

DOI: 10.26820/reciamuc/7.(2).abril.2023.252-265

URL: <https://reciamuc.com/index.php/RECIAMUC/article/view/1111>

EDITORIAL: Saberes del Conocimiento

REVISTA: RECIAMUC

ISSN: 2588-0748

TIPO DE INVESTIGACIÓN: Artículo de revisión

CÓDIGO UNESCO: 1203 Ciencia de Los Ordenadores

PAGINAS: 252-265



Generación del conocimiento en Instituciones de Salud mediante la aplicación de metadatos

Generation of knowledge in Health Institutions through the application of metadata

Geração de conhecimento em Instituições de Saúde através da aplicação de metadados

**César Antonio Bustamante Chong¹; Franklin Augusto Cabezas Galarza²;
Mariana Elizabeth Bustamante Chong³**

RECIBIDO: 23/02/2023 **ACEPTADO:** 12/03/2023 **PUBLICADO:** 15/05/2023

1. Magister en Administración de Empresas con Mención en Gestión En Mercadotecnia; Ingeniero en Sistemas Computacionales; Universidad de Guayaquil; Guayaquil, Ecuador; cesar.bustamantech@ug.edu.ec;  <https://orcid.org/0000-0003-0150-1088>
2. Magister en Administración de Empresas con Mención en Gestión en Mercadotecnia; Ingeniero en Sistemas Computacionales; Universidad de Guayaquil; Guayaquil, Ecuador; franklin.cabezasg@ug.edu.ec;  <https://orcid.org/0000-0001-5504-472X>
3. Magister en Administración Pública Mención Desarrollo Institucional; Magister en Docencia Universitaria e Investigación Educativa; Diplomado en Autoevaluación y Acreditación Universitaria; Economista; Universidad Estatal del Sur de Manabí; Jipijapa, Ecuador; mariana.bustamante@unesum.edu.ec;  <https://orcid.org/0000-0002-5815-3267>

CORRESPONDENCIA

César Antonio Bustamante Chong

cesar.bustamantech@ug.edu.ec

Guayaquil, Ecuador

RESUMEN

Las unidades hospitalarias del país carecen de médicos especialistas para atender el volumen intenso de pacientes que existen, muchas veces este número reducido de expertos se debe, entre otras cosas: ofertas de posgrado inexistentes, no existencia de políticas claras en la oferta y demanda. Por lo expuesto, se tiene como información, la demanda de las atenciones médicas a los pacientes, su posterior sintomatología y medicación; el número de profesionales y sus diferentes especialidades, etc.; lo cual nos conlleva a determinar cómo gestionar y aprovechar dicha información para impulsar mediante un enfoque predictivo, una herramienta de inteligencia de negocios Knime y ayudada por la metodología CRISP-DM propia de los proyectos que emplean minería de datos; el análisis y la toma de decisiones en seleccionar las mejores especialidades.

El objetivo es analizar la crisis de especialidades médicas mediante un enfoque predictivo de la información a través de herramientas y metodologías de minado de datos; los principales objetivos específicos son: identificar la información pertinente a partir de la información relevada; diseñar el enfoque predictivo a través de modelos y metodología de minería de datos y evaluar el modelo predictivo mediante variables en la toma de decisiones.

Finalmente, con el análisis predictivo de la información de datos hospitalarios, se pudo identificar, seleccionar y mejorar las variables que permitieron la toma de decisiones oportuna para la apertura de especialidades médicas necesarias y acertadas; identificar fuentes de trabajo; detectar las enfermedades catastróficas con un alto índice de mortalidad y poder mitigar su incidencia en la sociedad.

Palabras clave: Análisis Predictivo, Predicción, Demanda Especialidades.

ABSTRACT

The country's hospital units lack medical specialists to care for the intense volume of patients that exist, many times this reduced number of experts is due, among other things: non-existent postgraduate offers, non-existence of clear supply and demand policies. For the above, the information is the demand for medical attention to patients, their subsequent symptoms and medication; the number of professionals and their different specialties, etc.; which leads us to determine how to manage and take advantage of said information to promote, through a predictive approach, a Knime business intelligence tool and aided by the CRISP-DM methodology typical of projects that use data mining; analysis and decision-making in selecting the best specialties.

The objective is to analyze the crisis of medical specialties through a predictive approach to information through data mining tools and methodologies; The main specific objectives are: to identify the pertinent information from the collected information; design the predictive approach through models and data mining methodology and evaluate the predictive model through variables in decision making.

Finally, with the predictive analysis of hospital data information, it was possible to identify, select and improve the variables that allowed timely decision-making for the opening of necessary and correct medical specialties; identify job sources; detect catastrophic diseases with a high mortality rate and be able to mitigate their incidence in society.

Keywords: Predictive Analysis, Prediction, Specialties Demand.

RESUMO

As unidades hospitalares do país carecem de médicos especialistas para atender o intenso volume de pacientes que existem, muitas vezes este reduzido número de especialistas deve-se, entre outros factores: inexistência de ofertas de pós-graduação, inexistência de políticas claras de oferta e procura. Pelo exposto, temos como informação, a procura de assistência médica aos pacientes, a sua consequente sintomatologia e medicação; o número de profissionais e as suas diferentes especialidades, etc.; o que nos leva a determinar como gerir e explorar esta informação para promover através de uma abordagem preditiva, uma ferramenta de business intelligence Knime e auxiliada pela metodologia CRISP-DM dos projectos que utilizam data mining; análise e tomada de decisão na selecção das melhores especialidades.

O objectivo é analisar a crise das especialidades médicas através de uma abordagem preditiva da informação através de ferramentas e metodologias de data mining; os principais objectivos específicos são: identificar a informação relevante a partir da informação recolhida; desenhar a abordagem preditiva através de modelos e metodologia de data mining e avaliar o modelo preditivo através de variáveis na tomada de decisão.

Por fim, com a análise preditiva da informação dos dados hospitalares, foi possível identificar, seleccionar e melhorar as variáveis que permitiram a tomada de decisão atempada para a abertura de especialidades médicas necessárias e precisas; identificar fontes de trabalho; detectar doenças catastróficas com uma elevada taxa de mortalidade e mitigar a sua incidência na sociedade.

Palavras-chave: Análise Preditiva, Previsão, Procura de Especialidades.

Introducción

La salud en nuestro país está muy claramente diferenciada en cuanto se trata del sector público y del sector privado. En el sector público existen ciertas falencias, entre las principales:

- La falta de profesionales de la salud por paciente atendido y a su vez la cantidad de pacientes que demandan atención,
- En algunos casos la falta de una aplicación informática de salud que brinde la información oportuna y veraz del paciente.

Las unidades hospitalarias del país –en general, tanto públicas como privadas- carecen de médicos especialistas para atender el volumen intenso de pacientes que existen, muchas veces este número reducido de expertos se debe, entre otras cosas a: ofertas de posgrado inexistentes, la no existencia de políticas claras en la oferta y demanda, cupos para los programas de posgrados, pero sobre todo no existe un estudio real entre la necesidad y la oferta de las especialidades médicas.

Según CENEC (2019) nos informa que la tasa de personal médico en el país en el año 2010; es de 17,6 médicos por cada 10.000 habitantes; donde en las principales ciudades alcanzan los siguientes ratios: Guayaquil 33,7; Machala 33,3; Quito 28,9; Ambato 25,8. Ahora bien existen 24.960 médicos en general; de los cuales el porcentaje por las principales especialidades son: medicina general 24,0%; cirujanos generales 7,9%; ginecólogos/obstetras 7,8%; anestesiólogos 7,8%; pediatras 7,6%; traumatólogos 4,9%; medicina interna 3,9%; cardiólogos 3,3%; urólogos 2,5%; gastroenterólogos 2,1%; cirujanos plásticos 2,1%; otorrinos 2,1%; neurólogos 2,1%; entre los más importantes. Situándose en el extremo inferior los alergólogos, infectólogos y epidemiólogos con el 0,3%. Por otro lado, la Organización Mundial de la Salud (OMS) nos dice que deben existir 26 médicos por 10.000 habitantes.

Recientemente, según cifras emitidas por la Asociación de Facultades Ecuatorianas de Ciencias Médicas y Salud (AFECMS) a diciembre 2016 existen 29 mil médicos generales de los cuales 16 mil tienen especialidad. Mientras que a junio 2015 se acreditaron 10 instituciones de educación superior para seguir ofertando la carrera de medicina - 6 en la ciudad de Quito, 2 en Guayaquil y 2 en la ciudad de Cuenca – y existen 12 instituciones adicionales que no acreditaron, sin embargo, también la ofertan. Donde cada año se gradúan aproximadamente 3.200 estudiantes como médicos generales; mientras que el cupo para posgrados en las diferentes universidades llega a 500; esta cifra unida a la media de graduación de un especialista que es de 2 años; nos deja una clara inferioridad en cuanto a médicos especialistas se refiere.

Por otro lado, existe una red de instituciones públicas de salud; en las principales ciudades del país – las más importantes- que suman aproximadamente 10, entre estas tenemos: hospitales generales y los hospitales de especialidades médicas de adultos y de niños (pediátricos); estos generan información importante y valiosa; tales como: la demanda de las atenciones médicas a los pacientes, su posterior sintomatología y medicación; el número de profesionales y sus diferentes especialidades, etc.

Cada una de estas instituciones públicas de salud en general, tienen diferentes sistemas de información provistos en la mayor parte de los casos por el gobierno central o por el ministerio de salud pública. Como nos describe el informe de Corral Moscoso (2015) existen: sistema de administración financiera, integrado de talento humano, presupuestario de remuneraciones y nomina, contrataciones del estado, gestión de inventarios, gobierno por resultados en la parte administrativa, y en desde el punto de vista asistencial hospitalario: sistema de camas, registro diario automatizado de consultas y atenciones ambulatorias, vigilancia epidemiológica, evaluación y monitoreo VIH, infecciones asociadas a la atención de salud.

Sin embargo, en el momento de tener un sistema de información hospitalario (SIH) cada una de las casas asistenciales maneja un propio sistema.

Por lo expuesto, no se tiene una información claramente definida y unificada, lo cual nos conlleva a determinar cómo gestionar y generar dicha información y transformarla en un conocimiento que permita impulsar la toma de decisiones mediante el uso de un metadato, para posteriormente generar un modelo del conocimiento basado en una herramienta de inteligencia de negocios llamada K-nime y ayudada por la metodología CRISP-DM propia de los proyectos que emplean minería de datos.

Materiales y métodos

Metodología

Metadatos

Una definición que tiene su origen en las ciencias informáticas, proviene de la unión de la palabra griega *μετα*, meta, 'después de, más allá de' y el latín *datum*, 'lo que se da', dato; que literalmente se puede describir como –sobre datos-, es decir son datos que describen otros datos como nos señalaba Jack Myers en los años 60, estos metadatos muchas veces proporcionaban la información mínima para identificar un recurso o sobre los datos suministrados y a su vez suelen incluir información sobre su contexto, contenido y control así como todo lo que tenga que ver con el –dato- Pasquinelli (1997).

Según (NISO, 2018) (National Information Standards Organization, por sus siglas al inglés), define al metadato como la información estructurada que describe, explica, localiza o de otra manera hace de una forma fácil la recuperación, uso o manejo de los recursos de la información.

Este concepto de metadatos existía mucho antes del auge del internet y las páginas web, sin embargo, el creciente y exponencial aumento de la información y sobre todo

la búsqueda de la misma ha propiciado un interés por las políticas y el uso de los metadatos. Según Ercegovac (1999) nos afirma que, un metadato describe los atributos de un recurso, teniendo en cuenta que dicho recurso puede ser: objetos bibliográficos, objetos geoespaciales, recursos visuales, registros e inventarios de archivo, implementaciones de software, entre otros.

Existen muchas funciones que realiza un metadato, pero la mayoría las podemos encontrar en el trabajo realizado por Iannella y Waugh (1997), que nos manifiesta:

- Resumir el significado de los datos
- Permitir la búsqueda
- Determinar si el dato es el que se necesita
- Mostrar instrucciones de cómo interpretar el dato
- Aportar información acerca de la vida del dato
- Indicar relaciones con otros recursos, entre otros

Muchos autores nos presentan una clasificación o tipos de metadatos diferentes, según Gilliland-Swetland nos expresa que los tipos de metadatos se dividen en:

- administrativos (utilizados para la gestión y administración de los recursos de información),
- descriptivos (utilizados para describir o identificar recursos de información),
- de preservación (relacionados con la gestión de preservación de los recursos de información),
- técnicos (relacionados con el funcionamiento de los sistemas) y,
- de uso (relacionados con el nivel y el tipo de uso de los recursos de información).

Mientras que Tim Berners-Lee desde el punto de la web, los metadatos se dividen en:

1. Aquellos que se encuentran dentro del documento mismo (aquellos que se pueden encontrar en cualquier documento generado en un procesador de textos)
2. Los que se producen por una transferencia HTTP – cliente y servidor se envían información sobre el objeto que están transmitiendo por medio de los metadatos.
3. Y el último, es aquel metadato que se utiliza cuando se consulta en otro documento (para comprobar si se puede acceder a él – o al sitio web-, verificar derechos de autor).

Sistemas de información en salud

Llamados también Sistemas de Información Hospitalario (SIH O HIS por sus siglas en inglés); consisten en uno o un conjunto de programas instalados en una institución de salud, también se dice que es aquel SIH que apoya a las actividades en los niveles operativos, tácticos y estratégicos de un hospital. Según la Organización Mundial de la Salud (OMS) es también un mecanismo para la recolección, procesamiento, análisis y transmisión de la información que se requiere para la organización y el funcionamiento de los servicios sanitarios y también para la investigación y docencia.

El objetivo de un SIH es ser una herramienta sistemática y organizada con la que sea posible generar la información para el mejoramiento de la gestión y la toma de decisiones de los diferentes niveles organizativos de los sistemas de salud.

Por lo cual (Mozzheyko D.) nos manifiesta que: *“El tiempo que los médicos y las enfermeras consumen buscando y organizando información clínica es excesivo y aumenta cada vez que el volumen de información sobre de los pacientes se amplía. Como el campo de la medicina se desarrolla, la información relacionada con los exámenes, la gestión y el tratamiento de los pacientes aumenta a un ritmo acelerado. Una hospitalización puede producir diez mil observacio-*

nes, conteos de pruebas de laboratorio, aumentos del monitoreo fisiológico, pruebas de diagnóstico por imagen, estudios del paciente, evaluaciones clínicas y notas de visitas. Por tanto, es importante que un hospital tenga un sistema de información que organice la información médica que generan los pacientes y los servicios médicos.”, por lo cual se traduce en lentitud en la búsqueda de la información a través de registros de papel o al deterioro de la documentación de papel.

Un SIH debe ofrecer una información para las distintas necesidades en la toma de decisiones, tanto en el ámbito individual (historia clínica), como en la elaboración de estadísticas que soporten el desarrollo y evaluación de programas y formulación de políticas de salud a mediano y largo plazo.

Entre las funcionalidades que permite dicho sistema, se encuentran entre otras: Llevar una admisión a las diferentes unidades de atención, tener un control de todos los servicios prestados a los pacientes (incluidos los de imagenología y laboratorio clínico), poder obtener datos epidemiológicos, obtener los datos precisos de costos por cada paciente atendido, obtener un estricto control y gestión de los pacientes mediante su expediente clínico en forma electrónica, se puede obtener y facilitar el acceso de los datos del paciente acerca de su tratamiento en forma más segura, ágil y eficiente a través de la historia clínica electrónica.

Inteligencia de negocios

La definición de Inteligencia de Negocios, dentro de las organizaciones es relativamente nueva, muchas veces va relacionado con la toma de decisiones mediante el uso de información transformada en conocimiento del negocio.

Según (Parr, 2000) la extracción del conocimiento, es la aplicación de técnicas de extracción de la información; las mismas permiten reunir, refinar, transformar datos mediante el uso de metodologías, aplicaciones y tecnologías.

Este conocimiento puede estar estructurado para que sirvan e indiquen a un área de interés, el cual genera el conocimiento sobre problemas y oportunidades del negocio que pueden ser aprovechados, corregidos o repotenciados (Ballard et. al 2006).

Implementar herramientas de inteligencia, permite a las organizaciones a soportar las decisiones que se toman, desde dos puntos de vista: a nivel externo produce ventajas sobre sus competidores (Maureen 2009) y del nivel interno ayuda a la gestión del personal (Sharma et al. 2009); nos manifiestan (Rosado Gómez & Rico Bautista, 2010).

Minería de datos

Hoy en día, en la cual la tecnología del Internet ha evolucionado hasta poder decir que las organizaciones están hiper-conectadas con todo el mundo; podemos también anotar que la cantidad abrumadora de datos se ha multiplicado, ya no es –solamente- los datos sencillos explicados en tuplas son los que prevalecen, sino que ahora se combinan con datos multimedia (fotos, videos, audios, mapas geo-referenciales, etc.), generados por un sin número de dispositivos electrónicos.

Asimismo, el descubrir el conocimiento del enorme volumen de datos es un reto en sí; donde la minería de datos (MD) es el concepto que viene a darle sentido a esta realidad, porque según (Riquelme, Ruiz, & Gilbert, 2006) resumen en una frase una realidad –“los datos en bruto raramente son beneficiosos directamente, para lo cual se recurre a: a) la habilidad para extraer información útil la toma de decisiones o la exploración y b) la compresión del fenómeno gobernante en la fuente de datos; anteriormente el analista tenía un papel preponderante en la organización pues realizaba la tarea de procesador de información en la generación de reportes, resúmenes con la ayuda de herramientas estadísticas, pero con la ingente y creciente demanda de los datos esta tarea resulta difícil y complicada en la actualidad.

De esta necesidad surge el término KDD (Knowledge Discovery in Databases) que se refiere a todo el proceso de extracción del conocimiento a partir de una base de datos; este concepto todavía continúa desarrollándose; desde diferentes áreas como: las bases de datos, aprendizaje automático, reconocimiento de patrones, estadística, teoría de la información, inteligencia artificial, razonamiento de incertidumbre, visualización de datos y computación de altas prestaciones; estos sistemas KDD incorporan teorías, algoritmos y métodos de todos estos campos de acción como nos manifiesta (Hernández-Orallo, Ramírez-Quintana, & Ferri, 2004).

Modelos de proceso para proyectos de minado de datos

A lo largo del tiempo, las diferentes metodologías para el desarrollo de los sistemas (desde el ciclo de vida clásico hasta las metodologías ágiles) han evolucionado de la mano de áreas de conocimiento (sobre todo en lo que concierne a la ingeniería de software) y han logrado afianzar y estandarizar los procesos de desarrollo de software; por lo cual el minado de datos por poseer características muy particulares no puede tratarse con estas mismas metodologías. Los modelos de proceso para proyectos de minado de datos, es un campo que reúne varias áreas del conocimiento, como son la estadística, las ciencias de la computación que intentan descubrir patrones en volúmenes altos de conjuntos de datos; también involucra en algunos casos métodos u algoritmos de inteligencia artificial (es la inteligencia exhibida por máquinas, muchas veces se aplica cuando una máquina imita las funciones “cognitivas” que humanos asocian con otras mentes humanas), aprendizaje automático (“machine learning” por sus términos al inglés; son técnicas que permiten a las computadoras “aprender”) y sistemas gestores de bases de datos.

Los modelos de proceso para proyectos de minado de datos son varios, entre los más importantes podemos mencionar a

KDD (Knowledge Discovery in Databases); SEMMA (Sample, Explore, Modify, Model, Assess); DMAMC (Definir, Medir, Analizar, Mejorar, Controlar); P³TQ (Product, Place, Price, Time, Quantity); CRISP-DM (Cross Industry Standard process for Data Mining), muchos de ellos en su orden cronológico de creación, respectivamente.

Resultados y discusión

Según los autores Hernández, Fernández, & Baptista, (2010); el enfoque de este estudio es probabilístico y determinístico; por cuanto vamos a seleccionar una serie determinada de muestras en forma aleatoria para su revisión. Para determinar el diseño de estudio que se requiere se debe un enfoque cualitativo. Este generaliza los resultados de muestras probabilísticas, estas logran obtener los casos de personas, contextos y situaciones que interesan al investigador y que se llega a ofrecer una riqueza para la recolección y el análisis de datos.

Por otro lado, cuando se analizan datos de naturaleza médica, se utiliza para el mismo y con el fin de obtener un mejor resultado, una aplicación de regresión logística, la cual es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica (es aquella que puede adoptar un número limitado de categorías) en función de las variables independientes o predictivas.

Así también tenemos la aplicación de la metodología para estructurar proyectos de minado de datos con la finalidad de convertirlas en metadatos y generar el conocimiento necesario para la toma de decisiones; para lo cual se aplican las seis fases de la metodología CRISP-DM:

I. Comprensión del negocio.

En el país, en los últimos tiempos una de las prioridades como política de estado ha sido el impulso de la modernización de la infraestructura de salud, tanto en su parte de obra civil, en la dotación de insumos y medicamentos, y en la aplicación de diver-

sas tecnologías de la información que apoyen estos servicios de salud. Para esto en esta investigación se realizaron recolecciones de datos de diferentes fuentes y con variedad de repositorios que utilizan la información y la misma sirve como insumo para monitorear los diferentes casos de enfermedades, principalmente en dos regiones del Ecuador (Costa y Sierra).

El esquema organizacional de salud en el país (Ecuador, 2012) establecido por el Ministerio de Salud Pública del Ecuador (MSP) en su alcance sustitutivo publicado en Registro Oficial N.- 279 del 30 de Abril del 2012; establece que existen procesos gobernantes (ver gráfico N.-1); cuyo responsable es el Ministro de Salud, que a través del direccionamiento estratégico del Sistema Nacional de Salud Pública establece la rectoría, regulación, planificación, coordinación, control y gestión de la Salud Pública Ecuatoriana; este delega estas responsabilidades a procesos sustantivos, y de manera especial – para nuestro caso- a la provisión de servicios de salud; la cual maneja: Centros especializados, hospitales, primer nivel de atención de salud, discapacidades y atención pre-hospitalaria y unidades móviles. De este último grupo, los tres primeros centros de atención; son los generadores de datos y requerimientos; y presentan muchas formas de almacenar la información.



MAPA DE PROCESOS

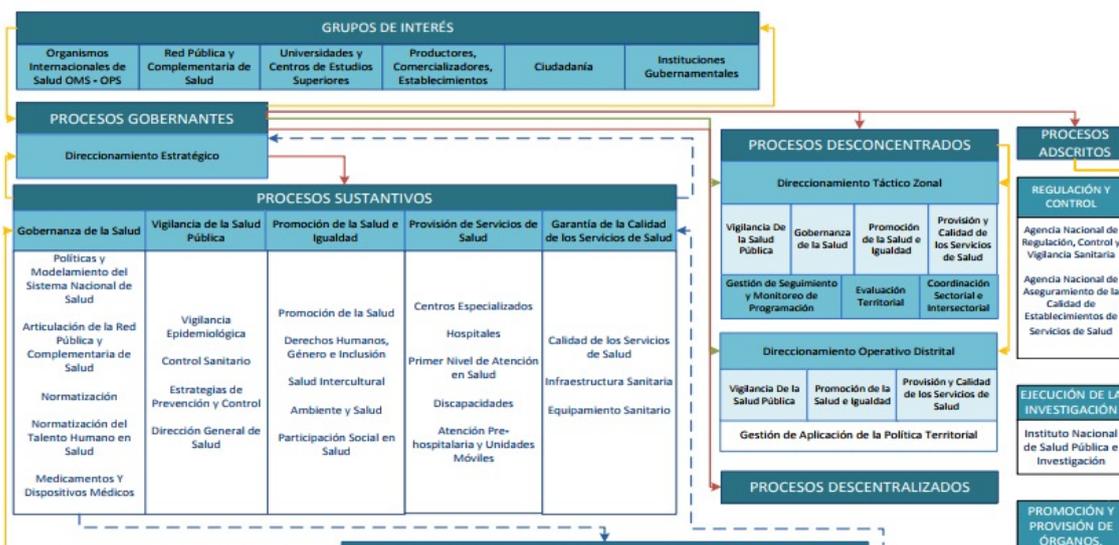


Figura 1. Procesos Gobernantes del M.S.P.

Elaboración: Tomado de la página del MSP – Estatuto Orgánico de Gestión Organizacional por procesos – pág. 13

En forma general; de los centros especializados y hospitales que se nombran en los procesos sustantivos; podemos señalar –entre los más importantes por su zona geográfica y tamaño de atención hospitalaria- que existen dos centros especializados emblemáticos que utilizan un HIS (región costa, ciudad de Guayaquil – un hospital de especialidades para adultos mayores y un hospital de especialidades pediátrico) cuyo repositorio de datos es una base de datos relacional Postgres; se tuvo acceso también a fuentes de datos realizadas en base de datos de “escritorio” Access e inclusive hojas de cálculo (región costa, ciudad de Portoviejo y Esmeraldas; región sierra, ciudad Latacunga) y también un sistema híbrido; es decir una parte de la información en un HIS y la otra en fuentes de datos diversas (región sierra, ciudad de Quito).

Del primer nivel de atención en Salud, se tuvo referencia que los mismos reportan a una gestión zonal su información realiza-

da en hojas de cálculo y en base de datos de escritorio “Access”; sin embargo, no se pudo obtener esta información.

II. Comprensión de los datos.

Del Sistema de Información de Salud (HIS), se pudo obtener el modelo inicial que consta de 788 entidades compuesto de 9753 campos; que entre sus principales esquemas –en forma general- nos muestra la información para usar en la parte clínica (citas médicas, historia clínica, consultas en emergencia, estadías en hospitalización, quirúrgica y de consultas a especialidades); financiera (pagos consultas, abonos, costos en insumos y materiales) y logística (manejo de bodegas, manejo de seguros médicos).

De la información que se puede obtener gracias a que se encuentra en bases de datos de tipo escritorio, está muy dispersa y en forma modular; en la mayoría de los casos desde Planta Central (Gobernanza de Salud) se ha enviado el aplicativo para



el uso del hospital; y en otros casos los departamentos de TI han desarrollado a la medida la base de datos de escritorio. Se encuentra principalmente las atenciones médicas y su sintomatología; las entidades guardan cierto grado de integridad en sus campos, sin embargo, por apuro o desconocimiento de los usuarios que operan las mismas; existen algunas incongruencias; alrededor de 30 entidades con 110 campos componen esta base.

La ultima fuente de datos que se encuentran en las instituciones son pequeñas hojas de cálculo; estas guardan la información más relevante en aproximadamente 10 pestañas, donde cada pestaña contiene de unos 30 a 40 campos en promedio por hoja. Se descubren principalmente los datos personales de pacientes, las fechas de atenciones, el motivo de la consulta y ciertas observaciones realizadas.

Cabe resaltar que toda esta información recopilada desde tres diferentes fuentes de datos (ver gráfico N.-2); tiene una época de aproximadamente 3 a 4 años; en el caso de los datos provenientes del HIS, existe un poco de más antigüedad en los datos; sin embargo, no se la considero por estar faltos de calidad, no íntegros por estabilización de la aplicación implementada. Adicionalmente podemos anotar que como se utilizara en el momento de predecir los datos, una regresión logística, las variables predictivas que se utilizarán serán: el estado socio-económico del paciente que está dividido en 5 categorías (alto, medio, alto bajo, medio bajo y bajo bajo); así como también por un estado en que se verifica si el paciente se pudo atender satisfactoriamente o si no lo pudo realizar sin importar la causa.

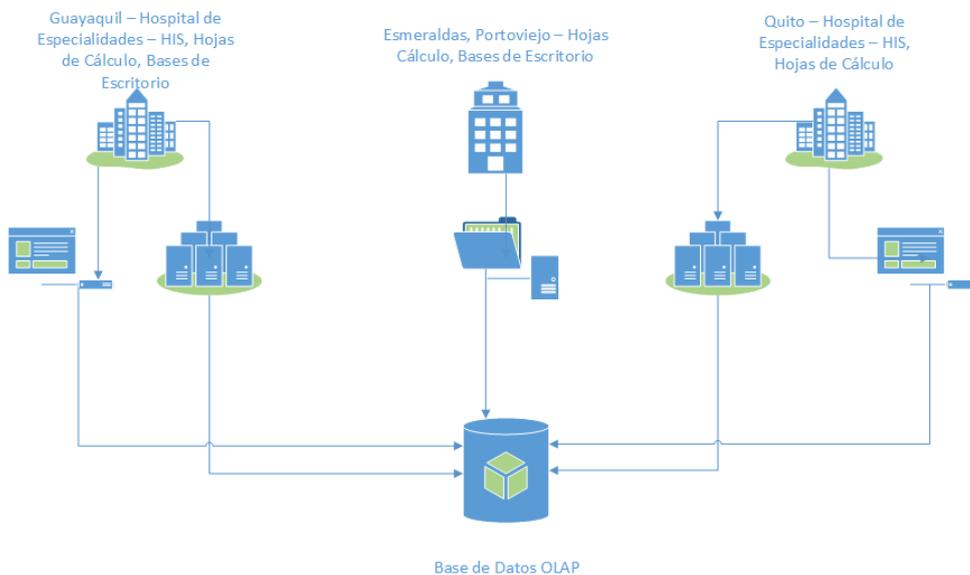


Figura 2. Procesos Gobernantes del M.S.P.

Elaboración: Autoría propia

III. Preparación de los datos.

En la selección de los datos de las tres diferentes fuentes de datos; se pudo notar una convergencia e integración de utilización de 3 entidades principales (pacientes, médicos (especialistas), atención médica)

en la base de datos de escritorio; de 4 entidades principales (pacientes, médicos (especialistas), atención médica (egresos en consulta externa) y atención médica (egresos en urgencia); y finalmente una pestaña en varias hojas de cálculo, seleccionando

las principales columnas (datos generales del paciente, fecha de atención, médico especialista, síntomas y observaciones) entre aproximadamente 40.

La mayoría de los datos obtenidos de las fuentes de datos arriba mencionadas, fueron ya validadas previamente; ya sea por el HIS o por la persona encargada de manejar la hoja de cálculo.

En el momento de la construcción del “dataset” o “vista minable” se añadió un campo para identificar la procedencia de la data (código del hospital que genero la información) y adicionalmente la zona geográfica

que se encuentra el mismo (en las tres fuentes de datos). En el momento de crear los nodos dentro del “workflow” de la herramienta analítica se procedieron a dividirlo en 3 cargas independientes: el primer nodo tiene una conexión de lectura a una base de datos postgres (ver gráfico N.- 3) cuyos datos provienen de la aplicación HIS; el segundo nodo es una conexión a un archivo “CSV” cuya información proviene de las hojas de cálculo tabuladas y finalmente una tercera conexión a una base de datos escritorio cuyos detalles pertenecen a las bases de datos aisladas de escritorio en “Access” que tienen algunos hospitales generales.

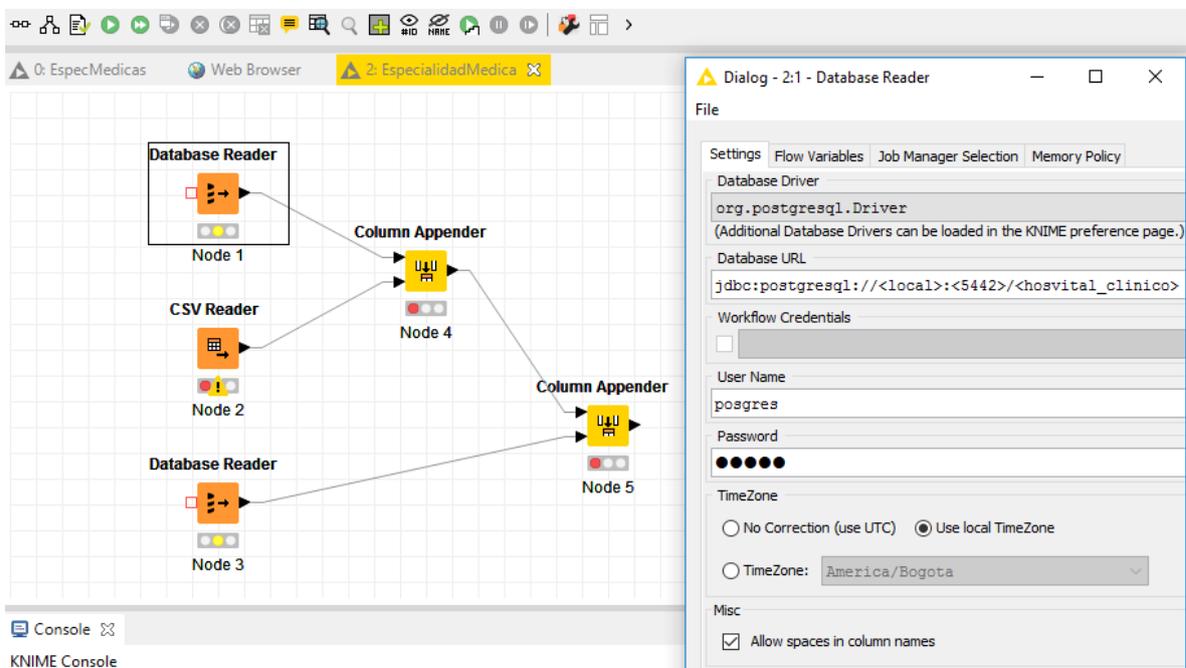


Figura 3. Construcción de datos en K-mine

Elaboración: Autoría propia

La integración y el formato de los datos se las realiza seleccionando el nodo de manipulación de datos, las columnas por las cuales se pretende realizar el patrón de la búsqueda son tres: columna que nos indica la edad del paciente en el momento de la atención, columna si fue atendido o no y la última columna es la enfermedad y su especialista que la reviso en cada unidad de atención pública

Para el respectivo formateo de los datos, estos van a ser visualizados desde el punto de vista de un histograma comparativo de las especialidades que más han sido requeridas a lo largo de un periodo de dos años y en seis entidades públicas, en conjunto con las edades que se han atendido en ese periodo.

IV. Modelado de los datos

En la selección de la técnica de modelado, podemos mencionar que una vez seleccionadas las columnas con el patrón de búsqueda arriba mencionada; se efectuó las reglas de clasificación a través del algoritmo seleccionado (Random Forest); el cual es un algoritmo predictivo que usa la técnica de Bagging para combinar diferentes árboles, donde cada árbol es construido con observaciones y variables aleatorias (la edad que fluctúa el paciente y la especialidad que ha sido atendido). Al realizar este modelado mediante el nodo propio de la herramienta analítica (ver gráfico N.-4); se puede decir que en forma interna se siguieron los siguientes pasos:

- Seleccionamos pacientes al azar (usando muestreos de reemplazo) para crear diferentes sets de datos.

- Se crea un árbol de decisión de cada set de datos, obteniendo como resultado diferentes arboles, donde cada set contiene diferentes pacientes y diferentes variables.
- Al crear los arboles se eligen variables al azar en cada nodo del árbol, dejando crecer el árbol en profundidad.
- Predice los nuevos datos usando el "voto mayoritario", donde clasificara como "positivo" si la mayoría de los árboles predicen la observación (una mayor cantidad de uso en especialidades) como positiva.

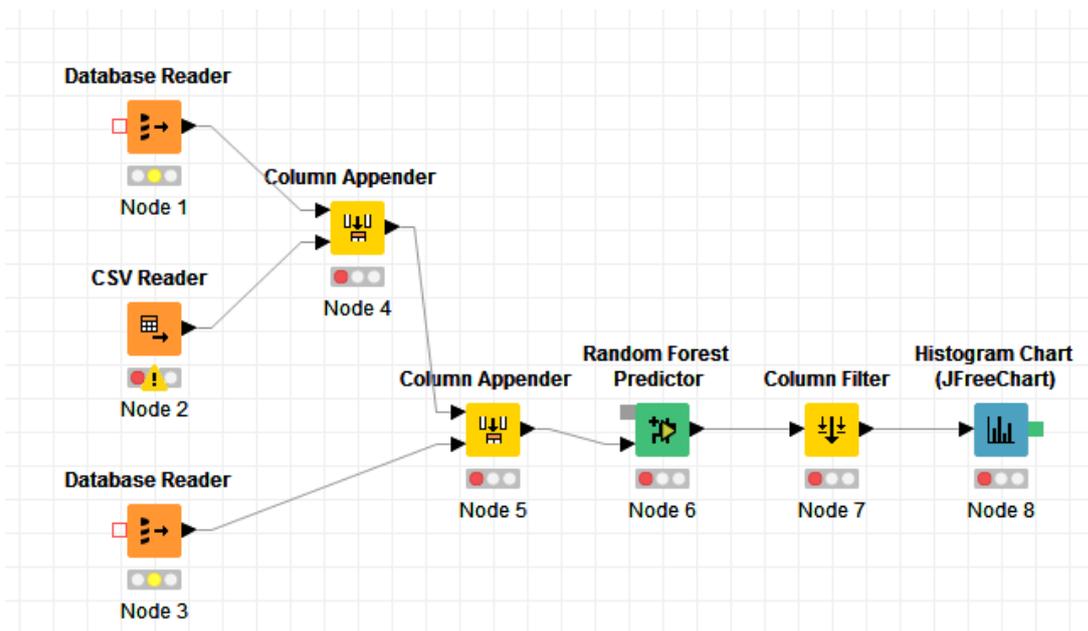


Figura 4. Modelado de datos en K-mine

Elaboración: Autoría propia

Finalmente, como resultado del nodo en que se muestra la información en forma de histogramas; podemos notar que cada una de las instituciones públicas tienden a tener un comportamiento binomial, por la cantidad de los datos –apenas dos años- no se pudo aplicar una regresión lineal, y mucho

menos logística. Por lo cual se tuvo que recurrir a una parte de los métodos numéricos para determinar su comportamiento.

En los comportamientos dados a partir de los histogramas; podemos notar que existen crestas altas y bajas; este tipo de

tendencia va muy acorde con las enfermedades cuando se presentan en periodos cíclicos; ya sea porque el periodo en que se evalúa la enfermedad depende de las estaciones climatológicas o mucho influye también la edad en la que se presentan las diferentes enfermedades.

Ahora bien, se realizó una tabulación porcentual; (ver tabla n.-1); que nos indica que sin importar las edades y las regiones; más

si por el mayor porcentaje de incidencia en el uso de las especialidades son: la neumología, la especialidad de emergenciología, la gastroenterología, la especialidad cirugía menor; poco después solo y exclusivamente en las mujeres –la ginecología- y finalmente en forma general la infectología.

Tabla 1. Mayor incidencia de especialidades

% de Especialistas por Hospital						
Especialista	Hospital Civil Delfina Torres Concha	Hospital De Niños Baca Ortiz	Hospital Del Niño Dr. Francisco Icaza Bustamante	Hospital Eugenio Espejo	Hospital Guayaquil Dr. Abel G. Pontón	Hospital Reg. Dr. Verdi Cevallos Balda
Neumología	6.27%	21.28%	14.48%	5.13%	2.55%	6.65%
Emergenciología	10.60%	12.64%	14.44%	12.55%	8.24%	9.74%
Gastrología	6.12%	3.81%	11.12%	7.49%	9.68%	6.49%
Cirugía	9.97%	5.18%	9.12%	2.96%	7.99%	8.60%
Ginecología	40.58%	3.18%	6.59%	1.78%	25.32%	36.54%
Infectología	3.25%	2.23%	4.77%	1.52%	1.59%	2.29%
Oncología	3.25%	5.02%	4.68%	17.44%	5.34%	3.44%
Otorrinología	0.24%	2.72%	3.99%	3.52%	1.17%	0.83%
Medicina General	2.78%	1.63%	3.80%	1.04%	1.33%	1.98%
Traumatología	0.38%	5.84%	3.41%	0.84%	0.50%	0.65%
Otros	16.55%	36.47%	23.59%	45.73%	36.28%	22.79%

Elaboración: Autoría propia

De la tabla arriba mencionada, se pudo deducir a partir de los datos logrados en los dos años realizar una proyección de atenciones por especialista, con la particularidad de el análisis se lo realizo en forma mensual durante los dos años de datos, logrando captar una distribución ajustada binomial; por lo que, para complementar la regresión se realizó mediante métodos numéricos, inferir que era una regresión polinómica de grado 4 (Mathews & Fink, 2012) (ver grafico n.- 5); cuyos valores que se muestran corresponden a los picos de atención que mucho tienen que ver con las épocas estacionales en que se comporta la data, así como también a las edades en que fluctúan las atenciones; como también nos manifiesta (Seminario Vásquez, 2012) en su estudio realizado para analizar funciones no lineales.

Se relacionó en algún momento a dicho análisis también como caso muy particular el campo “sexo” para cada atención, por cuanto existe un número muy importante en las estadísticas de atenciones para el sexo femenino en cuanto a la especialidad de ginecología-obstetricia que es exclusivo para este género.



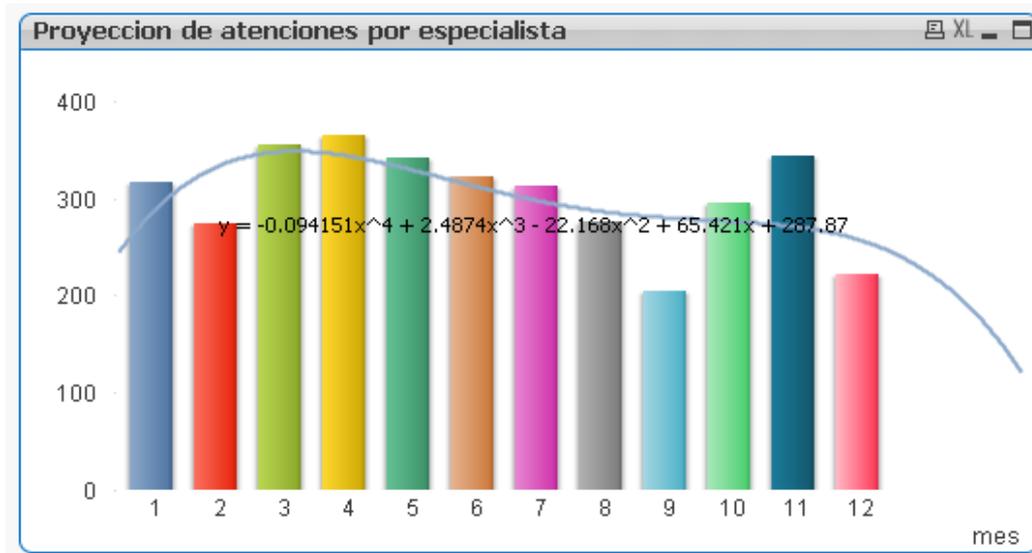


Figura 5. Proyección de atenciones por especialista

Elaboración: Autoría propia

Conclusión

En base al modelado de los datos extrapolados de diferentes repositorios de información y de diferentes centros de atención pública; podemos aplicar las fases de la metodología para un proceso de minería de datos; lo cual nos llevó después de aplicar un algoritmo predictivo a la consecución de las especialidades de mayor demanda en los años de estudio; teniendo como particularidad que se partió al comienzo del análisis, que los datos se podrían ajustar a una regresión lineal, pero después de aplicar los respectivos cálculos, la curva era imposible que se ajustara a este tipo de regresión.

Se procedió luego, a intentar a usar las regresiones no lineales más conocidas, como son una exponencial, logarítmica, etc., pero después de los consabidos análisis, tampoco se ajustan los datos a cualquiera de estos tipos de regresión.

En estos momentos, debido a lo expuesto anteriormente, a proceder el uso del análisis numérico, para tratar de ajustar los datos a una curva, y así obtener la función que define su comportamiento en el tiempo.

Consecuencia de este análisis, se determinó que el tipo de regresión que más se ajusta, es una regresión polinómica de grado 4. Además, no se puede generalizar que esta función sirva para más de un estudio o proyección por especialidad, ya que, para cada especialidad, a pesar que son funciones del mismo tipo, es decir, polinómicas de grado 4, difieren demasiado.

Esta dificultad tan elevada en los datos, de poder llegar a entregar un escenario predictivo, se debe en gran parte, a que el registro de la información no ha sido fiable todos los años debido a circunstancias políticas y técnicas, primero en la región en que se encuentra ubicado el hospital y luego a nivel general, en el país.

Se comprobó que el comportamiento de la información en un mismo hospital, difiere mucho por año, entre otras cosas, a que, en dichos hospitales, se ha cambiado de sistema de información hospitalario, generándose problemas de integridad, y fidelidad de los datos.

Pero, aun así, con todos estos inconvenientes a nivel de data, la metodología aplicada en la elaboración de este artículo ha cumplido las expectativas de los autores.

Sin embargo, este estudio se debe complementar con un estudio adicional acerca de las tendencias de selección de los diferentes profesionales al momento de seleccionar una especialidad dada, esto va muy de acuerdo a personalidades o rasgos básicos del profesional médico, así como también hasta cierto punto que tan rentable a lo largo del tiempo sea esta especialidad seleccionada.

Bibliografía

- Bitran, M., Zuñiga, D., Lafuente, M., Viviani, P., & Beltrán, M. (2015). Influencia de la personalidad y el estilo de aprendizaje en la elección de especialidad médica. *Revista Médica Chile* 2015, 133: 1191-1199.
- Cebron, N., & Berthold, M. (27 de 07 de 2008). Active learning for object classification: from exploration to exploitation. Obtenido de Active learning for object classification: from exploration to exploitation: http://140.123.102.14:8080/reportSys/file/paper/ahc/ahc_23_paper.pdf
- CENEC. (01 de 10 de 2010). Recursos y actividades de Salud 2010. Obtenido de Recursos y actividades de Salud 2010.
- Cubero, J., & Berzal, F. (30 de 08 de 2011). Sistemas Inteligentes de Gestión. Obtenido de Sistemas Inteligentes de Gestión: <http://elvex.ugr.es/decsai/intelligent/workbook/ai/PROLOG.pdf>
- Ecuador, M. d. (30 de 04 de 2012). Ministerio de Salud Pública. Obtenido de Ministerio de Salud Pública: <http://www.salud.gob.ec/wp-content/uploads/downloads/2014/09/ESTATUTO-SUSTITUTIVO-MSP-ALCANCE-REFORMA-ABRIL17.pdf>
- Hernández-Orallo, J., Ramírez-Quintana, M., & Ferri, C. (2004). *Introducción a la Minería de Datos*. Barcelona: Prentice Hall/Addison Wesley.
- Mathews, J. H., & Fink, K. D. (2012). *Métodos Numéricos con MatLab*. Ciudad de México DF: Pearson Prentice Hall.
- Montero Puñales, E., & Plasencia, A. (07 de 09 de 2017). www.researchgate.net/publication/282671158_Diseño_de_un_sistema_de_Vigilancia_Tecnologica_utilizando_tecnicas_de_mineria_web_para_identificar_las_lineas_investigativas_mas_prometedoras_para_los_cientificos_del_ICIMAF
- Moszheyko D., A. (s.f.). Diseño y desarrollo de un sistema de información de laboratorio en la atención pública de salud.
- Organization, W. H. (01 de 01 de 2000). Health Information systems development and strengthening. Obtenido de http://apps.who.int/iris/bitstream/10665/66203/1/WHO_EIP_OSD_00.6.pdf
- Organization, W. H. (01 de 01 de 2005). Improving the use of information for health care decision-making: what is needed. Obtenido de http://www.who.int/healthmetrics/library/issue_1_05apr.doc
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Inteligencia Artificial* Vol. 10 No 29, 11-18.
- Rosado Gómez, A., & Rico Bautista, D. (2010). *Inteligencia del Negocio: Estado del Arte*. Scientia Et Technica, 321-326.
- SAS Institute Inc. (09 de 08 de 2017). SAS The Power of Know. Obtenido de SAS The Power of Know: https://www.sas.com/content/dam/SAS/en_us/doc/factsheet/sas-enterprise-miner-101369.pdf
- Seminario Vásquez, R. (2012). *Metodos Numéricos para Ingeniería*. En R. Seminario Vásquez, *Metodos Numéricos para Ingeniería* (págs. 28-43). Bogotá.

CITAR ESTE ARTICULO:

Bustamante Chong, C. A., Cabezas Galarza, F. A., & Bustamante Chong, M. E. (2023). Generación del conocimiento en Instituciones de Salud mediante la aplicación de metadatos. *RECIAMUC*, 7(2), 252-265. [https://doi.org/10.26820/reciamuc/7.\(2\).abril.2023.252-265](https://doi.org/10.26820/reciamuc/7.(2).abril.2023.252-265)



CREATIVE COMMONS RECONOCIMIENTO-NOCOMERCIAL-COMPARTIRIGUAL 4.0.