primaria CAP III Metropolitano de Trujillo.

V. DISCUSIÓN

- Ante el resultado obtenido por, los diagnósticos con un grupo de enfermedades respiratorias en, el Centro de Atención Primaria CAP III Metropolitano de Trujillo al, implementar una aplicación de minería de datos donde, una de sus efectividades fue que se pudo disminuir el tiempo promedio necesario para reconocer la existencia de enfermedades respiratorias y costo promedio de diagnóstico.
- Con el primer indicador del nivel de morbilidad, se contaba con un 43.55% en el nivel de morbilidad antes de la implementación siendo este un nivel negativo, posterior a ello con la aplicación ya implementada el nivel aumento negativamente a un 48.51%, estos datos son comparados a base del estudio de Cordova et al. (2020), en el cual determina un nivel de 30.47% según los casos reportados en niños del centro materno-infantil, declarando un índice altamente negativo.
- Con el indicador tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias, previamente se contaba con 00:25 minutos siendo este un tiempo alto y después con 00:05 minutos. Se logró tener una reducción de 00:20 minutos, estos datos son respaldados por el estudio de Marimón y Navarro (2017), en el cual determina un tiempo de diagnóstico entre 15 a 30 min y cómo influye directamente en la detección de enfermedades a base de los antecedentes del paciente, declarándolo como tiempo deficiente.

- En el tercer indicador de costo promedio de diagnóstico de un grupo de enfermedades respiratorias, se contaba con un costo de S/334.32 siendo este un monto negativo y al implementarlo el costo fue de S/6.37, evidenciando una reducción en S/327.95, este indicador es respaldado por la investigación de Bernal, Carvajal y Alvis (2017), en el cual determina el impacto económico que ha generado las infecciones respiratorias agudas (IRAS) ante ello, su resultado con una población de 1576, llevó a un costo promedio de \$759.437, declarándolo como una elevada carga económica.

- Durante la investigación, se presentó diversos inconvenientes, entre los más destacados o los más conflictivos que daban bloqueos o un desvío general de la investigación, fue por el protocolo en base al proceso de trámite para la aceptación de desarrollo, debido al poco entendimiento y muchas observaciones declaradas en su mayoría como falsos positivos, generando un desvío al tiempo pactado en la investigación. Otro inconveniente fue la falta de organización por parte de la entidad, a la hora de la aceptación de desarrollo, debido a la falta de delegación de personal adecuado, generando un bloqueo en la aprobación de información a utilizar para la implementación. Además, tras el análisis de la información, se encontró falta de coherencia en los datos, perjudicando en su calidad, esto se relaciona ante el impacto ocasionado por la pandemia del COVID-19, en el cual generó un desvío grave en la información por el número de emergencias reportadas del mismo.

VI. CONCLUSIONES

- Se concluye satisfactoriamente que, al implementar una aplicación de minería de datos mejora, significativamente el diagnóstico de un grupo de enfermedades respiratorias, en el Centro de Atención Primaria CAP III Metropolitano de Trujillo. Al disminuir el tiempo promedio, los costos pero que al mismo tiempo aumenta de manera negativa la morbilidad.
- Se determinó que no existe una disminución significativa en el nivel de morbilidad, obteniendo un resultado de 43.55% (positivo) antes de la implementación, y un 48.51% (negativo) después de la implementación de la aplicación de minería de datos, consiguiendo un aumento del 4.96%, demostrado con la prueba estadística Wilcoxon, con un nivel de significación del 5% y un 95% de confianza, dando un valor de Z de -1.253 y sig. de 0.210.
- Se determinó que existe una disminución del tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias, obteniendo un resultado de 00:25 minutos antes de la implementación, y 00:05 minutos después de la implementación de la aplicación de minería de datos, consiguiendo una reducción de 00:20 minutos. Demostrado con la prueba estadística Wilcoxon, con un nivel de significación del 5% y un 95% de confianza, dando un valor de Z de -3.860 y sig. de 0.000.
- Se determinó que existe una disminución en los costos promedio de diagnóstico de un grupo de enfermedades respiratorias, obteniendo un resultado de S/334.32 antes de la implementación, y S/6.37 después de la implementación de la aplicación de minería de datos, consiguiendo una reducción de S/ 327.95. Demostrado con la prueba estadística Wilcoxon, con un nivel de significación del 5% y un 95% de confianza, dando un valor de Z de -3.981 y sig. de 0.00.

VII. RECOMENDACIONES

- Se recomienda al personal administrativo o técnico de la entidad, en centralizar la información para mejorar la calidad del dato, además realizar capacitaciones en resolución de problemas mediante ideas innovadoras, con el propósito de darle más aplicabilidad a la información para fines estadísticos.

- Se recomienda una mejor distribución del personal en seguimiento de los pacientes enfermos, además, durante cita, realizar orientaciones rápidas en revisiones a futuro, como medio preventivo para evitar rebrote de la enfermedad.

- Se recomienda al personal administrativo y técnico de la entidad, en tener una organización más centralizada de los datos de los pacientes, en el especial, los que tienen historial más de un año, con el propósito en reducir el tiempo de búsqueda en expedientes (actas medicas) de diversas fuentes.

- Se recomienda al personal administrativo de la entidad en tener una organización más centralizada entre los medicamentos y exámenes en diversos subsistemas, debido al cambio de nombre y código connotado entre ellos (se ve más perjudicado a los pacientes que tienen más de un año como historial), existiendo una poca precisión a la hora de realizar una receta de medicamentos a un paciente específico.

REFERENCIAS

- ÁGUILA, N., BRAVO, E., DELGADO, H.M., MONTENEGRO, T., HERRERA, L.R. y CENTENO, A., 2019. Algunas reflexiones sobre el análisis de la situación de salud. *MediSur*, vol. 17, no. 3, pp. 417-428.
- ALOMÍA, P., 2019. Infecciones respiratorias agudas en infantes menores de 5 años del Centro de Salud Javier Loyola, Ecuador. *Archivos Venezolanos de Farmacología y Terapéutica*, vol. 38, no. 6, pp. 758-761.
- ALVARADO, A.T. y FLORES, G., 2009. Errores médicos. *Acta Médica Costarricense*, vol. 51, no. 1, pp. 16-23. ISSN 0001-6002.
- ANDINA, A.P. de N., 2018. Contraloría encuentra deficiencias en Hospital Regional de Trujillo. [en línea]. [Consulta: 13 mayo 2020]. Disponible en: <https://andina.pe/agencia/noticia-contraloria-encuentra-deficiencias-hospital-regional-trujillo-703059.aspx>.
- BERNAL, C., CARVAJAL, H. y ALVIS, N.J., 2017. Costos económicos de la infección respiratoria aguda en un Municipio de Colombia. *Revista de la Universidad Industrial de Santander. Salud*, vol. 49, no. 3, pp. 470-477. ISSN 01210807, 21458464. DOI 10.18273/revsal.v49n3-2017005.
- BIBLIOTECA NACIONAL DE MEDICINA DE LOS ESTADOS UNIDOS, 2010. Rinofaringitis. *Emc. Pediatría*, vol. 45, no. 2, pp. 1-5. ISSN 1245-1789. DOI 10.1016/S1245-1789(10)70176-1.
- BOTERO, M., 2019. METODOLOGÍAS PARA EL DESARROLLO DE PROYECTOS DE MINERÍA DE DATOS. [en línea], [Consulta: 15 mayo 2020]. Disponible en: https://www.scipedia.com/public/Botero_Villada_2019a.
- CALABRESE, J., ESPONDA, S., PASINI, A.C., BORACCHIA, M. y PESADO, P.M., 2019. Guía para evaluar calidad de datos basada en ISO/IEC 25012. XXV Congreso Argentino de Ciencias de la Computación (CACIC) (Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019) [en línea]. S.l.: s.n., [Consulta: 16 mayo 2020]. ISBN 978-987-688-377-1. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/91086>.
- CARNOTA, O., 2010. El costo en salud y la corresponsabilidad clínica desde un enfoque gerencial. *Revista Cubana de Salud Pública*, vol. 36, pp. 222-232. ISSN 0864-3466, 0864-3466, 1561-3127. DOI 10.1590/S0864-34662010000300006.
- CARRASCO, S., 2005. Metodología de La Investigación Científica. [en línea]. [Consulta: 1 junio 2020]. Disponible en: https://kupdf.net/download/metodologia-de-la-investigacion-cientifica-carrasco-diaz_59065f94dc0d60a122959e9d_pdf#.
- COFZ, 2018. Reacciones Adversas a Medicamentos y la importancia de notificarlas. *Colegio Oficial Farmacéuticos de Zaragoza* [en línea]. [Consulta:

16 octubre 2020]. Disponible en: <https://cofzaragoza.org/reacciones-adversas-a-medicamentos-y-la-importancia-de-notificarlas/>.

COLINA, A.M., 2017. RETOS Y PERSPECTIVAS DE LAS TECNOLOGÍAS DE INFORMACIÓN. , pp. 117.

CORDOVA, D.A., CHÁVEZ, C.G., BERMEJO, E.W., JARA, X.N. y SANTA MARIA, F.B., 2020. Prevalencia de infecciones respiratorias agudas en niños menores de 5 años en un centro materno infantil de Lima. *Horizonte Médico (Lima)*, vol. 20, no. 1, pp. 54-60. ISSN 2227-3530. DOI 10.24265/horizmed.2020.v20n1.08.

CURIOSO, W.H., 2014. eSalud en Perú: implementación de políticas para el fortalecimiento de sistemas de información en salud. *Rev Panam Salud Publica*, pp. 5.

DATE, C.J., 2001. *Introducción a los sistemas de bases de datos*. S.I.: Pearson Educación. ISBN 978-968-444-419-5.

DÁVILA, F. y SÁNCHEZ, Y., 2012. Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas. *Revista Cubana de Informática Médica*, vol. 4, no. 2, pp. 174-183. ISSN 1684-1859.

DICCIONARIO DE CÁNCER, 2011. Definición de Diagnostico. *National Cancer Institute* [en línea]. [Consulta: 2 mayo 2020]. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionario.nciglobal>

DICCIONARIO DE CÁNCER, 2020. Definición de bronquitis crónica - Diccionario de cáncer del NCI - Instituto Nacional del Cáncer. [en línea]. [Consulta: 26 octubre 2020]. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionario/def/bronquitis-cronica.nciglobal,ncienterprise>

DUBÓN, M. del C., 2016. Bronquitis aguda. [en línea]. [Consulta: 3 abril 2021]. Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0026-17422016000100027.

DUQUE, A., 2008. Bronquitis aguda: diagnóstico y manejo en la práctica clínica. , pp. 10.

EL PERUANO, 2020. ¡Cuidado! No confunda las gripes comunes con el coronavirus. [en línea]. [Consulta: 16 julio 2020]. Disponible en: <http://elperuano.pe/noticia-¡cuidado-no-confunda-gripes-comunes-con-coronavirus-96661.aspx>.

EMPENDIUM, 2020. Neumonía adquirida en la comunidad (NAC). [en línea]. [Consulta: 26 octubre 2020]. Disponible en: <https://empendium.com/manualmibe/social/chapter/B34.II.3.11.1>.

- ESCURRA, L., 1988. Cuantificación de la validez de contenido por criterio de jueces. *Revista de Psicología*, vol. 6, no. 1-2, pp. 103-111. ISSN 2223-3733.
- EXPRESO, R., 2019. Expreso. [en línea]. [Consulta: 10 mayo 2020]. Disponible en: <https://www.expreso.com.pe/actualidad/neumonia-causa-575-muertes-en-el-peru/>.
- FERNÁNDEZ, R., 2014. Adecuación de la prescripción de antibióticos en un área de atención primaria: estudio descriptivo transversal. *Enfermedades Infecciosas y Microbiología Clínica*, vol. 32, no. 5, pp. 285-292. ISSN 0213-005X. DOI 10.1016/j.eimc.2013.05.004.
- GRAVITAR, 2014. Pentaho. *Business Intelligence, Data Warehouse, Monterrey, México: Gravitár* [en línea]. [Consulta: 10 mayo 2020]. Disponible en: <https://gravitar.biz/pentaho/>.
- HUAMAN, H.G., 2005. *MANUAL DE TECNICAS DE INVESTIGACION Conceptos y Aplicaciones*. S.I.: Hector Guillermo Huaman Valencia.
- IBM, 2020. Conceptos básicos de ayuda de CRISP-DM. [en línea]. [Consulta: 15 mayo 2020]. Disponible en: https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_crispdm_ddita/clementine/crisp_help/crisp_overview.html.
- INSIGHTSOFTWARE, 2019. ¿Cuál es la función de la minería de datos en la business intelligence? *insightsoftware Spain* [en línea]. [Consulta: 4 mayo 2020]. Disponible en: <https://insightsoftware.com/es/blog/what-role-does-data-mining-play-for-business-intelligence/>.
- INSTITUTO NACIONAL DEL CÁNCER, 2018. Definición de terapia farmacológica - Diccionario de cáncer del NCI -. [en línea]. [Consulta: 16 octubre 2020]. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionario/def/terapia-farmacologica.nciglobal,ncienterprise>
- JUY, E., CÉSPEDES, E., RUBAL, A. de la C., MAZA, A.M. y TERÁN, C.A., 2014. Morbilidad por infecciones respiratorias agudas en menores de 5 años. *MEDISAN*, vol. 18, no. 11, pp. 1490-1498. ISSN 1029-3019.
- LEONARD, E.I. y CASTRO, Y., 2013. Metodologías para desarrollar Almacén de Datos. *Revista de Arquitectura e Ingeniería*, vol. 7, no. 3, pp. 1-12. ISSN , 1990-8830.
- LÓPEZ, E. y MARTEL, P., 2001. *La escritura en uooh: una propuesta metodológica para el estudio de la escritura prehispánica maya-yucateca*. S.I.: UNAM. ISBN 978-968-36-9004-3.
- MARIMÓN, J.M. y NAVARRO, J.M., 2017. Métodos de diagnóstico rápido de las infecciones respiratorias. [en línea]. [Consulta: 22 mayo 2020]. Disponible en: <https://www.elsevier.es/es-revista-enfermedades-infecciosas-microbiologia-clinica-28-pdf-S0213005X16303871>.

- MICROSOFT, 2018. Procesamiento analítico en línea (OLAP) - Azure Architecture Center. [en línea]. [Consulta: 16 mayo 2020]. Disponible en: <https://docs.microsoft.com/es-es/azure/architecture/data-guide/relational-data/online-analytical-processing>.
- MICROSOFT, 2019. Online transaction processing (OLTP) - Azure Architecture Center. [en línea]. [Consulta: 16 mayo 2020]. Disponible en: <https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/online-transaction-processing>.
- MINISTERIO DE SALUD, 2018. Indicadores Básicos | Deis. [en línea]. [Consulta: 9 mayo 2020]. Disponible en: <http://www.deis.msal.gov.ar/index.php/indicadores-basicos/>.
- MINSALUD, 2015. Clasificación Internacional de Enfermedades (CIE): , pp. 8.
- OMS, 2020. OMS | Tuberculosis. *WHO* [en línea]. [Consulta: 26 octubre 2020]. Disponible en: <http://www.who.int/topics/tuberculosis/es/>.
- OMS, R., 2019. OMS | 10 datos sobre la seguridad del paciente. *WHO* [en línea]. [Consulta: 9 mayo 2020]. Disponible en: http://www.who.int/features/factfiles/patient_safety/es/.
- OPS, 2016. Situación de Salud en las Américas: Indicadores Básicos 2016. En: Accepted: 2016-10-21T14:56:14Z [en línea], [Consulta: 20 mayo 2020]. Disponible en: <https://iris.paho.org/handle/10665.2/31288>.
- ORDÓÑEZ, L.A., 2019. Situación epidemiológica de las infecciones respiratorias agudas (IRA) en el Perú. *Centro Nacional de Epidemiología, Prevención y Control de Enfermedades* [en línea], [Consulta: 15 mayo 2020]. Disponible en: https://www.dge.gob.pe/portal/index.php?option=com_content&view=article&id=666:bol2018-2&catid=55:bolepi.
- ORTIZ, 1996. Introducción a la Medicina Clínica. [en línea]. [Consulta: 16 octubre 2020]. Disponible en: https://sisbib.unmsm.edu.pe/bibvirtual/libros/medicina/intro_medi_clin/caratula.htm.
- PAVEZ, D., PÉREZ, R., COFRÉ, J., RODRÍGUEZ, J., PAVEZ, D., PÉREZ, R., COFRÉ, J. y RODRÍGUEZ, J., 2019. Recomendaciones para el diagnóstico y tratamiento etiológico de la faringoamigdalitis aguda estreptocócica en pediatría. *Revista chilena de infectología*, vol. 36, no. 1, pp. 69-77. ISSN 0716-1018. DOI 10.4067/S0716-10182019000100069.
- PERU21, R., 2019. Minsa refuerza sus acciones en zonas de alto riesgo para prevenir casos de neumonía en menores de 5 años. *Peru21* [en línea]. [Consulta: 10 mayo 2020]. Disponible en: <https://www.gob.pe/institucion/minsa/noticias/27255-minsa-refuerza-sus-acciones-en-zonas-de-alto-riesgo-para-prevenir-casos-de-neumonia-en-menores-de-5-anos>.

- RACERO, L., 2018. Epidemiología de la faringitis aguda bacteriana en un hospital general de agudos. [en línea]. [Consulta: 3 abril 2021]. Disponible en: http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S0325-29572018000100011&lang=es.
- RAMOS, B.N., 2006. Enfoque conceptual y de procedimiento para el diagnóstico o análisis de la situación de salud. *Revista Cubana de Salud Pública*, vol. 32, no. 2, pp. 0-0. ISSN 0864-3466.
- REAL ACADEMIA ESPAÑOLA, 2020a. enfermedad | Diccionario de la lengua española. «*Diccionario de la lengua española*» - Edición del Tricentenario [en línea]. [Consulta: 9 mayo 2020]. Disponible en: <https://dle.rae.es/enfermedad>.
- REAL ACADEMIA ESPAÑOLA, 2020b. morbilidad | Diccionario de la lengua española. «*Diccionario de la lengua española*» - Edición del Tricentenario [en línea]. [Consulta: 2 mayo 2020]. Disponible en: <https://dle.rae.es/morbilidad>.
- REAL ACADEMIA ESPAÑOLA, 2020c. patogenia | Diccionario de la lengua española. «*Diccionario de la lengua española*» - Edición del Tricentenario [en línea]. [Consulta: 5 mayo 2020]. Disponible en: <https://dle.rae.es/patogenia>.
- REUS, C.M.T. y ORTIZ, E.A.B., 2013. Morbilidad por infecciones respiratorias agudas en pacientes menores de 5 años. *MEDISAN*, vol. 17, no. 12, pp. 9075-9080. ISSN , 1029-3019.
- RIVO, E., DE LA FUENTE, J., RIVO, Á., GARCÍA, E., CAÑIZARES, M.Á. y GIL, P., 2012. Cross-Industry Standard Process for data mining is applicable to the lung cancer surgery domain, improving decision making as well as knowledge and quality management. *Clinical and Translational Oncology*, vol. 14, no. 1, pp. 73-79. ISSN 1699-3055. DOI 10.1007/s12094-012-0764-8.
- ROMERO, M., 2016. Pruebas de bondad de ajuste a una distribución normal. *Revista Enfermería del Trabajo*, vol. 6, no. 3, pp. 114. ISSN 2174-2510.
- SALDAÑA, 2016. Modelo predictivo de minería de datos de apoyo a la gestión hospitalaria sobre la morbilidad de pacientes hospitalizados. En: Accepted: 2016-10-11T22:51:45Z, *Universidad Privada Antenor Orrego* [en línea], [Consulta: 16 octubre 2020]. Disponible en: <http://repositorio.upao.edu.pe/handle/upaorep/1927>.
- TAFFAREL, O., 2009. *La creación del conocimiento* | ISBN 978-612-00-0017-5 - Libro [en línea]. S.l.: s.n. [Consulta: 13 junio 2020]. Disponible en: <https://isbn.cloud/9786120000175/la-creacion-del-conocimiento/>.
- TO ERR IS HUMAN, 2016. To err IS human. *Nursing2020 Critical Care*, vol. 11, no. 5, pp. 4. ISSN 1558-447X. DOI 10.1097/01.CCN.0000490961.44977.8d.

VELARDE, A., 2003. Minería de Datos. Una Introducción. *Conciencia Tecnológica* [en línea], no. 23. [Consulta: 25 abril 2020]. ISSN 1405-5597,. Disponible en: <https://www.redalyc.org/articulo.oa?id=94402303>.

ANEXOS

Anexo 1. Variables de investigación e indicadores

Anexo 1.1 Matriz de operacionalización de variables

VARIABLES DE ESTUDIO	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	INDICADORES	ESCALA DE MEDICIÓN
Aplicación de minería de datos	La minería de datos se presenta como una herramienta de mucha aplicabilidad para la exploración de grandes almacenes de base de datos, utilizando modelos predictivos, de segmentación, agrupación sobre el conjunto existente para presentar datos visuales en la toma de decisiones tanto en instituciones como	Se determinará la calidad mediante el modelo ISO 25012.	<ul style="list-style-type: none">• Exactitud Sintáctica• Exactitud Semántica	De razón.

	empresas (Velarde, 2003).			
Diagnóstico de un grupo de enfermedades respiratorias	El diagnóstico de la situación de salud (DSS) es un procedimiento aplicado en una población, grupo o comunidad el poder conocer y analizar la situación de salud, con la finalidad de identificar o diagnosticar problemas y necesidades de salud, el proceso no concluye hasta el control o solución del problema o problemas detectado (Ramos, 2006)	Los indicadores nivel de morbilidad y costo promedio de diagnóstico de un grupo de enfermedades respiratorias se medirán a través de una ficha de registro y el costo promedio de diagnóstico de un grupo de enfermedades respiratorias a través de una ficha de observación.	<ul style="list-style-type: none"> • Nivel de morbilidad. • Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias. • Costo promedio de diagnóstico de un grupo de enfermedades respiratorias. 	De razón.

Fuente: elaboración propia de los autores.

Anexo 1.2 Indicadores de variables

OBJETIVO ESPECÍFICO	INDICADOR	DESCRIPCIÓN	TÉCNICA / INSTRUMENTO	TIEMPO EMPLEADO	MODO DE CÁLCULO
Disminuir el nivel de morbilidad.	Nivel de morbilidad.	La morbilidad es la cantidad de personas que se enferman en un lugar y tiempo específico (Real Academia Española, 2020).	Fichaje / Ficha de Registro	Mensual	$NM = \frac{TC \times 100}{TCE}$ <p>NM: Nivel de morbilidad. TC: Total de casos de un grupo de enfermedades respiratorias según la edad. TCE: Total de casos egresados.</p>
Disminuir el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias.	El diagnóstico es el proceso en el que se identifica una enfermedad, afección o lesión por sus signos y síntomas	Fichaje / Ficha de observación	Mensual	$TPIEGER = \frac{\sum_{i=1}^n (TIEER)_i}{n}$ <p>TPIEGER: Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias TIEER: Tiempo para identificar la existencia de un grupo de enfermedades respiratorias</p>

		(Diccionario de cáncer, 2011).			n: número de casos egresados.
Disminuir el costo promedio de diagnóstico de un grupo de enfermedades respiratorias	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias.	El abordar los problemas de salud a partir de los síntomas y no de sus determinantes, implica mayor consumo de medicamentos y por tanto incremento de los costos (Carnota, 2010).	Fichaje / Ficha de Registro	Mensual	$CPDGER = \frac{\sum_{i=1}^{n_{ce}} ((NE \times CE) + (NT \times CT))}{NCE}$ <p>CPDGER: Costo promedio de diagnóstico de un grupo de enfermedades respiratorias</p> <p>NE: número de exámenes</p> <p>CE: costo de exámenes</p> <p>NT: número de tratamientos</p> <p>CT: costo de tratamientos</p> <p>NCE: número de casos de una enfermedad específica.</p>

Fuente: elaboración propia de los autores.

Anexo 2. Instrumentos de recolección de datos

Ficha de Registro			
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe	Tipo de prueba	
Empresa investigada	CAP III Metropolitano Trujillo		
Motivo de investigación	Nivel de Morbilidad		
Fecha de inicio		Fecha de termino	

Objetivo	Indicador	Medida	Formula
Disminuir el nivel de morbilidad	Nivel de morbilidad	Porcentaje	$NM = \frac{TC \times 100}{TCE}$ <p>NM: Nivel de morbilidad. TC: Total de casos de un grupo de enfermedades respiratorias según edad. TCE: Total de casos egresados.</p>

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Fecha de nacimiento	Fecha de diagnóstico	Enfermedad diagnosticada
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						

INSTRUMENTO

Ficha de Observación			
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe	Tipo de prueba	
Empresa investigada	CAP III Metropolitano Trujillo		
Motivo de investigación	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias		
Fecha de inicio		Fecha de termino	

Objetivo	Indicador	Medida	Formula
Disminuir el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias	Horas	$TPIEGER = \frac{\sum_{i=1}^n (TIEER)_i}{n}$ <p>TPIEGER: Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias</p> <p>TIEER: Tiempo para identificar la existencia de un grupo de enfermedades respiratorias</p> <p>n: número de casos egresados.</p>

INSTRUMENTO

Ficha de Registro			
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe	Tipo de prueba	
Empresa investigada	CAP III Metropolitano Trujillo		
Motivo de investigación	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias		
Fecha de inicio		Fecha de termino	

Objetivo	Indicador	Medida	Formula
Disminuir el costo promedio de diagnóstico de un grupo de enfermedades respiratorias	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias	Unidad	$CPDGER = \frac{\sum_{i=1}^{nce} ((NE \times CE) + (NT \times CT))i}{NCE}$ <p> CDGER: Costo promedio de diagnóstico de un grupo de enfermedades respiratorias. NE: Numero de exámenes CE: Costo de exámenes NT: Numero de tratamientos CT: Costo de tratamientos NCE: Número de casos de una enfermedad específica </p>

Anexo 3. Validez para el indicador del nivel de morbilidad

TABLA DE VALIDACIÓN DE INSTRUMENTOS

DATOS DEL EXPERTO	
Apellidos y nombres	Icanaque Ordinola, Juan Damian
Centro laboral	Universidad Cesar Vallejo - Filial Piura
Grado obtenido	Magister en Ciencias mención Biotecnología
Fecha de validación	17/06/2020

DATOS DEL TRABAJO DE INVESTIGACIÓN	
Título de la tesis	Aplicación De Minería De Datos Para Mejorar El Diagnóstico De Un Grupo De Enfermedades Respiratorias En Un Hospital De Trujillo
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe
Tipo de instrumento	Ficha de registro
Nombre del indicador	Nivel de Morbilidad

RANGOS DE EVALUACIÓN	Deficiente 0 – 20%	Regular 21 – 50%	Bueno 51 – 70%	Muy bueno 71 – 80 %	Excelente 81 – 100 %

Nº	PREGUNTAS	CALIFICACIÓN	OBSERVACIÓN
1	¿El instrumento de medición cumple con el diseño adecuado?	90%	
2	¿El instrumento de recolección de datos tiene relación con el título de la investigación?	90%	
3	¿El instrumento de recolección de datos tiene la relación con las variables de investigación?	90%	
4	¿El instrumento de recolección de datos facilitará el logro de los objetivos de investigación?	90%	
5	¿El instrumento analiza los datos de la organización?	90%	
6	¿El instrumento de medición explica en forma precisa y clara el grado de cumplimiento de la meta o resultado?	90%	
7	¿El resultado del instrumento es entendible para ser correctamente analizado?	100%	

Fuente: Adaptado de Cesar Robledo Mérida

El instrumento puede ser aplicado: SI (x) NO ()

Sugerencias:


FIRMA DEL EXPERTO

TABLA DE VALIDACIÓN DE INSTRUMENTOS

DATOS DEL EXPERTO	
Apellidos y nombres	Moreno Zavaleta, Hilda Edith
Centro laboral	Hospital I de la Esperanza
Grado obtenido	Químico Farmacéutico
Fecha de validación	27/06/2020

DATOS DEL TRABAJO DE INVESTIGACIÓN	
Título de la tesis	Aplicación de minería de datos para mejorar el diagnóstico de un grupo de enfermedades respiratorias en un Hospital de Trujillo
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe
Tipo de instrumento	Ficha de registro
Nombre del indicador	Nivel de Morbilidad

RANGOS DE EVALUACIÓN	Deficiente 0 – 20%	Regular 21 – 50%	Bueno 51 – 70%	Muy bueno 71 – 80 %	Excelente 81 – 100 %
----------------------	-----------------------	---------------------	-------------------	------------------------	-------------------------

Nº	PREGUNTAS	CALIFICACIÓN	OBSERVACIÓN
1	¿El instrumento de medición cumple con el diseño adecuado?	100%	
2	¿El instrumento de recolección de datos tiene relación con el título de la investigación?	100%	
3	¿El instrumento de recolección de datos tiene la relación con las variables de investigación?	100%	
4	¿El instrumento de recolección de datos facilitará el logro de los objetivos de investigación?	100%	
5	¿El instrumento analiza los datos de la organización?	100%	
6	¿El instrumento de medición explica en forma precisa y clara el grado de cumplimiento de la meta o resultado?	100%	
7	¿El resultado del instrumento es entendible para ser correctamente analizado?	100%	

Fuente: Adaptado de Cesar Robledo Mérida

El instrumento puede ser aplicado: SI (x) NO ()

Sugerencias:



FIRMA DEL EXPERTO

TABLA DE VALIDACIÓN DE INSTRUMENTOS

DATOS DEL EXPERTO	
Apellidos y nombres	Leopoldo Segura Murga
Centro laboral	Vircatex International Trading SAC
Grado obtenido	Ingeniero Industrial
Fecha de validación	01-07-2020

DATOS DEL TRABAJO DE INVESTIGACIÓN	
Título de la tesis	Aplicación de minería de datos para mejorar el diagnóstico de un grupo de enfermedades respiratorias en un Hospital de Trujillo
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe
Tipo de instrumento	Ficha de registro
Nombre del indicador	Nivel de Morbilidad

RANGOS DE EVALUACIÓN	Deficiente 0 – 20%	Regular 21 – 50%	Bueno 51 – 70%	Muy bueno 71 – 80 %	Excelente 81 – 100 %
----------------------	-----------------------	---------------------	-------------------	------------------------	-------------------------


Nº	PREGUNTAS	CALIFICACIÓN	OBSERVACIÓN
1	¿El instrumento de medición cumple con el diseño adecuado?	Muy bueno	
2	¿El instrumento de recolección de datos tiene relación con el título de la investigación?	Muy bueno	
3	¿El instrumento de recolección de datos tiene la relación con las variables de investigación?	Muy bueno	
4	¿El instrumento de recolección de datos facilitará el logro de los objetivos de investigación?	Muy bueno	
5	¿El instrumento analiza los datos de la organización?	Muy bueno	
6	¿El instrumento de medición explica en forma precisa y clara el grado de cumplimiento de la meta o resultado?	Muy bueno	
7	¿El resultado del instrumento es entendible para ser correctamente analizado?	Muy bueno	

Fuente: Adaptado de Cesar Robledo Mérida

El instrumento puede ser aplicado: SI (x) NO ()

Sugerencias: _____ Mejora en la toma de datos y determinar el tamaño muestra referente a que población y tiempo de aplicación.

**VIRCATEX INTERNATIONAL
TRADING S.A.C.**


Ing. Leopoldo Segura Murga
Gerente de Planeamiento
FIRMA DEL EXPERTO

Anexo 4. Validez para el indicador del tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias

TABLA DE VALIDACIÓN DE INSTRUMENTOS

DATOS DEL EXPERTO	
Apellidos y nombres	Icanaque Ordinola, Juan Damian
Centro laboral	Universidad Cesar Vallejo - Filial Piura
Grado obtenido	Magister en Ciencias mención Biotecnología
Fecha de validación	11/06/2020

DATOS DEL TRABAJO DE INVESTIGACIÓN	
Título de la tesis	Aplicación De Minería De Datos Para Mejorar El Diagnóstico De Un Grupo De Enfermedades Respiratorias En Un Hospital De Trujillo
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe
Tipo de instrumento	Ficha de observación
Nombre del indicador	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias

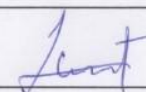
RANGOS DE EVALUACIÓN	Deficiente 0 – 20%	Regular 21 – 50%	Bueno 51 – 70%	Muy bueno 71 – 80 %	Excelente 81 – 100 %
----------------------	-----------------------	---------------------	-------------------	------------------------	-------------------------

Nº	PREGUNTAS	CALIFICACIÓN	OBSERVACIÓN
1	¿El instrumento de medición cumple con el diseño adecuado?	90%	
2	¿El instrumento de recolección de datos tiene relación con el título de la investigación?	90%	
3	¿El instrumento de recolección de datos tiene la relación con las variables de investigación?	90%	
4	¿El instrumento de recolección de datos facilitará el logro de los objetivos de investigación?	90%	
5	¿El instrumento analiza los datos de la organización?	90%	
6	¿El instrumento de medición explica en forma precisa y clara el grado de cumplimiento de la meta o resultado?	90%	
7	¿El resultado del instrumento es entendible para ser correctamente analizado?	100%	

Fuente: Adaptado de Cesar Robledo Mérida

El instrumento puede ser aplicado: SI (X) NO ()

Sugerencias:



FIRMA DEL EXPERTO

TABLA DE VALIDACIÓN DE INSTRUMENTOS

DATOS DEL EXPERTO	
Apellidos y nombres	Moreno Zavaleta, Hilda Edith
Centro laboral	Hospital 1 de la Esperanza
Grado obtenido	Químico Farmacéutico
Fecha de validación	27/06/2020

DATOS DEL TRABAJO DE INVESTIGACIÓN	
Título de la tesis	Aplicación de minería de datos para mejorar el diagnóstico de un grupo de enfermedades respiratorias en un Hospital de Trujillo
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe
Tipo de instrumento	Ficha de observación
Nombre del indicador	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias

RANGOS DE EVALUACIÓN	Deficiente 0 – 20%	Regular 21 – 50%	Bueno 51 – 70%	Muy bueno 71 – 80 %	Excelente 81 – 100 %
----------------------	-----------------------	---------------------	-------------------	------------------------	-------------------------

Nº	PREGUNTAS	CALIFICACIÓN	OBSERVACIÓN
1	¿El instrumento de medición cumple con el diseño adecuado?	100%	
2	¿El instrumento de recolección de datos tiene relación con el título de la investigación?	100%	
3	¿El instrumento de recolección de datos tiene la relación con las variables de investigación?	100%	
4	¿El instrumento de recolección de datos facilitará el logro de los objetivos de investigación?	100%	
5	¿El instrumento analiza los datos de la organización?	100%	
6	¿El instrumento de medición explica en forma precisa y clara el grado de cumplimiento de la meta o resultado?	100%	
7	¿El resultado del instrumento es entendible para ser correctamente analizado?	100%	

Fuente: Adaptado de Cesar Robledo Mérida

El instrumento puede ser aplicado: SI (x) NO ()

Sugerencias:



FIRMA DEL EXPERTO

TABLA DE VALIDACIÓN DE INSTRUMENTOS

DATOS DEL EXPERTO	
Apellidos y nombres	Leopoldo Segura Murga
Centro laboral	Vircatex International Trading
Grado obtenido	Ingeniero Industrial
Fecha de validación	01-07-2020

DATOS DEL TRABAJO DE INVESTIGACIÓN	
Título de la tesis	Aplicación de minería de datos para mejorar el diagnóstico de un grupo de enfermedades respiratorias en un Hospital de Trujillo
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe
Tipo de instrumento	Ficha de observación
Nombre del indicador	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias

RANGOS DE EVALUACIÓN	Deficiente 0 – 20%	Regular 21 – 50%	Bueno 51 – 70%	Muy bueno 71 – 80 %	Excelente 81 – 100 %
----------------------	-----------------------	---------------------	-------------------	------------------------	-------------------------

Nº	PREGUNTAS	CALIFICACIÓN	OBSERVACIÓN
1	¿El instrumento de medición cumple con el diseño adecuado?	Muy bueno	
2	¿El instrumento de recolección de datos tiene relación con el título de la investigación?	Muy bueno	
3	¿El instrumento de recolección de datos tiene la relación con las variables de investigación?	Muy bueno	
4	¿El instrumento de recolección de datos facilitará el logro de los objetivos de investigación?	Muy bueno	
5	¿El instrumento analiza los datos de la organización?	Muy bueno	
6	¿El instrumento de medición explica en forma precisa y clara el grado de cumplimiento de la meta o resultado?	Muy bueno	
7	¿El resultado del instrumento es entendible para ser correctamente analizado?	Muy bueno	

Fuente: Adaptado de Cesar Robledo Mérida

El instrumento puede ser aplicado: SI (x) NO ()

Sugerencias: ___ detallar tiempo de análisis del laboratorio de cada examen, tiempo de consulta médica.

**VIRCATEX INTERNATIONAL
TRADING S.A.C.**

Ing. Leopoldo Segura Murga
 Gerente de Planeamiento
FIRMA DEL EXPERTO

Anexo 5. Validez para el indicador del costo promedio de diagnóstico de un grupo de enfermedades respiratorias



TABLA DE VALIDACIÓN DE INSTRUMENTOS

DATOS DEL EXPERTO	
Apellidos y nombres	Icanaque Ordinola, Juan Damian
Centro laboral	Universidad Cesar Vallejo - Filial Piura
Grado obtenido	Magister en Ciencias mención Biotecnología
Fecha de validación	11/06/2020

DATOS DEL TRABAJO DE INVESTIGACIÓN	
Título de la tesis	Aplicación De Minería De Datos Para Mejorar El Diagnóstico De Un Grupo De Enfermedades Respiratorias En Un Hospital De Trujillo
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe
Tipo de instrumento	Ficha de registro
Nombre del indicador	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias

RANGOS DE EVALUACIÓN	Deficiente 0 – 20%	Regular 21 – 50%	Bueno 51 – 70%	Muy bueno 71 – 80 %	Excelente 81 – 100 %
----------------------	-----------------------	---------------------	-------------------	------------------------	-------------------------

Nº	PREGUNTAS	CALIFICACIÓN	OBSERVACIÓN
1	¿El instrumento de medición cumple con el diseño adecuado?	90%	
2	¿El instrumento de recolección de datos tiene relación con el título de la investigación?	90%	
3	¿El instrumento de recolección de datos tiene la relación con las variables de investigación?	90%	
4	¿El instrumento de recolección de datos facilitará el logro de los objetivos de investigación?	90%	
5	¿El instrumento analiza los datos de la organización?	90%	
6	¿El instrumento de medición explica en forma precisa y clara el grado de cumplimiento de la meta o resultado?	90%	
7	¿El resultado del instrumento es entendible para ser correctamente analizado?	100%	

Fuente: Adaptado de Cesar Robledo Mérida

El instrumento puede ser aplicado: SI (X) NO ()

Sugerencias:


FIRMA DEL EXPERTO

TABLA DE VALIDACIÓN DE INSTRUMENTOS

DATOS DEL EXPERTO	
Apellidos y nombres	Moreno Zavaleta, Hilda Edith
Centro laboral	Hospital 1 de la Esperanza
Grado obtenido	Químico Farmacéutico
Fecha de validación	27/06/2020

DATOS DEL TRABAJO DE INVESTIGACIÓN	
Título de la tesis	Aplicación de minería de datos para mejorar el diagnóstico de un grupo de enfermedades respiratorias en un Hospital de Trujillo
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe
Tipo de instrumento	Ficha de registro
Nombre del indicador	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias

RANGOS DE EVALUACIÓN	Deficiente 0 – 20%	Regular 21 – 50%	Bueno 51 – 70%	Muy bueno 71 – 80 %	Excelente 81 – 100 %
----------------------	-----------------------	---------------------	-------------------	------------------------	-------------------------

N.º	PREGUNTAS	CALIFICACIÓN	OBSERVACIÓN
1	¿El instrumento de medición cumple con el diseño adecuado?	100%	
2	¿El instrumento de recolección de datos tiene relación con el título de la investigación?	100%	
3	¿El instrumento de recolección de datos tiene la relación con las variables de investigación?	100%	
4	¿El instrumento de recolección de datos facilitará el logro de los objetivos de investigación?	100%	
5	¿El instrumento analiza los datos de la organización?	100%	
6	¿El instrumento de medición explica en forma precisa y clara el grado de cumplimiento de la meta o resultado?	100%	
7	¿El resultado del instrumento es entendible para ser correctamente analizado?	100%	

Fuente: Adaptado de Cesar Robledo Mérida

El instrumento puede ser aplicado: SI (x) NO ()

Sugerencias:



FIRMA DEL EXPERTO

TABLA DE VALIDACIÓN DE INSTRUMENTOS

DATOS DEL EXPERTO	
Apellidos y nombres	Leopoldo Segura Murga
Centro laboral	Vircatex International Trading
Grado obtenido	Ingeniero Industrial
Fecha de validación	01-07-2020

DATOS DEL TRABAJO DE INVESTIGACIÓN	
Título de la tesis	Aplicación de minería de datos para mejorar el diagnóstico de un grupo de enfermedades respiratorias en un Hospital de Trujillo
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe
Tipo de instrumento	Ficha de registro
Nombre del indicador	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias

RANGOS DE EVALUACIÓN	Deficiente 0 – 20%	Regular 21 – 50%	Bueno 51 – 70%	Muy bueno 71 – 80 %	Excelente 81 – 100 %
----------------------	-----------------------	---------------------	-------------------	------------------------	-------------------------

Nº	PREGUNTAS	CALIFICACIÓN	OBSERVACIÓN
1	¿El instrumento de medición cumple con el diseño adecuado?	Muy bueno	
2	¿El instrumento de recolección de datos tiene relación con el título de la investigación?	Muy bueno	
3	¿El instrumento de recolección de datos tiene la relación con las variables de investigación?	Muy bueno	
4	¿El instrumento de recolección de datos facilitará el logro de los objetivos de investigación?	Muy bueno	
5	¿El instrumento analiza los datos de la organización?	Muy bueno	
6	¿El instrumento de medición explica en forma precisa y clara el grado de cumplimiento de la meta o resultado?	Muy bueno	
7	¿El resultado del instrumento es entendible para ser correctamente analizado?	Muy bueno	

Fuente: Adaptado de Cesar Robledo Mérida

El instrumento puede ser aplicado: SI (x) NO ()

Sugerencias: detallar nombre del laboratorio y/o área que hace el análisis y tipo de análisis, también el resultado. _____

**VIRCATEX INTERNACIONAL
TRADING S.A.C.**

 Ing. Leopoldo Segura Murga
 Gerente de Planeamiento
 FIRMA DEL EXPERTO

Anexo 6. Confiabilidad de los instrumentos mediante la V de Aiken

La V de Aiken

$$V = \frac{S}{(n(c - 1))}$$

Dónde:

S: sumatoria de sí

sí: valor asignado por el juez i

n: número de expertos

c: número de valores de la escala de valoración

Intervalos de confianza para la V de Aiken

Límite inferior

$$L = \frac{2nkV + Z^2 - Z\sqrt{4nkV(1 - V) + Z^2}}{2(nk + Z^2)}$$

Límite superior

$$U = \frac{2nkV + Z^2 + Z\sqrt{4nkV(1 - V) + Z^2}}{2(nk + Z^2)}$$

Dónde:

Z: valor en distribución estándar correspondientes al 90, 95 o 99%

n: número de expertos

K: rango de calificaciones posibles

V: valor de la V de Aiken

Confiabilidad - Nivel de morbilidad

Ítem	Calificaciones de los Jueces			Promedio	V	I.C. Al 95%, Z: 1.96	
	Juez 1	Juez 2	Juez 3			Límite Inferior	Límite Superior
1	5	5	4	4.67	0,92	0.65	0.99
2	5	5	4	4.67	0.92	0.65	0.99
3	5	5	4	4.67	0.92	0.65	0.99
4	5	5	4	4.67	0.92	0.65	0.99
5	5	5	4	4.67	0.92	0.65	0.99
6	5	5	4	4.67	0.92	0.65	0.99
7	5	5	4	4.67	0.92	0.65	0.99
V de Aiken General					0.92		

Fuente: elaboración propia de los autores.

Análisis: según la tabla mostrada anteriormente se muestran los 7 ítems que se usaron para la validez del instrumento 1 (Anexo 3), el resultado obtenido de la V de Aiken general es 0.92 ($V > 0.65$), con un nivel de confianza del 95%, lo cual significa que la validación es casi perfecta.

Confiabilidad - Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias

Ítem	Calificaciones de los Jueces			Promedio	V	I.C. Al 95%, Z: 1.96	
	Juez 1	Juez 2	Juez 3			Límite Inferior	Límite Superior
1	5	5	4	4.67	0,92	0.65	0.99
2	5	5	4	4.67	0.92	0.65	0.99
3	5	5	4	4.67	0.92	0.65	0.99
4	5	5	4	4.67	0.92	0.65	0.99
5	5	5	4	4.67	0.92	0.65	0.99
6	5	5	4	4.67	0.92	0.65	0.99
7	5	5	4	4.67	0.92	0.65	0.99
V de Aiken General					0.92		

Fuente: elaboración propia de los autores.

Análisis: según la tabla mostrada anteriormente se muestran los 7 ítems que se usaron para la validez del instrumento 2 (Anexo 4), el resultado obtenido de la V de Aiken general es 0.92 ($V > 0.65$), con un nivel de confianza del 95%, lo cual significa que la validación es casi perfecta.

Confiabilidad - Costo promedio de diagnóstico de un grupo de enfermedades respiratorias.

Ítem	Calificaciones de los Jueces			Promedio	V	I.C. Al 95%, Z: 1.96	
	Juez 1	Juez 2	Juez 3			Límite Inferior	Límite Superior
1	5	5	4	4.67	0,92	0.65	0.99
2	5	5	4	4.67	0.92	0.65	0.99
3	5	5	4	4.67	0.92	0.65	0.99
4	5	5	4	4.67	0.92	0.65	0.99
5	5	5	4	4.67	0.92	0.65	0.99
6	5	5	4	4.67	0.92	0.65	0.99
7	5	5	4	4.67	0.92	0.65	0.99
V de Aiken General					0.92		

Fuente: elaboración propia de los autores.

Análisis: según la tabla mostrada anteriormente se muestran los 7 ítems que se usaron para la validez del instrumento 3 (Anexo 5), el resultado obtenido de la V de Aiken general es 0.92 ($V > 0.65$), con un nivel de confianza del 95%, lo cual significa que la validación es casi perfecta.

Anexo 7. Elección de metodología para desarrollo de minería de datos



TABLA DE VALIDACIÓN DE METODOLOGÍAS

DATOS DEL EXPERTO					
Apellidos y nombres	MENDOZA RIVERA RICARDO DARIO				
Centro laboral	UNIVERSIDAD NACIONAL DE TRUJILLO				
Grado obtenido	DOCTOR				
Fecha de validación	05/06/2020				
DATOS DEL TRABAJO DE INVESTIGACIÓN					
Título de la tesis	Aplicación De Minería De Datos Para Mejorar El Diagnóstico De Un Grupo De Enfermedades Respiratorias En Un Hospital De Trujillo				
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe				
PUNTAJE DE EVALUACIÓN	Deficiente	Regular	Bueno	Muy bueno	Excelente
	1	2	3	4	5
N°	Criterios	METODOLOGÍAS			
		CRISP-DM	SEMMA	KDD (Knowledge Discovery Databases)	
1	La metodología se enfoca en el escenario de esta investigación.	5	4	5	
2	La metodología es flexible y se puede personalizar fácilmente.	5	4	4	
3	La metodología permite crear modelos que se adapten a las necesidades concretas.	5	4	4	
4	Las fases de la metodología son detalladas y se distinguen fácilmente.	5	4	4	
5	La metodología es rápida ante su implementación.	5	4	4	
6	La metodología es base los objetivos y los resultados esperados se distinguen fácilmente.	5	4	5	
TOTAL		30	24	26	

Adaptado de Villegas La Torre; IBM; Brizuela, Castro Blanco; Moine, Haedo, Gordillo.

El instrumento puede ser aplicado: SI (X) NO ()

Sugerencias: TENER CUIDADO CON DATOS

FIRMA DEL EXPERTO

CIP: 51622

TABLA DE VALIDACIÓN DE METODOLOGÍAS

DATOS DEL EXPERTO	
Apellidos y nombres	GOMEZ AVILA JOSE ALBERTO
Centro laboral	UCV – UPN – UNT – DAVOS SALUD
Grado obtenido	DOCTOR EN CIENCIAS E INGENIERIA
Fecha de validación	13/06/2020

DATOS DEL TRABAJO DE INVESTIGACIÓN	
Título de la tesis	Aplicación De Minería De Datos Para Mejorar El Diagnóstico De Un Grupo De Enfermedades Respiratorias En Un Hospital De Trujillo
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe

PUNTAJE DE EVALUACIÓN	Deficiente 1	Regular 2	Bueno 3	Muy bueno 4	Excelente 5
-----------------------	-----------------	--------------	------------	----------------	----------------

N°	Criterios	METODOLOGÍAS		
		CRISP-DM	SEMMA	KDD (Knowledge Discovery Databases)
1	La metodología se enfoca en el escenario de esta investigación.	5	4	4
2	La metodología es flexible y se puede personalizar fácilmente.	5	5	4
3	La metodología permite crear modelos que se adapten a las necesidades concretas.	5	5	5
4	Las fases de la metodología son detalladas y se distinguen fácilmente.	5	5	5
5	La metodología es rápida ante su implementación.	4	5	4
6	La metodología es base los objetivos y los resultados esperados se distinguen fácilmente.	5	4	4
TOTAL		29	28	26

Adaptado de Villegas La Torre; IBM; Brizuela, Castro Blanco; Moine, Haedo, Gordillo.

El instrumento puede ser aplicado: SI (X) NO ()

Sugerencias: _____



José A. Gómez Avila
ING. DE SISTEMAS
R. CIP. 116475

FIRMA DEL EXPERTO

TABLA DE VALIDACIÓN DE METODOLOGÍAS

DATOS DEL EXPERTO	
Apellidos y nombres	Maco Elera, Jimmy Willer
Centro laboral	DevTech Industries
Grado obtenido	Maestría en Ingeniería de Sistemas
Fecha de validación	15/06/2020

DATOS DEL TRABAJO DE INVESTIGACIÓN	
Título de la tesis	Aplicación De Minería De Datos Para Mejorar El Diagnóstico De Un Grupo De Enfermedades Respiratorias En Un Hospital De Trujillo
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe

PUNTAJE DE EVALUACIÓN	Deficiente	Regular	Bueno	Muy bueno	Excelente
	1	2	3	4	5

N°	Criterios	METODOLOGÍAS		
		CRISP-DM	SEMMA	KDD (Knowledge Discovery Databases)
1	La metodología se enfoca en el escenario de esta investigación.	3	3	4
2	La metodología es flexible y se puede personalizar fácilmente.	2	2	2
3	La metodología permite crear modelos que se adapten a las necesidades concretas.	3	3	4
4	Las fases de la metodología son detalladas y se distinguen fácilmente.	3	3	4
5	La metodología es rápida ante su implementación.	3	4	3
6	La metodología es base los objetivos y los resultados esperados se distinguen fácilmente.	4	4	4
TOTAL		3	3	4

Adaptado de Villegas La Torre; IBM; Brizuela, Castro Blanco; Moine, Haedo, Gordillo.

El instrumento puede ser aplicado: SI (X) NO ()

Sugerencias: _____



FIRMA DEL EXPERTO

Anexo 8. Tabla resumen elección de metodología

Expertos	Metodologías		
	CRISP-DM	SEMMA	KDD
Doc. Mendoza Rivera, Ricardo Dario	30	24	26
Doc. Gomez Avila, José Alberto	29	28	26
Mtro. Maco Elera, Jimmy Willer	18	18	21
Total	77	70	73

Fuente: elaboración propia de los autores

En la tabla se visualiza metodologías para el desarrollo de la aplicación de minería de datos que fueron elegidas por los tres expertos y el puntaje obtenido de la elección para las metodologías (Anexo 7), CRISP-DM tiene un total de 77 puntos, SEMMA tiene un total de 70 puntos y KDD tiene un total de 73

Anexo 9. Fotos y documentos

Anexo 9.1 Documento de CO-Asesoría



"Año de la universalización de la Salud"

CONSTANCIA

Por la presente se extiende la Constancia de co-asesoría para realizar el trabajo de investigación "APLICACIÓN DE MINERÍA DE DATOS PARA MEJORAR EL DIAGNÓSTICO DE UN GRUPO DE ENFERMEDADES RESPIRATORIAS EN UN HOSPITAL DE TRUJILLO" de los estudiantes Ronald Segundo Alva Mariños con DNI 74972186 y Luis Felipe Cruz Isla con DNI 73451110 de la carrera profesional de Ingeniería de Sistemas de la Universidad Cesar Vallejo, que se realizará en la entidad CAP III Metropolitano Trujillo.

Se extiende la constancia a solicitud de los interesados, para fines académicos.


Trujillo, 01 de Julio del 2020.





Ing. Margarita Santapac Paredes
CIP 84809
JEFE ADMINISTRATIVO
CAP III METROPOLITANO
RED ASISTENCIAL LA LIBERTAD



Anexo 9.2. Correo de validaciones para los instrumentos


Re: validacion de instrumentos

 CABURE EIRL Leo Segura <lsegura@polojean...>
Para Luis Felipe Cruz Isla
CC ronaldalvamarinos@outlook.com.pe
jueves 2/07

 Respondió a este mensaje el 2/07/2020 07:47.

 UCV formato 1.pdf
397 KB

 UCV formato 3.pdf
392 KB

 UCV formato 2.pdf

atte
Leopoldo Segura
EL CABURE EIRL
www.polojeanscbr.com

El lun., 29 de jun. de 2020 a la(s) 19:24, Luis Felipe Cruz Isla (luisfelipe.cruzisla@hotmail.com) escribió:

Buen día Mg. Leopoldo Segura, Envió el siguiente adjunto con contiene los formatos para las validaciones de instrumentos que se va usar en mi investigación.

cualquier observación comunicarme por medio de correo o por mi número personal (+ 51 986 675 458)

Gracias y un saludo.

RE: Validación de Instrumentos



Hilda Edith Moreno Zavaleta <edithzavaleta2

Para Luis Felipe Cruz Isla



domingo 28/06

De: Luis Felipe Cruz Isla <luisfelipe.cruzisla@hotmail.com>

Enviado: sábado, 27 de junio de 2020 23:47

Para: Hilda Edith Moreno Zavaleta <edithzavaleta2012@hotmail.com>

Asunto: RE: Validación de Instrumentos

Buen día, le envió este correo con razón de agradecimiento por haber participado como experto en mi investigación.

Muchas gracias y un saludo.

De: Hilda Edith Moreno Zavaleta <edithzavaleta2012@hotmail.com>

Enviado el: sábado, 27 de junio de 2020 22:13


Para: Luis Felipe Cruz Isla <luisfelipe.cruzisla@hotmail.com>


Asunto: RE: Validación de Instrumentos


FIRMA DEL EXPERTO

Anexo 9.3. Correo de validaciones para metodología

Re: Formato de validación de experto

 gavalaj@ucvirtual.edu.pe
Para Luis Felipe Cruz Isla 13/06/2020

 Respondió a este mensaje el 13/06/2020 22:46.
Si hay problemas con el modo en que se muestra este mensaje, haga clic aquí para verlo en un explorador web.

 Formato para validación de metodologías (1).pdf
91 KB

El 12/06/20 a las 19:11, Luis Felipe Cruz Isla escribió:

Se me olvido el instrumento, Adjunto documento.

De: Luis Felipe Cruz Isla
Enviado: sábado, 13 de junio de 2020 00:10
Para: gavalaj@ucvirtual.edu.pe <gavalaj@ucvirtual.edu.pe>
Cc: ronaldalvamarinos@outlook.com.pe <ronaldalvamarinos@outlook.com.pe>
Asunto: Formato de validación de experto

Buen día Ing. José Gómez Avila, Envió el siguiente documento con contiene las posibles metodologías a usar en mi investigación.

cualquier observación comunicarme por medio de correo o por mi número personal (+ 51 986 675 458)

Gracias y un saludo.

RE: Formato de validación de experto



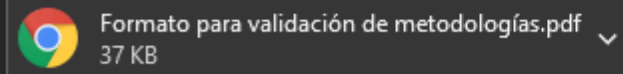
Ricardo Mendoza <rimenri@hotmail.com>

Para Luis Felipe Cruz Isla



5/06/2020

Respondió a este mensaje el 5/06/2020 18:09.



Luis

Envío formato validado

Saludos

Dr. Ing. Ricardo D. Mendoza Rivera
IT ESAN - Ms. Ing. de Sistemas
Cel: (051)949511552 · rimenri@premiunnet.com
www.premiunnet.com

[Premiun.NET BlogPersonal](#)

De: Luis Felipe Cruz Isla <luisfelipe.cruzisla@hotmail.com>

Enviado: miércoles, 3 de junio de 2020 0:04

Para: Rimenri@hotmail.com <Rimenri@hotmail.com>

Asunto: Formato de validación de experto

Buen día Ing Rimenri. Envío el siguiente documento con contiene las posibles

Anexo 10. Desarrollo de la Metodología

Fase 1: Comprensión del negocio

1.1 Determinar los objetivos del negocio

- Recopilar información de la empresa:

En el Centro de Atención Primaria CAP III Metropolitano Trujillo se detectó un incremento de casos, por lo que, en la mayoría de estos fueron enfermedades respiratorias, los cuales afectaron a niños entre 1 a 5 años, viéndose reflejado la poca eficacia en el proceso de diagnóstico médico y tratamientos en el departamento de Pediatría de dicho conjunto de enfermedades, ocasionando un aumento en el nivel de morbilidad, tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias y costo promedio de diagnóstico del dicho grupo.

- Definir los objetivos del negocio:

Brindar una atención con calidad y calidez a miles de asegurados.

Atender a pacientes con enfermedades de especialidades complejas.

- Definir los criterios de rendimiento comercial:

La posibilidad de realizar diagnósticos a pacientes con enfermedades respiratorias con un elevado porcentaje de fiabilidad.

1.2 Evaluación de la situación

- Analizar terminología

Ver anexo: Glosario de terminología del negocio

- Definir requisitos

Reporte de Historias clínicas de pacientes

Reporte de costos

Reporte de atenciones médicas

- Definir riesgos

Acceso a la base de datos

- Definir contingencias

Solicitar reportes de historias clínicas

- Elaborar análisis de costes/beneficios

Los datos de este proyecto tienen un costo de 100 dólares.

1.3 Determinar los objetivos de minería de datos

- Definir objetivos de minería de datos
 - ✓ Nivel de morbilidad
 - ✓ Disminuir el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias.
 - ✓ Disminuir el costo promedio de diagnóstico de un grupo de enfermedades respiratorias.
- Definir criterios de rendimiento de minería de datos

Se establece como criterio la posibilidad de realizar predicciones sobre la enfermedad respiratoria que sufre un paciente con un elevado porcentaje de fiabilidad.

1.4 Realizar el plan del proyecto

El proyecto se dividirá en las siguientes etapas para facilitar su organización y estimar el tiempo de realización del mismo

Etapas 1: Análisis de la estructura de los datos y la información de la base de datos. Tiempo estimado: 2 semanas

Etapas 2: Ejecución de consultas para tener muestras representativas de los datos. Tiempo estimado: 1 semana

Etapas 3: Preparación de los datos (selección, limpieza, conversión y formateo, si fuera necesario) para facilitar la minería de datos sobre ellos. Tiempo estimado: 3 semanas

Etapas 4: Elección de las técnicas de modelado y ejecución de las mismas sobre los datos. Tiempo estimado: 1 semana

Etapas 5: Análisis de los resultados obtenidos en la etapa anterior, si fuera necesario repetir la etapa 4. Tiempo estimado: 1 semana

Etapas 6: Producción de informes con los resultados obtenidos en función de los objetivos de negocio y los criterios de éxito establecidos. Tiempo estimado: 1 semana

Etapas 6: Producción de informes con los resultados obtenidos en función de los objetivos de negocio y los criterios de éxito establecidos. Tiempo estimado: 1 semana

Fase 2: Comprensión de los datos

2.1 Recolectar los datos iniciales

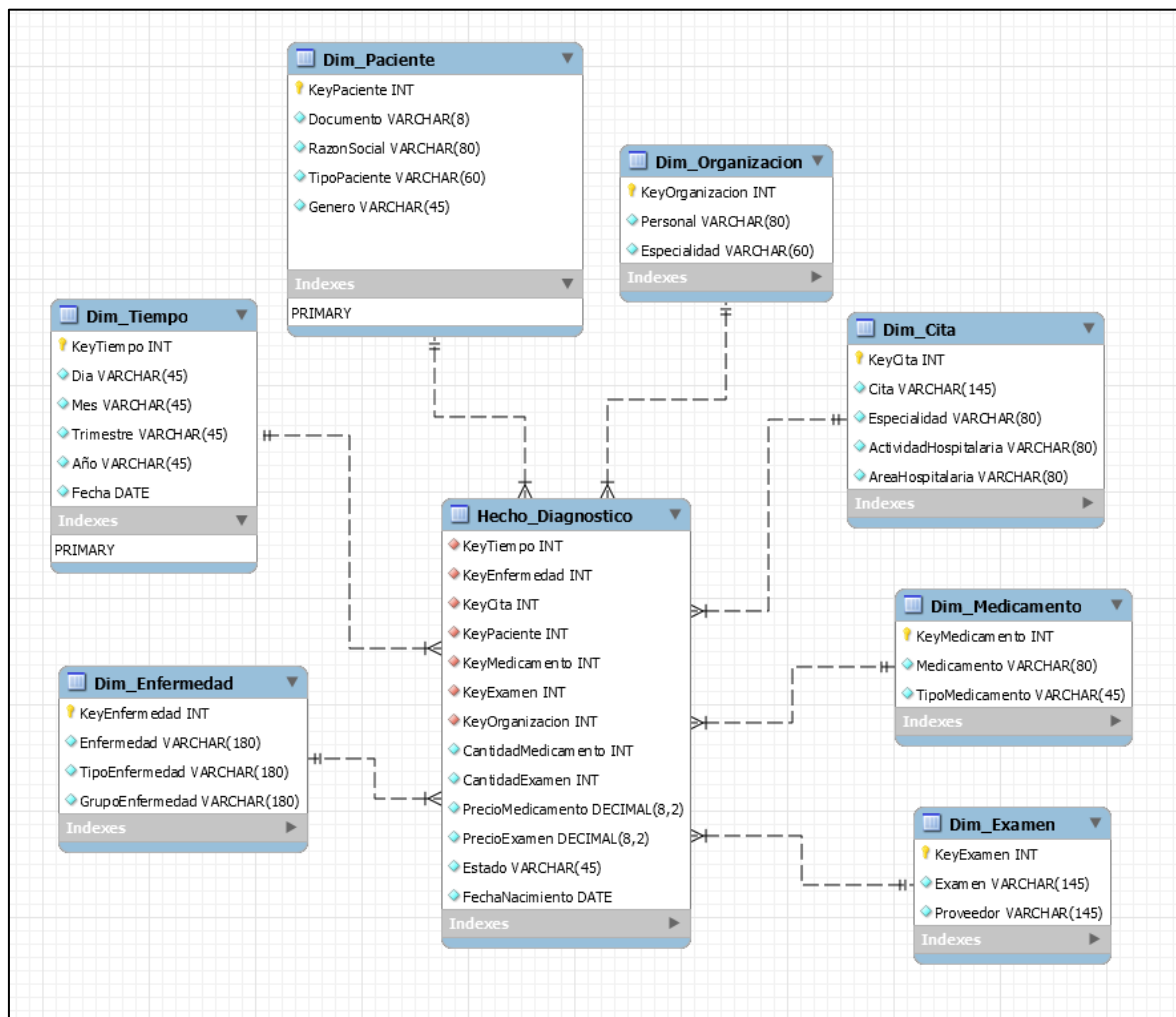
Los datos empleados en este proyecto de investigación son de historias medicas de pacientes del Centro de Atención Primaria Metropolitano CAP III Trujillo que incluyen información personal sobre ellos como: nombres, apellidos, DNI, fecha de nacimiento, entre otros.

Listado de los datos adquiridos:

- Paciente
- Enfermedad
- Cita
- Examen
- Medicamento
- Personal

2.2 Descripción de los datos

Los datos se encuentran almacenados en un almacén de datos con esquema dimensional en estrella. Podemos observar el esquema relaciona de la base de datos, la cual fue generada por la herramienta MySQL Workbench.



Se puede observar que el almacén de datos consta de ocho tablas: Hecho_Diagnostico, Dim_Paciente, Dim_Organizacion, Dim_Cita, Dim_Medicamento, Dim_Examen, Dim_Tiempo y Dim_Enfermedad, a continuación, describiremos cada una de ellas detallando cada uno de sus campos.

Tabla Hecho_Diagnostico

Esta tabla es la tabla central del almacén de datos, también llamada tabla de hechos. Al ser la tabla central tiene como clave primaria una combinación de cada una de las claves principales del resto de las tablas, estas claves son: KeyTiempo, KeyEnfermedad, KeyCita, KeyPaciente, KeyMedicamento, KeyExamen y KeyOrganizacion. Los campos de cada registro de esta tabla son:

- KeyTiempo: Tipo entero. Este campo es un número que identifica a cada fecha y que es único.
- KeyEnfermedad: Tipo entero. Este campo es un número que identifica a cada enfermedad y que es único.
- KeyCita: Tipo entero. Este campo es un número que identifica a cada cita y que es único.
- KeyPaciente: Tipo entero. Este campo es un número que identifica a cada paciente y que es único.
- KeyMedicamento: Tipo entero. Este campo es un número que identifica cada medicamento y que es único.
- KeyExamen: Tipo entero. Este campo es un número que identifica cada examen y que es único.
- KeyOrganizacion: Tipo entero. Este campo es un número que identifica cada personal y que es único.
- CantidadMedicamento: Tipo entero. Este campo es un número que identifica cada la cantidad de medicamentos.
- CantidadExamen: Tipo entero. Este campo es un número que identifica cada la cantidad de exámenes.
- PrecioMedicamento: Tipo decimal. Este campo representa el precio de cada medicamento.
- PrecioExamen: Tipo decimal. Este campo representa el precio de cada examen.
- Estado: Tipo alfanumérico. Este campo representa el estado del diagnóstico.

- FechaNacimiento: Tipo fecha. Este campo representa la fecha de nacimiento del paciente.

TABLA Dim_PACIENTE

KeyPaciente: Tipo entero. Este campo es un número que identifica a cada paciente y que es único.

Documento: tipo alfanumérico. Este campo representa el documento de identificación del paciente.

RazonSocial: tipo alfanumérico. Este campo representa el nombre completo de un paciente.

TipoPaciente: Tipo alfanumérico. Este campo representa el tipo de paciente.

Género: Tipo alfanumérico. Este campo describe el género del paciente.

Tabla Dim_ENFERMEDAD

KeyEnfermedad: Tipo entero. Este campo es un número que identifica a cada enfermedad y que es único

Enfermedad: Tipo alfanumérico. Este campo describe la enfermedad que tiene el paciente.

TipoEnfermedad: Tipo alfanumérico. Este campo representa el tipo de enfermedad.

GrupoEnfermedad: Tipo alfanumérico. Este campo representa el grupo al que pertenece una enfermedad.

Tabla Dim_CITA

KeyCita: Tipo entero. Este campo es un número que identifica a cada cita y que es único.

Cita: Tipo alfanumérico. Este campo describe la cita que tiene el paciente.

Especialidad: tipo alfanumérico. Este campo representa la especialidad en la cita médica.

ActividadHospitalaria: tipo alfanumérico. Este campo representa la actividad solicitada en la cita médica.

Areahospitalaria: tipo alfanumérico. Este campo representa el área donde se realizará la cita médica.

Tabla Dim_EXAMEN

KeyExamen: Tipo entero. Este campo es un número que identifica cada examen y que es único.

Examen: Tipo alfanumérico. Este campo describe el examen que se ha realizado.

Proveedor: Tipo alfanumérico. Este campo representa el proveedor de un examen.

Tabla Dim_Medicamento

KeyMedicamento: Tipo entero. Este campo es un número que identifica cada medicamento y que es único.

Medicamento: Tipo alfanumérico. Este campo describe el medicamento recetado.

TipoMedicamento: Tipo alfanumérico. Este campo representa el tipo de medicamento

Tabla Dim_Organizacion

KeyOrganizacion: Tipo entero. Este campo es un número que identifica cada personal y que es único.

Personal: tipo alfanumérico. Este campo representa el nombre completo del profesional asistencial.

Especialidad: tipo alfanumérico. Este campo representa la especialidad del personal.

Tabla Dim_Tiempo

KeyTiempo: Tipo entero. Este campo es un número que identifica a cada fecha y que es único.

Dia: Tipo fecha. Este campo representa el día que se realizó el diagnostico.

Mes: Tipo fecha. Este campo representa el mes que se realizó el diagnostico.

Trimestre: Tipo fecha. Este campo representa el trimestre que se realizó el diagnostico.

Año: Tipo fecha. Este campo representa el año que se realizó el diagnostico.

Fecha: Tipo fecha. Este campo representa la fecha completa que se realizó el diagnostico.

2.3 Exploración de los datos

Una vez que se han descrito los datos, se procede a explorarlos, esto implica aplicar pruebas estadísticas básicas que revelarán propiedades de los datos, y crear tablas de frecuencia y gráficos de distribución de los datos. Este informe sirve principalmente para determinar la consistencia y completitud de los datos.

En la tabla 1 muestra los costos de los diagnósticos realizados en el año, esta tabla se aprecia las diferencias de costo entre meses.

Quiero ver 2019 ▾

Costo promedio Anual de diagnóstico

Mes actual	Costo promedio del mes actual	Mes Anterior	Costo promedio del Mes siguiente	Diferencia de costos
Julio	S/. 10,946.78	-	S/. 0.00	S/. 10,946.78(-)
Agosto	S/. 11,121.65	Julio	S/. 10,946.78	S/. 174.87(-)
Septiembre	S/. 10,636.66	Agosto	S/. 11,121.65	S/. 484.99(+)
Octubre	S/. 10,545.49	Septiembre	S/. 10,636.66	S/. 91.17(+)
Noviembre	S/. 7,136.96	Octubre	S/. 10,545.49	S/. 3,408.54(+)
Diciembre	S/. 10,332.11	Noviembre	S/. 7,136.96	S/. 3,195.16(-)

En la tabla 2 muestra los costó de medicamentos según el grupo de enfermedades (CIE 10), esta tabla se aprecia el código de la enfermedad, su descripción de la enfermedad, el costo del mismo, la cantidad de pacientes involucrados y costo promedio por paciente, definido en un rango de fechas.

Rango de Fechas

2019-01-01 a 2021-06-27

Costo de medicamentos según grupo de enfermedades

Search:

Código de Enfermedad	Grupo de Enfermedad	Costo de la enfermedad	Cantidad de pacientes	Costo promedio por paciente
N17-N19	Insuficiencia renal	7561.69931	125	60.49359448
J00-J06	Infecciones agudas de las vías respiratorias superiores	3924.4247732061	413	9.50223916030534
J30-J39	Otras enfermedades de las viías respiratorias superiores	1372.0105709524	159	8.6289973015873
J40-J47	Enfermedades crónicas de las vías respiratorias inferiores	1052.286535	59	17.835365
J20-J22	Otras infecciones agudas de las vías respiratorias inferiores	352.4739957143	29	12.15427571428571
N10-N16	Enfermedad renal tubulointersticial	34.229958	21	1.629998
J20-J22	Otras infecciones agudas de las vías respiratorias inferiores	65.086	8	8.13575
J95-J99	Otras	91.348205	13	7.026785
H65-H75	Enfermedades del oído medio y de la mastoides	428.124645	39	10.977555
A00-A99	Enfermedades infecciosas intestinales	84.194	44	1.9135

Showing 1 to 10 of 16 entries

Previous

1

2

Next

Rango de Fechas

2019-01-01 a 2021-04-02

Costo de medicamentos según grupo de enfermedades

Search:

Código de Enfermedad	Grupo de Enfermedad	Costo de la enfermedad	Cantidad de pacientes	Costo promedio por paciente
N17-N19	Insuficiencia renal	7561.69931	125	60.49359448
J00-J06	Infecciones agudas de las vías respiratorias superiores	3924.4247732061	413	9.50223916030534
J30-J39	Otras enfermedades de las vías respiratorias superiores	1372.0105709524	159	8.6289973015873
J40-J47	Enfermedades crónicas de las vías respiratorias inferiores	1052.286535	59	17.835365
J20-J22	Otras infecciones agudas de las vías respiratorias inferiores	352.4739957143	29	12.15427571428571
N10-N16	Enfermedad renal tubulointersticial	34.229958	21	1.629998
J20-J22	Otras infecciones agudas de las vías respiratorias inferiores	65.086	8	8.13575
J95-J99	Otras	91.348205	13	7.026785
H65-H75	Enfermedades del oído medio y de la mastoides	428.124645	39	10.977555
A00-A99	Enfermedades infecciosas intestinales	84.194	44	1.9135

Showing 1 to 10 of 16 entries

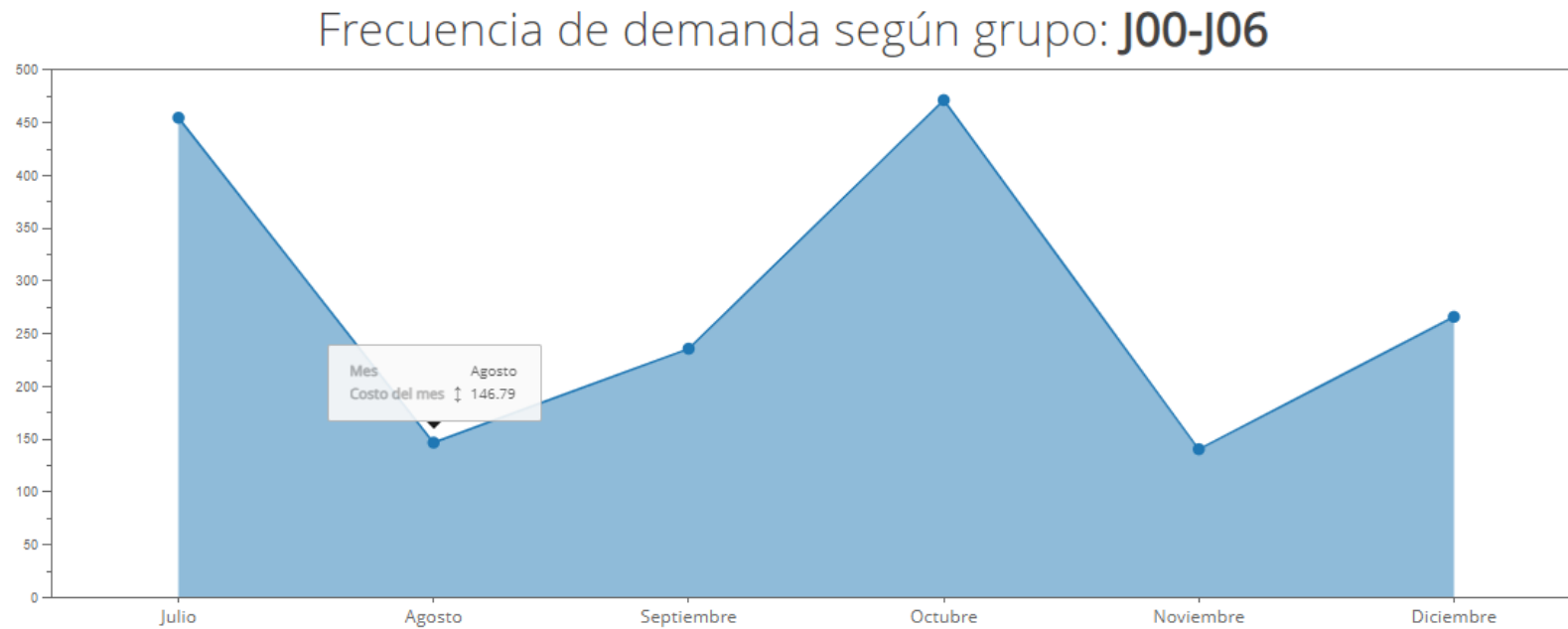
Previous

1

2

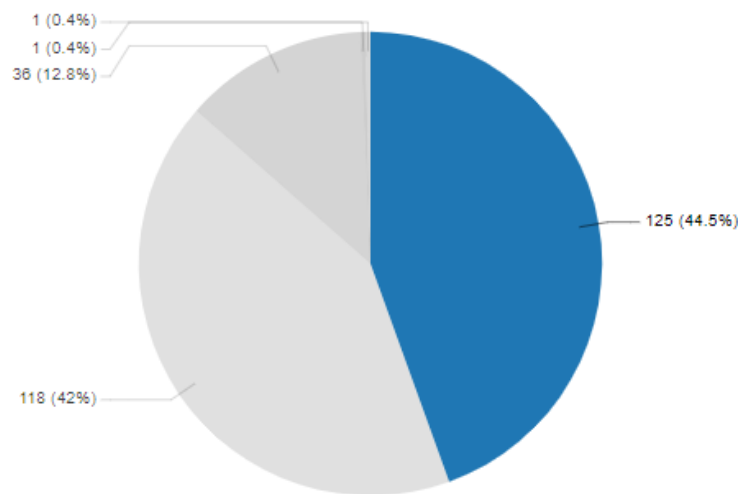
Next

En la figura 3 muestra la frecuencia de demanda según el grupo de enfermedades seleccionado en los costos de medicamentos de la tabla anterior, la gráfica se aprecia los costos de mes a mes (según el rango de fechas).



En la figura 4 muestra el top de tipo de enfermedades con más atenciones del grupo seleccionado, la gráfica se aprecia la cantidad de atenciones y el tipo de enfermedad (CIE10) según el rango de fechas.

Top tipos de enfermedades con más atenciones del grupo **J30-J39**



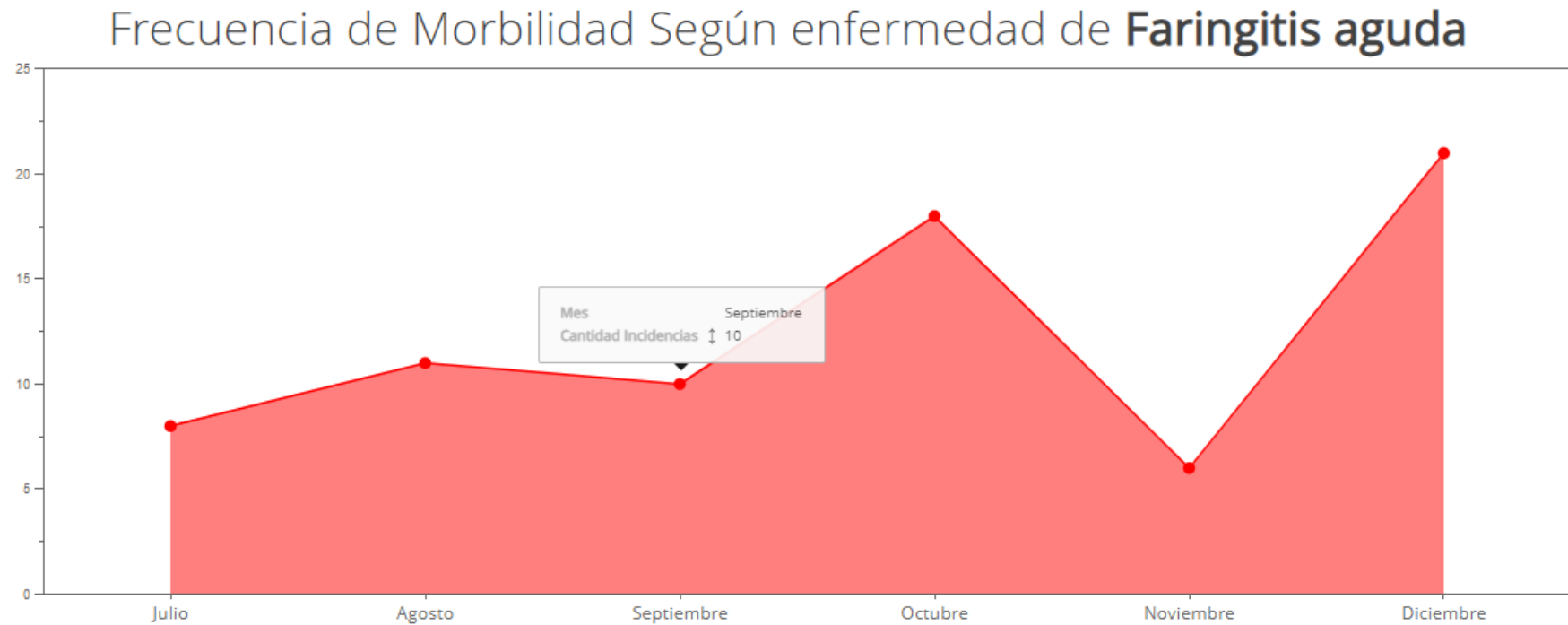
Search:

Enfermedad	Cantidad de Atenciones
Rinitis alérgica y vasomotora	125
Otras enfermedadaes de las vías respiratorias superiores	118
Enfermedades crónicas de las amígdalas y de las adenoides	36
Laringitis y laringotraqueítis crónicas	1
Rinitis rinofaringitis y faringitis crónicas	1

Showing 1 to 5 of 5 entries

Previous Next

En la figura 5 muestra la frecuencia de Morbilidad según enfermedad seleccionada en la gráfica anterior, la gráfica describe la cantidad de casos según de forma de tendencia a lo largo de los meses.



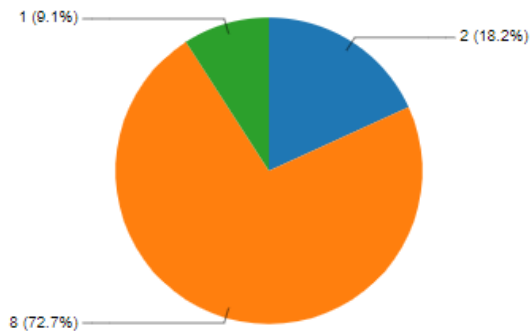
En la figura 6 muestra la cantidad de enfermedades de atenciones relacionadas que medico lo atendió de un paciente especifico, se aprecia en la gráfica y en tabla la cantidad de atenciones hechas por médico y la cantidad de atenciones por enfermedad.

Paciente: [REDACTED]
Genero: Fememino
Documento: [REDACTED]

Especialidad: Pediatría
Edad: 5 Años con 6 Meses y 1 Día

Rango de Fechas 2019-01-01 a 2021-04-02

Atenciones por medico



Search:

Enfermedad	Cantidad de atenciones relacionadas
J20.9/Bronquitis aguda, no especificada	9
J30.4/Rinitis alérgica, no especificada	2
U07.2/COVID-19, virus no identificado	2
N39.0/Infección de vías urinarias, sitio no especificado	1
B34.9/Infección viral, no especificada	1

Showing 1 to 5 of 8 entries

Previous **1** 2 Next

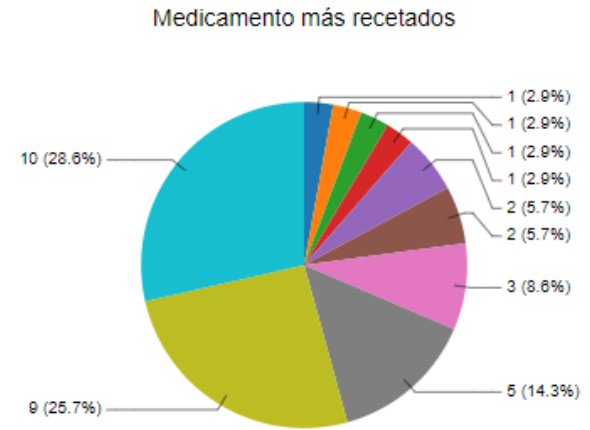
En la figura 7 muestra la cantidad de medicamentos recetados por el paciente seleccionado.

Search:

Medicamento	Cantidad de Medicamentos
AEROCAMARA PEDIATRICO	1
FLUTICASONA 0.125 MG P/INHAL AEROSOL	1
AZITROMICINA 200 MG/5 ML X 30 ML + DOSIFICADOR SUSPENSION ORAL	1
SALBUTAMOL 100 MG P/INHAL AEROSOL X 200 A 250 DOSIS	1
DEXTROMETORFANO 15 MG/5 ML JARABE X 100 A 120 ML	2

Showing 1 to 5 of 10 entries

Previous 1 2 Next



2.4 Verificar la calidad de los datos

Después de hacer la exploración inicial de los datos se puede afirmar que estos son completos. Los datos cubren los casos requeridos para la obtención de los resultados necesarios para poder cumplir los objetivos del proyecto. Los datos no contienen errores, estos son datos propios de las actas médicas. Tampoco se encuentran valores fuera de rango, dado que los datos son controlados por medio de la integración, por lo que no hay riesgo de ruido en el proceso de la minería de datos.

3. Preparación de los datos

3.1 Seleccionar los datos

En términos de registros, se van a utilizar todos los registros dentro de cada tabla que compone la base de datos, puesto que al ser ésta una base de datos específicamente creada para este proyecto. Sin embargo, hay campos dentro de estos registros que no son necesarios para nuestros objetivos de minería de datos, por lo que se puede prescindir de algunos de ellos.

TABLA Dim_PACIENTE

- Documento.
- RazonSocial.
- Género

Tabla Dim_ENFERMEDAD

- Enfermedad.
- TipoEnfermedad.
- GrupoEnfermedad.

Tabla Dim_EXAMEN

- KeyExamen.
- Examen

Tabla Dim_Medicamento

- Medicamento.

Tabla Dim_Organizacion

- Personal.
- Especialidad.

Tabla Dim_Tiempo

- Mes.
- Año.
- Fecha.

Tabla Hecho_Diagnostico

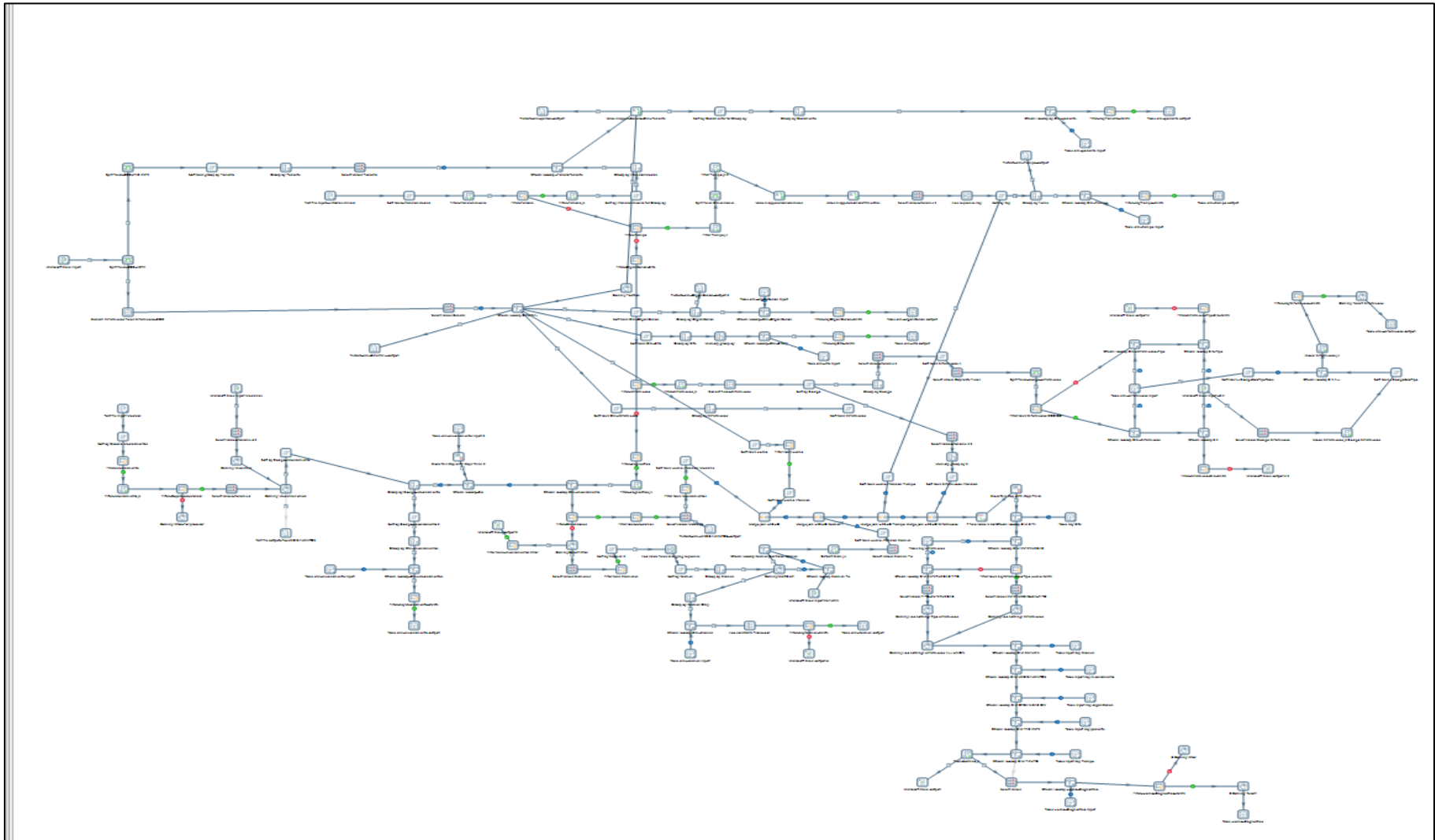
- KeyTiempo.
- KeyEnfermedad.
- KeyPaciente.
- KeyMedicamento.
- KeyExamen.
- KeyOrganizacion.
- CantidadMedicamento.
- CantidadExamen.
- PrecioMedicamento.
- PrecioExamen.
- ActaMedica.
- Paciente_Anual.
- Paciente_Mes.
- Paciente_Dia.
- Estado.
- FechaNacimiento.

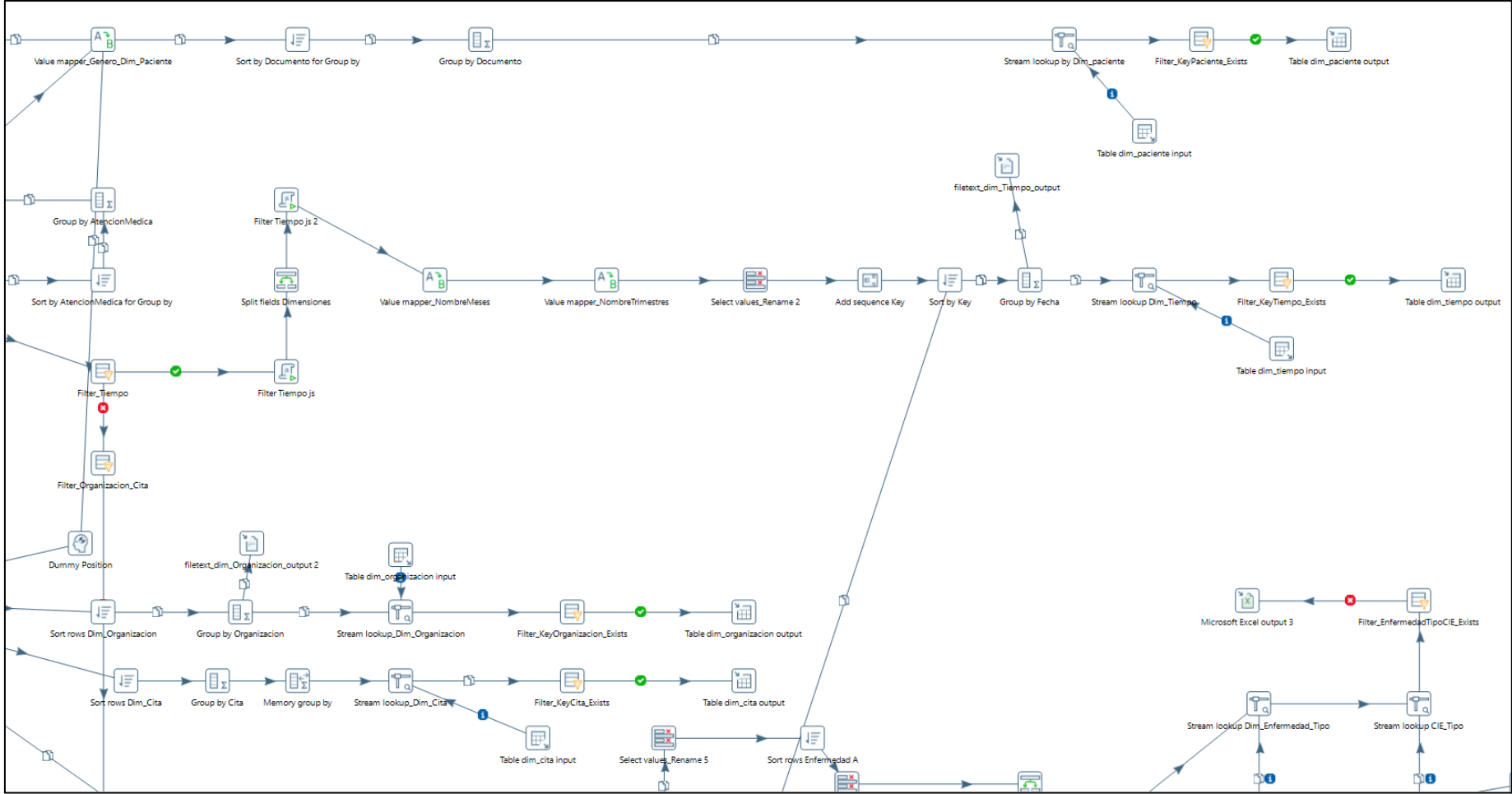
3.2 Limpiar los datos

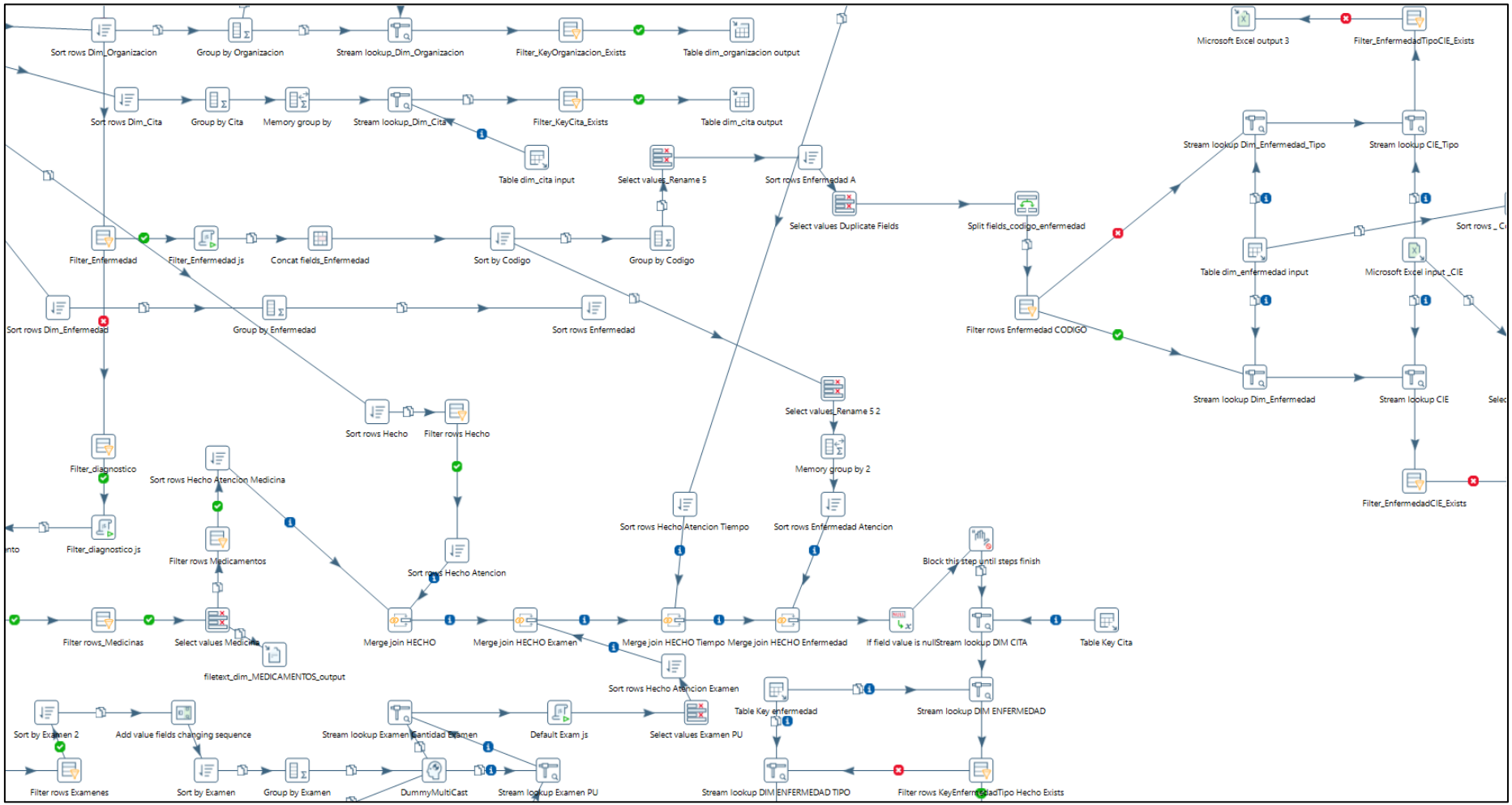
La base de datos con la que se cuenta para el proyecto contiene toda la información necesaria para poder cumplir los objetivos, además, estos datos son obtenidos por atenciones médicas para el caso que se presenta, son datos limpios y por lo tanto no hay necesidad de hacer una limpieza más profunda sobre ellos.

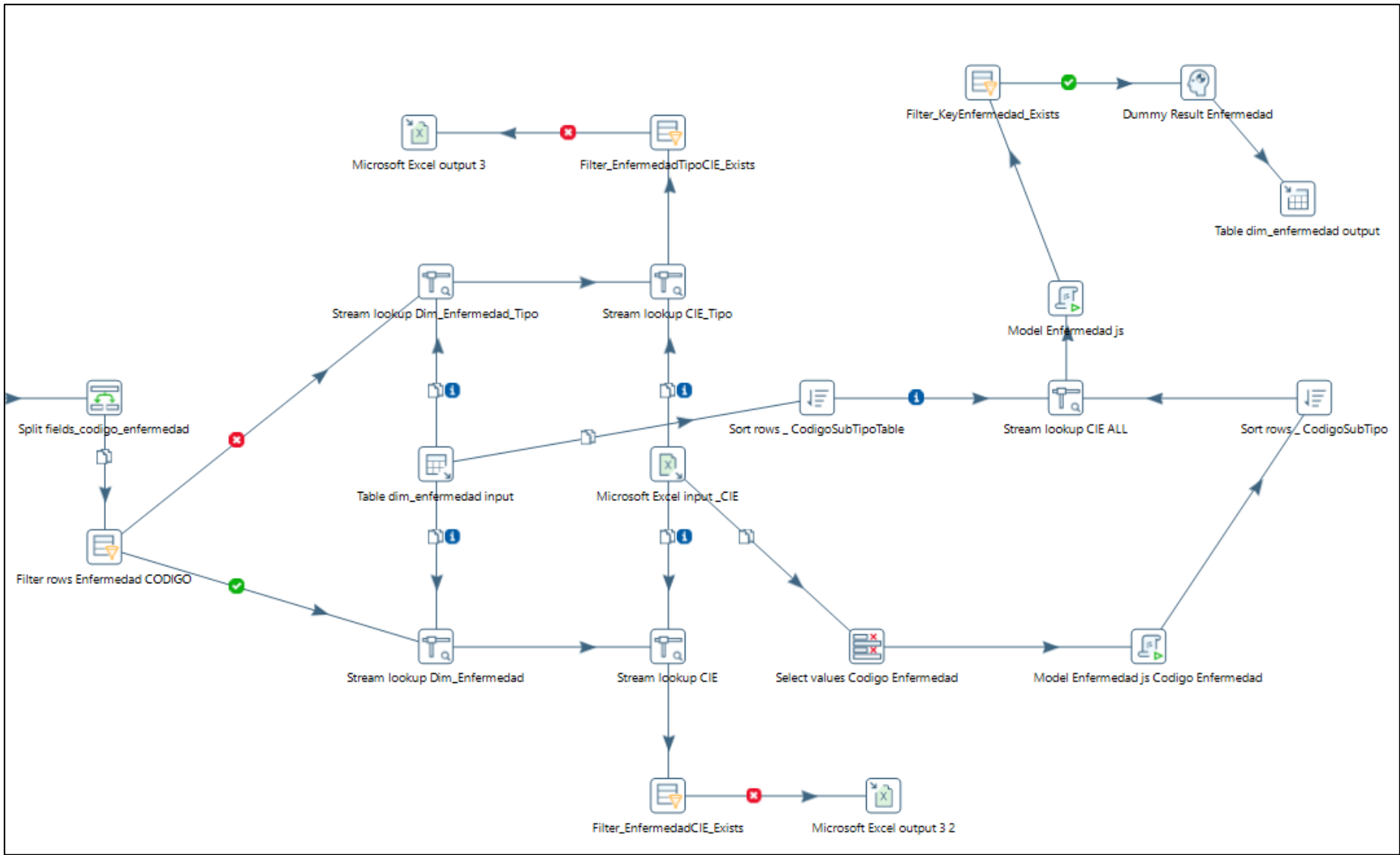
Tampoco tenemos campos en los que falten valores, más allá de los valores nulos que aparecen cuando la información que se quiere representar no existe, y por lo tanto no se consideran como datos faltantes, por lo que no es necesario realizar ningún tipo de estimación de valores faltantes. Estos valores nulos se tratarán a la hora de hacer la minería de datos simplemente ignorándolos porque no aportan ninguna información adicional al estudio.

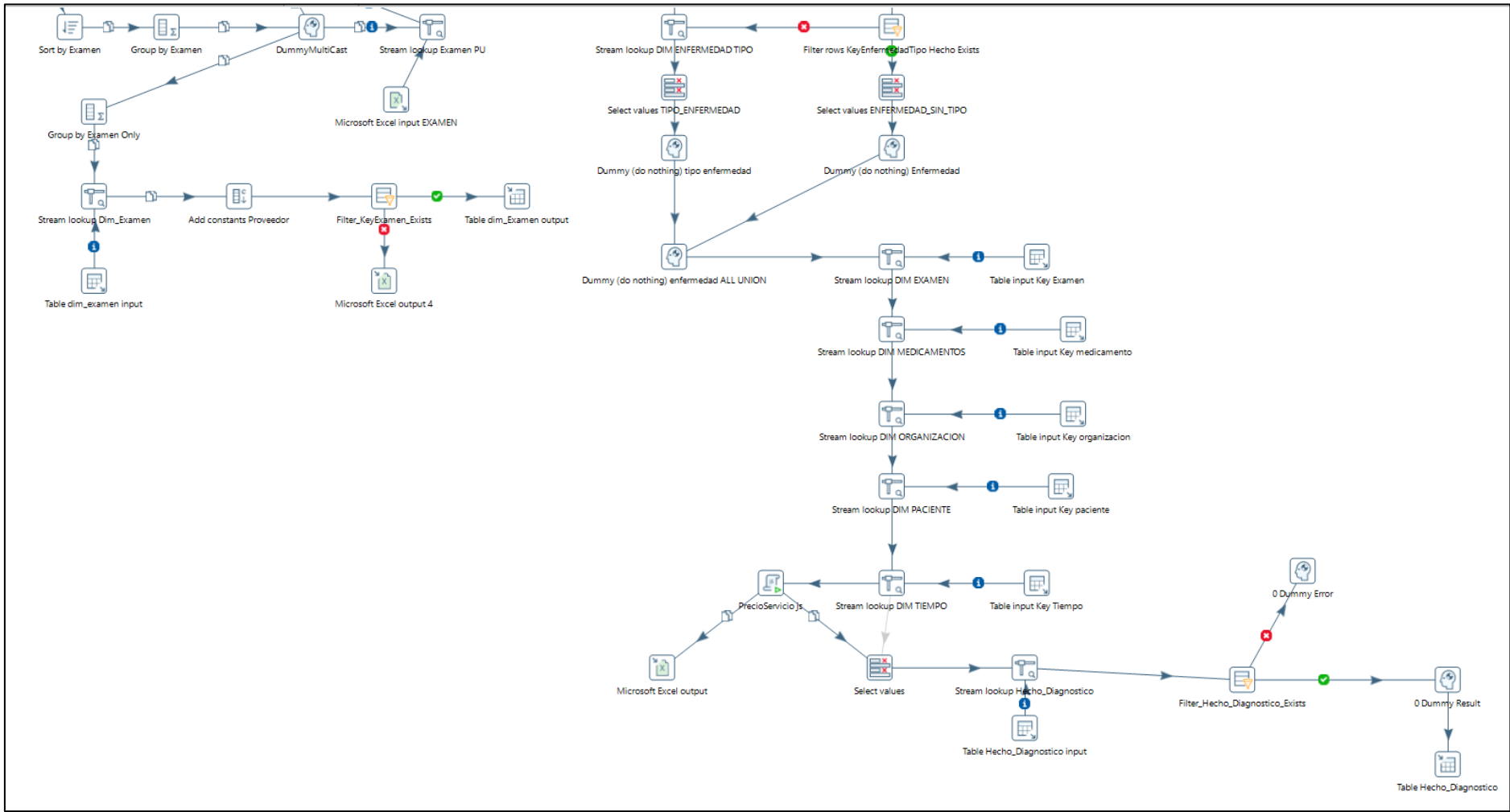
Para cargar el modelo de datos se usó la herramienta Pentaho data integration(PDI).











3.3 Construir los datos

En este apartado sólo se puede destacar la transformación del campo Fecha de la tabla Dim_Tiempo. Dicha transformación consistió en codificar numéricamente los valores del campo que inicialmente contenía caracteres de tipo fecha con el nombre del tiempo según su tipo de granularidad (Año, Mes y Día).

3.4 Integrar los datos

No ha sido necesaria la creación de nuevas estructuras (campos, registros, etc.), ni la fusión entre distintas tablas de la base de datos, el programa Pentaho data integration(PDI) se encargará de realizar estas tareas automáticamente sin que el usuario tenga que crear nuevas tablas, registros o campos manualmente.

3.5 Formateo de los datos

No es necesario cambiar el orden de ningún campo dentro de los registros, ni tampoco la reordenación de los registros dentro de las tablas. Tampoco es necesario cambiar el formato de ninguno de los campos que se van a utilizar para la minería de datos.

4. Modelado

4.1. Escoger la Técnica de Modelado

Se va a utilizar el software Pentaho Data Integration para realizar la minería de datos, deberemos utilizar alguna de las técnicas de modelado que nos ofrece esta herramienta de acuerdo con los objetivos de nuestro proyecto que están reflejados en el apartado 1.3 (Objetivos de la minería de datos).

4.2 Construir el modelo

Como principios se usó el modelado de base de datos tipo estrella, porque es el más práctico a la hora de diseñar un datawarehouse y se ajusta a un mejor rendimiento en el filtrado y agrupamiento de información.

5. Evaluación

En esta fase de la metodología se intentan evaluar los modelos generados, pero en esta ocasión la evaluación se hace desde el punto de vista de los objetivos de negocio en lugar de los objetivos de minería de datos. Una vez realizada esta evaluación, se debe decidir si los objetivos han sido cumplidos y de ser así se puede avanzar a la fase de implantación, de lo contrario se tendría que identificar cualquier factor que se haya podido pasar por alto y hacer una revisión del proceso.

5.1 Determinar los próximos pasos

El siguiente paso a realizar en el proyecto es el de ejecutar la etapa de implantación.

6. Implantación.

Esta es la última fase de la metodología CRISP-DM y el objetivo de la misma es el de explicar al cliente como poner en funcionamiento el proyecto que se ha construido en las fases anteriores, así como exponer los resultados obtenidos al cliente de forma que lo pueda entender fácilmente. Otro objetivo de esta fase es el de crear una estrategia para el mantenimiento del proyecto y producir un informe en el que se incluyan posibles mejoras para el futuro y un listado de las dificultades encontradas a la hora de realizarlo.

6.1 Plantear la implantación

Para poder implantar este proyecto en el negocio sería necesario en primer lugar tener acceso a la base de datos del negocio, es decir la base de datos que contiene toda la información de los pacientes. A partir de ahí, los pasos a seguir serían los mismos que se han seguido en este documento desde la comprensión del negocio hasta la implantación. Si bien, cabe decir que habrá algunas fases, como la de comprensión y preparación de los datos, que en el negocio real probablemente sean más complejas y llevarán más tiempo que en este proyecto se puede esperar que en la base de datos se tengan muchos más registros y estos mismos contengan más ruido que en nuestra base de datos basado en algunas actas de pacientes.

6.2. Planear la Monitorización y Mantenimiento

La supervisión y mantenimiento de la implementación del presente proyecto es una fase importante del mismo debido a que los datos que se procesan con mucha frecuencia pueden ser modificados por el personal médico. Los datos pueden ser modificados por diferentes motivos como haber realizado una codificación incorrecta, haber asignado un diagnóstico incorrecto al paciente, etc. El volumen de estos datos en movimiento es grande motivo por el cual la extracción de las muestras debe ser realizada cuidadosamente y realizando siempre backups de los datos explotados en cada proceso. La minería de datos debería ser realizada en periodos mensual, sin embargo, esta medida podría variar en cualquier momento en función de la necesidad que esté vigente en cada momento.

Anexo 11. Solicitud de asignación de responsable para acceso a la información para desarrollo de Proyecto de Investigación



"Año de la Universalización de la Salud"

CARTA N° 50 - DIR-CAP-III-MET-RALL-ESSALUD-2020.

Trujillo, 20 de Noviembre de 2020

**Señores
LUIS FELIPE CRUZ ISLA
RONALD SEGUNDO ALVA MARIÑOS
Estudiantes de la UCV
Presente.-**

ASUNTO : Solicitud de asignación de responsable para acceso a la información para desarrollo de Proyecto de Investigación

REFER : Solicitud de fecha 20/11/2020

Sirva la presente para saludarlos cordialmente y al mismo tiempo hacer de conocimiento de acuerdo a su solicitud que el responsable con acceso a la información para desarrollo del Proyecto de Investigación descrita, correspondería al Licenciado Jaime Risco Mozo, profesional Estadístico del Centro de Atención Primaria III Metropolitano, quien proporcionaría la información disponible de los sistemas institucionales, previa autorización obtenida de la Oficina de Capacitación Docencia e Investigación de la Red Asistencial La Libertad y de acuerdo a la normatividad vigente para el desarrollo de trabajos de investigación en Essalud.

Sin otro particular.

Atentamente,



Firmado digitalmente por:
CAVIPOS REYNA Jorge Luis
FAU 20131257750 hard
Motivo: Soy el autor del documento
Fecha: 23/11/2020 09:32:05-0500

NIT: 6861-2020-563
Folios (02)

www.essalud.gob.pe

Parque Industrial Av. 5 s/n.
Distrito La Esperanza
Trujillo- Perú
T: 044-480860 Anexo 1238

Anexo 12. Autorización de aplicación del instrumento firmado por la entidad



"Año de la Universalización de la Salud"

RED ASISTENCIAL LA LIBERTAD
OFICINA DE CAPACITACIÓN DOCENCIA E INVESTIGACIÓN
COMITÉ DE INVESTIGACIÓN Y ÉTICA



PI N° 58 CIYE- O.C.I.YD-RALL-ESSALUD-2020

CONSTANCIA N.º 82

El presidente del Comité de Investigación de la Red Asistencial La Libertad – ESSALUD, ha aprobado el Proyecto de Investigación Titulado:

**“Aplicación De Minería De Datos Para Mejorar El
Diagnóstico De Un Grupo De Enfermedades Respiratorias En
Un Hospital De Trujillo”**

**ALVA MARIÑOS, RONALD SEGUNDO
CRUZ ISLA, LUIS FELIPE**

Al finalizar el desarrollo de su proyecto deberá alcanzar un ejemplar del trabajo desarrollado vía virtual al email (capacitacionrall@gmail.com), según Directiva N° 04-IETSI-ESSALUD-2016, a la Oficina de Capacitación, Investigación y Docencia - GRALL y ser remitido a la Biblioteca de la RALL, caso contrario la información del Trabajo de Investigación no será avalada por ESSALUD.

Trujillo, 22 de diciembre 2020

Andrés Sánchez Reyna
Dr. Andrés Sánchez Reyna
PRESIDENTE
Comité de Investigación
de Capacitación Docencia
e Investigación - RALL
EsSalud

NIT: 1319-2020-2817

Anexo 13. Acta de conformidad



"Año del Bicentenario del Perú: 200 años de Independencia"

RED ASISTENCIAL LA LIBERTAD
OFICINA DE CAPACITACION DOCENCIA E INVESTIGACION
COMITÉ DE INVESTIGACIÓN Y ÉTICA

TI N° 01 CIYE- O.C.I.y D.-RALL-ESSALUD-2021

CONSTANCIA N° 11

El presidente del Comité de Investigación y Ética de la Red Asistencial La Libertad – ESSALUD, ha aprobado el trabajo de Investigación ejecutado Titulado:

**" APLICACIÓN DE MINERÍA DE DATOS PARA MEJORAR EL
DIAGNÓSTICO DE UN GRUPO DE ENFERMEDADES RESPIRATORIAS EN
UN HOSPITAL DE TRUJILLO".**

ALVA MARIÑOS, RONALD SEGUNDO

CRUZ ISLA, LUIS FELIPE

Dejamos constar que la Oficina de Capacitación, Investigación y Docencia –GRALL, esta recepcionado 1 ejemplar de INFORME FINAL del Trabajo de Investigación Ejecutado y un CD del mismo, para ser remitido a la Biblioteca de la RALL, ESSALUD.

Trujillo, 08 de abril del 2021

Dr. Daniel Górriz Koonzi
Médico

Dr. Andrés Sánchez Reyna
PRESIDENTE
Comité de Investigación
y Docencia
e Investigación - R.A.L.L.
EsSalud

NIT: 1319-2020-2817

Anexo 14. Resultados antes de la implementación y después de la implementación

Ficha de registro – Nivel de morbilidad

PRETEST

Ficha de Registro			
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe	Tipo de prueba	
Empresa investigada	CAP III Metropolitano Trujillo		
Motivo de investigación	Nivel de Morbilidad		
Fecha de inicio	1/11/2020	Fecha de termino	30/11/2020

Objetivo	Indicador	Medida	Formula
Disminuir el nivel de morbilidad	Nivel de morbilidad	Porcentaje	$NM = \frac{TC \times 100}{TCE}$ <p>NM: Nivel de morbilidad. TC: Total de casos de un grupo de enfermedades respiratorias según edad. TCE: Total de casos egresados.</p>

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Fecha de nacimiento	Fecha de diagnóstico	Enfermedad diagnosticada
1	Robles Sanchez Luis	Julia Codezido 644 S To D Omin	-	2014-10-03	2019-09-26	Bronquitis aguda, no especificada
2	Gil Acosta Yanelí	Inca Paulo 292 Santa	-	2014-01-01	2019-07-26	Bronquitis aguda, no especificada
3	Asmat Leyva Nathan	Teofilo Castillo 910 S/N El Bosque	-	2014-09-02	2019-08-28	Bronquitis aguda, no especificada
4	Rafael Calderon Braulio	Carlos Baca Flor 331 San To D	-	2018-03-03	2019-10-14	Bronquitis aguda, no especificada
5	Vega Graus Erick	Gonzales Prada F-16 Urb Santa Maria	-	2017-11-01	2019-10-14	Bronquitis aguda, no especificada
6	Abarca Chilcon Bruno	Victor Fajardo 381 S S/N Anta Maria	-	2015-02-01	2019-12-13	Faringitis aguda, no especificada
7	Aguirre Castillo Arturo	Julia Codesido 590 S S/N	-	2014-10-03	2019-12-06	Faringitis aguda, no especificada
8	Amaral Bernaque Jamín	Av. Los Colibries Dp To. 300b	-	2014-07-01	2019-08-26	Faringitis aguda, no especificada
9	Arando Jibaja John	Las Casuarinas G5	-	2014-10-02	2019-08-13	Faringitis aguda, no especificada
10	Argomedo Vargas Ariana	David Lozano El Bosq	-	2014-09-03	2019-11-02	Faringitis aguda, no especificada
11	Armas Valverde Yoshiro	Arequipa 370 S/N	-	2013-10-02	2019-07-15	Faringitis aguda, no especificada
12	Avila Plasencia Maria	Santa Cruz 362	-	2016-03-03	2019-10-04	Faringitis aguda, no especificada

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Fecha de nacimiento	Fecha de diagnóstico	Enfermedad diagnosticada
13	Pajares Cisneros Nelson	Cv. Covidunt Mz. C L T. 1 3	-	2015-06-03	2019-12-16	Faringitis aguda, no especificada
14	Barrios Gamboa Fabio	Blas Pascal 167 Urb La Noria	-	2019-08-30	2019-10-03	Faringitis aguda, no especificada
15	Benites Llaros Rosmery	Gonzales Prada 1090	-	2017-10-01	2019-10-05	Faringitis aguda, no especificada
16	Burgos Camasca Nikólas	Calle Renato Descart Es 1 71 U	-	2014-01-02	2019-12-09	Faringitis aguda, no especificada
17	Cabrera Mestanza Leonel	Luis Valle Goicochea 647 Urb	-	2014-05-31	2019-07-18	Faringitis aguda, no especificada
18	Cañote Morales Bruno	Luis Valle Goicochea 316 Pal	-	2015-05-04	2019-10-03	Faringitis aguda, no especificada
19	Namay Caruapoma Alexis	Urb Los Laureles Mz . F Lt.	-	2014-04-03	2019-07-12	Rinofaringitis aguda [resfriado común]

POSTEST

Ficha de Registro			
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe	Tipo de prueba	
Empresa investigada	CAP III Metropolitano Trujillo		
Motivo de investigación	Nivel de Morbilidad		
Fecha de inicio	1/12/2020	Fecha de termino	31/12/2020

Objetivo	Indicador	Medida	Formula
Disminuir el nivel de morbilidad	Nivel de morbilidad	Porcentaje	$NM = \frac{TC \times 100}{TCE}$ <p>NM: Nivel de morbilidad. TC: Total de casos de un grupo de enfermedades respiratorias según edad. TCE: Total de casos egresados.</p>

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Fecha de nacimiento	Fecha de diagnóstico	Enfermedad diagnosticada
1	Sipiran Saavedra Angela	Jr. Miguel Iglesias 188 Dpto	-	2015-04-04	2020-10-02	Bronquitis aguda, no especificada
2	Sipiran Saavedra Antuanett	Jr. Miguel Iglesias 188 Dpto	-	2017-01-01	2020-07-16	Bronquitis aguda, no especificada
3	Zavaleta Rodriguez Fabiana	Calle Descartes 0192 Urb . La	-	2019-07-01	2020-08-18	Rinofaringitis aguda [resfriado común]
4	Quezada Cancino Yamila	Calle Las Fabricas U Rb. Semi	-	2016-08-04	2020-11-24	Rinofaringitis aguda [resfriado común]
5	Cubas Ruiz Rihana	104 Santa Maria Etap A V Mz.	-	2015-04-04	2020-10-16	Rinofaringitis aguda [resfriado común]
6	Zavaleta Blas Dhalia	Av. Jesus De Nazaret S/N H 108 Asent.H.	-	2016-04-03	2020-10-29	Rinofaringitis aguda [resfriado común]
7	Bendezu Mendoza Leonardo	Calcuchimac 185 Urb San Ta M	-	2017-03-04	2020-10-23	Rinofaringitis aguda [resfriado común]
8	Cubas Ruiz Astrid	104 Santa Maria Etap A V Mz.	-	2017-05-04	2020-10-16	Rinofaringitis aguda [resfriado común]
9	Navarro Diaz Viveka	Jr. Victor Fajardo 4 S/N 55 Urb. Santa M	-	2018-02-02	2020-11-03	Faringitis aguda, no especificada
10	Daza Amaya Adriano	Mz. P' Lt. 0 1 Int. 0 3 Ur B. S	-	2018-03-04	2020-11-24	Faringitis aguda, no especificada
11	Rabelo Lecca Nicóle	Servulo Gutierrez 60 9 St O Do	-	2015-06-05	2020-11-24	Rinofaringitis aguda [resfriado común]

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Fecha de nacimiento	Fecha de diagnóstico	Enfermedad diagnosticada
12	Rodriguez Quiroz Luana	Lora Y Lora 446	-	2015-07-05	2020-11-06	Rinofaringitis aguda [resfriado común]
13	Sipiran Saavedra Angela	Jr. Miguel Iglesias 188 Dpto	-	2015-04-04	2020-07-16	Bronquitis aguda, no especificada
14	Sipiran Saavedra Antuanett	Jr. Miguel Iglesias 188 Dpto	-	2017-01-01	2020-07-21	Bronquitis aguda, no especificada
15	Sipiran Saavedra Angela	Jr. Miguel Iglesias 188 Dpto	-	2015-04-04	2020-08-21	Bronquitis aguda, no especificada
16	Sipiran Saavedra Angela	Jr. Miguel Iglesias 188 Dpto	-	2015-04-04	2020-07-02	Rinofaringitis aguda [resfriado común]
17	Cubas Ruiz Astrid	104 Santa Maria Etap A V Mz.	-	2017-05-04	2020-10-22	Rinofaringitis aguda [resfriado común]
18	Cubas Ruiz Rihana	104 Santa Maria Etap A V Mz.	-	2015-04-04	2020-10-22	Rinofaringitis aguda [resfriado común]
19	Aponte Acapana Ruth	Calle. Francisco Adr lanz En N	-	2016-09-03	2020-11-02	Faringitis aguda, no especificada

Ficha de observación – Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias

PRETEST

Ficha de Observación			
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe	Tipo de prueba	
Empresa investigada	CAP III Metropolitano Trujillo		
Motivo de investigación	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias		
Fecha de inicio	1/11/2020	Fecha de termino	30/11/2020

Objetivo	Indicador	Medida	Formula
Disminuir el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias	Horas	$TPIEGER = \frac{\sum_{i=1}^n (TIEER)_i}{n}$ <p>TPIEGER: Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias TIEER: Tiempo para identificar la existencia de un grupo de enfermedades respiratorias n: número de casos egresados.</p>

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Hora Ingreso (hh:mm)	Hora Salida (hh:mm)	Fecha de diagnostico	Enfermedad diagnosticada	Numero de análisis adicionales	Días de tratamiento farmacológico
1	Robles Sanchez Luis	Julia Codezido 644 S To D Omin	-	16:12	17:00	2019-09-26	Bronquitis aguda, no especificada	3	0
2	Gil Acosta Yanelí	Inca Paulo 292 Santa	-	16:48	17:24	2019-07-26	Bronquitis aguda, no especificada	4	0
3	Asmat Leyva Nathan	Teofilo Castillo 910 S/N El Bosque	-	17:36	18:36	2019-08-28	Bronquitis aguda, no especificada	-	0
4	Rafael Calderon Braulio	Carlos Baca Flor 331 San To D	-	08:00	08:24	2019-10-14	Bronquitis aguda, no especificada	-	0
5	Vega Graus Erick	Gonzales Prada F-16 Urb Santa Maria	-	09:00	09:36	2019-10-14	Bronquitis aguda, no especificada	-	0
6	Abarca Chilcon Bruno	Victor Fajardo 381 S S/N Anta Maria	-	15:12	15:24	2019-12-13	Faringitis aguda, no especificada	-	0
7	Aguirre Castillo Arturo	Julia Codesido 590 S S/N	-	16:12	16:24	2019-12-06	Faringitis aguda, no especificada	-	0
8	Amaral Bernaque Jamín	Av. Los Colibries Dp To. 300b	-	08:00	08:12	2019-08-26	Faringitis aguda, no especificada	-	0
9	Arando Jibaja John	Las Casuarinas G5	-	16:00	16:12	2019-08-13	Faringitis aguda, no especificada	-	0
10	Argomedo Vargas Ariana	David Lozano El Bosq	-	08:00	09:00	2019-11-02	Faringitis aguda, no especificada	-	0
11	Armas Valverde Yoshiro	Arequipa 370 S/N	-	08:00	08:24	2019-07-15	Faringitis aguda, no especificada	3	0
12	Avila Plasencia Maria	Santa Cruz 362	-	15:00	15:12	2019-10-04	Faringitis aguda, no especificada	-	0
13	Pajares Cisneros Nelson	Cv. Covidunt Mz. C L T. 1 3	-	11:24	11:48	2019-12-16	Faringitis aguda, no especificada	5	0
14	Barrios Gamboa Fabio	Blas Pascal 167 Urb La Noria	-	11:12	11:24	2019-10-03	Faringitis aguda, no especificada	-	0

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Hora Ingreso (hh:mm)	Hora Salida (hh:mm)	Fecha de diagnóstico	Enfermedad diagnosticada	Numero de análisis adicionales	Días de tratamiento farmacológico
15	Benites Llaros Rosmery	Gonzales Prada 1090	-	11:48	12:00	2019-10-05	Faringitis aguda, no especificada	-	0
16	Burgos Camasca Nikólas	Calle Renato Descart Es 1 71 U	-	08:24	08:36	2019-12-09	Faringitis aguda, no especificada	11	0
17	Cabrera Mestanza Leonel	Luis Valle Goicochea 647 Urb	-	15:12	16:12	2019-07-18	Faringitis aguda, no especificada	-	0
18	Cañote Morales Bruno	Luis Valle Goicochea 316 Pal	-	11:00	11:12	2019-10-03	Faringitis aguda, no especificada	-	0
19	Namay Caruapoma Alexis	Urb Los Laureles Mz . F Lt.	-	09:36	09:48	2019-07-12	Rinofaringitis aguda [resfriado común]	-	-

POSTEST

Ficha de Observación			
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe	Tipo de prueba	
Empresa investigada	CAP III Metropolitano Trujillo		
Motivo de investigación	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias		
Fecha de inicio	1/12/2020	Fecha de termino	31/12/2020

Objetivo	Indicador	Medida	Formula
Disminuir el tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias	Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias	Horas	$TPIEGER = \frac{\sum_{i=1}^n (TIEER)_i}{n}$ <p>TPIEGER: Tiempo promedio para identificar la existencia de un grupo de enfermedades respiratorias</p> <p>TIEER: Tiempo para identificar la existencia de un grupo de enfermedades respiratorias</p> <p>n: número de casos egresados.</p>

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Hora Ingreso (hh:mm)	Hora Salida (hh:mm)	Fecha de diagnostico	Enfermedad diagnosticada	Numero de análisis adicionales	Días de tratamiento farmacológico
1	Sipiran Saavedra Angela	Jr. Miguel Iglesias 188 Dpto	-	10:57	11:00	2020-10-02	Bronquitis aguda, no especificada	0	0
2	Sipiran Saavedra Antuanett	Jr. Miguel Iglesias 188 Dpto	-	09:44	09:46	2020-07-16	Bronquitis aguda, no especificada	0	0
3	Zavaleta Rodriguez Fabiana	Calle Descartes 0192 Urb . La	-	12:55	13:01	2020-08-18	Rinofaringitis aguda [resfriado común]	0	0
4	Quezada Cancino Yamila	Calle Las Fabricas U Rb. Semi	-	11:44	11:52	2020-11-24	Rinofaringitis aguda [resfriado común]	0	0
5	Cubas Ruiz Rihana	104 Santa Maria Etap A V Mz.	-	13:05	13:08	2020-10-16	Rinofaringitis aguda [resfriado común]	0	0
6	Zavaleta Blas Dhalia	Av. Jesus De Nazaret S/N H 108 Asent.H.	-	09:25	09:33	2020-10-29	Rinofaringitis aguda [resfriado común]	0	0
7	Bendezu Mendoza Leonardo	Calcuchimac 185 Urb San Ta M	-	09:06	09:12	2020-10-23	Rinofaringitis aguda [resfriado común]	0	0
8	Cubas Ruiz Astrid	104 Santa Maria Etap A V Mz.	-	13:00	13:13	2020-10-16	Rinofaringitis aguda [resfriado común]	0	0
9	Navarro Diaz Viveka	Jr. Victor Fajardo 4 S/N 55	-	16:34	16:38	2020-11-03	Faringitis aguda, no especificada	0	0

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Hora Ingreso (hh:mm)	Hora Salida (hh:mm)	Fecha de diagnóstico	Enfermedad diagnosticada	Numero de análisis adicionales	Días de tratamiento farmacológico
		Urb. Santa M							
10	Daza Amaya Adriano	Mz. P' Lt. 01 Int. 0 3 Ur B. S		10:01	10:11	2020-11-24	Faringitis aguda, no especificada	0	0
11	Rabelo Lecca Nicóle	Servulo Gutierrez 60 9 St O Do	-	12:00	12:15	2020-11-24	Rinofaringitis aguda [resfriado común]	0	0
12	Rodriguez Quiroz Luana	Lora Y Lora 446	-	12:45	12:48	2020-11-06	Rinofaringitis aguda [resfriado común]	0	0
13	Sipiran Saavedra Angela	Jr. Miguel Iglesias 188 Dpto	-	09:36	09:43	2020-07-16	Bronquitis aguda, no especificada	0	0
14	Sipiran Saavedra Antuanett	Jr. Miguel Iglesias 188 Dpto	-	13:46	13:55	2020-07-21	Bronquitis aguda, no especificada	0	0
15	Sipiran Saavedra Angela	Jr. Miguel Iglesias 188 Dpto	-	09:16	09:20	2020-08-21	Bronquitis aguda, no especificada	0	0
16	Sipiran Saavedra Angela	Jr. Miguel Iglesias 188 Dpto	-	13:09	13:14	2020-07-02	Rinofaringitis aguda [resfriado común]	0	0
17	Cubas Ruiz Astrid	104 Santa Maria Etap A V Mz.	-	09:01	09:07	2020-10-22	Rinofaringitis aguda [resfriado común]	0	0
18	Cubas Ruiz Rihana	104 Santa Maria Etap A V Mz.	-	09:08	09:16	2020-10-22	Rinofaringitis aguda [resfriado común]	0	0
19	Aponte Acapana Ruth	Calle. Francisco Adr lanz En N	-	15:53	16:01	2020-11-02	Faringitis aguda, no especificada	0	0

Ficha de registro – Costo promedio de diagnóstico de un grupo de enfermedades respiratorias

PRETEST

Ficha de Registro			
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe	Tipo de prueba	
Empresa investigada	CAP III Metropolitano Trujillo		
Motivo de investigación	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias		
Fecha de inicio	1/11/2020	Fecha de termino	30/11/2020

Objetivo	Indicador	Medida	Formula
Disminuir el costo promedio de diagnóstico de un grupo de enfermedades respiratorias	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias	Unidad	$CPDGER = \frac{\sum_{i=1}^{n_{ce}} ((NE \times CE) + (NT \times CT))_i}{NCE}$ <p>CDGER: Costo promedio de un grupo de enfermedades respiratorias. NE: Numero de exámenes CE: Costo de exámenes NT: Numero de tratamientos CT: Costo de tratamientos NCE: Número de casos de una enfermedad específica</p>

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Numero de exámenes	Costo de exámenes	Numero de tratamiento	Costo de tratamientos	Fecha de diagnostico	Enfermedad diagnosticada	Costo total de exámenes	Costo total de tratamientos
1	Robles Sanchez Luis	Julia Codezido 644 S To D Omin	-	3	20.33	11	1.17	2019-09-26	Bronquitis aguda, no especificada	61	12.87
2	Gil Acosta Yanelí	Inca Paulo 292 Santa	-	5	22.20	4	4.10	2019-07-26	Bronquitis aguda, no especificada	111	16.40
3	Asmat Leyva Nathan	Teofilo Castillo 910 S/N El Bosque	-	0	0	8	1.11	2019-08-28	Bronquitis aguda, no especificada	0	8.88
4	Rafael Calderon Braulio	Carlos Baca Flor 331 San To D	-	0	0	4	2.72	2019-10-14	Bronquitis aguda, no especificada	0	10.88
5	Vega Graus Erick	Gonzales Prada F- 16 Urb Santa Maria	-	0	0	8	1.85	2019-10-14	Bronquitis aguda, no especificada	0	14.80
6	Abarca Chilcon Bruno	Victor Fajardo 381 S S/N Anta Maria	-	0	0	12	1.31	2019-12-13	Faringitis aguda, no especificada	0	15.72
7	Aguirre Castillo Arturo	Julia Codesido 590 S S/N	-	0	0	14	0.92	2019-12-06	Faringitis aguda, no especificada	0	12.88
8	Amaral Bernaque Jamín	Av. Los Colibries Dp To. 300b	-	0	0	13	1.66	2019-08-26	Faringitis aguda, no especificada	0	21.58

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Numero de exámenes	Costo de exámenes	Numero de tratamiento	Costo de tratamientos	Fecha de diagnostico	Enfermedad diagnosticada	Costo total de exámenes	Costo total de tratamientos
9	Arando Jibaja John	Las Casuarinas G5	-	0	0	6	2.93	2019-08-13	Faringitis aguda, no especificada	0	17.58
10	Argomedo Vargas Ariana	David Lozano El Bosq	-	0	0	5	2.62	2019-11-02	Faringitis aguda, no especificada	0	13.10
11	Armas Valverde Yoshiro	Arequipa 370 S/N	-	3	21.20	12	1.58	2019-07-15	Faringitis aguda, no especificada	63.60	18.96
12	Avila Plasencia Maria	Santa Cruz 362	-	0	0	6	3.37	2019-10-04	Faringitis aguda, no especificada	0	20.22
13	Pajares Cisneros Nelson	Cv. Covidunt Mz. C L T. 1 3	-	5	25.25	1	4.88	2019-12-16	Faringitis aguda, no especificada	126.25	4.88
14	Barrios Gamboa Fabio	Blas Pascal 167 Urb La Noria	-	0	0	3	1.94	2019-10-03	Faringitis aguda, no especificada	0	5.82
15	Benites Llaros Rosmery	Gonzales Prada 1090	-	0	0	15	2.44	2019-10-05	Faringitis aguda, no especificada	0	36.60
16	Burgos Camasca Nikólas	Calle Renato Descart Es 1 71 U	-	11	19.13	9	2.86	2019-12-09	Faringitis aguda, no especificada	210..43	13.92

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Numero de exámenes	Costo de exámenes	Numero de tratamiento	Costo de tratamientos	Fecha de diagnostico	Enfermedad diagnosticada	Costo total de exámenes	Costo total de tratamientos
17	Cabrera Mestanza Leonel	Luis Valle Goicochea 647 Urb	-	0	0	13	1.58	2019-07-18	Faringitis aguda, no especificada	0	20.54
18	Cañote Morales Bruno	Luis Valle Goicochea 316 Pal	-	0	0	14	0.41	2019-10-03	Faringitis aguda, no especificada	0	5.74
19	Namay Caruapoma Alexis	Urb Los Laureles Mz . F Lt.	-	0	0	14	0.41	2019-07-12	Rinofaringitis aguda [resfriado común]	0	5.74

POSTEST

Ficha de Registro			
Investigadores	Alva Mariños, Ronald Segundo Cruz Isla, Luis Felipe	Tipo de prueba	
Empresa investigada	CAP III Metropolitano Trujillo		
Motivo de investigación	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias		
Fecha de inicio	1/12/2020	Fecha de termino	31/12/2020

Objetivo	Indicador	Medida	Formula
Disminuir el costo promedio de diagnóstico de un grupo de enfermedades respiratorias	Costo promedio de diagnóstico de un grupo de enfermedades respiratorias	Unidad	$CPDGER = \frac{\sum_{i=1}^{nce} ((NE \times CE) + (NT \times CT))i}{NCE}$ <p> CDGER: Costo promedio de un grupo de enfermedades respiratorias. NE: Numero de exámenes CE: Costo de exámenes NT: Numero de tratamientos CT: Costo de tratamientos NCE: Número de casos de una enfermedad específica </p>

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Numero de exámenes	Costo de exámenes	Numero de tratamiento	Costo de tratamientos	Fecha de diagnostico	Enfermedad diagnosticada	Costo total de exámenes	Costo total de tratamientos
1	Sipiran Saavedra Angela	Jr. Miguel Iglesias 188 Dpto	-	0	0	6	1.38	2020-10-02	Bronquitis aguda, no especificada	0	8.28
2	Sipiran Saavedra Antuanett	Jr. Miguel Iglesias 188 Dpto	-	0	0	11	0.72	2020-07-16	Bronquitis aguda, no especificada	0	7.92
3	Zavaleta Rodriguez Fabiana	Calle Descartes 0192 Urb . La	-	0	0	4	1.38	2020-08-18	Rinofaringitis aguda [resfriado común]	0	5.52
4	Quezada Cancino Yamila	Calle Las Fabricas U Rb. Semi	-	0	0	2	1.55	2020-11-24	Rinofaringitis aguda [resfriado común]	0	3.10
5	Cubas Ruiz Rihana	104 Santa Maria Etap A V Mz.	-	0	0	5	1.38	2020-10-16	Rinofaringitis aguda [resfriado común]	0	6.09
6	Zavaleta Blas Dhalia	Av. Jesus De Nazaret S/N H 108 Asent.H.	-	0	0	1	0.24	2020-10-29	Rinofaringitis aguda [resfriado común]	0	0.24

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Numero de exámenes	Costo de exámenes	Numero de tratamiento	Costo de tratamientos	Fecha de diagnostico	Enfermedad diagnosticada	Costo total de exámenes	Costo total de tratamientos
7	Bendezu Mendoza Leonardo	Calcuchimac 185 Urb San Ta M	-	0	0	4	1.65	2020-10-23	Rinofaringitis aguda [resfriado común]	0	6.60
8	Cubas Ruiz Astrid	104 Santa Maria Etap A V Mz.	-	0	0	3	1.38	2020-10-16	Rinofaringitis aguda [resfriado común]	0	4.14
9	Navarro Diaz Viveka	Jr. Victor Fajardo 4 S/N 55 Urb. Santa M	-	0	0	3	0.96	2020-11-03	Faringitis aguda, no especificada	0	2.88
10	Daza Amaya Adriano	Mz. P' Lt. 01 Int. 0 3 Ur B. S	-	0	0	2	1.44	2020-11-24	Faringitis aguda, no especificada	0	2.88
11	Rabelo Lecca Nicóle	Servulo Gutierrez 60 9 St O Do	-	0	0	7	1.12	2020-11-24	Rinofaringitis aguda [resfriado común]	0	7.84
12	Rodriguez Quiroz Luana	Lora Y Lora 446	-	0	0	5	1.38	2020-11-06	Rinofaringitis aguda [resfriado común]	0	6.90

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Numero de exámenes	Costo de exámenes	Numero de tratamiento	Costo de tratamientos	Fecha de diagnostico	Enfermedad diagnosticada	Costo total de exámenes	Costo total de tratamientos
13	Sipiran Saavedra Angela	Jr. Miguel Iglesias 188 Dpto	-	0	0	11	1.85	2020-07-16	Bronquitis aguda, no especificada	0	20.35
14	Sipiran Saavedra Antuanett	Jr. Miguel Iglesias 188 Dpto	-	0	0	1	3:30	2020-07-21	Bronquitis aguda, no especificada	0	3:30
15	Sipiran Saavedra Angela	Jr. Miguel Iglesias 188 Dpto	-	0	0	9	1.69	2020-08-21	Bronquitis aguda, no especificada	0	15.21
16	Sipiran Saavedra Angela	Jr. Miguel Iglesias 188 Dpto	-	0	0	6	2.25	2020-07-02	Rinofaringitis aguda [resfriado común]	0	13.50
17	Cubas Ruiz Astrid	104 Santa Maria Etap A V Mz.	-	0	0	1	0.24	2020-10-22	Rinofaringitis aguda [resfriado común]	0	0.24
18	Cubas Ruiz Rihana	104 Santa Maria Etap A V Mz.	-	0	0	12	0.59	2020-10-22	Rinofaringitis aguda [resfriado común]	0	7.08

N°	Paciente	Procedencia del paciente	RAM (Reacciones alérgicas al medicamento)	Numero de exámenes	Costo de exámenes	Numero de tratamiento	Costo de tratamientos	Fecha de diagnostico	Enfermedad diagnosticada	Costo total de exámenes	Costo total de tratamientos
19	Aponte Acapana Ruth	Calle. Francisco Adr Ianz En N	-	0	0	0	0	2020-11-02	Faringitis aguda, no especificada	0	0



UNIVERSIDAD CÉSAR VALLEJO

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

Declaratoria de Autenticidad del Asesor

Yo, CIEZA MOSTACERO SEGUNDO EDWIN, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - TRUJILLO, asesor de Desarrollo de Proyecto de Tesis titulada: "APLICACIÓN DE MINERÍA DE DATOS PARA MEJORAR EL DIAGNÓSTICO DE UN GRUPO DE ENFERMEDADES RESPIRATORIAS EN UN HOSPITAL DE TRUJILLO", cuyos autores son CRUZ ISLA LUIS FELIPE, ALVA MARIÑOS RONALD SEGUNDO, constato que la investigación cumple con el índice de similitud establecido, y verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Desarrollo de Proyecto de Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

TRUJILLO, 11 de Noviembre del 2021

Apellidos y Nombres del Asesor:	Firma
CIEZA MOSTACERO SEGUNDO EDWIN DNI: 45434553 ORCID 0000-0002-3520-4383	Firmado digitalmente por: OALCANTARA el 11-11- 2021 09:22:08

Código documento Trilce: TRI - 0194594