



UNAPÉC
UNIVERSIDAD APEC

DECANATO DE INGENIERÍA E INFORMÁTICA
ESCUELA DE INFORMÁTICA

Trabajo de Grado para Optar por el Título de:

Ingeniero (a) de Sistemas de Computación

Título:

**Propuesta de diseño de un sistema de información para la toma de decisiones
médicas y administrativas en el Hospital Regional Taiwán 19 de Marzo en Azua,
República Dominicana, aplicando analítica de Big Data sobre el historial médico
de los pacientes.**

Sustentado por:

María José Gutiérrez Pérez	2016-1918
Wilkin Alexander Santana González	2014-0320
Manuel Ariel Galva Fernández	2012-1458

Asesor:

Ing. Eddy Alcántara, MSc

Distrito Nacional, República Dominicana

Marzo 2020



DECANATO DE INGENIERÍA E INFORMÁTICA
ESCUELA DE INFORMÁTICA

Trabajo de Grado para Optar por el Título de:

Ingeniero (a) de Sistemas de Computación

Título:

Propuesta de diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de Marzo en Azua, República Dominicana, aplicando analítica de Big Data sobre el historial médico de los pacientes.

Sustentado por:

María José Gutiérrez Pérez	2016-1918
Wilkin Alexander Santana González	2014-0320
Manuel Ariel Galva Fernández	2012-1458

Asesor:

Ing. Eddy Alcántara, MSc

“Los conceptos expuestos en esta investigación son de la exclusiva responsabilidad de sus autores”.

Distrito Nacional, República Dominicana

Marzo 2020

**“PROPUESTA DE DISEÑO DE UN SISTEMA DE INFORMACIÓN
PARA LA TOMA DE DECISIONES MÉDICAS Y
ADMINISTRATIVAS EN EL HOSPITAL REGIONAL TAIWÁN 19
DE MARZO EN AZUA, REPÚBLICA DOMINICANA, APLICANDO
ANALÍTICA DE BIG DATA SOBRE EL HISTORIAL MÉDICO DE
LOS PACIENTES**

ÍNDICE

AGRADECIMIENTOS	i
DEDICATORIA	iv
RESUMEN EJECUTIVO	vii
GLOSARIO DE TÉRMINOS	viii
INTRODUCCIÓN	xii
CAPÍTULO I: ASPECTOS INTRODUCTORIOS DEL PROYECTO	1
INTRODUCCIÓN	2
1.1 Selección del tema.....	3
1.2 Planteamiento del problema.....	3
1.3 Formulación del problema	4
1.4 Sistematización del problema	4
1.5 Justificación	4
1.6 Objeto de Estudio.....	5
1.7 Objetivos	5
1.7.1 Objetivo General	5
1.7.2 Objetivos Específicos.....	6
1.8 Campo de acción.....	6
1.9 Alcance.....	6
1.10 Metodología de investigación	7
1.10.1 Métodos de investigación.....	7
1.10.2 Técnicas de investigación	7
1.10.3 Técnicas de recolección	7
1.11 Población de la investigación.....	8
1.11.1 Determinación de la muestra.....	8
1.12 Delimitación de tema	9
CONCLUSIÓN.....	10
CAPÍTULO II: SISTEMA DE GESTIÓN DE INFORMACIÓN E HISTORIALES MÉDICOS ELECTRÓNICOS	11
INTRODUCCIÓN	12
2.1 Sistema de gestión de información	13
2.1.1 Historia.....	13
2.1.2 Elementos.....	14
2.1.3 Características	15
2.1.4 Tipologías.....	15

2.1.5	Ciclo de vida	16
2.1.6	Estándares internacionales	17
2.2	Historiales Médicos.....	17
2.2.1	Funciones	18
2.2.2	Características	19
2.2.3	Clasificación.....	20
2.2.4	Importancia	22
2.3	Historiales Médicos Electrónicos (HME)	22
2.3.1	Evolución de los Historiales Médicos Electrónicos	23
2.3.2	Componentes de los HME	25
2.3.3	Características de HME.....	25
2.3.4	Beneficios de HME.....	26
2.3.5	Servicios de Admisiones y Documentación Clínica	26
2.3.6	E-Salud.....	27
	CONCLUSIÓN.....	28
	CAPÍTULO III: ANALÍTICA DE BIG DATA.....	29
	INTRODUCCIÓN	30
3.1	Ciencia de los datos.....	31
3.1.1	Antecedentes	31
3.1.2	Disciplinas (Núcleos o Áreas).....	33
3.1.3	Procesos de análisis de los datos	33
3.1.4	Aplicaciones.....	34
3.1.5	Importancia	35
3.2	Análisis de datos (Data Analytics).....	35
3.2.1	Antecedentes	36
3.2.2	Análisis avanzado de datos	37
3.2.3	Algoritmo para el análisis de datos	38
3.2.4	Marco de referencia para el análisis de datos.....	40
3.2.5	Ventajas y limitaciones	41
3.2.6	El binomio Big Data -Data Analytics	41
3.3	Big Data	42
3.3.1	Antecedentes	43
3.3.2	Evolución	44
3.3.3	Características	45
3.3.4	Tipos de Datos del Big Data	47
3.3.5	Fuentes de del Big Data	48
3.3.6	Seguridad	50

3.4	Sistemas gestores de bases de datos.....	51
3.4.1	Evolución	51
3.4.2	Base de Datos Relacional.....	51
3.4.3	Base de datos no relacional.....	53
3.5	Analítica de Big Data.....	53
3.5.1	Antecedentes	55
3.5.2	Características	56
3.5.3	Tipos de Analítica de Big Data	57
3.5.4	Arquitecturas de la analítica de Big Data.....	58
3.5.4.1	Arquitectura Krishnan.....	59
3.5.4.2	Arquitectura Bob Marcus	60
3.5.4.3	Arquitectura de Microsoft	61
3.5.5	Ciclo de la analítica de Big Data.....	62
3.5.5.1	Capa de Recolección de Datos	63
3.5.5.1.1	Herramientas de recolección	63
3.5.5.2	Capa de agregación de datos	63
3.5.5.2.1	Adquisición	64
3.5.5.2.2	Transformación	64
3.5.5.2.3	Almacenamiento	64
3.5.5.3	Capa de procesamiento y análisis.....	64
3.5.5.4	Capa de visualización de los datos	65
3.5.5.5	Capa de gobierno de datos	65
3.5.6	Actores principales en una arquitectura de Big Data	66
3.5.7	Infraestructura de Big Data	66
3.5.7.1	Plataformas.....	66
3.5.7.2	Herramientas de recolección de datos.....	68
3.5.7.3	Tecnologías para el almacenamiento	70
3.5.7.4	Lenguaje de programación.....	72
3.5.7.5	Herramientas de visualización	73
	CONCLUSIÓN.....	74
	CAPÍTULO IV: CARACTERIZACIÓN Y DIAGNÓSTICO DE LA SITUACIÓN ACTUAL DEL SISTEMA DE GESTIÓN DE INFORMACIÓN ADMINISTRATIVA Y DEL HISTORIAL MÉDICO DE LOS PACIENTES DEL HOSPITAL REGIONAL TAIWÁN 19 DE MARZO DE AZUA, REPÚBLICA DOMINICANA	75
	INTRODUCCIÓN	76
4.1.	Antecedentes históricos del Hospital Taiwán	77
4.1.1	Misión	78
4.1.2	Visión.....	78

4.1.3	Valores	78
4.1.4	Organigrama.....	79
4.2	Procesos para la Creación y/o Gestión de Historiales Médicos	79
4.3	Procesos para la planificación del presupuesto	84
4.4	Sistema de Gestión de Información del Taiwán.....	88
4.4.1	Infraestructura de Hardware y Software del Centro de Salud.....	90
4.4.2	Entorno y plataforma de desarrollo.....	91
4.4.3	Versiones del sistema.....	91
4.4.4	Seguridad del sistema.....	92
4.4.5	Copias de Seguridad.....	92
4.5	Operación y manejo para pacientes Emergencia.....	92
4.6	Operación y manejo de pacientes referidos.....	93
4.7	Análisis FODA para la analítica de Big Data	93
4.8	Presentación y Análisis de la información recopilada a través de encuestas	96
4.9	Resumen del analisis e interpretacion de los resultados de las encuestas	105
	CONCLUSIÓN.....	107
	CAPÍTULO V: PROPUESTA DE UN DISEÑO DE UN SISTEMA DE INFORMACIÓN PARA LA TOMA DE DECISIONES MÉDICAS Y ADMINISTRATIVAS EN EL HOSPITAL REGIONAL TAIWÁN 19 DE MARZO DE AZUA, R.D. APLICANDO ANALÍTICA DE BIG DATA SOBRE EL HISTORIAL MÉDICO DE LOS PACIENTES.	
	108
	INTRODUCCIÓN	109
5.1	Fundamentación de la propuesta	110
5.2	Presentación de la Propuesta.....	112
5.2.1	Documento Visión del Proyecto	112
5.2.1.1	Propósito	112
5.2.1.2	Alcance.....	113
5.2.1.3	Posicionamiento	113
5.2.1.4	Descripción del personal involucrado	113
5.2.1.5	Entorno de Usuario	114
5.2.2	Especificaciones del diseño del Sistema	114
5.2.2.2	Requisitos Funcionales.....	115
5.2.2.3	Requisitos no Funcionales.....	116
5.2.2.4	Diagramas y Especificaciones de Casos de Uso	116
5.2.2.3	Diagrama de Modelado del Sistema.....	128
5.2.2.3.1	Diagramas de Secuencia	128
5.2.2.3.2	Diagrama de clases.....	132
5.2.3	Propuesta de solución Sistema de Gestión de Información de Analítica de Big Data	132

5.2.3.1	Diseño de la Arquitectura del Sistema de Información de Analítica de Big Data	133
5.2.3.1.1	Base de Datos Relacional.....	133
5.2.3.1.2	Arquitectura del sistema de información de Analítica de Big Data	138
5.2.3.1.2.1	Tecnologías de Desarrollo Sistema de Analítica de Big Data.....	139
5.2.3.1.2.1.1	Componentes de las bases de datos relacional	139
5.2.3.1.2.1.2	Componentes del Sistema de Información de Analítica de Big Data.....	140
5.2.4	Prototipo de Interfaz de Usuario	150
5.2.5	Estudio de Factibilidad de la Propuesta	162
5.2.5.1	Factibilidad Técnica.....	163
5.2.5.2	Factibilidad Operacional	163
5.2.5.3	Factibilidad Económica.....	164
5.2.6	Escalabilidad	165
5.2.7	Cronograma de Implementación	165
	CONCLUSIÓN	168
	CONCLUSIONES	xvi
	RECOMENDACIONES	xviii
	REFERENCIAS	xix
	ANEXOS	xxii

ÍNDICE DE FIGURAS

Figura 1- Elementos de un sistema de información.....	14
Figura 2- Ciclo de Vida de Desarrollo de un SGI.....	16
Figura 3- Formato de una historia clínica.....	20
Figura 4- Ejemplo de un historial médico electrónico.....	23
Figura 5- Las 11 E de la E-salud.....	27
Figura 6- Un Enfoque Clásico Para Enfrentarse A Un Problema De Extracción De Conocimiento.....	37
Figura 7- 7v Big Data.....	46
Figura 8- Tipos de datos y su ejemplo.....	47
Figura 9- Kapow Software Infographic.....	48
Figura 10- Retos de privacidad y seguridad en Big Data.....	50
Figura 11- Beneficios de la analítica de Big Data.....	54
Figura 12- Arquitectura de Big Data Analytics.....	55
Figura 13- Tipos de análisis de Big Data.....	58
Figura 14- Arquitectura Big Data por capas.....	58
Figura 15- Arquitectura de Krishnan.....	59
Figura 16- Arquitectura Bob Marcus.....	61
Figura 17- Capas de una arquitectura de Analítica de Big Data.....	62
Figura 18 Módulos de Hadoop.....	67
Figura 19- Arquitectura de Chukwa.....	68
Figura 20- La arquitectura de Chukwa.....	68
Figura 21- Arquitectura de Flume Fuente Big Data Análisis de herramientas y soluciones.....	69
Figura 22- Componentes de un agente.....	69
Figura 23- Ejemplo de un espacio de nombres de Cassandra. Fuente Big Data Análisis de Herramientas y Soluciones.....	70
Figura 24- Arquitectura de Hive elaborada en fases o etapas.....	71
Figura 25- Hospital Regional Taiwán 19 de Marzo.....	77
Figura 26- Organigrama del Hospital Taiwán.....	79
Figura 27- Doctora completando los formularios de los pacientes.....	80
Figura 28- Gabinetes donde son guardados los historiales médicos.....	83
Figura 29- Libro de Banco.....	85
Figura 30- Esquematización del Análisis FODA.....	93
Figura 31- Resumen del análisis FODA.....	95
Figura 32- Personal Involucrado.....	113
Figura 33- Definición del personal involucrado.....	114
Figura 34- Diagrama de caso de uso general.....	117
Figura 35- Diagrama de caso de uso autenticación de usuario.....	117
Figura 36- Diagrama de caso de uso realizar diagnostico.....	120
Figura 37- Diagrama de Caso de uso para consultar paciente.....	122
Figura 38- Diagrama de Caso de uso para consultar fármaco.....	124
Figura 39- Diagrama de Caso de uso para consultar producto de cocina.....	126
Figura 40- Diagrama de secuencia para autenticación.....	128
Figura 41- Diagrama de Secuencia Realizar Diagnostico.....	129
Figura 42- Diagrama de Secuencia para consultar paciente.....	130
Figura 43- Diagrama de Secuencia para consultar producto cocina.....	130
Figura 44- Diagrama de Secuencia para consultar fármaco.....	131
Figura 45- Diagrama de Clase.....	132
Figura 46- Diagrama Entidad-Relación.....	133
Figura 47 Arquitectura de Analítica de Big Data.....	138

Figura 48- Fuente de Datos	141
Figura 49- Estructura de la herramienta Sqoop.....	143
Figura 50- Infraestructura tecnológica	144
Figura 51- Ecosistema Hadoop	147
Figura 52- Análisis.....	148
Figura 53- Herramientas de seguridad	149
Figura 54- Pantalla inicial del sistema	150
Figura 55- Prototipo de Interfaz Gráfica (Autenticación 1).....	151
Figura 56- Prototipo de Interfaz Gráfica (Autenticación 2).....	152
Figura 57- Prototipo de Interfaz Gráfica (Autenticación 3).....	152
Figura 58- Pantalla que muestra el menú del sistema	153
Figura 59- Prototipo de la creación de historial médico	154
Figura 60- Listado de Pacientes	155
Figura 61- Pantalla principal del Historial Médico.....	155
Figura 62- Análisis del historial médico	156
Figura 63- Crear un producto de cocina.....	156
Figura 64- Listado de productos de la cocina	157
Figura 65- Análisis del producto.....	158
Figura 66- Plantilla para registrar un fármaco	159
Figura 67- Listado de fármacos.....	160
Figura 68- Análisis del fármaco.....	160
Figura 69- Predicción enfermedad	161
Figura 70- Prototipo Móvil	161
Figura 71- Predicción de productos de cocina	162
Figura 72- Diagrama de Gantt Cronograma de Implementación	167

ÍNDICE DE TABLAS

Tabla 1- Valores de nivel de confianza.....	9
Tabla 2- Frameworks y su descripción	40
Tabla 3- Actores principales en una arquitectura de Big Data.....	66
Tabla 4- Procesos y Servicios que ofrece el Hospital Taiwán	78
Tabla 5- Prioridad según el estado del paciente.....	82
Tabla 6- Sistema de Triage Manchester.....	82
Tabla 7- Caso de uso para autenticarse	118
Tabla 8- Caso de uso para realizar diagnóstico.....	120
Tabla 9- Caso de uso para consultar paciente	122
Tabla 10- Caso de uso para consultar fármaco	124
Tabla 11- Caso de uso para consultar producto cocina.....	126
Tabla 12- Paciente.....	135
Tabla 13- Doctor	135
Tabla 14- Personal	135
Tabla 15- Enfermedad.....	135
Tabla 16- Enfermero	135
Tabla 17- Departamento.....	136
Tabla 18- Cita.....	136
Tabla 19- Prescripción	136
Tabla 20- Producto.....	136
Tabla 21- Proveedor.....	136
Tabla 22- Diagnostico.....	137
Tabla 23- Historial	137
Tabla 24- Síntoma.....	137
Tabla 25- ARS	137

ÍNDICE DE GRÁFICOS

Gráfico 1- Tiempo para recibir atención médica	81
Gráfico 2- Interés en predecir enfermedades	84
Gráfico 3- Necesidad de un sistema para generar el presupuesto	86
Gráfico 4- Predicción de eventos médicos/administrativos en el hospital	86
Gráfico 5- Definir proveedores confiables.....	87
Gráfico 6- Conocimiento del SISGEHOSPI	90
Gráfico 7- Conocimiento sobre los HME	96
Gráfico 8- Conocimiento sobre la Analítica de Big Data	96
Gráfico 9- Adopción de nuevas tecnologías al sector salud.....	97
Gráfico 10- Acceso inmediato a los Historiales Médicos	97
Gráfico 11- Adopción de tecnologías para predicciones	98
Gráfico 12- Uso de tecnologías para el trabajo	98
Gráfico 13- Sistema que genere estadísticas	99
Gráfico 14- Enfermedades más frecuentes en el Hospital	99
Gráfico 15- Beneficios de la propuesta.....	100
Gráfico 16- Tiempo en otorgar turnos	100
Gráfico 17- Demora en canalizar y legalizar admisión.....	101
Gráfico 18- Ingresos por día	101
Gráfico 19- Proceso de control estadístico.....	102
Gráfico 20- Sistema para arrojar estadísticas.....	102
Gráfico 21- Desarrollo de nuevos productos.....	103
Gráfico 22- Asistencia amable y oportuna.....	103
Gráfico 23- Atención en el centro de salud.....	104
Gráfico 24- Predicción a tiempo de virus	104
Gráfico 25- Desarrollo de vacunas o acciones preventivas.....	105
Gráfico 26 Recibimiento de un diagnóstico inmediato	105

AGRADECIMIENTOS

Esta página me queda corta para mencionar a todas esas personas que aportaron un granito de arena para hoy orgullosamente ser la Ingeniera Gutiérrez Pérez.

A Dios en quien deposito cada día de mi vida, aunque muchas veces sentí su abandono, aprendí a buscarlo más y a incrementar mi fe y a mi madre celestial la virgen María que siempre me cubre con su manto e intercede por mi cada momento.

A mi madre, **Belki Pérez**, quien es el motor de mi vida y a quien le dedico cada uno de mis logros. Madre GRACIAS por tu entrega y dedicación hacia mí. ERES LA MEJOR, nunca lo dudes.

A mi padre, **Guillermo Gutiérrez**, eres mi ángel guardián, aunque te fuiste antes de lo esperado sé que estás muy orgulloso de los lejos que he llegado. Te extraño y gracias por tu protección

A mi tío, **Ramon Melo**, me acogiste como una hija, fuiste y eres esa luz que me guía por los mejores senderos, pero siempre con humildad. Me enseñaste que el sacrificio tiene su recompensa. Vamos por más.

A mis hermanos Lupita, Ángel, Noel, Melvin, José Fabian, Wanda Guillermina, y a ustedes dos que están junto a nuestro padre Luis y Daniel.

A mis tíos Maira, Nilda, Soraida, Pablo, Pedro, Carlos, Mariano, Valentín, Altagracia, Frank, José, Antonio, Amantina, Susana, Isabel, Ignacio, Alejandro, Iván, Henar, Rosili, María y Joaquín, pero de manera muy especial a ti Tía Ana que me abriste las puertas de tu casa y fuiste mi apoyo incondicional en esos momentos cuando no tenía fuerzas para continuar.

A mis abuelos Domingo Pérez y Manuela Marte y a todos mis primos y primas.

A mis mejores amigos **Emmy De Los Santos**, quien desempeñó un papel crucial en mi carrera, **Soleidi Aragonés** quien a pesar de la distancia me ha apoyado, **Félix Hernández** por soportarme y amarme tal y como soy, **Carina Beltré** gracias mi bruja por acompañarme cada día que lo necesitaba. A esos que se fueron por una razón u otra pero que estuvieron ahí para mí, y a los que siguen José M., Freisis 005, mi teamRD, Katerin M. Cristian T.

A mi grupo de oración Movi2 X el Amor por su apoyo y acogimiento.

A mis compañeros de tesis Wilkin y Ariel, por su entrega y soportar mis exigencias.

A la Universidad APEC por abrirme sus puertas y acogerme, dicen que lo mejor se queda de último, a mi estimado asesor quien hoy en día lo considero como mi mentor, gracias por todo Ing. Eddy Alcántara (SI) y siempre recordare que “La Calidad es un estilo de vida”.

María José Gutiérrez Pérez

AGRADECIMIENTOS

Aunque siempre se inicia dando gracias, me acuerdo del primer sermón que prediqué en la iglesia “La gratitud”, no es lo mismo dar gracias que ser agradecidos. Por esto me es necesario poner a mi Redentor por darme gracia y estar conmigo siempre. Él me ayuda a poner todo lo que ya dio de antemano a sus pies. De manera que pueda utilizarlo no para ser orgulloso, darlo en servicio para su obra, mi familia, su iglesia y la patria.

Agradezco a mi madre Fiordaliza González y a mi abuela Eulogia González, dos tesoros del Señor que me han criado y dado ánimo para alcanzar cada meta. Qué Él me ayude a honrar les como buen hijo.

No puedo dejar de agradecer a Dios por la hermosa iglesia donde me ha colocado, la Iglesia de Convertidos a Cristo en Azua, con todos los hermanos que no han desmayado orando por sabiduría durante mi carrera universitaria. Tampoco puedo pasar por alto a la Iglesia Bautista Cristiana, por su preocupación por mí aún en esta Tesis, porque Dios es mi testigo muchos se han preocupado con la frase “Y la universidad”.

Doy muchas gracias a Dios por Francisco Solano y su familia, él ha sido un pastor ejemplar y un amigo, velando y preguntando cómo voy cómo mis estudios.

No puedo dejar de dar gracias a mi Padre celestial por mis compañeros de Tesis, sé que muchas veces me enojé y los exasperé en algún momento, todo con el objetivo de que dieran un esfuerzo para que las cosas sean excelentes, pero han sido gran apoyo. Gracias María Gutiérrez, gracias, Ariel Galva.

Gracias a Dios por nuestro tutor Eddy Alcántara, por guiarnos y esforzarnos a innovar y alcanzar metas no comunes, es de los profesores que te dejan una enseñanza.

Ciertamente no ha sido fácil trabajar y estudiar, pero Dios ha sido maravillosamente bueno, que Él me ayude a traer mis talentos como corona a sus pies, porque un día él me pedirá cuenta, usándolo darle honor. Estoy seguro de que hay muchísimas personas que Dios ha usado tanto en la Universidad APEC, como en todo el trayecto recorrido en esta ciudad de Santo Domingo. Profesores, maestros compañeros de estudio, hermanos en la fe.

Por otro lado, también doy gracias a Dios por todas las vicisitudes por las que hemos cruzado porque también hay que dar gracias a Dios por lo bueno y por lo malo. Si de verdad creemos que Él es soberano, entonces estamos seguros de que orquesta toda nuestra vida.

Finalmente me acuerdo de lo que dice Dios a través de Esdras, en mis palabras serían: Este logro no es mi logro, sino que la buena mano de Dios ha sido conmigo. Él puso los recursos, el ánimo y las fuerzas, la gracia y las personas en su tiempo. Gratitud hay en mi alma hoy.

Wilkin Alexander Santana González

AGRADECIMIENTOS

En primer lugar, quiero agradecer con todo mi corazón a Dios, por permitirme tener esta gran oportunidad que tanto soñé, que, a pesar de todos los obstáculos, de tantas complicaciones de salud y de otras índoles, me dio la sabiduría y los medios para poder llegar hasta el final.

A mi padre **Carlos Manuel Galva**, Quien a pesar de todas las dificultades siempre me brindó su apoyo y su preocupación para que no me rindiera y nunca me faltara lo esencial y sin ese apoyo no hubiese sido capaz de llegar hasta aquí, Muchas gracias por todo, este logro no es solo mío, también es tuyo.

A mi madre **Griselda Fernández**, tú que siempre has estado ahí al pendiente de todo, brindando esa motivación en los momentos difíciles para que nunca me rinda. Muchas gracias por todo y por darme la crianza para ser una persona de bien es que he podido llegar hasta aquí y por eso y más te dedico este logro.

A mis hermanas Carys Galva y Catia Galva, de que alguna forma u otra siempre presentaron su preocupación por mi bienestar y desempeño, y siempre estuvieron ahí aportando ese granito de arena que fueron parte del combustible que impulsó mi motor de vida.

A tías y tíos Albida, Adelina, Elsa, Bolívar, Celeste, Evangelina, y a todos en general, que de una forma u otra siempre me extendieron la mano cuando lo necesité, siempre me motivaron a crecer, siempre me aconsejaron y reprocharon cuando fue necesario y eso es algo que cabe destacar y agradecer.

A mis primos que siempre me vieron como ejemplo a seguir, que me motivaron a ser cada día mejor y a aquellos que a su vez fueron inspiración para mí, muchas gracias.

A nuestro maestro de carrera, mentor y asesor de nuestro trabajo de grado, quien brindó su apoyo de manera excepcional y estuvo disponible en todo momento e hizo posible lograr culminar este proyecto, muchas gracias.

Desde el inicio de esta travesía, desde mi primer día de universidad, pensaba en cómo sería mi último día de carrera y que tipo de personas conocería y con quien realizaría mi trabajo final, y agradezco grandemente a Dios por poner en mi camino a dos excelentísimas personas para que pudiéramos trabajar juntos y compartir este gran logro, en el cual yo fui quien les dio dolores de cabeza, pero al final sin ellos esto no hubiese sido posible. De manera muy especial y de todo corazón Muchas gracias, María José Gutiérrez y Wilkin Alexander Santana.

A mis compañeros de trabajo Massiel, Luis, Francia, Esperanza, Santa, Ana, Alberto, Ybernia, Arles, Lucrecia, Pedro y demás quienes cada uno a su manera siempre me brindaron su apoyo y motivación incondicional las cuales agradezco y resalto. Y a todas aquellas personas que a su manera deseaban verme lograr esta meta, esto también es para ustedes, muchas gracias.

Manuel Ariel Galva Fernández

DEDICATORIA

Culminar mi carrera es uno de mis logros más preciados, y al mismo tiempo el primer escalón para seguir creciendo y quiero dedicárselo al Todopoderoso por ser mi sustento en esos días grises y a la Virgen de Schoenstatt por su protección durante todo este camino.

A cada una de esas personas que físicamente se marcharon de mi vida pero que sus buenos deseos viven en mí, a ti abuela en especial **Juliana Guillén** y a todas esas que me han dado unas palabras de alivio.

Pero de manera super especial a ti padre mío **Guillermo Gutiérrez**, aunque no estuviste ahí para ver mi progreso durante estos 3 años y 5 meses, estoy segura de que siempre cuidabas de mí, cada mañana al salir de casa, cada noche al regresar, cada día que me enfermaba, pero también esos días cuando era reconocida como estudiante meritoria por eso a ti viejo mío te dedico este gran logro.

María José Gutiérrez Pérez

DEDICATORIA

Personalmente quiero dedicar el logro de culminar este trabajo de grado como uno de los requisitos fundamentales para optar por el título de Ingeniería de Sistemas de Computación a Dios, porque reconozco su ayuda en todo tramo de mi vida. Luego lo dedico a mi mamá Fiordaliza y a mi abuela Eulogia porque ellas dos se han esforzado con todos los recursos que poseen a dar todo su amor y trabajo para que pudiera concluir mis estudios de grado. Es innegable el hecho de que estas tres personas han sido mis soportes.

Ser ingeniero no es colocar una abreviatura y presumir con orgullo un anillo o bien un título; es ingeniárselas para prestar ayuda ante los problemas que se presentan en la familia, la sociedad, la patria, el mundo. Debemos ser humildes y gozarnos con los triunfos que Dios en su gracia nos permite conquistar.

Wilkin Alexander Santana González

DEDICATORIA

Culminar este recorrido de mi vida, este logro de poder ser llamado Ingeniero de Sistemas es solo el inicio de un largo trayecto que me espera. Me siento muy agradecido con mi Dios todo poderoso, a quien le antepongo todas mis hazañas y quien fue mi guía en todo este trayecto, este logro es para ti.

De igual manera quiero dedicar los frutos de este trayecto al cual dediqué parte importante de mi vida, y que con gran sacrificio seguí adelante sin darme por vencido a mis padres, Carlos Manuel Galva y Griselda Fernández quienes lucharon y se esmeraron para que todo esto sea posible y llegado el día les entrego el resultado de todo ese sacrificio.

Manuel Ariel Galva Fernández

RESUMEN EJECUTIVO

El presente trabajo de investigación aborda la problemática actual por la cual está atravesando el Hospital Regional Taiwán 19 de Marzo, donde se presenta una gestión precaria al momento de tomar decisiones, debido a que carecen de un sistema que en base a los historiales médicos proporcione informaciones precisa y veraz que incidan en las mejoras del centro.

Los avances tecnológicos de este siglo han influido significativamente en la forma como se gestiona el sector salud en términos investigativos y de atención primaria. Hoy día no solo se habla de salud predictiva, sino que como resultado de la automatización de los datos personales clínicos de los pacientes se ha evolucionado a historiales médicos electrónicos, se habla también de telemedicina.

Estos historiales clínicos generan millones de datos que las herramientas tecnológicas, actuales no son capaces de manejar, debido a la rapidez con la que los datos crecen. Como consecuencia de esto, la ciencia de la salud ha tenido que valerse de la ciencia de datos, el análisis de Big Data y por su puesto el análisis en sus diferentes tipos: descriptivos, predictivos, diagnóstico y prescriptivo para extraer el mayor beneficio de estos.

Uno de los propósitos principales que persigue esta investigación es tomar como herramienta el análisis de Big Data a través de un diseño de un sistema de información para analizar los historiales médicos y extraer conocimiento útil, tanto en el área clínica como en el área administrativa. Para lo cual, se toma como lugar de aplicación el Hospital Regional Taiwán 19 de Marzo. Por otro lado, que este trabajo de grado pueda trazar un referente en el marco de la salud de la República Dominicana y alinear las metas de la estrategia nacional y objetivos del milenio de las Naciones Unidas en el marco de la salud.

Palabras clave: Big Data, Análisis de Big Data, Ciencia de Datos, Historiales Médicos Electrónicos, Analítica Predictiva.

GLOSARIO DE TÉRMINOS

- Amenazas:** Son sucesos y situaciones del entorno que pueden impedir la implementación de una propuesta, reducir su efectividad, alargar los tiempos de las etapas y en el peor de los casos podría ser sinónimo del fracaso.
- ARS:** Las administradoras de riesgos de salud, son entidades públicas, privadas o mixtas, con patrimonio propio y personería jurídica, autorizada por la superintendencia de salud y riesgos laborales (SIALRIL), para asumir y administrar el riesgo de la provisión del Plan Básico de Salud a una determinada cantidad de afiliados y beneficiarios.
- Back Up:** Es una copia de seguridad o copia de respaldo. En ciencias de la información e informática es una copia de los datos originales que se realiza con el fin de disponer de un medio para recuperarlos en caso de su pérdida.
- C#:** Es un lenguaje de programación diseñado por la conocida compañía Microsoft. Fue estandarizado en hace un tiempo por la ECMA e ISO dos de las organizaciones más importantes a la hora de crear estándares para los servicios o productos.
- Dashboard:** Un tablero de instrumentos es un tipo de interfaz gráfica de usuario que a menudo proporciona vistas de los indicadores clave de rendimiento.
- Deep Learning:** Es un tipo de machine learning que entrena a una computadora para que realice tareas como las hacemos los seres humanos, como el reconocimiento del habla, la identificación de imágenes o hacer predicciones. En lugar de organizar datos para que se ejecuten a través de ecuaciones predefinidas, el Deep Learning configura parámetros básicos acerca de los datos y entrena a la computadora para que aprenda por cuenta propia reconociendo patrones mediante el uso de muchas capas de procesamiento.
- Debilidades:** Son factores que provocan un obstáculo para el avance y el desarrollo efectivo, representa una amenaza frente a la competencia ya que dependiendo del grado de importancia de dicha debilidad podría significar el fracaso del proyecto.

Diagnóstico:	Es el proceso de reconocimiento, análisis y evaluación de una cosa o situación para determinar sus tendencias, solucionar un problema o remediar un mal.
Fortalezas:	Son las capacidades que constituyen una herramienta, empresa, proyecto, como también cualquier invención, que le permiten alcanzar una posición o destacar y ser competitivo abriendo paso a nuevas oportunidades.
Microsoft SQL Server:	Es un sistema para la gestión de bases de datos creado por Microsoft, el mismo se basa en el modelo relacional.
Módulo:	Es una estructura o bloque de piezas que, en una construcción, se en cantidad a fin de hacerla más sencilla.
Oportunidades:	Son todos aquellos elementos que representan un camino o una posibilidad positiva y explotable que otorga algunas ventajas competitivas.
OMS:	La Organización Mundial de la Salud es el organismo de la Organización de las Naciones Unidas (ONU) especializado en gestionar políticas de prevención, promoción e intervención en salud a nivel mundial. Inicialmente fue organizada por el Consejo Económico y Social de las Naciones Unidas que impulsó la redacción de los primeros estatutos de la OMS.
(PROMESE/CAL):	El Programa de Medicamentos Esenciales Central de Apoyo Logístico, fue creado en el año 1984 por decreto de la Presidencia de la República como respuesta a los altos costos de los medicamentos que se definían como esenciales para la atención básica de los segmentos más vulnerables de la sociedad dominicana.
Responsive:	El diseño web adaptable, es una filosofía de diseño y desarrollo cuyo objetivo es adaptar la apariencia de las páginas web al dispositivo que se esté utilizando para visitarlas.
SNS:	El Servicio Nacional de Salud y tiene como propósito asegurar la efectividad, técnica, administrativa y financiera de los Servicios Regionales de Salud.
SGH:	El Sistema de gestión hospitalaria es una solución TIC orientada a dar una gestión integral del proceso asistencial hospitalario.

TIC:	Las Tecnologías de la Información y la Comunicación, son todos aquellos recursos, herramientas y programas que se utilizan para procesar, administrar y compartir la información mediante diversos soportes tecnológicos
UNAP:	Es una unidad Funcional y Operativa básica y la puerta de entrada al sistema de salud dominicano en el primer nivel de atención, estas están ubicadas en zonas estratégicas del país para dar asistencia preventiva médicas, atender caso de urgencia y hacer referencias a otros centros médicos más desarrollados dependiendo el caso.
XLS:	Es una extensión para los archivos de hoja de cálculo utilizados en la aplicación Microsoft Excel. Actualmente la extensión más usada de este formato es .xlsx y es compatible con la mayoría de los manejadores de datos.
XML:	El lenguaje de marcado es un conjunto de códigos que se pueden aplicar en el análisis de datos o la lectura de textos creados por computadoras o personas. El lenguaje XML proporciona una plataforma para definir elementos para crear un formato y generar un lenguaje personalizado.
Windows Server:	Es un sistema operativo de servidor que permite a un equipo manejar funciones de red como servidor de impresión, controlador de dominio, servidor web y servidor de archivos.
M2M:	El concepto de “machine to machine” se refiere al intercambio de información o comunicación en forma de datos entre dos máquinas remotas.
Ecléctico:	Es un proceso mediante el cual se presentan rasgos o características diferentes entre sí, que normalmente no se combinarían pero que igualmente pueden dar un estilo, fenómeno o realidad nueva y diferente al resto.
MDM:	La gestión de datos maestros es un método que permite a una organización relacionar todos sus datos críticos con un solo archivo llamado archivo maestro, de forma que se obtiene un punto de referencia común para los datos más importantes, simplificando además el intercambio de datos entre personal y departamentos.
Triaje:	El triaje es un proceso que permite una gestión del riesgo clínico para poder manejar adecuadamente y con seguridad los flujos de pacientes cuando la demanda y las necesidades clínicas superan a los recursos.

CNA: Consejo Nacional de Acreditación. Organismo de naturaleza académica que depende del Consejo Nacional de Educación Superior (CESU), integrado por personas de las más altas calidades científicas y profesionales, cuya función esencial es la de promover y ejecutar la política de acreditación adoptada por el CESU y coordinar los respectivos procesos.

ETL: Es un tipo de integración de datos que hace referencia a tres procesos (extraer, transformar, cargar) que se utilizan para mezclar datos de varias fuentes. Se utiliza con frecuencia para construir almacenes de datos.

INTRODUCCIÓN

En las últimas décadas se han suscitado nuevas tecnologías de la información y la comunicación (TIC), que han revolucionado la manera de vivir de los seres humanos y consigo aspectos remarcables, tanto en los niveles económico, social y político. El sector salud no es la excepción, es notable que este se ha visto influenciado por la incorporación de herramientas tecnológicas para el procesamiento y análisis de los datos, con el propósito de facilitar su uso y manejo en procesos como, la gestión de historiales médicos, la toma de decisiones médicas o administrativas, procesos quirúrgicos entre otros. A esto se suma la acumulación del conocimiento en paralelo con los avances tecnológicos que han incidido en la gestión médica, permitiendo a los profesionales de la salud acceder de manera directa a los datos que posean mayor relevancia en el historial médico de pacientes y logrando un trato personalizado y especializado.

En sus inicios la gestión de la información era totalmente análoga, únicamente se hacía uso de papel y lápiz, los registros estaban almacenados en archiveros. Esto cambió con el advenimiento del ordenador, el internet y otras tecnologías como las bases de datos.

En la actualidad, nos encontramos frente a la llamada cuarta revolución tecnológica, enfocada en las nuevas Tics, las cuales han transformado nuestra forma de vivir en prácticamente todos los aspectos. En lo relativo a la salud, el papel de las Tics es decisivo, permitiendo que la población no atendida en zonas remotas de los países reciba servicios médicos, y que estos sean más eficientes. A raíz de esto surge lo que se conoce como e-salud o cibersalud que puede reducir la exclusión de los servicios de salud de personas en la base de la pirámide, pues ayuda a proporcionar a los pacientes comunicación constante, acceso a la información, nuevas interfaces de consulta, acerca el conocimiento de médicos especialistas a zonas remotas a través de la telesalud (atención médica remota) y genera eficiencia en la provisión de los servicios relacionados. (Mariscal, Rentería, & Arteaga, 2014)

Por tanto, como consecuencia de todas las actividades que se realizan a través de procesos digitalizados y dispositivos electrónicos, el volumen de los datos aumenta con una velocidad sin precedentes, ahí es cuando nace la necesidad de desarrollar una tecnología con capacidad para almacenar y tratar enormes cantidades de datos y es en el año 2003 cuando Francis Diebold utiliza el término de Big Data para explicar el fenómeno del crecimiento de datos. (Diebold, 2012)

Antes de hablar de Big Data, hay que resaltar un concepto que en los últimos años ha dado mucho de qué hablar: Ciencia de Datos que según José Moreno: *“La ciencia de datos es una disciplina emergente y de gran pertinencia para todas las organizaciones que deseen codificar el valor oculto e intangible de sus datos”* (Moreno, 2017). La ciencia de datos surgió como una disciplina nueva e importante, se puede ver como una fusión de disciplinas clásicas como la estadística, minería de datos, bases de datos y sistemas

distribuidos. Los enfoques existentes deben combinarse para convertir los datos disponibles, en valor para los individuos, las organizaciones y la sociedad.

Las técnicas de minería de procesos utilizan datos de eventos para descubrir procesos, verificar su cumplimiento, analizar cuellos de botella, comparar variantes de procesos y sugerir mejoras.

Por su parte, Big Data es el conjunto de datos que no posee un sólo formato estándar, si bien, estos pueden ser estructurados, semiestructurados y no estructurados. Esta tecnología se caracteriza por cinco cualidades: volumen, velocidad, veracidad, valor y variedad estas las distinguen del resto de las tecnologías de su clasificación y por lo cual será utilizada en la presente propuesta. Big Data permite el análisis de los datos por lo que puede ser aplicada en la toma de decisiones, análisis de datos, historiales médicos e investigaciones científicas. Dentro de la ciencia de datos se encuentra la **Analítica de Big Data** la cual se refiere al almacenamiento, administración y análisis de grandes volúmenes de datos a través de métodos estadísticos o científicos para descubrir relaciones entre los datos, esta se ha aplicado recientemente al campo de la salud. (Reyes, Bautista, Thompson, y Penna, 2018)

Por otra parte, es necesario destacar el reto para los centros de salud de garantizar un uso óptimo de toda la información que es recopilada por medio de los historiales médicos electrónicos, reconociendo que, en la mayoría de los casos se subvalora la importancia que radica en esos datos. Por esto, se considera la Analítica de Big Data como una de las tecnologías que ha impactado significativamente en la salud. Esta tiene como propósito permitir la toma de las mejores decisiones y estrategias de negocios, como es el caso de los centros de salud. Esta herramienta ofrece optimización en el almacenamiento de datos, reduce costos, favorece a la investigación, provee posibles problemas y mejora de la asistencia a los ciudadanos.

Aplicar Analítica de Big Data a los historiales médicos electrónicos constituye una oportunidad, para investigadores y especialistas de la salud, debido a la adquisición de información precisa, gracias a características como la granularidad de los datos, que brinda niveles de detalle más completos. Todo esto hace posible que se puedan predecir, prevenir y proyectar posibles eventualidades como enfermedades, epidemias, brotes y otras patologías. Esto permitirá que la gestión clínica incorpore avances desde el punto de vista de atención al paciente, como también determinar factores que incidan directamente en la planificación anual de los centros de salud.

En el ámbito internacional son varios los países que han asumido el reto de implementar sistemas para el análisis de los datos utilizando estas herramientas. Entre los países que se han destacado por la implementación de esta tecnología está España, que resalta lo siguiente: *“su principal objetivo es desarrollar metodologías y herramientas de apoyo a la toma de decisiones clínicas”*. Asimismo, el doctor Bernardo Valdivieso, director del Área de Planificación del Hospital la Fe y director del Área de Atención Domiciliaria y Telemedicina del Departament de Salut La Fe, responsable de la nueva Plataforma *“todo esto nos permite generar más y mejor información para caracterizar de manera más*

eficiente tanto en las enfermedades como la provisión de servicios en salud lo que nos permite conocer el “valor de nuestras intervenciones” – resultados clínicos- e impulsar la medicina de precisión” (Arcángel J. 2018).

Según la Organización Mundial de la Salud sobre el desarrollo de las Tics y salud pública en América Latina destacó que “El avance de los países del continente americano en materia de e-salud es diverso. Si bien el 61% de ellos ya tiene una estrategia nacional de e-salud, muchos aún deben pasar de la fase de la formulación de políticas y estrategias de e-salud a la de implementación.” Asimismo, hicieron énfasis en que los macrodatos se encuentran en una etapa inicial. Salud, O. P. (2016).

Según el portal oficial del Estado Dominicano: La República Dominicana tiene una Estrategia y Plan de Acción de e-Salud. Esta es una iniciativa liderada por el Ministerio de Salud, aprobada en Washington en la quincuagésima primera sesión del Consejo Directivo de la OPS y de la Organización Mundial de la Salud.

En la actualidad se ha creado un comité estratégico y una agenda nacional de e-Salud de manera que se pueden mencionar:

- El Comité Estratégico Nacional e-Salud fue constituido por el Ministerio de Salud Pública y la Organización Panamericana de la Salud (OPS) y se llamó: Comité Estratégico Nacional e-Salud.
- La aplicación de la Estrategia y Plan de Acción ayudará a mejorar el acceso a los servicios de salud, mediante el uso de la tecnología de la información y la comunicación (TIC). De esa manera se dará una respuesta más rápida a los usuarios de los servicios de salud.
- Integran además el Comité Estratégico; el Ministerio de Planificación y Desarrollo, el Instituto Dominicano de las Telecomunicaciones, el Seguro Nacional de Salud, el Ministerio de la Presidencia y la Oficina Presidencial de Tecnologías de la Información y la Comunicación (OPTIC), la Universidad Autónoma de Santo Domingo y la Superintendencia de Salud y Riesgos Laborales.

A pesar de los esfuerzos de organismos para el desarrollo de estrategias que contemplen el uso de la tecnología en el sector salud, no se han diseñado sistemas en materia de e-salud, haciendo uso de herramientas para el análisis de grandes volúmenes de datos, por tanto, la propuesta que se presenta en este trabajo forma parte de las iniciativas que deben tomarse en la agenda nacional de e-salud.

En la actualidad, una de las problemáticas que afecta algunos centros de salud públicos de la Republica Dominicana como es el Hospital Regional Taiwán es el proceso de registro de los datos médicos de los pacientes y el seguimiento de estos para su tratamiento.

El trabajo de investigación está contiene y desarrolla cinco capítulos, los cuales se detallan a continuación:

En un primer capítulo, se desarrollan los aspectos introductorios del proyecto, con la finalidad de brindar un panorama general del sistema a proponer y lo que se persigue con este.

En el capítulo dos, definen los sistemas de gestión de información, sus características y otros conceptos relacionados. De igual manera se detallan los historiales médicos tradicionales como los historiales médicos electrónicos.

En el capítulo tres, se explicarán los conceptos fundamentales de la analítica de Big Data con el objetivo de ofrecer un entendimiento general que permita identificar eficazmente la importancia de los historiales clínicos en el manejo de los datos de los pacientes, los recursos económicos.

En un cuarto capítulo, conociendo todos los elementos necesarios para la elaboración de la propuesta, se realiza un diagnóstico y caracterización de la situación actual del Hospital Regional Taiwán 19 de Marzo Azua, en materia de historia médica y sistemas.

Finalmente, un quinto capítulo que establece la propuesta de un diseño de un sistema de información para la toma de decisiones médicas y administrativas en dicha institución de salud, aplicando analítica de Big Data sobre los historiales médicos de los pacientes.

CAPÍTULO I: ASPECTOS INTRODUCTORIOS DEL PROYECTO

INTRODUCCIÓN

En el marco de la agenda 2030 propuesta por la Organización de las Naciones Unidas (ONU) se definieron 17 objetivos de los cuales el número 3 “*Garantizar una vida sana y promover el bienestar para todos en todas las edades*” se enfoca en el fortalecimiento de la salud y el bienestar a través del establecimiento de metas y/o políticas que contribuyan con los desafíos del milenio.

En este sentido, se destacan dos metas fundamentales para alcanzar el objetivo: “*Aumentar sustancialmente la financiación de la salud y la contratación, el desarrollo, la capacitación y la retención del personal sanitario en los países en desarrollo, especialmente en los países menos adelantados y los pequeños Estados insulares en desarrollo*”. Asimismo, “*Reforzar la capacidad de todos los países, en particular los países en desarrollo, en materia de alerta temprana, reducción de riesgos y gestión de los riesgos para la salud nacional y mundial*”.

Por esta razón, la presente propuesta en su primer capítulo pretende abordar los aspectos metodológicos que fungirán como base para su desarrollo. Dentro de estos encontramos la selección del tema que corresponde al nombre de esta

Asimismo, el planteamiento y formulación del problema que atraviesa el Hospital Taiwán Regional 19 de Marzo, lo cual ayuda a inferir sobre el panorama en la región sur del país en el sector salud. Así como la sistematización de este, los objetivos tanto generales como específicos y el objeto de estudio en lo que se refiere a la analítica de Big Data. Por último, el campo de acción, el alcance, las metodologías de investigación, como la población y delimitación del tema.

1.1 Selección del tema

Propuesta de diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de Marzo en Azua, República Dominicana, aplicando analítica de Big Data sobre el historial médico de los pacientes.

1.2 Planteamiento del problema

El Hospital Regional Taiwán 19 de Marzo inaugurado en el año 2005, es un centro dotado con los mejores equipamientos, incluidos un tomógrafo y un mamógrafo, fue construido con el propósito de ser el centro principal de salud de la región sur de la República Dominicana. Lamentablemente este no ha cumplido su objetivo.

Desde el año 2009 el Hospital Taiwán localizado en la provincia de Azua, ha presentado inconvenientes y anomalías tanto con los recursos médicos como los recursos económicos, lo cual implica que no puede ofrecer un servicio de calidad a sus pacientes. En este centro, los doctores deben completar los formularios de los pacientes utilizando lápiz y papel, esto representa una situación de riesgo, debido a que existe la posibilidad de que estos sufran accidentes, es decir, que factores como la humedad deterioren el material.

Por otra parte, se ve afectado el proceso de atención al paciente, por desconocimiento de sus datos, lo cual se debe a que no existe una historia médica actualizada. Asimismo, se emplea mucho tiempo en la recepción de los pacientes y proceder con su evaluación. Además, carecen de un sistema de información que les permita tomar decisiones en base a los registros médicos de los pacientes.

Otro de los inconvenientes que afecta a esta institución es el bajo presupuesto que recibe por parte del Gobierno Dominicano, según María Teresa Morel (2019) equivale a 3.4 millones que se reduce a la mitad de los descuentos en medicamentos e insumos que corresponde aproximadamente a RD\$ 3 millones por la venta de servicios al SENASA subsidiado.

Por tanto, con el diseño de un sistema que permita conocer informaciones como: la cantidad de usuarios recibidos, enfermedades, brotes, recursos fármacos, demandas y otros datos, se facilitará el proceso de toma de decisiones en el área médica y administrativas.

1.3 Formulación del problema

¿Cómo el diseño e implementación de una arquitectura de análisis de Big Data permitirá la toma de decisiones médica y administrativas en el Hospital regional Taiwán 19 de Marzo en Azua?

1.4 Sistematización del problema

¿Cuál es el proceso para realizar el registro del historial médico de los pacientes en el centro de salud?

¿Cuál es el proceso para la elaboración del presupuesto del centro de salud?

¿Cuál es el proceso para la toma de decisiones médicas y administrativas que es llevado a cabo por la directiva del Hospital?

¿Es factible desarrollar un sistema de información para la toma de decisiones?

¿Cuáles son las herramientas y técnicas de la Analítica de Big Data propicias para el diseño de un sistema de información para la toma de decisiones?

¿Qué tipo de sistema de gestión de información es necesario para automatizar los historiales médicos de los pacientes?

1.5 Justificación

La presente propuesta tiene como objetivo principal diseñar un sistema de información para la toma de decisiones médicas y administrativas en el hospital Regional Taiwán 19 de Marzo en Azua, R.D. aplicando Analítica de Big Data sobre el historial médico de los pacientes.

El Hospital Regional Taiwán 19 de Marzo de Azua es el principal centro de salud público de la zona sur de la República Dominicana, a pesar de ser un centro que no tiene tantos años de funcionamiento como otros en el país este no ha logrado cumplir a cabalidad la misión para la que fue concebido.

De tal modo, que se ha visto reducir progresivamente su cartera de servicios y a la fecha no ha podido poner en funcionamiento áreas tan vitales como la Unidad de Cuidados Intensivos ni el Banco de Sangre, los cuales están pendientes de equipamiento por parte del Servicio Nacional de Salud (SNS). Tampoco ha logrado mantener en funcionamiento

el sistema de climatización constante, al punto que las áreas de Internamiento no tienen aire acondicionado, como tampoco las salas de Pediatría y Ginecoobstetricia.

Como resultado de las situaciones planteadas, todos los pacientes con traumas craneoencefálicos y otras patologías deben ser trasladados a los hospitales del Gran Santo Domingo, esta situación ocurre en gran parte de los hospitales de la región sur.

Por tanto, con el diseño de este sistema de información aplicando Analítica de Big Data podrán tomarse decisiones médicas y administrativas, que mejorarían considerablemente la situación en la que se encuentre sumergida esta institución y por ende enfrentar los desafíos de salud por los que atraviesa la región sur del país, que están incluidos en la agenda 2030 de las Naciones Unidas. De igual manera, este sistema permitirá cuantificar los recursos económicos referentes a las áreas operativas y de fármacos, es decir, preparar un presupuesto que pueda mantener el funcionamiento óptimo en según los estándares de calidad.

Por otra parte, este sistema ofrecerá proyecciones de las enfermedades que frecuentan en la región. También, las temporadas en que con mayor probabilidad se manifiestan, los lugares de mayor propagación y áreas de brotes.

En este mismo sentido, cabe destacar que los usuarios serán beneficiados con un sistema de calidad que garantice que los servicios ofrecidos por este centro estén de acuerdo con los estándares de salud.

1.6 Objeto de Estudio

El **objeto de estudio** es la analítica de Big Data sobre los historiales médicos electrónicos dentro del área de las Tecnologías de la Información.

1.7 Objetivos

1.7.1 Objetivo General

Diseñar de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de Marzo de Azua, R.D. aplicando Analítica de Big Data sobre el historial médico de los pacientes durante el período enero-abril del 2020.

1.7.2 Objetivos Específicos

- Analizar la información procedente de los historiales médicos electrónicos.
- Evaluar las tecnologías más eficientes para el diseño de una arquitectura de analítica de Big Data.
- Determinar los beneficios del análisis de los datos que son recolectados a través de los historiales médicos electrónicos.
- Identificar los problemas que pueden ser solucionados con la implementación de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de Marzo en Azua, República Dominicana, aplicando Analítica de Big Data sobre el historial médico de los pacientes.
- Examinar el impacto de una propuesta de diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de Marzo en Azua.
- Evaluar la viabilidad técnica, operativa y económica de la propuesta en el sector salud de la República Dominicana.

1.8 Campo de acción

El **campo de acción** es el almacenamiento, administración y análisis de los historiales médicos electrónicos de los pacientes para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de Marzo, República Dominicana.

1.9 Alcance

El sistema propuesto se puede aplicar a todos los centros de salud de la República Dominicana (públicos y privados), no obstante, en esta ocasión se desarrolla una propuesta de investigación de un sistema de información para la toma de decisiones aplicando analítica de Big Data sobre los historiales médicos de los pacientes para el Hospital Regional Taiwán 19 de Marzo en Azua, R.D. durante el período enero-abril del 2020. Desde el punto de vista administrativo se trabajará el presupuesto del centro y la predicción de posibles enfermedades.

1.10 Metodología de investigación

1.10.1 Métodos de investigación

El método de investigación a implementarse para la realización este trabajo de grado es el analítico y deductivo, el primero permitirá realizar un examen profundo de la problemática que nos concierne. De tal manera, que se puedan identificar las posibles causas que generan que el Hospital Taiwán no funcione como se espera. Por su parte, el método deductivo permitirá abordar la situación desde lo general hasta lo particular.

1.10.2 Técnicas de investigación

Los tipos de investigaciones que se realizan son exploratoria, descriptiva y por observación, debido a que el tema se aborda desde el punto de vista científico. Por tanto, se comprende que llevar a cabo una investigación exploratoria garantizará poder indagar lo suficiente y alcanzar el éxito esperado en la misma.

Además, se empleará la observación porque esta permitirá recolectar directamente los datos, aplicando las técnicas adecuadas y sin la manipulación de las posibles variables. Mientras que por su parte la investigación descriptiva ofrecerá una visión más amplia y detallada del objeto de estudio. También, en el proceso investigativo se tomarán en cuenta informes, trabajos de investigación y revistas científicas obtenidas a través de internet y bibliotecas.

1.10.3 Técnicas de recolección

Las técnicas de recolección que se aplican son las encuestas y entrevistas, estas últimas se realizarán a profesionales tanto en el área médica, administrativa como a los expertos en Big Data y cuestionarios que permitirán la recolección de información desde los diferentes puntos de vista.

Como fuentes de carácter primario:

- ❖ Libros de texto
- ❖ Estudios de investigaciones relacionadas al tema
- ❖ Tesis
- ❖ Ensayos
- ❖ Revistas científicas
- ❖ Sitios Web

Como fuentes de carácter secundario:

- ❖ Entrevistas
- ❖ Encuestas
- ❖ Enciclopedias
- ❖ Artículos de revistas y periódicos
- ❖ Diccionarios

1.11 Población de la investigación

La población o universo de la investigación incluye especialistas de la salud, personal administrativo y pacientes del Hospital Taiwán.

El hospital cuenta con un personal integrado por 812 empleados que están distribuidos en operaciones médicas como administrativas. De esta cantidad el 23% representa a los especialistas de salud, y el 77% el personal administrativo. Los doctores se distribuyen durante toda la semana en diversos turnos, con el fin de atender a los pacientes que usualmente son 400 por día. Por tanto, se estima que alrededor de 725 personas frecuentan cada día el Taiwán, lo que haría un total de 65,250 durante el período de investigación.

1.11.1 Determinación de la muestra

La muestra está conformada por las personas de la población o universo capaces de entender la problemática y emitir su juicio de valor sobre el mismo.

Según (feedbacknetworks, s.f.), La muestra es calculada con la formula siguiente:

$$n = \frac{k^2 * p * q * N}{(e^2 * (N - 1)) + k^2 * p * q}$$

Donde:

N:es el tamaño de la población o universo (número total de posibles encuestados).

k:es una constante que depende del nivel de confianza que asignemos. El nivel de confianza indica la probabilidad de que los resultados de esta investigación sean ciertos: un 95,5% de confianza es lo mismo que decir que se puede equivocar con una probabilidad del 4,5%. Los valores de k se obtienen de la tabla de la distribución normal estándar N (0,1).

Los valores de k más utilizados y sus niveles de confianza son:

Tabla 1- Valores de nivel de confianza

Valor de <i>k</i>	1.28	1.65	1.69	1.75	1.81	1.88	1.96
Nivel de confianza	80%	90%	91%	92%	93%	94%	95%

Fuente: (feedbacknetworks, s.f.)

p: es la proporción de individuos que poseen en la población la característica de estudio. Este dato es generalmente desconocido y se suele suponer que $p=q=0.5$ que es la opción más segura.

q: es la proporción de individuos que no poseen esa característica, es decir, es $1-p$.

n: es el tamaño de la muestra (número de encuestas que vamos a hacer).

$$n = \frac{k^2 * p * q * N}{(e^2 * (N - 1)) + k^2 * p * q}$$

$N = 725$ personas por día

$k = 1.96$ es igual a un 95% de confianza

$p = q = 0.5$

$n = \frac{(1.96)^2 * 0.5 * 0.5 * 725}{(e^2 * (N-1)) + ((1.96)^2 * 0.5 * 0.5)}$

$n = \frac{3.8416 * 0.5 * 0.5 * 725}{(2.71828^2 * (725-1)) + (1.96)^2 * 0.5 * 0.5}$

$n = \frac{3.8416 * 0.5 * 0.5 * 725}{(2.71828^2 * (725-1)) + (1.96)^2 * 0.5 * 0.5}$

$n = \frac{696.29}{5,250.63}$

$n = 696.29/5,250.63$

$n = 132$

Esta investigación toma como muestra 132 personas, a ser evaluados y encuestados, que frecuentan las instalaciones del hospital.

1.12 Delimitación de tema

El diseño del sistema propuesto se estará realizando en el periodo enero-abril del 2020 en el Hospital Regional Taiwán 19 de Marzo localizado en la provincia de Azua, República Dominicana.

CONCLUSIÓN

En este capítulo se definieron todos los aspectos metodológicos que sustentan la propuesta que será trabajada, de manera clara, concisa y precisa. Esta primera parte se enfocó en la delimitación del problema que se persigue resolver con el diseño de un sistema de información utilizando la Analítica de Big Data que es el objeto de estudio.

Otros aspectos tomados en cuenta fueron el objetivo general de la investigación que se pretende alcanzar en el proceso de estudio. Asimismo, el conjunto de objetivos específicos que complementan a este, abarcando de forma individual los diversos subtemas que sustentan de manera conjunta el capítulo. Además, el campo de acción y el alcance del proyecto, como el diseño de la metodología donde se detallan los diferentes tipos de investigaciones a abordar, los tipos de métodos y las técnicas que empleadas para la recolección de la información.

Por último, la población de la investigación que mediante la fórmula expuesta anteriormente se identificaron que deben ser encuestadas 132 personas de las que frecuentan el hospital, como la delimitación del tema que son las instalaciones del Hospital.

**CAPÍTULO II: SISTEMA DE GESTIÓN DE
INFORMACIÓN E HISTORIALES MÉDICOS
ELECTRÓNICOS**

INTRODUCCIÓN

En la República Dominicana es común que los pacientes conozcan lo que coloquialmente se llama ‘tarjeta amarilla’, que consiste en una tarjeta generalmente de color amarillo donde se registran las visitas médicas, medicamentos aplicados a los pacientes y otros indicadores de salud del paciente. Este método provoca que cada vez que se realiza una cita médica o la persona se somete a un procedimiento, de manera previa, se le cuestione por la misma información, esto con el objetivo de saber todos los procesos y procedimientos a los cuales se ha sometido un individuo.

Observando esta realidad y el concepto de registro personal conocido como historial médico, no es igual como era veinte años atrás, pues no existían ordenadores, tampoco los sistemas. Con el desarrollo e incorporación de nuevas tecnologías en la actualidad existe lo que se conoce como historial médico electrónico, un documento que lleva un registro de todos los aspectos, procesos, procedimientos, consultas, revisiones relativos a la salud de un individuo de manera digital. Esto comprende cualquier factor que pudiera ser catalogado como registrable y relevante para trazar la condición de un paciente antes, durante o después de cualquier patología.

En el capítulo II, Sistema de Gestión de Información e Historiales Médicos Electrónicos se estarán desarrollando tres ideas fundamentales para la elaboración de la propuesta: Sistema de Gestión de Información (SGI), en el cual tomará en cuenta su concepto, su historia, elementos y características que lo componen. También su tipología, el ciclo de vida de este y los estándares internacionales de mismo.

En segundo lugar, se abordan los historiales médicos, concepto, funciones características y su clasificación y la importancia de estos.

Por último, se exponen los Historiales Médicos Electrónicos, su evolución componentes, características, beneficios, servicios, documentación clínica y se desarrolla el concepto de E-Salud.

2.1 Sistema de gestión de información

La información es un conjunto de datos ordenados que construyen un mensaje, este puede ser utilizado tanto para la resolución de problemas, como para la toma de decisiones. La misma, es materia prima en el área empresarial, una vez generada y organizada puede adquirir un gran valor y en esa vía es vendida y comprada para ser explotada con el fin de obtener beneficios y mejoras a través de ella. Existen un conjunto de procesos para el control, la gestión y la determinación del ciclo de vida de la información, desde el momento en que es capturada, almacenada procesada y depurada. Este proceso en general se denomina gestión de la información.

Un SGI funciona bajo un conjunto de acciones que se efectúan de forma sistemática sobre la materia prima para cumplir con los lineamientos y expectativas del cliente, empresa y colaboradoras al menor costo posible. Estos mantienen un enfoque en la gestión de la calidad, inocuidad, seguridad de la información, entre otros.

Las funciones de la Gestión Información según González (2007) son:

1. Determinar las necesidades de información en correspondencia a sus funciones y actividades.
2. Mejora de los canales de comunicación y acceso a la información.
3. Mejora de los procesos informativos.
4. Empleo eficiente de los recursos.

2.1.1 Historia

Los sistemas de información han ido evolucionando durante los últimos años hasta constituir los denominados sistemas de información estratégicos. En principio, los sistemas de información eran considerados como un instrumento simplificador de las distintas actividades de la empresa, una herramienta con la cual se facilitaban los trámites y se reducía la burocracia. Su finalidad era básicamente llevar la contabilidad y el procesamiento de los documentos a nivel operativo.

Posteriormente, el desarrollo de la informática y las telecomunicaciones permitieron incrementar la eficacia en la realización de las tareas, ahorrar tiempo en el desarrollo de las actividades y almacenar la mayor cantidad de información en el menor espacio posible, lo cual aumentó en las organizaciones el interés en los sistemas de información. Con el transcurrir del tiempo las empresas fueron observando como las tecnologías y sistemas de información permitían a la empresa obtener mejores resultados que sus competidores, constituyéndose por sí mismas como una fuente de ventaja competitiva y una poderosa arma que permitía diferenciarse de sus competidores y obtener mejores

resultados que estos. De este modo, los sistemas de información se constituyeron como una de las herramientas estratégicas de la empresa, que han de considerarse siempre en todo proceso de planificación empresarial. (Hernández, 2013, p. 4)

2.1.2 Elementos

Elementos de un sistema de información:

(1) **Base de Datos:** Es donde se almacena toda la información que se requiere para la toma de decisiones. La información se organiza en registros específicos e identificables.

(2) **Transacciones:** Corresponde a todos los elementos de interfaz que permiten al usuario: consultar, agregar, modificar o eliminar un registro específico de Información.

(3) **Informes:** Corresponden a todos los elementos de interfaz mediante los cuales el usuario puede obtener uno o más registros y/o información de tipo estadístico (contar, sumar) de acuerdo con criterios de búsqueda y selección definidos.

(4) **Procesos:** Corresponden a todos aquellos elementos que, de acuerdo con una lógica predefinida, obtienen información de la base de datos y generan nuevos registros de información.

(5) **Usuario:** Identifica a todas las personas que interactúan con el sistema, esto incluye el máximo nivel ejecutivo que recibe los informes de estadísticas procesadas.

(6) **Procedimientos Administrativos:** Corresponde al conjunto de reglas y políticas de la organización, que rigen el comportamiento de los usuarios frente al sistema. Particularmente, debieran asegurar que nunca, en ninguna circunstancia un usuario tenga acceso directo a la Base de Datos.

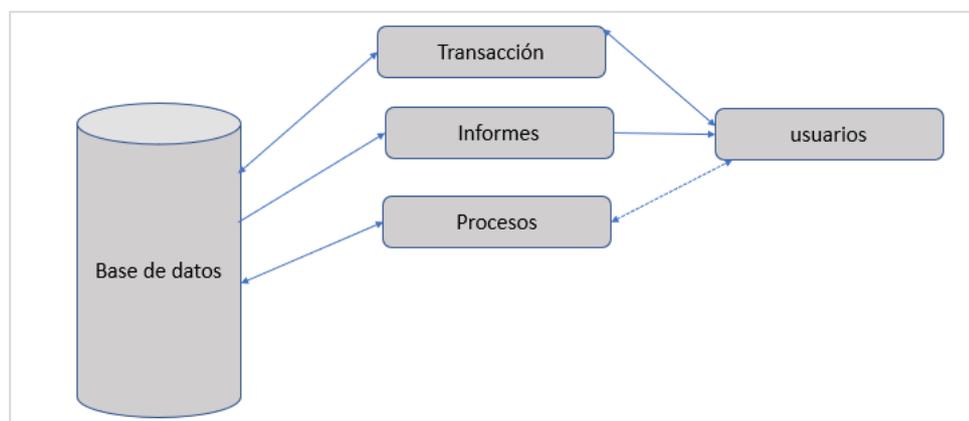


Figura 1- Elementos de un sistema de información.

Fuente: (Peña, 2006)

2.1.3 Características

La característica principal de un sistema de gestión de información es su capacidad para almacenar datos y facilitar información para ser recuperada por los usuarios de este. El tipo de base de datos utilizada determina cómo el sistema de gestión de información responde a las peticiones o consultas de información. Los sistemas que utilizan una base de datos relacional almacenan los datos en tablas separadas en vez de una tabla grande.

Las características de un sistema de gestión de información son:

Informes: otra característica importante de un sistema de gestión de información viene en forma de informes. Un SGI es sólo bueno en la manera como genera informes. La capacidad de producir información que ayuda en el proceso de toma de decisiones es un atributo clave para este tipo de sistema. La mayoría de SGI proporciona varias plantillas de informes mientras que otras ofrecen la posibilidad de crear informes específicos y guardarlos como una plantilla para que otros los utilicen. (IBM, 2012)

Acceso abierto: esta característica permite el acceso abierto a la arquitectura del sistema lo que facilita a la empresa cumplir más fácilmente con las regulaciones externas y requisitos internos. El acceso abierto significa que la empresa puede integrar más fácilmente el SGI con los sistemas existentes. Esta habilidad reduce la necesidad de que la gente realice cambios de servicio interno y reduce los gastos de mantenimiento debido a los recursos internos pueden gestionar el mantenimiento del sistema. (IBM, 2012)

Integración: permite que los SGI se integran con los sistemas existentes de la empresa y estos a otros que con frecuencia crean muchos desafíos para los recursos internos y externos. (IBM, 2012)

Escalabilidad: Debido a que no todas las empresas requieren la oferta completa de algunos SGI, la escalabilidad se convierte en una contraprestación clave. Las empresas más pequeñas pueden requerir una versión reducida un SGI al momento, pero dentro de unos años necesitan características adicionales y funciones mayores de administración de bases de datos. (IBM, 2012)

2.1.4 Tipologías

Desde un punto de vista empresarial u organizativo, los sistemas de información pueden clasificarse según Estala (2020):

- **Sistemas de Procesamiento de Transacciones (TPS).** También conocidos como sistemas de gestión operativa, recopilan la información pertinente a las transacciones de la organización, es decir, de su funcionamiento.

- **Sistemas de Información Ejecutiva (EIS).** Monitoriza las variables gerenciales de un área específica de la organización, a partir de la información interna de esta.
- **Sistemas de Información Gerencial (MIS).** Contemplan la información general de la organización y la comprenden como un todo.
- **Sistemas de soporte de decisiones (DSS).** Orientados al procesamiento de información intra y extra organizacional, para el apoyo en la empresa.

2.1.5 Ciclo de vida

El ciclo de vida de un SGI tiene como propósito brindar las pautas para el desarrollo e implementación de sistemas. De igual manera, este permite a los administradores de proyectos reunir los requisitos de usuarios, mediante la planificación de lo que sería el desarrollo y la puesta en marcha del sistema completado dentro del tiempo y el presupuesto establecido.

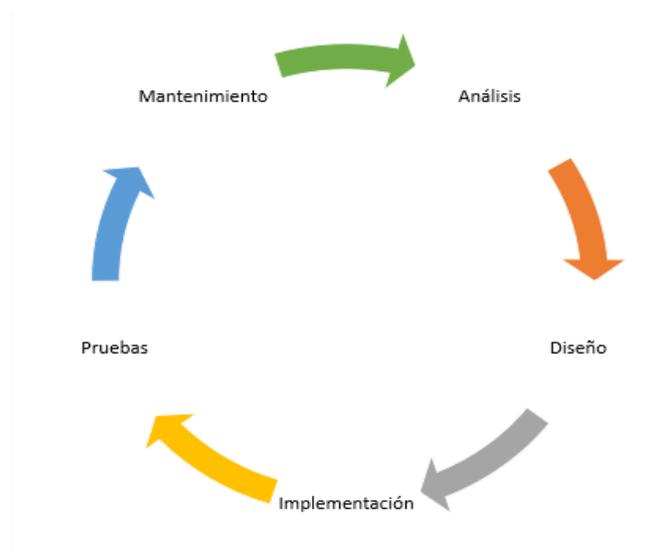


Figura 2- Ciclo de Vida de Desarrollo de un SGI

Fuente: (Autores)

Las fases de este ciclo son las siguientes:

- **Planificación Conceptual:** en esta fase los interesados en el desarrollo del sistema definen la propuesta y el alcance de este. Asimismo, se determinan los aspectos fundamentales y limitantes como presupuestos, recursos y tiempo.
- **Análisis de Requisitos:** el personal de tecnología de la información define los requisitos no funcionales según las necesidades de los usuarios. Luego se verifica y valida con los usuarios finales si los mismos responden a los requerimientos realizados.

- **Diseño:** se elabora un diseño desde el punto de vista técnico que describa las especificaciones de todas las funcionalidades para la implementación del sistema.
- **Desarrollo y Pruebas:** el sistema se encuentra en desarrollo y se pueden hacer las pruebas y validaciones necesarias por parte del personal de calidad para que verifiquen que se cumplan los requisitos previamente definidos
- **Operaciones y Mantenimiento:** el sistema ya está en funcionamiento, está en un ambiente de producción donde ya el usuario está haciendo uso de este.

2.1.6 Estándares internacionales

Algunos estándares que comprenden este sistema, incluidos los lineamientos del Consejo Nacional de Acreditación (CNA), para la acreditación institucional de gestión son:

- **ISO 30301:** Sistema de Gestión para los Documentos, establece los requisitos para la implantación de un Sistema de Gestión de los documentos.
- **ISO 14001:** Sistema de Gestión Ambiental, consigue que las empresas puedan demostrar que son responsables y están comprometidas con la protección del medio ambiente. Anteriormente hemos mencionado que lo consiguen a través de la gestión de los riesgos medioambientales que puedan surgir del desarrollo de la actividad empresarial.
- **ISO 9001:** Sistema de Gestión de la Calidad, reconocida internacionalmente, esta norma demuestra que se deben emplear controles y procedimientos adecuados en producción para proporcionar un servicio de calidad de TI coherente y que el mismo sea a un costo efectivo.
- **ISME:** Información de Gestión de Seguridad Modelo de Madurez, está basado en un conjunto de estándares como ISO 2000, CMMI e ITIL. Se puede utilizar como base de una ISO 9001 y se concentra en procesos, incluyendo sus las métricas.
- **ISO 27001:** Sistema de Gestión de la Seguridad de la información, es una norma internacional que permite el aseguramiento, la confidencialidad e integridad de los datos y de la información, así como de los sistemas que la procesan. Esta permite a las organizaciones la evaluación del riesgo y la aplicación de los controles necesarios para mitigarlos o eliminarlos.

2.2 Historiales Médicos

Antes de hablar de Historial médico, es necesario introducir el concepto de *registro personal de salud* que no es más que una recopilación de datos sobre la salud, de manera que, si un paciente posee un registro de vacunas o bien una carpeta con documentación

relativa a los aspectos médicos, ya se tiene un registro de salud básico (Mayo Clinic, 2018).

Los historiales médicos están constituidos de documentos médicos-legales que se recopilan mediante el trato directo del paciente y el médico a través y la colaboración que estos hacen en las diferentes visitas que un paciente hace durante su vida. En el expediente médico se detallan diversas enfermedades padecidas por el paciente y la medicación que se le ha otorgado. Este antecedente es el que hace la función de constituir un plan de evaluación, terapéutico y asistencial del paciente.

El historial médico consiste en la agrupación de los documentos correspondientes al desarrollo asistencial de cada paciente atendido, con el reconocimiento de los médicos y demás profesionales que han participado en el proceso, con la finalidad de adquirir la máxima incorporación posible en la credencial clínica de cada paciente, al menos, en el entorno de cada promedio.

Uno de los roles clave que desempeñan los historiales médicos es el de servir de objeto para facilitar todo lo relacionado a la asistencia en el sector de salud. El levantamiento de información inicia desde que el paciente solicita la prestación de algún servicio sanitario hasta que el médico a cargo inicia el historial y le da seguimiento al mismo a lo largo del tiempo. Los historiales médicos son indispensables dentro del campo de la medicina debido a que esta es la que permite a este tener una visión global y completa del paciente para posteriormente poder asistir lo de la mejor manera. (Pimentel G, Reyes E., 2018)

2.2.1 Funciones

La función de más peso por la que es justificable un historial clínico es la asistencial ya que permite dar continuidad a los servicios de atención brindados al paciente. Otras funcionalidades son: instrucción, realización de estudios de investigación y epidemiología, apreciación de la calidad asistencial, la planificación y gestión sanitaria, así como también su empleo en casos reglamentarios en aquellas circunstancias jurídicas en que se solicita.

Para satisfacer estas funciones, el historial médico debe ser elaborado con rigurosidad, puntualizando todos los detalles necesarios y suficientes que demuestren el diagnóstico y el tratamiento con una escritura comprensible.

La información adquirida y organizada en el historial médico es una constancia de datos vital para el incremento de las actividades profesionales de los médicos:

- ✓ *Docencia:* permite aprender tanto de los aciertos como de la falsedad de las labores desarrolladas.
- ✓ *Medicolegal:* es un documento lícito, que es empleado asiduamente para dictaminar la relación médico-paciente.

- ✓ *Investigación:* a partir de la información que aporta el historial médico se puede plantear un proceso de preguntas para la investigación sanitaria, con la finalidad de localizar respuestas científicas razonables.
- ✓ *Clínica o asistencial:* da coherencia a la producción y a la utilización continua en la relación médico-paciente.
- ✓ *Mejora continua de la calidad:* es estimado por las normas deontológicas y legítimas como un derecho del paciente, elaborado del derecho a un auxilio médico de calidad; su investigación y evaluación accede a establecer el nivel de índole asistencial prestada.
- ✓ *Epidemiología:* la información aglomerada en el historial médico, se puede extrapolar perfiles e información sanitaria local, nacional e internacional.
- ✓ *Gestión y administración:* el historial médico es el elemento esencial para el control y gestión de los servicios médicos de los establecimientos sanitarios (Pimentel G, Reyes E., 2018).

2.2.2 Características

Los historiales médicos presentan una serie de características entre las que se encuentran:

Confidencialidad: el secreto médico, la confidencialidad e intimidad de la historia clínica, son tres cuestiones que se implican recíprocamente y se relacionan.

Disponibilidad: aunque debe preservarse la confidencialidad y la intimidad de los datos debe ser así mismo un documento disponible, facilitándose en los casos legalmente contemplados, su acceso y disponibilidad.

Única: la historia clínica debe ser única para cada paciente por la importancia de cara a los beneficios que ocasiona al paciente la labor asistencial y la gestión.

Legible: una historia clínica desordenada y difícilmente legible perjudica a todos, a los médicos, porque dificulta su labor asistencial y a los pacientes por los errores que pueden derivarse de una inadecuada interpretación de los datos contenidos en la historia clínica.

Veracidad: la historia clínica debe caracterizarse por ser un documento veraz, constituyendo un derecho del usuario. El no cumplir tal requisito puede incurrirse en un delito tipificado en el actual Código Penal, corresponde a falsedad documental.

Sincronización en los registros: la historia clínica debe realizarse de forma simultánea y coetánea con la asistencia prestada al paciente.

Completa: debe contener datos suficientes y sintéticos sobre la patología del paciente, debiéndose reflejar en ella todas las fases médico-legales que comprenden todo acto clínico-asistencial. Así mismo, debe contener todos los documentos integrantes de la

Historial Médico Procesal (Narrativo o Cronológico)

Este modelo es utilizado en la primera intervención con el paciente, la entrevista debe ir enfocada a responder una serie de preguntas donde se indagan las siguientes informaciones:

I) Datos demográficos: Nombre, sexo, edad, estado civil, nombre de pareja, años de casado, cantidad de hijos, ciudad donde reside, enfermedades físicas o de salud mental, abuso de sustancias, etc.

II) Problemática: se indaga con el paciente el motivo por el cual se está consultando o presentándose en el centro médico (síntomas, personas involucradas en el problema, medicamentos indicados, etc.) se escribe tal y como lo describe el paciente.

III) Acciones a tomar: esta es el área donde se detallan los procedimientos seguidos, se concluye la labor, se emiten referidos y se orienta al paciente lo que se estará llevando a cabo.

IV) Peligrosidad: se señala como se encontraba el paciente al momento del ingreso en la unidad de salud, orientado en el tiempo, lugar y persona, pensamientos suicidas-homicidas, disturbios perceptuales, etc.

Historial Resumido (No Detallado)

Es utilizado cuando se ha efectuado el historial médico narrativo, que cumple con las siguientes funciones:

1. Se le estará dando continuidad al suceso.
2. Se pide un resumen del caso para analizarlo con el personal multidisciplinario.
3. Se realizarán una serie de cambios al plan del historial resumido.
4. La información recolectada se mantendrá almacenada para cuando el personal del recinto la requiera.

Historial Codificado (SOAP)

Este historial es utilizado por especialistas de la conducta, consta de los siguientes incisos:

1. Notas del progreso médico.
2. Evaluación psiquiátrica.
3. Evaluación psicológica.

Como muy bien detalla (Pimentel G, Reyes E., 2018), el significado de cada una de las siglas que componen el nombre de este historial:

(S) - Subjetivo: da referencia a los datos subjetivos o sentimentales del paciente.

(O) - Objetivo: registro de datos reales, objetivos y verificables.

(A) - Análisis: es la apreciación o el análisis profesional.

(P) – Plan: Es necesario establecer un alternativa o plan de tratamiento que debe seguir el paciente (prescripción de medicamentos, procedimientos, cirugía).

2.2.4 Importancia

Haciendo una observación simple se puede destacar que un historial no sólo son datos, debido a que este representa la trayectoria médica de una persona lo cual incluye, análisis, consultas, procesos e intervenciones clínicas, afecciones y patologías que tiene un individuo desde el momento que es ingresado a una institución del sector salud, por tanto la importancia de estos radica en que tener información precisa, organizada y disponible de una persona puede incluso salvarle la vida, hacerse predicciones y por ende estudios científicos.

Si se resumieran algunos de los aspectos más relevantes de la misma se podría decir que:

- Son documentos que ayudan a los médicos a informarse sobre el estado de salud de un paciente.
- Permiten garantizar el adecuado diagnóstico de un paciente en todo el proceso clínico.
- Ayudan a los especialistas a identificar los tratamientos que ha recibido y se aplicarán adecuados para un paciente.

2.3 Historiales Médicos Electrónicos (HME)

Los historiales médicos electrónicos son un conjunto de informaciones de carácter médico de una persona, que son recolectados, almacenado y manipulados digitalmente mediante un sistema electrónico diseñado para tales fines. Estos permiten identificar la trayectoria de un paciente y sirven de base para prevenir, predecir o disminuir futuras enfermedades.

Este avance en la medicina ha revolucionado la forma en la cual los pacientes son atendidos y permiten que los profesionales ofrezcan un mejor servicio, debido a que tienen la información disponible, segura, íntegra y asequible desde cualquier dispositivo que contengan un sistema de gestión de información médica.

Figura 4- Ejemplo de un historial médico electrónico

Fuente: (Reyes, 2016)

A pesar de que los mismos ofrecen una visualización ordenada sistemáticamente del registro de las informaciones, la parte más valiosa para los profesionales es que los HME sirven para mejorar los conocimientos relacionados a la prestación de servicios de salud mediante la integración de protocolos de acción, guías médicas, bases bibliográficas y su uso como herramienta para ayudar a tomar decisiones. (Pimentel y Reyes. 2016, p.09)

2.3.1 Evolución de los Historiales Médicos Electrónicos

Tradicionalmente los historiales médicos eran plasmados en un papel, lo que representaba un riesgo en cuanto a disponibilidad y accesibilidad se refería. Asimismo, la integridad de los datos la cual es considerada un factor clave debido a que la importancia que radica en estos, debido a que son datos personales con un alto nivel de criticidad. A esto se le suma que, por los diversos procedimientos que un paciente puede estar involucrado, es decir, emergencia, ambulatorio, internación u otro nivel de atención, su información debe estar disponible en todos los lugares al mismo tiempo. De igual manera, la confidencialidad de este tipo de información pueda afectarse debido a que se debe permitir el acceso a otros profesionales.

En la década de los sesenta con el desarrollo de las computadoras surgió un nuevo método para el almacenamiento, recuperación y visualización de la información contenida en los historiales médicos, sustituyendo el papel por un formato electrónico. Los historiales médicos electrónicos iniciaron como parte exclusiva del ambiente académico y experimental y consistían en sistemas de información centrado con el fin de dar soporte a los procesos administrativos. Con el avance tecnológico y el surgimiento de las

computadoras portátiles dio inicio al desarrollo enfocado a las áreas clínicas y de exámenes complementarios, pero sin centralización ya que cada sistema era independiente.

A raíz de la necesidad de interconectar los sistemas para integrar la información de los departamentos mediante el uso de un repositorio de datos clínico (Clinical Data Repository o CDR), se facilitó la creación de sistemas de información clínicos basados en componentes. Carnicero y Fernández (2012), afirman que: “Una de las premisas de estos nuevos sistemas fue respetar los procesos asistenciales y tener al acto médico como eje central de su modelo de información”. Desde entonces, la descentralización de la atención médica en redes asistenciales generó nuevamente la necesidad de conectar múltiples sistemas, más allá de los muros de una institución, que posibiliten la fluida comunicación de la información clínica.

Para el siglo XXI, los historiales médicos electrónicos cobran mucha más fuerza y se integran a lo que son los servicios electrónicos de la salud. En esta etapa comienza su integración creándose repositorios que permitían la movilidad, seguridad y más amplia infraestructura a nivel del registro de los pacientes. Dentro de los países que poseen la mayor cantidad de hospitales con sistemas que sirven para la gestión de los HME se encuentra Noruega, este país tiene el 96% de sus hospitales con esta propiedad. (Laerum, 2004)

Con el tiempo, como una nueva estrategia para los historiales médicos electrónicos, se han agregado algunas funcionalidades y datos correspondientes a registros del estado del paciente como son: peso, nivel de colesterol, presión arterial, medicamentos que debe tomar el paciente, cantidad de alcohol que toma y la fecha y hora de las citas médicas que el mismo ha tenido. (National Health Service, 2009)

En el 2001, Canadá crea una estrategia llamada Inforway con la finalidad de crear un repositorio contenido de la información del sector de salud y que la misma se gestione de forma centralizada para que pueda ser accedida por los médicos, pacientes, farmacias y laboratorios. (Canada Infoway, 2001)

Por otro lado, en México, el IMSS desde el 2004 ha impulsado los avances tanto en los HME como también en asuntos que impactan la informática en la salud, incorporando conceptos que integran teleconsulta y telediagnóstico mediante medios de transmisión a grandes velocidades que sirven como herramientas de apoyo para el manejo de indicadores de salud. Como se puede ver en el transcurso de las últimas décadas, la utilización de los sistemas de gestión para los historiales médicos electrónicos se ha vuelto mucho más popular en los centros de salud en todo el mundo, brindando una información mucho más detallada y organizada al personal de salud y médicos, lo cual ha facilitado la interacción en tiempo real entre las diversas entidades de área de salud.

2.3.2 Componentes de los HME

Un sistema que tiene como tarea fundamental la gestión de los historiales médicos electrónicos puede tener distintas arquitecturas a nivel informático, pero en general tienen los siguientes componentes:

- ✓ Aplicación médica
- ✓ Almacén de datos
- ✓ Aplicaciones complementarias
- ✓ Software base
- ✓ Plataforma de interoperabilidad e información
- ✓ Equipo de computo
- ✓ Periféricos
- ✓ Conectividad
- ✓ Servidores

2.3.3 Características de HME

Los historiales médicos electrónicos son un elemento fundamental para la gestión de información en los centros de salud, debido a las facilidades que ofrece en cuanto al almacenamiento y recolección de los datos se refiere.

A continuación, se describen brevemente algunas cualidades importantes de los HME:

Disponibilidad: estos estarán a disposición de los profesionales de la salud y las personas correspondientes los 365 días del año.

Seguro: los datos estarán almacenados de forma digital, lo que minimizara el riesgo de que se corrompan, y que se extravíen.

Integridad: debido a que la existencia de controles de acceso que posee un sistema de información permite que la información se mantenga en su estado original y no sea alterado por personas no autorizada.

Confidencialidad: esto permite que solo las personas indicadas tengan acceso a la información.

Accesibilidad: a diferencia de los historiales médicos tradicionales que se almacenaban en archivero y dificultaba su búsqueda, los HME están en un sistema que desde un dispositivo electrónico podría acceder a este a través de ciertos parámetros.

2.3.4 Beneficios de HME

Todo avance tecnológico tiene beneficios que brindar y los HME no son la excepción. Según la Organización Panamericana de la Salud dentro de los beneficios de los HME se destacan:

- Con la ayuda de la tecnología adecuada se puede asegurar la identificación precisa de los pacientes.
- Salvaguardar la seguridad y confidencialidad de los datos personales del paciente.
- Facilitan la integración con sistemas administrativos lo que puede agilizar el agendamiento y programación de consultas médicas.
- Si están correctamente implementados pueden ayudar a los profesionales a que dispongan menor tiempo dedicado a tareas administrativas.
- Permite compartir información clínica entre profesionales sanitarios tanto de la unidad como de diferentes hospitales.
- Mejoran la seguridad jurídica del profesional por el registro electrónico de la información del paciente.
- Incrementan la calidad asistencial al disponer de información de la paciente inmediata y en tiempo real y proporcionan indicadores de salud en tiempo real, para la toma de decisiones.
- Mejoran la eficiencia del proceso asistencial al evitar desperdicio de medicamentos y estudios de diagnóstico innecesarios.
- Mejoran el proceso de toma de decisiones en todos los niveles.
- Disminuyen los errores por ilegibilidad en las recetas de los médicos.
- Disminuyen del gasto en papel y ayudan a conservar el medio ambiente.
- Facilitan la vigilancia epidemiológica.

2.3.5 Servicios de Admisiones y Documentación Clínica

En la actualidad es indispensable para los centros de salud disponer de un sistema informático que permita gestionar todos los procesos clínicos y administrativos de este. Dicho tipo de sistema es conocido como HIS que significa ‘Hospital Information System’ y está compuesto por uno o varios elementos de software y subsistemas. Uno de los componentes esenciales de un HIS es el proceso administrativo que consiste en la identificación del paciente, la solicitud de prestación de servicios, la cita y su realización.

De igual manera, están los servicios de admisión que son todos los que están relacionado al proceso de recibimiento de los pacientes, es decir, desde que llegan al centro, son atendidos y son dados de alta. Este consiste en completar el formulario de admisión con sus datos personales, estado del paciente y otros. Luego se procede a realizar las evaluaciones de lugar y dar seguimiento al estado de este hasta que pueda darse de alta. Por su parte, la documentación clínica es toda la información que se obtiene desde el momento que la persona llega al centro de salud hasta que está en condiciones apta para retornar a su hogar.

2.3.6 E-Salud

El término e-salud fue adoptado por la Organización Panamericana de la Salud (OPS) y la Organización Mundial de la Salud, a partir de los principios de la Cumbre Mundial sobre la Sociedad de la Información, celebrada en 2003 en Ginebra, Suiza (UIT, 2017). En el documento CD51/13 del primero de agosto del 2011, la OPS y la OMS establecieron su “Estrategia y plan de acción sobre e-Salud” en la que definen la ciber salud o e-Salud como “el apoyo que la utilización costo eficaz y segura de las TIC ofrece a la salud y a los ámbitos relacionados con ella, con inclusión de los servicios de atención de salud, la vigilancia y la documentación sanitarias, así como la educación, los conocimientos y las investigaciones en materia de salud”. (Oficina Regional de la UIT para las Américas, 2018)

La e-salud consiste en la incorporación de la tecnología en los procesos que se realizan en el sector salud, con el propósito mejorar la experiencia del paciente como el trabajo de los profesionales de la salud. De igual manera, prevenir, predecir, diagnosticar, tratar, dar seguimiento al conjunto de factores que inciden en el bienestar de las personas.

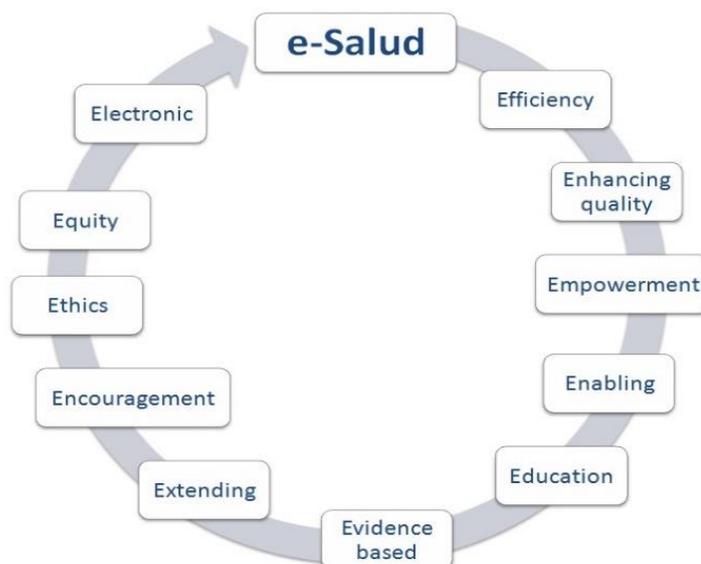


Figura 5- Las 11 E de la E-salud

Fuente: (Eysenbach,2015)

CONCLUSIÓN

En la primera sección de este capítulo se desarrolló a nivel conceptual los sistemas de gestión de información y aspectos relevantes como su historia, sus elementos principales, sus características, sus tipologías. De igual manera, fue abarcado el ciclo de vida y los estándares internacionales que se han desarrollado con el propósito de que estos cumplan ciertos requisitos.

Por otra parte, se habló sobre los historiales médicos que es un conjunto de documentos de carácter medicolegal que describen el estado de salud de una paciente. Dentro de los temas tratados referente a estos se destacan: sus funciones, características y clasificación.

Como parte de la presenta propuesta se ha desarrollado un apartado para definir los Historiales Médicos Electrónicos (HME) que han sido los sucesores de los historiales médicos tradicionales. Se describe su evolución, sus componentes, características, beneficios. Asimismo, se detallaron servicios de admisiones y documentación clínica y por último un concepto que a pesar de ser nuevo influye día a día en las labores médicas y es la denominada e-salud que consiste en la aplicación de la tecnología en el campo de la salud.

CAPÍTULO III: ANALÍTICA DE BIG DATA

INTRODUCCIÓN

Hoy en día es una necesidad que los profesionales de la informática y específicamente aquellos cuya área de especialidad guarda relación con los datos, comprendan la importancia que radica en estos. De igual manera, los efectos positivos que ha dejado la gestión de estos en el siglo XXI para cualquier tipo de entidad. A raíz de lo anterior surge la importancia que los profesionales y las entidades conozcan con exactitud conceptos como, Ciencia de Datos, Análisis de Datos, Big Data, Análisis de Big Data, Inteligencia de Negocios, Estadística, Maching Learning y no menos importante el concepto de Deep Learning.

Actualmente autores de renombre en el mundo de los negocios y la tecnología defienden la idea que los datos, que antes recibían escasa importancia, se constituyen como el principal activo de las empresas. De ahí que estas destinen fondos a la adquisición de recursos económicos y no económicos con el objetivo asegurar, manejar y obtener beneficios de esos datos.

Por tanto, no es difícil darse cuenta de que nos encontramos en la era de la información, razón principal por la cual todo dato es importante por sencillo o complejo que sea, esto hace que organizaciones obtengan conocimiento lo cual se traduce en beneficios.

En este capítulo se describen los conceptos fundamentales que sustentaran la teoría de la presente investigación, concentrándose en cuatro aspectos esenciales que son la Ciencia de Datos, Big Data, el Análisis de Datos y la Analítica de Big Data.

3.1 Ciencia de los datos

Existen muchas definiciones y puntos de vista sobre la ciencia de los datos, debido a que autores están de acuerdo con la idea de extracción de conocimiento, mientras que otros difieren. A continuación, se detallan algunas definiciones para afianzar los conocimientos con relación a esta:

La ciencia de los datos se trata del estudio de la extracción generalizada de conocimiento a partir de información. Así, como técnicas y teorías conocidas en diversas áreas como la Matemática, la Estadística y las Tecnologías de la Información. De igual manera, procesamiento de señales, modelos probabilísticos, machine learning, aprendizaje estadístico, programación, ingeniería de datos, reconocimiento de patrones, visualización, data warehousing, y computación de altas prestaciones.

La ciencia de los datos es un área de trabajo interdisciplinar que incluye procesos para recopilar, preparar, analizar, visualizar y modelizar datos que permitan generar conocimiento útil para comprender problemas complejos y ayudar en la toma de decisiones. Estos datos con frecuencia son no estructurados y heterogéneos. *“Es la ingeniería civil de los datos. Sus acólitos poseen un conocimiento práctico de herramientas y materiales, junto con una comprensión teórica de lo que es posible”* (Driscoll, 2010).

3.1.1 Antecedentes

La ciencia de datos como tal se desempeña como una profesión que opera en las bases de Big Data. Esta es una mezcla que incorpora aspectos y herramientas de informática y computación con estudio estadístico.

En el año de 1962 el Sr John W. Tukey escribe su artículo El futuro del análisis de datos, donde expresa: “Realmente pensé que era un estadístico interesado en inferencias de lo particular a lo general. Pero a medida que observé la evolución de la estadística y las matemáticas, en realidad llegué a sentir que mi interés era el análisis de datos.”. Más tarde en el año 1974 Peter Naur publica una encuesta concisa de métodos computacionales en Suecia y los Estados Unidos. El libro es una encuesta de los métodos contemporáneos de procesamiento de datos que se utilizan en una amplia gama de aplicaciones.

Luego en el año 1977, Tukey publicó un artículo llamado “Análisis Exploratorio de Datos”, donde destacaba que era necesario hacer más hincapié en el uso de datos para sugerir hipótesis para probar y que el análisis exploratorio de datos y el análisis confirmatorio de datos pueden y deben “Procede lado a lado”.

En septiembre de 1994 BusinessWeek publica un artículo de portada sobre “Marketing de base de datos”, en el cual resaltaban lo siguiente: “Las empresas recopilan una gran cantidad de información sobre usted, lo analizan para predecir la probabilidad de que

compre un producto y utilizan ese conocimiento para elaborar un mensaje de marketing exactamente calibrado para obtenerlo”.

Ya para el año 1996 los miembros de la Federación Internacional de Sociedades de Clasificación (IFCS) se reúnen en Kobe, Japón, para su conferencia bienal. Por primera vez, el término “ciencia de datos” se incluye en el título de la conferencia. (“Ciencia de datos, clasificación y métodos relacionados”).

Más adelante en el año 1997 en la conferencia inaugural para la Cátedra de Estadística HC Carver en la Universidad de Michigan, el Profesor CF Jeff Wu (actualmente en el Instituto de Tecnología de Georgia), pide que las estadísticas pasen a llamarse ciencia de datos y estadísticos a los que se les cambie el nombre de científicos de datos. En este año también se lanza la revista *Data Mining and Knowledge Discovery* en cuyo título se refleja el ascenso de la “minería de datos” como la forma más popular de designar “la extracción de información de grandes bases de datos”.

Para el año 2001 William S. Cleveland publica “Ciencia de datos: un plan de acción para expandir las áreas técnicas del campo de la estadística”, es un plan para ampliar las áreas principales del trabajo técnico del campo de las estadísticas. En el mes de enero de 2003 se realizó el lanzamiento de *Journal of Data Science*: por Data Science donde se expresaba: “entendemos casi todo lo que tiene que ver con datos: recopilación, análisis, modelado etc., pero la parte más importante son sus aplicaciones.

En mayo de 2005 Thomas H. Davenport, Don Cohen y Al Jacobson publican “Competing on Analytics”, un informe del Babson College Working Knowledge Research Center, que describe “la aparición de una nueva forma de competencia basada en el uso extensivo de análisis, datos y toma de decisiones basada en hechos”.

Ya en el año 2007 El Centro de Investigación de Dataología y Ciencia de Datos se estableció en la Universidad de Fudan, Shanghai, China. En el año 2009, dos de los investigadores del centro, Yangyong Zhu y Yun Xiong, publicaron “Introducción a la Dataología y Ciencia de Datos”, en el que afirman “Diferentes a las ciencias naturales y sociales, la Dataología y la Ciencia de datos toman datos en el ciberespacio como su objeto de investigación”.

En el año 2009, Hal Varian, economista jefe de Google, le dice al McKinsey Quarterly: “Sigo diciendo que el trabajo sexy en los próximos diez años será estadístico”, y en junio del mismo año Troy Sadkowsky crea el grupo de científicos de datos en LinkedIn como complemento de su sitio web. Un año después Kenneth Cukier escribe en “The Economist Special Report Data”, “Data Everywhere” donde comenta: “surgió un nuevo tipo de profesional, el científico de datos, que combina las habilidades del programador de software, estadístico y narrador artista para extraer las pepitas de oro escondido bajo montañas de datos”.

Para septiembre de 2011 Harlan Harris escribe en “Data Science, Moore’s Law, and Moneyball”: ‘Data Science’ se define como lo que hacen los ‘Data Scientists’.

Finalmente, en septiembre de 2012 Tom Davenport y DJ Patil publican “Data Scientist: El trabajo más sexy del siglo XXI” en Harvard Business Review. (Ortiz, 2019)

Después de ese año han surgido diversos métodos para trabajar lo que hoy es denominado como “Ciencia de Datos”, y que es parte de esta propuesta.

3.1.2 Disciplinas (Núcleos o Áreas)

Como la ciencia de datos es interdisciplinar es necesario que el profesional de esta, es decir, el científico de datos posea las siguientes competencias:

Matemática (Estadística). La ciencia de datos posee enfoque matemático específicamente con la estadística, la cual consiste en el estudio de la recolección, análisis, interpretación, presentación y organización de datos.

Informática (Programación). También guarda una estrecha relación con la informática, la cual se conoce como la ciencia que estudia el tratamiento automático de la información.

Conocimiento (Negocios). La ciencia de datos también se aplica y guarda relación en el mundo de los negocios. Las empresas tienen datos, necesitan hacer negocios por lo cual deben administrar esos datos para generar valor.

3.1.3 Procesos de análisis de los datos

En las últimas décadas los procesos de análisis de datos han cambiado puesto que se realiza de acuerdo con el modelo utilizado. Esto también va de la mano con la necesidad y la arquitectura diseñada.

En la ciencia de datos tenemos varios procesos que en una cantidad considerable son comunes. A continuación, se listan algunos de estos:

Almacenamiento de datos: consiste en persistir los datos que y colocarlos en un destino.

Selección de datos: trata de determinar el tipo de conocimiento a extraer u objetivo.

Preprocesamiento de datos: limpiar completar, ordenar y reducir con el objetivo de obtener calidad para etapas posterior.

Transformación de datos: convertir los datos brutos en estructuras y formatos adecuados a los objetos establecidos.

Minería de datos: consiste en la extracción de patrones de conocimiento a partir de los datos estructurados.

Interpretación o Evaluación de datos: se trata de poner en valor los resultados y conocimientos obtenidos, de manera que sea más legible, entendible y sencilla para el usuario. La ciencia de datos y los procesos afines a la misma, lo suelen supervisar tres tipos de administradores:

Gerentes comerciales: aquellos que trabajan con el equipo de ciencia de datos para definir el problema y desarrollar una estrategia para el análisis. Pueden ser los jefes de una línea de negocios como marketing, finanzas o ventas y contar con un equipo de ciencia de datos que les informe. Trabajan en fuerte colaboración con el gerente de ciencia de datos y tecnología para garantizar que se entreguen los proyectos.

Gerentes de Tecnología Informática: Esos son responsables de la planificación de la infraestructura y de la arquitectura que brindará soporte y asistencia a las operaciones de ciencia de datos.

Gerentes de ciencia de datos: Supervisan el equipo de ciencia de datos y el flujo de trabajo de manera diaria. Son creadores de equipos que pueden equilibrar el desarrollo de este con la planificación y el monitoreo del proyecto.

3.1.4 Aplicaciones

Spotify, Google, Facebook, Twitter, Netflix o Amazon tienen muchas cosas en común, el empleo de la ciencia de datos es una de ellas; emplean sistemas para recolectar y analizar información sobre sus usuarios, a fin de ofrecerles una experiencia más personalizada, según explicó el maestro en Ciencia de Datos, Eduardo Blancas Reyes. En este mismo sentido, el investigador de la Universidad de Columbia (Nueva York) recalcó que la ciencia de datos está comenzando a cobrar importancia.

La ciencia de los datos es usada por los diversos motores de búsqueda a fin de mostrar los mejores resultados, la mercadotecnia digital se basa enteramente en ella, básicamente todo sistema de recomendaciones con el que interactúan los cibernautas, el mundo de los videojuegos la emplea para ofrecer una experiencia personalizada, portales de ventas por internet la usan en la creación de los comparativos de precios, aerolíneas hacen uso de la ciencia de datos en la planeación de rutas aéreas y otros aspectos. Por su parte, en el campo de la finanza se ha convertido en aliada para detección de fraudes y riesgos, mientras que las compañías de paquetería se apoyan en ella con el propósito de mejorar su logística.

También cabe destacar algunas aplicaciones futuras en la salud como es la predicción del porqué los pacientes serán readmitidos o cuándo podrían necesitar terapias preventivas (contra sobrepeso, diabetes o hipertensión, por ejemplo), incluso con el objetivo de detectar mejores prácticas que reduzcan costos. (Ruiz, 2018)

3.1.5 Importancia

Como se ha citado una frase que suena un poco cliché, pero los datos se han constituido el activo más importante de las empresas y corporaciones del siglo XXI. Ya no es el petróleo por el carácter perecedero que este posee, al contrario de los datos que crecen de manera exponencial y por ende son el objeto y la razón de ser de la ciencia de datos.

Una plataforma de ciencia de datos reduce considerablemente la redundancia e impulsa la innovación al permitir que los equipos compartan código, resultados e informes. Elimina los cuellos de botella en el flujo de trabajo al simplificar la administración y al utilizar herramientas, marcos e infraestructura de código abierto.

Por ejemplo, una plataforma de ciencia de datos podría permitir a los científicos de datos implementar modelos, APIs, estas últimas facilitan la integración con diferentes aplicaciones en cuanto a datos.

La demanda de plataformas de ciencia de datos ha explotado en el mercado. De hecho, se espera que la plataforma del mercado crezca a una tasa anual compuesta de más del 39 % en los próximos años y se proyecta que alcance los \$ 385 mil millones para el 2025. (Oracle, n.d.)

3.2 Análisis de datos (Data Analytics)

La analítica de datos es la ciencia de examinar datos en bruto con el propósito de sacar conclusiones sobre esa información. Implica realizar procesos algorítmicos y/o mecánicos para obtener conocimiento; por ejemplo, aplicar un proceso para buscar correlaciones significativas entre varias series de datos. Las técnicas usadas tradicionalmente en productos BI (Inteligencia Empresarial) para analizar y generar conocimiento de los datos en bruto están pensadas para trabajar con datos estructurados. (Junta de Galicia, 2017)

El término análisis de datos se refiere a un conjunto de aplicaciones que van desde la inteligencia empresarial básica o Business Intelligence, la elaboración de informes y el procesamiento analítico en línea (OLAP, del inglés Online Analytical Processing) hasta diversas formas de análisis avanzado como machine learning o data mining. Sin embargo, en la literatura, es común encontrar el término analítica de datos referido específicamente a análisis avanzado de datos, tratando las técnicas de BI como una categoría separada. (EnergyLab, 2017)

La analítica de datos en sus tres vertientes, descriptiva, predictiva y prescriptiva permite detectar patrones y comportamientos de los clientes para predecir situaciones y que las empresas puedan anticiparse a la toma de decisiones y junto con la capacidad de

recomendar productos y servicios personalizados al cliente, permiten una mejor experiencia de este. (Porrás, J., 2018)

En definitiva, se puede decir que la analítica de datos consiste en aplicar procesos algorítmicos o mecánicos con el propósito de obtener información, como puede ser seleccionar un conjunto de datos y buscar un factor común entre todos. El enfoque de la analítica de datos radica en la inferencia, que es el proceso de derivar conclusiones que se basan únicamente en lo que el investigador ya sabe.

3.2.1 Antecedentes

En el 1962 John W. Tukey escribe "El futuro del análisis de datos" donde expresa: Durante mucho tiempo pensé que era un estadístico, interesado en inferencias de lo particular a lo general. Pero a medida que observé la evolución de las estadísticas matemáticas, tuve motivos para preguntarme y dudar ... Llegué a sentir que mi interés central está en el análisis de datos.... El análisis de datos, y las partes de las estadísticas que se adhieren a él, deben ... asumir las características de la ciencia en lugar de las matemáticas ... el análisis de datos es intrínsecamente una ciencia empírica ... Qué vital y qué importante ... es el auge del programa almacenado computadora electrónica En muchos casos, la respuesta puede sorprender a muchos por ser "importante pero no vital", aunque en otros no hay duda, pero la computadora ha sido "vital". (Tukey, 1962)

Más tarde el mismo autor en 1977 publica el artículo "Exploratory Data Analysis", en el que se argumenta que la importancia radica en el uso de los datos para sugerir hipótesis que permitan testear y explorar los mismos, permitiendo extraer conclusiones veraces. Ese mismo año la International Association for Statistical Computing (IASC) establece que su misión principal consistirá en unir la estadística tradicional, la tecnología informática y el conocimiento experto, para convertir los datos en información y conocimiento. (Junta de Galicia, 2017)

En el año 1989 se celebra la primera edición de la Conferencia Anual sobre Descubrimiento y minería de datos que es conocida por sus siglas en inglés como KDD (Knowledge Discovery in Database). Siete años más tarde 1996 se realiza la Cumbre Bienal de la Federación Internacional de Clasificación de Sociedades donde por primera vez se hace uso del término Data Science.

En el año de 1997 es publicada la revista Data Mining and Knowledge Discovery y como se puede observar los nombres con relación a 1989, destacando la importancia de la minería de datos como la técnica más popular cuando se habla de la extracción de la información.

En septiembre de 2005, The National Science Board publica "Long lived Digital Data Collections: Enabling Research and Education in the 21st" En el informe se define a los Científicos de Datos como: "informáticos, ingenieros y programadores de bases de datos

y de software, expertos en estadística, bibliotecarios, y otros, cruciales para el éxito de la gestión de una colección de datos digitales”. (JUNTA DE GALICIA, 2017)

En el año 2009 se publica el informe de la Interagency Working Group on Digital “Harnessing the power of Digital Data for Science and Society”. En él se establece que se necesita identificar y promover la aparición de nuevas disciplinas y especialistas expertos en abordar los retos complejos y dinámicos de la preservación digital, el acceso sostenido a los datos, y la reutilización de datos. (JUNTA DE GALICIA, 2017)

3.2.2 Análisis avanzado de datos

Según la consultora Gartner, “El análisis avanzado es el examen autónomo o semiautónomo de datos a través de técnicas y herramientas sofisticadas más allá del Business Intelligence (BI) tradicional, con el objetivo de descubrir conocimiento más detallado, hacer recomendaciones y generar predicciones. Las técnicas de análisis avanzado incluyen minería de datos, aprendizaje de máquina, reconocimiento de patrones, predicción, visualización, análisis semántico, análisis de sentimientos, análisis de redes y clústeres, estadística multivariante, análisis de gráficos, simulación, procesamiento de eventos complejos y redes neuronales.”

Para entender que es el análisis avanzado de datos, se debe comprender el proceso de extracción e interpretación de los datos. Lo cual puede observarse en la figura no. 6, en la cual se divide tres pasos fundamentales: entrada de información, análisis de datos y salida de información

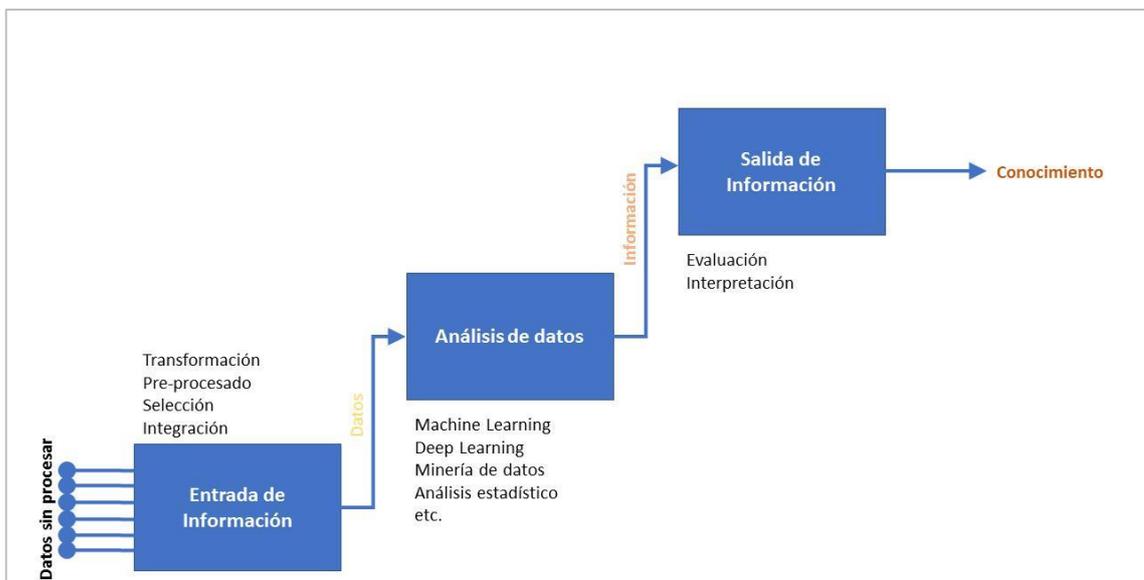


Figura 6- Un Enfoque Clásico Para Enfrentarse A Un Problema De Extracción De Conocimiento

Fuente: (Junta Galicia, 2017)

Este proceso inicia con la integración de los datos que tienen originadas por una o varias fuentes, para identificar y tratar la información para una aplicación previamente definida. Hay que tomar en cuenta que estos datos por su naturalidad pueden tener formatos diferentes por lo cual se deben transformar y luego proceder a su almacenamiento.

Luego de que los datos son recopilados, se procede a la filtración de estos, es decir, eliminar duplicidad, errores, datos que no aportan y que pueden afectar el análisis, culminada esta fase se comienza con el análisis de los datos utilizando técnicas como el machine learning o Deep learning. Para lo cual, se elabora un modelo analítico usando herramientas específicas para el análisis predictivo u otro software, seleccionando una muestra de los datos, los cuales son la base fundamental para realizar las pruebas antes de aplicarlo a toda la información.

3.2.3 Algoritmo para el análisis de datos

Los algoritmos son técnicas metódicas y sistemáticas que permiten realizar una tarea, como también solucionar problemas y lograr a un fin deseado por medio de patrones de comportamiento de los datos. Para el análisis de estos datos existen una gran variedad de algoritmos para las diferentes áreas saber, cada una orientadas a resolver necesidades específicas.

Según (SENIE, 2017), destaca 3 grupos de estas técnicas según su propósito:

Filtrado Colaborativo:

Basado en el usuario: consiste en generar un perfil de usuario tomando como base su comportamiento digital y luego encontrar otros usuarios con gustos similares.

Basado en elemento: consiste en que un usuario encuentre elementos similares que gustaron a otro usuario.

Algoritmo de Clasificación:

Naive Bayes. - es un clasificador estadístico, basado en el teorema de Bayes. Ayuda a predecir si una entidad pertenece a una clase. Puede manejar grandes cantidades de datos con rapidez y precisión.

Bosque aleatorio. Clasificador para grandes volúmenes de datos, introduce la aleatoriedad en cada clasificador individual. Parte de lo que se conoce como árbol de decisión.

Modelos ocultos de Markov. Permite encontrar parámetros desconocidos (ocultos) a partir de parámetros observables. Se puede ver como un doble proceso estocástico.

Perceptron multicapa. Se considera un algoritmo que realiza un aprendizaje por corrección de error. Permite un alto grado de conectividad en sus capas o neuronas. Puede usarse para predecir estancias hospitalarias de pacientes.

Agrupamiento (clustering)

K-Means clustering. Permite dividir un conjunto de datos en subconjuntos (clústeres). Generalmente, se emplea la distancia Euclidiana para establecer la cercanía del clúster a un centroide y de esta manera identificar a qué clúster pertenece un elemento.

Fuzzy k-Means. Los elementos de un conjunto de datos tienen un grado de pertenencia a un clúster. Los elementos no son dicotómicos al no estar sujetos a pertenecer de forma definitiva a un clúster.

Spectral Clustering. Usa la matriz de similitud de datos para reducir dimensionalidad. Los datos son representados como un grafo, la unión de los vértices tiene un peso. A mayor peso, se interpreta como mejor similitud, menor peso, menor similitud.

Según el blog PowerData (PowerData,2017), Entre las **técnicas de análisis de datos que mejor contribuyen a ampliar la visión empresarial**, al aportar conocimiento de calidad al negocio se encuentran:

1. **Aprendizaje automático:** Esta subespecialidad de la informática se engloba en el campo de la inteligencia artificial y es conocida también como machine learning. Está relacionada con el diseño y desarrollo de algoritmos que permiten a las computadoras promover la acción en base a datos empíricos. Su objetivo es aprender a reconocer automáticamente patrones complejos y tomar decisiones inteligentes basadas en datos
2. **Redes neuronales:** Este tipo de técnicas de análisis de datos consiste en modelos computacionales, inspirados por la estructura y el funcionamiento de redes neuronales biológicas. De la misma forma que las células y conexiones trabajarían y se establecerían dentro del cerebro, estas redes permiten encontrar patrones en los datos. Su especialidad son los patrones no lineales y están muy recomendadas tanto en aplicaciones que implican aprendizaje supervisado como en las que implican aprendizaje no supervisado. Un ejemplo de esta clase de técnicas sería la identificación de clientes en riesgo de abandono.
3. **Aprendizaje de reglas de asociación:** Se trata de un conjunto de técnicas de análisis de datos empleadas para descubrir relaciones interesantes entre variables en grandes bases de datos. La generación y prueba de posibles reglas es el resultado de la aplicación de algoritmos y, en la práctica, uno de sus usos más comunes es el análisis de la cesta de la compra, que permite a los minoristas determinar qué productos se compran con mayor y menor frecuencia para optimizar su planificación y decisiones de abastecimiento.

4. **Algoritmos genéticos:** Características más significativas de la información. Podría decirse que se trata del uso de un modelo para predecir valores futuros de una serie temporal en base a valores pasados conocidos de la misma u otras series. El pronóstico de cifras de ventas sería una de sus aplicaciones en la empresa.

3.2.4 Marco de referencia para el análisis de datos

Actualmente, se puede encontrar en el mercado una gama de marcos de referencias o como popularmente son conocidos framework, y biblioteca que facilitan y ofrecen la analítica de datos y aprendizaje en general. A continuación, se presentan los de mayor relevancia en el mercado:

Tabla 2- Frameworks y su descripción

Framework	Descripción
Apache Singa	Se trata de una plataforma distribuida de Deep Learning sobre grandes conjuntos de datos. Está diseñado con un modelo de programación intuitivo basado en la abstracción de capas. Soporta una amplia variedad de modelos de Deep Learning como son las redes convolución (CNN), modelos de energía como Restricted Boltzman Machine (RBM), y redes recurrentes (RNN).
Apache Mahout	Este proyecto ha nacido con la intención de construir un entorno para facilitar la creación de aplicaciones de aprendizaje maquina escalables basadas fundamentalmente en clustering, clasificación y filtrado colaborativo. (técnica usada por sistemas recomendadores). Apache Mahout ofrece tres características principales: <ul style="list-style-type: none"> • Un entorno de programación simple y extensible y un framework para la construcción de algoritmos escalable. • Una Ampla variedad de algoritmos predefinidos para Scala + Apache Spark, H2O y apache Flink. • Samara, un entorno de experimentación matemática vectorial con una sintaxis similar al lenguaje del entorno R8 y que funciona a escala.
Amazon Machine Learning	Es un servicio pensado para facilitar a los desarrolladores el uso de la tecnología de aprendizaje automático. Amazon Machine Learning provee herramientas de visualización y asistentes para guiar el desarrollo durante el proceso de creación de los modelos Machine Learning. Utiliza datos almacenados en Amazon S3, Redshift, o RDS, y puede ejecutar clasificación binaria, multiclase o regresión.
Azure Machine Learning Studio	Machine Learning Studio permite a los usuarios de Microsoft Azure probar e implementar soluciones de análisis predictivo, para posteriormente publicarlos como servicios web y poder utilizarlos en otras aplicaciones o herramientas de BI como por ejemplo Excel.
Caffe	Es una Framework de deep Learning desarrollado por el Berkeley Vision and Learning Center y por los contribuidores de la comunidad. Caffe ha sido liberado bajo la licencia BSD 2-Clause. Una de sus principales ventajas es su sencillez puesto que los modelos se definen mediante configuración, esto es, sin necesidad de codificación. Además, ofrece al usuario la posibilidad de elegir de forma sencilla donde realizar el procesamiento, en la CPU p la GPU. Otra ventaja de Caffe es la velocidad de procesamiento, puede procesar más de 60 millones de imágenes por día con una sola GPU Nvidia K40.

Fuente: (Autores)

3.2.5 Ventajas y limitaciones

La posibilidad de extraer conocimiento oculto de un conjunto de datos es la ventaja de mayor relevancia de la analítica de datos. Asimismo, facilita la exploración de información que antes era imposible procesar para las empresas y que ahora representan ventajas competitivas. También, en lo que concierne en la analítica de datos se encuentra la tecnología de Business Intelligence que es fundamental para los negocios del siglo XXI. La analítica de datos ofrece una serie de ventajas que las organizaciones pueden usar a su favor según (Junta de Galicia, 2017):

- Segmentación del cliente.
- Ahorro de costos
- Mejoras en la fabricación.
- Realización de previsiones y planificación.
- Cuantificación del riesgo.
- Detección del fraude.
- Mejora en la identificación del cliente objetivo.
- Mejora de la experiencia del cliente.
- Ventas personalizadas.
- Establecimiento de precios dinámicos

Así como existen los beneficios de la analítica de datos, se presentan inconvenientes como pueden ser:

- Personal no capacitado en el área de Big Data.
- Costo en la implementación de nuevas tecnologías e infraestructura.
- Ineficiencia en las habilidades o falta de información
- Presupuestos elevados para adecuar el ambiente a las nuevas tecnologías

3.2.6 El binomio Big Data -Data Analytics

Tanto Big Data como la analítica de datos pueden trabajarse de manera separada, sin embargo, es su unión la que presenta un potencial realmente no común y es lo que se conoce como la analítica de Big Data que es el objeto de estudio.

Según el estudio realizado por la consultora Gartner en 2015 “Big Data Industry Insight” las aplicaciones de Big Data Analytics son múltiples y diferentes para cada tipo de sector.

De esta forma podremos encontrar aplicaciones en los siguientes sectores:

- Industria
- Media y Comunicaciones
- Servicios
- Administración pública
- Educación
- Venta al por menor
- Banca
- Seguros
- Salud
- Transporte

3.3 Big Data

En principio Big Data tiene sus orígenes en el uso y manipulación de datos por los diferentes manejadores de base de datos. Con el avance de los años y de la tecnología, los tamaños de los archivos han ido creciendo exponencialmente y de forma paralela surgieron nuevas estructuras, las cuales no pueden ser trabajadas con los métodos y software tradicionales. Esta es la razón por la cual nace lo que hoy se conoce como Big Data, que son datos complejos con volúmenes masivos y creciente de datos a gran velocidad y no pueden ser gestionados con las herramientas convencionales. Esta tecnología llegó al sector salud cuando se identificó la necesidad de digitalizar récords médicos y los historiales de los pacientes, puesto que se visualizó la oportunidad de que analizando estos datos se podían hacer predicciones.

El concepto de “Big Data” o los datos a gran escala hacen referencia al almacenamiento de grandes cantidades de datos que pueden estar organizados o no organizados y en conjunto a estos, los distintos procedimientos usados con el propósito de encontrar patrones repetitivos dentro de esos datos. Cuando dichos datos son recogidos en el ámbito de la salud usualmente son denominados Salud Masiva o Big Health. (OMS, 2016)

Big Data son datos que contienen una mayor variedad y que se presentan en volúmenes crecientes y a una velocidad superior. Esto se conoce como "las tres V". Dicho de otro modo, el Big Data está formado por conjuntos de datos de mayor tamaño y más complejos, especialmente procedentes de nuevas fuentes de datos. Estos conjuntos de datos son tan voluminosos que el software de procesamiento de datos convencional sencillamente no puede gestionarlos. Sin embargo, estos volúmenes masivos de datos

pueden utilizarse para abordar problemas empresariales que antes no hubiera sido posible solucionar (Oracle, 2019).

Esta tecnología es considerada como un paradigma que permite la recopilación, el almacenamiento, el análisis y la visualización, potencialmente en condiciones de tiempo real, de grandes conjuntos de datos con características heterogéneas.

3.3.1 Antecedentes

A medida que han pasado los años los datos han tenido un gran desarrollo y crecimiento que ha hecho necesario buscar nuevos métodos y nuevas tecnologías para manejarlos. Un vivo ejemplo de esto son los millones de transacciones bancarias que se realizan cada segundo, cuando los métodos tradicionales ya no fueron eficientes para manejar esta gran cantidad de información, es cuando se piensa en una nueva alternativa, dando paso a Big Data.

Si bien el concepto "Big Data " en sí mismo es relativamente nuevo, los orígenes de los grandes conjuntos de datos se remontan a las décadas de 1960 y 1970, cuando el mundo de los datos acababa de iniciar con los primeros centros de datos y el desarrollo de las bases de datos relacionales.

Según la página de Oracle (Oracle, s.f), alrededor de 2005, la gente empezó a darse cuenta de la cantidad de datos que generaban los usuarios a través de Facebook, YouTube y otros servicios online. Ese mismo año, se desarrollaría Hadoop, un marco de código abierto creado específicamente para almacenar y analizar grandes conjuntos de datos. En esta época, también empezaría a adquirir popularidad NoSQL.

El desarrollo de marcos de código abierto tales como Hadoop (y, más recientemente, Spark) sería esencial para el crecimiento del Big Data, pues estos hacían resultase más fácil de usar y más barato de almacenar. En los años transcurridos desde entonces, el volumen de Big Data se ha disparado. Los usuarios continúan generando enormes cantidades de datos, pero ahora los humanos no son los únicos que lo hacen.

Con la llegada del Internet de las cosas (IoT), hay un mayor número de objetos y dispositivos conectados a Internet que generan datos sobre patrones de uso de los clientes y el rendimiento de los productos. El surgimiento del aprendizaje automático ha producido aún más datos. (Oracle, s.f)

3.3.2 Evolución

Sus orígenes no son precisos, (Wang, Kung y Byrd, 2018).

La historia de Big Data está inextricablemente vinculada con la de la ciencia de datos, este término fue escrito por primera vez en 1997 por Michel Cox y David Ellsworth en un artículo presentado en una conferencia de IEEE para explicar la visualización de datos y los desafíos que presentó para los sistemas informáticos (Cox y Ellsworth, 1997), por A fines de la década de 1990, las mejoras tecnológicas y de innovación de TI habían permitido generar una gran cantidad de datos, pero poca comparación de información utilizable. Los conceptos de inteligencia Negocios (BI) creados para enfatizar la importancia de la recopilación, integración, análisis e interpretación de la información empresarial y cómo este proceso puede ayudar a las empresas a tomar decisiones más apropiadas y obtener una mejor comprensión de los comportamientos y tendencias del mercado.

El período de 2001 a 2008 fue una etapa evolutiva para el desarrollo de Big Data. esta se definió por primera vez en términos de volumen, velocidad y variedad (3V). después de lo cual se hizo posible desarrollar un software más sofisticado para satisfacer las necesidades de manejar la explosión de información en consecuencia. Desarrollos de software y aplicaciones como servicios web de lenguaje de marcado extensible (XML), sistemas de administración de bases de datos y Hadoop agregaron módulos y funciones de análisis a los módulos centrales que se enfocaron en mejorar la usabilidad para los usuarios finales, y permitieron al usuario procesar grandes cantidades de datos en y dentro de las organizaciones en colaboración y en tiempo real. En ese momento, las organizaciones de atención médica estaban comenzando a digitalizar sus registros médicos y agregar datos clínicos en enormes bases de datos electrónicos. Este desarrollo hizo que los datos de salud fueran almacenables, utilizables, investigables y procesables. Este avance ayudó a los proveedores de atención médica a practicar una medicina más efectiva.

A principios de 2009, el análisis de Big Data entró en la etapa revolucionaria (Bryant et al., 2008). No solo la informática de Big Data se convirtió en una innovación innovadora para la inteligencia empresarial, sino que los investigadores predijeron que la gestión de datos y sus técnicas estaban a punto de pasar de datos estructurados a datos no estructurados, y de un entorno terminal estático a un entorno ubicuo de bases de poder. El análisis de datos de licitación de las industrias pioneras de la informática, como los bancos y el comercio electrónico, comenzaba a tener un impacto en la mejora de los procesos comerciales y la eficacia de la fuerza de trabajo, reduciendo los costos empresariales y atrayendo nuevos clientes. En lo que respecta a la industria del cuidado de la salud, a partir de 2011, los datos almacenados de atención de salud habían alcanzado 150 exabytes (1 EB = 10 a los 18 bytes) en todo el mundo, principalmente a partir de registros electrónicos de salud (Institute for Health Technology Transformation, 2013). Sin embargo, la mayoría de la creación de valor potencial todavía está en su infancia,

porque las técnicas de modelado y simulación predictivas para analizar los datos de atención médica en su conjunto aún no se han desarrollado adecuadamente.

La tendencia más reciente de la tecnología de análisis de datos de ofertas ha sido hacia el uso de cloud (información en la nube). Las empresas han adoptado cada vez más una solución de esta innovadora tecnología del software como servicio (SaaS) que ofrece una alternativa atractiva con un costo menor. Según la predicción de tendencias de TI 2013 de Gartner, aprovechar la ventaja de los servicios de computación en la nube para sistemas de análisis de Big Data que brindan una capacidad de análisis en tiempo real y un almacenamiento rentable se convertirá en una solución preferida para 2019.

3.3.3 Características

Big Data es una tecnología que en la actualidad está tomando mucha importancia en diversos sectores entre ellos la salud y posee tres características fundamentales que sustentan su objetivo y que son conocidas como las 3Vs:

Volumen: la característica principal que define Big Data es la gran cantidad de volumen de información que maneja (Ishwarappa, 2015). Esta masiva cantidad de datos se acumula con un crecimiento exponencial, ampliando continuamente las estructuras de almacenamiento de datos existentes.

Velocidad: otra de las características fundamentales de Big Data es la enorme velocidad en la generación, recogida y proceso de la información. Lo que permite que la tecnología Big Data sea capaz de almacenar y trabajar en tiempo real con las fuentes generadoras de información como sensores, cámaras de videos, redes sociales, blogs, páginas webs y otras fuentes que generan millones de datos por segundo.

Variación: por último, la tercera de las “v” que explica Big Data es la elevada capacidad de agregar información procedente de una amplia variedad de fuentes de información independientes, como redes sociales, sensores, máquinas o personas individuales.

La empresa Oracle sustenta la existencia de las tres V, sin embargo, destacan la inclusión de dos V más, valor y veracidad.

Valor: una gran cuantía de datos frecuentemente extrae pequeñas informaciones de valor. Cómo conseguir dicha información de manera eficiente es uno de los retos que afronta día a día el área de la inteligencia de datos. El valor es sin duda una cualidad fundamental en el análisis.

Veracidad: saber la fiabilidad de la información recogida es importante para obtener unos datos de calidad e, incluso, dependiendo de las aplicaciones que se le vaya a dar a misma, se convierte en fundamental. Es un factor que puede influir mucho en conseguir una ventaja competitiva en la explotación del Big Data.

Por su parte, la Universidad de Complutense de Madrid agrega dos V a las anteriormente mencionadas para un total de 7, y estas son

Variabilidad: en un entorno tan cambiante como el de los macro datos, la información varía mucho. Y también han de hacerlo los modelos o tratamientos que se aplican en torno a esta, pues no son fijos en el tiempo y requieren de un control periódico.

Visualización: Convertir cientos de hojas de información en un único gráfico que muestre claramente unas conclusiones predictivas es un ejemplo de cómo mostrar los resultados de forma clara y sencilla en un evidente ejercicio de sinterización.



Figura 7- 7v Big Data

Fuente: (Autores)

Se apoya la definición de la Universidad de Madrid, ya que esta propone las 3 ‘Vs’ principales, y la inclusión de cuatro más las cuales consideramos que deben agregarse, siendo estas: la veracidad, visualización, variabilidad y la valorización. A pesar de no ser medidas de magnitud son habitualmente aspectos considerados intrínsecos al Big Data.

3.3.4 Tipos de Datos del Big Data

Cuando se hace referencia a los datos y a su clasificación se habla de un conjunto de la unidad mínima de la información agrupados con un fin determinado. Estos datos pueden ser almacenados de diversas maneras, de acorde a su estructura, su naturaleza, su uso e incluso su origen, pero haciendo enfoque al punto de vista de este tema de estudio, estos se pueden clasificar de la siguiente manera:

- **Datos estructurados:** son datos que constan de una estructura, que ya han sido procesados y responden a un formato definido, logrando que sean más fáciles de trabajar y procesar. Dependen de la existencia de un modelo de datos, en el que cada campo es accesible y está relacionado con los demás.
- **Datos semiestructurados:** Este tipo de dato es un conglomerado de los datos estructurados y no estructurados. No corresponden a un tipo de estructura en específico, pero es una información ya procesada. Es un tipo de dato estructurado que no está asociado a ningún modelo de datos pero que, sin embargo, tiene etiquetas para separar elementos semánticos y hacer cumplir las jerarquías de campos y registros. Los más comunes son XML y JSON.
- **Datos no estructurados:** Estos son aquellos datos que no se corresponden con ninguna estructura homogénea, dígame que son un conjunto de datos de cualquier formato y de manera desorganizada.



Figura 8- Tipos de datos y su ejemplo

Fuente: (Autores)

No tienen etiquetas ni campos que estén definidos por un modelo de datos, normalmente se trata de texto plano con fechas y números. Su uso está extendiéndose cada vez más debido al análisis de imágenes, vídeo o PDF. Estos tipos de archivos suelen provenir de fuentes como las redes sociales, multimedia o archivos y documentos. Este tipo de datos son los más difíciles de procesar.

3.3.5 Fuentes de del Big Data

Las fuentes de información hoy siguen ampliándose, llegando a ofrecer un gran abanico de alternativas para las empresas que desean explotar datos masivamente. Las diferentes clasificaciones según Kapow Software se puede otorgar a los datos basándose en su naturaleza son las siguientes:

Archivos. Este tipo de archivos pueden ser registros médicos, documentos escaneados, archivos en papel o correspondencia de usuarios. Suelen tener una baja velocidad, variedad y también volumen, y no son fácilmente integrables en el sistema de almacenamiento de la empresa ya que los métodos de consulta son muy variados y no todos los archivos están digitalizados.



Figura 9- Kapow Software Infographic

Fuente: (Column Five, 2020)

Documentos. Quedan incluidos todo tipo de documentos como XML, JSON, CSV, etc., así como también de tipo ofimático como Word, PowerPoint y PDF. También, se encuentran correos electrónicos y archivos de texto plano. La diversidad de tipos de documentos indica que se puede observar una notoria variedad.

Multimedia. La generación de este tipo de fuente suele ser muy alta, implicando una gran volumen y velocidad de creación de los datos. Imágenes, audios, vídeo, como los generados a partir de una radiografía, una ecografía, o los registros de audios de las llamadas de los call centers, son los que se incluyen en este tipo de fuente de información.

Almacenamiento de datos. Aquí quedarían recogidos los datos más comunes de las organizaciones, como son las bases de datos, ficheros de sistema, repositorios o HDFS. Éstos suelen ocupar un alto volumen del total de la información almacenada por las empresas y están totalmente implementados para otorgar rapidez al momento de proporcionar consultas e informe.

Aplicaciones de negocio. Las empresas suelen tener a menudo diversas aplicaciones para la gestión de proyectos, recursos humanos, CRM, ERP o programas financieros. Todas estas aplicaciones generan una cantidad media de datos, que puede tener un valor significativo para la empresa. Suelen ser fáciles de explotar debido a su estructuración.

Web. Hace referencia a la web pública, a todos esos datos que se pueden extraer de páginas de bases de datos como Wikipedia o IMDb, así como también datos de índole pública como los gubernamentales, tráfico, tiempo, economía, etc. Éstos suelen ser totalmente externos a las organizaciones, y se contempla una gran variedad y volumen de información.

Redes Sociales. Todas las redes sociales son recogidas en este bloque, las más conocidas como Facebook, LinkedIn o YouTube, y se destaca la velocidad con la que se generan los datos debido a su gran número de usuarios a nivel mundial. El análisis en este tipo de datos suele ser más complicado debido a la implicación del factor del lenguaje natural y de sentimientos, para los cuales deben llevarse a cabo procesos específicos que clasifiquen dichos datos correctamente para su posterior uso.

Registros de máquinas. Se incluyen todas las comunicaciones entre servidores, aplicaciones y máquinas, así como los registros que generan las mismas máquinas como, por ejemplo, registros de eventos, localización móvil o clickstream. Este tipo de datos tienen una muy alta variedad, volumen y velocidad de creación de estos, suelen ser fáciles de implementar en este sistema de almacenamiento para posteriormente realizar consultas o informes sobre los mismos.

Sensores. Aquí encontramos todo tipo de sensores como cámaras de tráfico, dispositivos médicos, etc. Podríamos incluir en este grupo todos los dispositivos relacionados con Internet of Things y wearables, los cuales están teniendo una creciente expansión en todos ocho tipos de datos masivos de las organizaciones. Debido a esto la velocidad, el volumen y la variedad de los datos suele ser muy alta, pero a pesar de ello, tienen una muy buena estructuración que nos otorga un acceso más fácil a los datos. A parte de dichas clasificaciones también encontramos otras que pueden englobar diversos grupos dentro de ellas, como son las de datos estructurados, semiestructurados o datos no estructurados.

3.3.6 Seguridad

Uno de los casos de uso más comunes y que las nuevas tecnologías de Big Data han facilitado y mejorado su práctica es el de seguridad. Cada vez hay más empresas que utilizan estas tecnologías para implementar soluciones para este campo. Un análisis completo de seguridad puede llegar a involucrar muchas fuentes de información como logs (de servidores o web, por ejemplo), transacciones con los clientes o distintos sensores. Las tecnologías Big Data permiten almacenar todo el contenido –no estructurado- ofrecido por estas fuentes sin tener que hacer una preselección o cualquier tipo de proceso. Luego pueden distribuir el análisis de la información en distintos nodos agilizando el proceso y obteniendo un resultado de manera más rápida que con los procedimientos convencionales existentes.

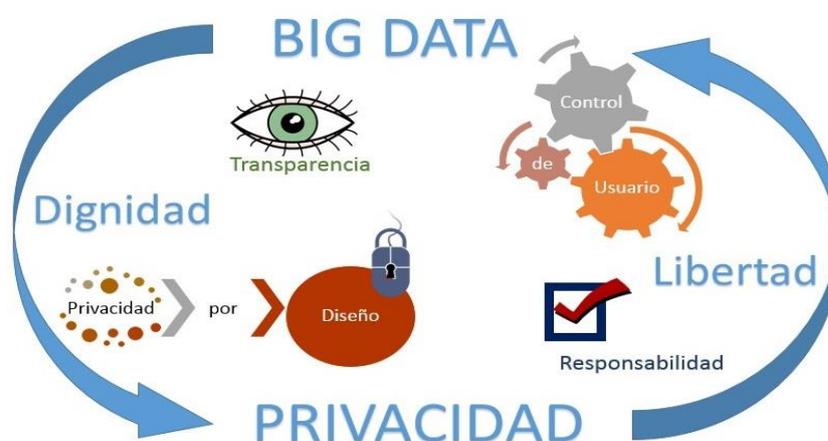


Figura 10- Retos de privacidad y seguridad en Big Data

Fuente (INCIBE, 2016)

Algunos de los ejemplos más específicos son la detección de ataques cibernéticos -como malware en el sistema o phishing- en, por ejemplo, entidades bancarias para detectar intentos de fraude o localizar a una persona o su teléfono mediante la información generada por las torres de comunicación inalámbrica -a través de sus llamadas o conexiones a internet. (Morros y Picaño, 2013)

3.4 Sistemas gestores de bases de datos

Un Sistema Gestor de Base de Datos (SGBD) o DataBase Management System (DBMS) es un sistema que permite la creación, gestión y administración de bases de datos, así como la elección y manejo de las estructuras necesarias para el almacenamiento y búsqueda de información del modo más eficiente posible. (Martín, 2019)

3.4.1 Evolución

Con el tiempo, las bases de datos han pasado de gestionar datos, a gestionar Big Data, la gestión de datos ha evolucionado de un sistema de gestión de base de datos relacional a un sistema de gestión de datos de Big Data que integra Hadoop, noSQL y Data Warehouse relacional e incluso posiblemente otras fuentes de datos.

En la década de los sesenta, las aplicaciones informáticas a nivel de gestor de datos eran trabajadas por lotes (batch) y estaban pensadas para una tarea muy específica relacionada con muy pocas entidades. A finales de esta década se incorporan los discos fijos, estos cambiaron en gran medida el escenario del procesamiento de datos, ya que los discos fijos permitieron el acceso directo a los datos.

Durante los años ochenta aparecen y se extienden muy rápido los ordenadores personales, debido al surgimiento del software para estos equipos monousuario (por ejemplo, dBase y sus derivados, Access), con los cuales era muy fácil crear y utilizar conjuntos de datos, que se denominan Personal Data Bases. Más tarde surgen los SGBD relacionales y pronto se utilizarían prácticamente en todas las empresas. Aún a mitad de los noventa cuando sería necesario un rendimiento elevado.

En la actualidad, los SGBD relacionales están en plena transformación para adaptarse a las nuevas tecnologías, por lo que ha surgido la necesidad de clasificarlos de acuerdo con la forma como administran según (Martín, 2019):

- Relacionales (SQL)
- No relacionales (NoSQL)

3.4.2 Base de Datos Relacional

Una base de datos relacional o Relational Database Management System (RDMS) es una recopilación de elementos de datos con relaciones predefinidas entre ellos. Estos elementos se organizan como un conjunto de tablas con columnas y filas. Las tablas se utilizan para guardar información sobre los objetos que se van a representar en la base de datos. Cada columna de una tabla guarda un determinado tipo de datos y un campo almacena el valor real de un atributo. (Amazon, s.f.)

Es un elemento fundamental a la hora de estructurar los datos procedentes de Internet, lo cual se consigue gracias a su arquitectura considerada un estándar para la gestión de bases de datos durante décadas, y todavía necesaria para poder trabajar con grandes volúmenes de datos. No obstante, a medida que el tamaño de la información contenida en ella crece, su rendimiento disminuye, al carecer de esa escalabilidad que le permitiría satisfacer el elevado nivel de exigencia del trabajo actual con Big Data.

Existen diversos tipos de RDMS:

Sistemas de Ficheros

Los sistemas de ficheros son una parte fundamental de la arquitectura Big Data ya que es por encima de ellos que el resto de las herramientas están construidas. Además, el hecho de querer trabajar con datos no estructurados los hace aún más importante ya que son el medio principal para trabajar con este tipo de información. Adicionalmente, un objetivo que buscan los sistemas Big Data es la escalabilidad, es decir, un sistema que pueda variar su tamaño (ya sea aumentándolo o disminuyéndolo) según las necesidades y que esto no afecte al rendimiento general de todo el sistema. Esta necesidad fue la que motivó la aparición de los sistemas de ficheros distribuidos, que consisten en una red o clúster de ordenadores -o nodos- interconectados que están configurados para tener un sólo sistema de ficheros lógico. (Morros, BIG DATA- ANÁLISIS DE HERRAMIENTAS Y SOLUCIONES , 2013)

Data Warehouse

Un data Warehouse (almacén de datos) es un tipo de sistema de gestión de datos diseñado para permitir y respaldar actividades de inteligencia empresarial, especialmente análisis. Los almacenes de datos se han diseñado exclusivamente para realizar consultas y análisis y, a menudo, contienen grandes cantidades de datos históricos. Los datos de un almacén de datos se obtienen normalmente de una amplia variedad de fuentes, como archivos de registro de las aplicaciones y aplicaciones de transacciones. Un almacén de datos centraliza y consolida grandes cantidades de datos de varias fuentes. Sus capacidades analíticas permiten que las organizaciones obtengan información empresarial valiosa de sus datos para mejorar la toma de decisiones. (Oracle, 2018)

Data Mart

Un Data Mart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Este puede ser alimentado desde los datos de un data Warehouse o integrar por sí mismo un compendio de distintas fuentes de información. (Oracle, 2018)

Data Mining

La minería de datos (Data Mining) busca los patrones ocultos en los datos que pueden utilizarse para predecir el comportamiento futuro. Las empresas, los científicos y los gobiernos han utilizado este enfoque por años para transformar los datos en conocimientos proactivos. Este es el proceso que analiza e intenta descubrir patrones entre los diferentes grandes volúmenes de datos. La analítica predictiva ayuda a evaluar lo que sucederá en el futuro. (Oracle, 2018)

3.4.3 Base de datos no relacional

Una base de datos no relacional (NoSQL) es aquella base de datos que no requiere de estructuras de datos fijas como tablas, no garantiza completamente las características ACID, escala muy bien horizontalmente. Estas se utilizan en entornos distribuidos que han de estar siempre disponibles y operativos y que gestionan un importante volumen de datos. Para la administración de este tipo de bases de datos, actualmente los principales sistemas gestores de bases de datos (SGBD NoSQL) son: MongoDB, Cassandra, Apache Hbase, entre otros. (Martín, 2019)

Entre los sistemas que administran datos no estructurados esta:

Lago de datos (Data lake)

Un lago de datos es un repositorio centralizado que le permite almacenar todos sus datos estructurados y no estructurados a cualquier escala. Puede almacenar sus datos tal cual, sin tener que estructurar primero los datos, y ejecutar diferentes tipos de análisis, desde paneles y visualizaciones hasta procesamiento de grandes datos, análisis en tiempo real y aprendizaje automático para guiar mejores decisiones. (Amazon, 2020)

3.5 Analítica de Big Data

Analítica de Big Data es una agrupación de diferentes tipos de tecnologías de manipulación de datos muy grandes desde terabytes hasta zettabytes en sus diferentes tipos ya sean estructurados, semiestructurados y no estructurados, combinados con un enfoque al conocimiento matemático y estadísticos con el fin de cruzar informaciones con el objetivo de encontrar relaciones, tendencias y similitudes lógicas entre ellas buscando patrones de comportamiento para luego ser analizada y poder resolver interrogantes de la empresa por medio de conclusiones viables para la toma de decisiones.



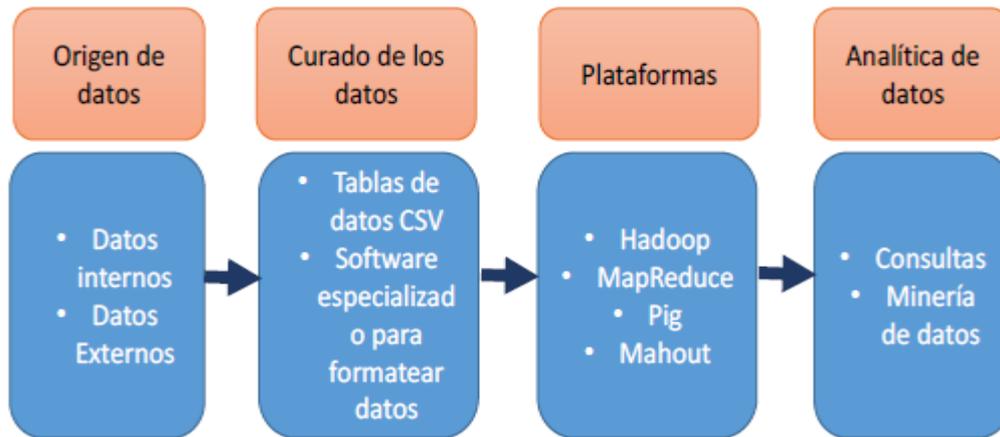
Figura 11- Beneficios de la analítica de Big Data

Fuente: (SAS Institute, 2019)

Con la analítica Big Data se pueden analizar enormes volúmenes de datos que las herramientas tradicionales de inteligencia de negocio no podrían abordar. Se necesita una analítica de alto desempeño para procesar todos estos datos: eso es la analítica Big Data (ibertech,2016).

Frente a las necesidades que han surgido en el mercado y con el crecimiento exponencial de la cantidad de información que deben de manejar los sistemas Big Data y analítica de datos han puesto en marcha un trabajo conjunto para satisfacer las necesidades de conjuntos de datos y técnicas para el análisis efectivo de esas voluminosas masas de datos y a su vez asegurar el almacenamiento y administración de estos.

La arquitectura de Big analítica de Big Data sigue un proceso de cómo debe ser el tratamiento de esas grandes cantidades de datos, Primero los datos son adquiridos de diversas fuentes, ya sean internos o externos a la entidad de análisis, luego son convertidos a un formato compatible ya sea en XML, Excel, CVS entre otros formatos. Estos datos que son recibidos de diferentes fuentes luego de ser recibidas pasan por alguna herramienta de manejo de datos y finalmente se hace efectivo el proceso de análisis de datos. Estos resultados pueden ser presentados por diagramas, gráficos, tablas organizadas, entre otros.



Arquitectura de *Big Data Analytics*.

Figura 12- Arquitectura de Big Data Analytics

Fuente: (Universidad Popular Autónoma del Estado de Puebla, 2018)

Con el tiempo, los datos relacionados con la salud se crearán y acumularán continuamente, lo que dará como resultado un volumen de datos no creíble. El volumen ya desalentador de datos de salud existentes incluye registros médicos personales, imágenes de radiología, datos de ensayos clínicos, presentaciones de la FDA, secuencias genómicas de datos y genéticos humanos, etc. Se presentan nuevas formas de datos grandes, como imágenes en 3D, lecturas genómicas y sensores biométricos. alimentando también este crecimiento exponencial.

3.5.1 Antecedentes

Desde décadas cercanas a 1950 la gran mayoría de las empresas ya almacenaba sus datos y aplicaban analítica básica en los números que se adquirían de las hojas de cálculo de manera manual para encontrar tendencias dentro de estos datos, incluso mucho antes de que se mencionara el término de “Big Data”.

Los primeros análisis de datos se ejecutaron con las primeras hojas de cálculo en los años 50 para apoyar la toma de decisiones en las empresas. Estas hojas de cálculo van evolucionando de la mano de las compañías de IT y SAS, pero trabajando siempre con datos estructurados. Hace una década cuando, en Silicon Valley, empiezan a emerger las aplicaciones para tratar información desestructurada y poder manejar los enormes volúmenes de datos existentes hoy en día.

La historia del análisis de Big Data está indisolublemente unida a la de Ciencia de los datos. El término "Big Data " se utilizó por primera vez en 1997 por Michael Cox y David Ellsworth en un artículo presentado en una conferencia de IEEE para explicar la visualización de datos y los desafíos que plantea para sistemas informáticos. (Cox y Ellsworth, 1997)

El período de 2001 a 2008 fue la etapa evolutiva de Big Data desarrollo. Big Data se definió por primera vez en términos de su volumen, velocidad y variedad (3V), después de esto, fue posible desarrollar software más sofisticado para satisfacer las necesidades de manejar la explosión de información, en consecuencia, desarrollos de software y aplicaciones como servicios web de lenguaje de marcado extensible (XML), base de datos sistemas de gestión, y Hadoop agregó módulos de análisis y funciones para módulos centrales que se centraron en mejorar la usabilidad para usuarios finales y permitieron a los usuarios procesar grandes cantidades de datos dentro de las organizaciones en colaboración.

Las empresas cada vez más adoptan una solución de "Big Data en la nube", es decir, software como servicio (SaaS) que ofrece una alternativa atractiva con menor costo. De acuerdo con la predicción de tendencias de TI de Gartner, 2013, aprovechando los servicios de computación en la nube para sistemas de análisis de Big Data que admiten un tiempo real, se preferirá la capacidad analítica y el almacenamiento rentable.

Según Solución de TI para 2016, la principal tendencia en la industria de la salud es cambio en el tipo de datos estructurados a semiestructurado (por ejemplo, monitoreo doméstico, telesalud, dispositivos inalámbricos basados en sensores) y datos no estructurados (por ejemplo, notas transcritas, imágenes y video).

El uso creciente de sensores y monitores remotos es un factor clave que respalda el aumento de los servicios de atención médica a domicilio, lo que significa que la cantidad de datos a ser generado a partir de sensores continuará creciendo significativamente. Esta a su vez, mejorará la calidad de los servicios de salud a través del análisis y predicciones más precisos. (Elsevier, 2016)

3.5.2 Características

Para el año 2017, autores concordaban que el 90% de los datos existentes hasta ese momento en el mundo eran producto de los dos últimos años. Esto denota que la humanidad inmersa en la era del Big data. De manera que el análisis de Big Data significa el análisis de montones de datos con el fin de extraer de ellos patrones de conocimiento. Según ACCIONA (2017) estas son algunas de sus características:

- **Velocidad.** Estos datos se caracterizan por la gran velocidad a la que se generan
- **Volumen.** Corresponde al enorme volumen que representan.
- **Variedad.** Hace referencia a la inmensa variedad de tipologías que engloban.

- **Veracidad.** Tiene que ver con el grado de veracidad que los datos poseen.

3.5.3 Tipos de Analítica de Big Data

Descriptivos

Se trata del tipo de analítica más simple. El que permite condensar el Big Data en datos más pequeños, con piezas de información más manejable. La mayoría de los datos en brutos no son aptos para el “consumo” humano, pero la información derivada de los mismos sí lo es. Por ello, el propósito de la analítica descriptiva es resumir lo que ha pasado. Y es eso precisamente lo que analizan la mayoría de las empresas.

Diagnóstico

Si se busca las causas de una situación concreta, el análisis de diagnóstico nos aclarará por qué ha sucedido algo. Esto está relacionado con el análisis antes mencionado. Sin embargo, en este caso se busca con minuciosidad las causas de una situación actual. De esta manera, estudiará y relacionará todos los datos disponibles para hallar patrones de comportamiento que expliquen los resultados obtenidos hasta ahora. Este tipo de análisis de diagnóstico es vital para solucionar problemas y evitar que sucedan de nuevo en un futuro. (Universidad de Alcalá, 2019)

Predictivos

Es una forma de analítica avanzada que usa datos nuevos e históricos para predecir una actividad, comportamiento o tendencia. El análisis predictivo implica realizar análisis estadísticos y/o aplicar técnicas de machine learning para realizar modelos predictivos que permitan valorar la posibilidad de que un determinado evento ocurra (ESTADO DEL ARTE DE BIG DATA, 2017). La analítica predictiva se materializa mediante la creación de modelos de conocimiento predictivos. Siegel define el modelo predictivo como un “mecanismo que predice un comportamiento de un individuo, como un clic, una compra, una muerte, o una mentira.

Prescriptivos

Es aquella tecnología que va más allá de los modelos descriptivos y predictivos, recomendando uno o varios itinerarios de acción y mostrando las posibles consecuencias de cada decisión. Desde cierto punto de vista, también es un tipo de analítica predictiva pero sólo en el caso del autor de la toma de decisiones en el momento en que le corresponde usar la información y tomar acción. En este caso no se dice que predice lo que va a ocurrir, sino un conjunto de escenarios futuros dependiendo del curso o las acciones tomadas previamente. (Wu, 2013)



Figura 13- Tipos de análisis de Big Data

Fuente: (Álvarez, 2019)

3.5.4 Arquitecturas de la analítica de Big Data

Para alcanzar el objetivo de este estudio, que es definir una arquitectura de análisis de Big Data, es necesario conocer y comprender en detalles cada una de las partes que la componen y no obstante a eso, diversas arquitecturas que se han desarrollado y se adaptan a diferentes entornos para dar una solución a problemas de índole distinto.

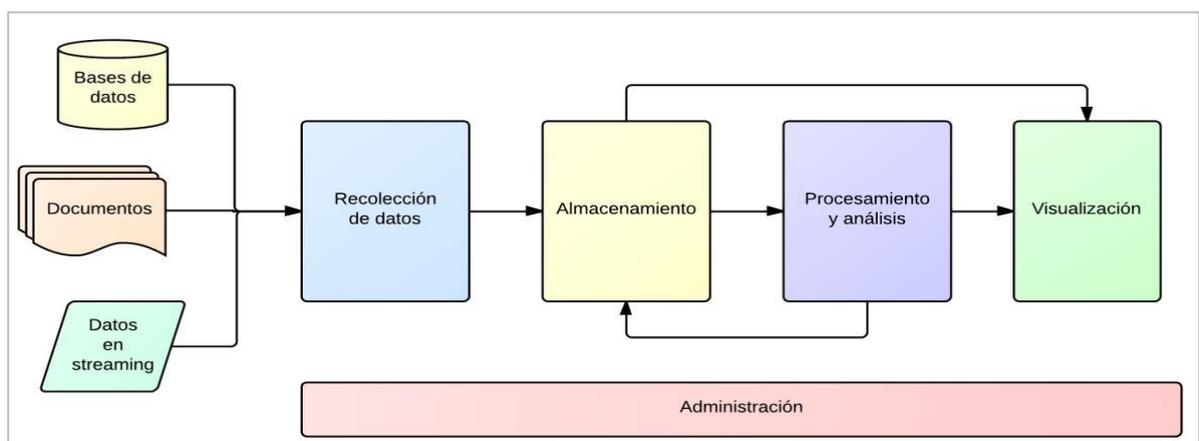


Figura 14- Arquitectura Big Data por capas.

Fuente: (Serrat, 2013)

3.5.4.1 Arquitectura Krishnan

La arquitectura propuesta por el señor Krishna Gurmmadi está basada en la típica gestión de datos conocida de antaño. Esta consiste en cuatro etapas como se puede ver en la figura 15, recopilación de datos, carga, transformación y extracción de datos.

Este plantea que en la primera etapa los datos provienen de diversas fuentes como pueden ser historiales médicos, análisis clínicos, imágenes clínicas, informaciones procedentes de dispositivos médicos como un marcapasos en un sistema destinado al sector salud. Asimismo, redes sociales, páginas web, M2M, transacciones, biometría, o generados por el ser humano.

En la segunda etapa, se cargan los datos utilizando sus metadatos y son estructurados. Es de aclarar que los metadatos son información que describe características de cualquier dato, como el nombre, la ubicación, la importancia percibida, la calidad y sus relaciones con otros objetos de datos que la empresa considere dignos de la gestión. (Soares, 2012)

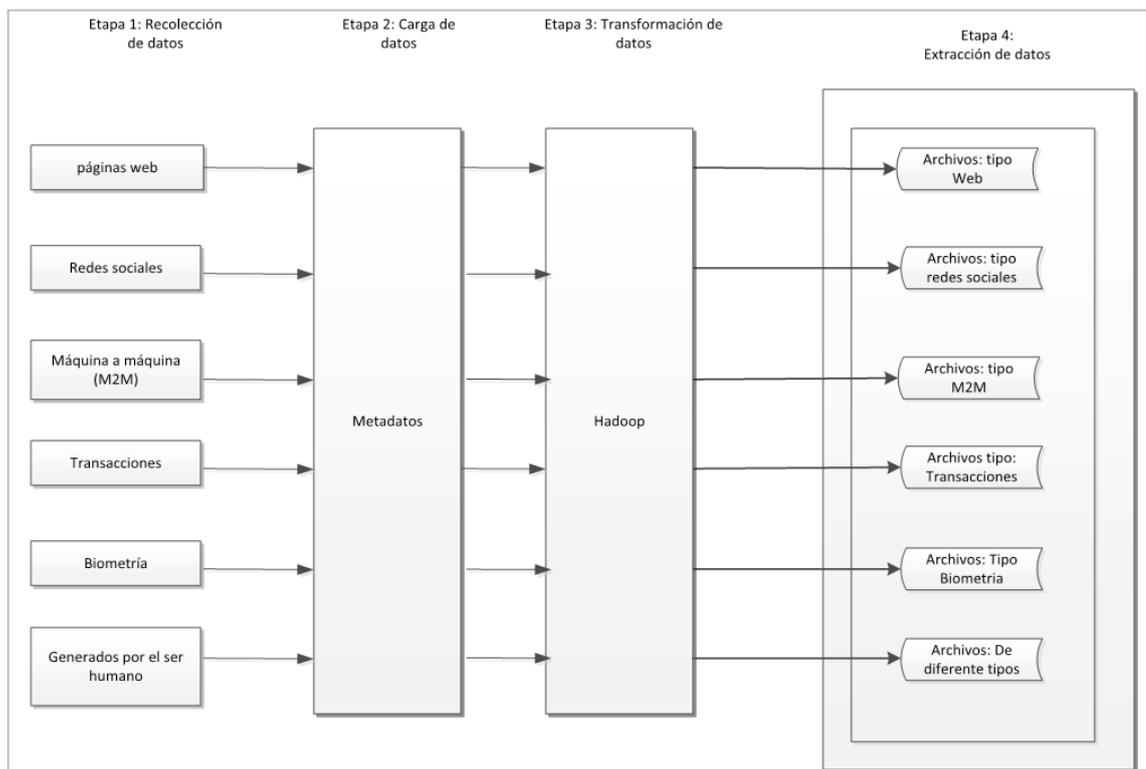


Figura 15- Arquitectura de Krishnan

Fuente: Camargo, J. J. Camargo, J. F. y Aguilar, L. (2015)

En la tercera etapa los datos se transforman mediante la aplicación de las reglas del negocio y el procesamiento de los datos. Respecto al procesamiento, en cada etapa producen resultados intermedios que se pueden almacenar para un posterior examen. El resultado de esta etapa son unas cuantas claves de metadatos con modelo clave-valor. (Krishnan, 2013)

En la última, etapa el objetivo es la extracción de los datos para obtener datos para su posterior análisis, generar informes operativos y su posible visualización, por último, no el menos importante, es su almacenamiento (Krishnan, 2013).

3.5.4.2 Arquitectura Bob Marcus

A diferencia de la arquitectura anterior que trabaja en etapas, el modelo propuesto Bob Marcus se define por niveles.

En el nivel 1 fuentes de datos externos se suministran las entradas de los datos y la producción de componentes internos. Seguido del nivel 2 cuyas tareas fundamentales son la filtración y transformación del flujo de dato que provienen de las fuentes externas y se realiza el procesamiento de los datos que se denomina movimiento entre los almacenes de datos.

Por su parte, el nivel 3 corresponde a la escalabilidad de la arquitectura y define tres tipos de escalonamiento según: (Camargo y Joyanes, 2014).

- El primero, a nivel de la infraestructura, existe con el fin de poder atender el almacenamiento y procesamiento de grandes volúmenes de datos.
- El segundo se refiere a los almacenes de datos. Tal como lo menciona Marcus, “es la esencia de la arquitectura Big Data”, la cual sucede en forma de “escalabilidad horizontal que usando componentes menos caros puede apoyar el crecimiento ilimitado de almacenamiento de datos”. Sin embargo, debe haber capacidades de tolerancia a fallas disponibles para manejar los fallos de sus componentes.

El nivel 4 está compuesto por las clases de bases de datos y el autor define tres de ellas base de datos analíticas, bases de datos operacionales y memoria de datos en forma de grilla. Mientras que en el nivel 5 se describen tres tipos de análisis: análisis de interfaces de procesos en lotes, análisis de interfaces interactivas y análisis de interfaces en tiempo real. El nivel 6 “se refiere a las aplicaciones e interfaces de usuario, las cuales no deben ser algoritmos complejos, al usar grandes cantidades de datos distribuidos”. (Camargo y Joyanes, 2014)

Por último, están los servicios de apoyos que son transversales a los niveles anteriormente mencionados, estos son necesarios para la implementación y gestión de los sistemas de Big Data y dentro de los cuales se pueden destacar:

- Diseñar, desarrollar e implementar herramientas
- Seguridad
- Gestión de procesos

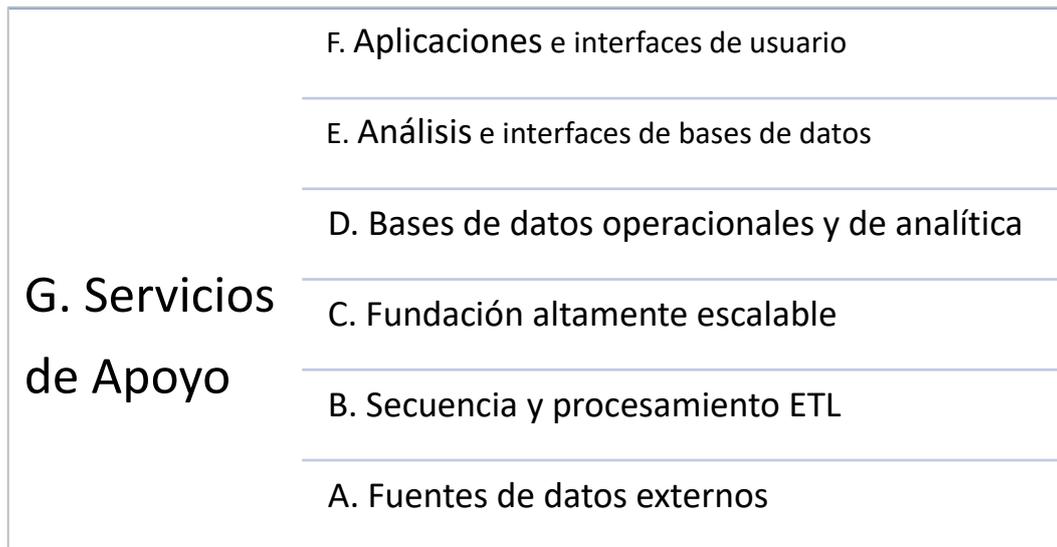


Figura 16- Arquitectura Bob Marcus

Fuente: Camargo, J. J. Camargo, J. F. y Aguilar, L. (2015)

3.5.4.3 Arquitectura de Microsoft

Por su parte, la arquitectura propuesta por Microsoft consta de 4 componentes que se presentan a continuación:

Fuente de Datos: los datos que provienen de diversas fuentes deben poseer tres características fundamentales: volumen, velocidad y variedad que son las que definen al Big Data. Estos datos “*son independientes de su contenido o del contexto. “Por lo general, los datos de ‘Big Data’ se recogen para un propósito específico”* (Camargo y Joyanes, 2014).

Transformación de Datos: este componente consta de 4 subetapas destacando que cada una puede aplicar procedimientos diferentes para alcanzar su propósito. Según Camargo y Joyanes (2014) estas son:

Colección. En la transformación de datos aparece un proceso llamado colección, donde se recogen datos en diferentes tipos y formas, se pueden allegar datos de fuentes y estructuras similares o iguales, o combinadas. Además, se crean metadatos para que sea más fácil hacer una búsqueda de los datos.

Agregación. Consiste en adicionar datos a una colección más grande, cuando uno o varios metadatos tengan claves iguales. “Como resultado, la información acerca de cada objeto se enriquece o el número de objetos en la colección crece”.

Congruencia. En esta etapa se recogen datos con metadatos sin importar si son diferentes y se unen a una colección más grande. Al final se obtiene que cada objeto sea enriquecido.

Minería de datos. La minería o data mining es una extracción de datos para luego poder hallar relaciones entre ellos. Existen “dos tipos de minería de datos: descriptivo, que proporcione información sobre los datos existentes, y predictivo, lo que hace que los pronósticos basados en los datos”.

Infraestructura de Big Data: según Microsoft la arquitectura es definida como “un paquete de almacenamiento de datos o software de base de datos, servidores, almacenamiento y redes utilizados en apoyo de las funciones de transformación de datos y de almacenamiento de datos según sea necesario”.

Uso de los datos: el último componente es el más importante para el usuario debido a que este podrá tomar decisiones en base a esas informaciones que pueden presentarse en diversos formatos.

3.5.5 Ciclo de la analítica de Big Data

La arquitectura de análisis de Big Data se basa en el concepto de marco de datos del ciclo de vida que comienza con la captura de datos, continúa a través de la transformación de datos y culmina con el consumo de datos. Como pueden ver la figura 17 representa la mejor secuencia de procesos para una propuesta de arquitectura de análisis de Big Data que se compone de cuatro capas arquitectónicas principales: La recolección de datos, almacenamiento, procesamiento y análisis y visualización que serán descritas más adelante. Estas capas lógicas hacen los componentes de análisis de Big Data que realizan funciones específicas, y por lo tanto permitirá a los gerentes de atención médica comprender cómo Transformar los datos de salud de varias fuentes en significativos información clínica a través de implementaciones de Big Data. Wang, Kung, y Byrd (2018).



Figura 17- Capas de una arquitectura de Analítica de Big Data

Fuente: (Autores)

3.5.5.1 Capa de Recolección de Datos

Esta primera fase del ciclo de análisis de Big Data se basa en el proceso de recolección de los datos provenientes de repositorios, fuentes y archivos. Estos datos se recopilan de varias ubicaciones internas o externas, y se almacenarán de inmediato en las bases de datos apropiadas, según el formato del contenido como por ejemplo en Hadoop que es un sistema de archivos distribuidos de los tantos que existen que serán desarrollados más adelante. Los datos pueden estar estructurados como pueden ser los registros electrónicos en el campo de la medicina, otros de origen semiestructurado como los registros de dispositivos de monitoreo y no estructurado como las imágenes y los videos.

3.5.5.1.1 Herramientas de recolección

Como se explicó en el apartado anterior esta etapa el sistema se conecta a sus fuentes de información con el fin de extraer las. Existen dos herramientas de recolección de estados según el sistema al origen de los datos:

- **Batch o por lotes:** se conectan de manera periódica a la fuente de datos buscando nueva información. Generalmente se usan para conectarse a sistemas de ficheros o bases de datos, buscando cambios desde la última vez que se conectaron. Una herramienta para migrar datos periódicamente -una vez al día, por ejemplo- de una base de datos a otra es un ejemplo de recolección de datos por lotes (Morros, BIG DATA- ANÁLISIS DE HERRAMIENTAS Y SOLUCIONES, 2013).
- **Transmisión en tiempo real:** este método consiste que el sistema se conecta de manera continua a las fuentes de datos, es decir, actualizando cada vez que transmiten datos. Este es factible para actividades de monitoreo con el propósito de detectar fallos y aumentar la seguridad.

3.5.5.2 Capa de agregación de datos

Esta capa es la responsable de manejar los datos que fueron recolectados en la fase I, dirigiéndolos de forma inteligente a través de tres pasos adquisición de datos, transformación y almacenamiento, los cuales se encargarán de la validación de los datos, limpieza, normalización y otros procesos.

3.5.5.2.1 Adquisición

La finalidad de la adquisición de datos es leer la información que fue proporcionada desde varios canales de comunicación, frecuencias, tamaños y formatos. En muchas ocasiones este paso es considerado como un gran obstáculo en las primeras etapas de la implementación de análisis de Big Data, debido a que las características de los datos entrantes pueden variar considerablemente. Una variable fundamental es el costo ya que este puede exceder y perjudicar el presupuesto disponible para establecer nuevos almacenes de datos y ampliar la capacidad para evitar el cuello de botella.

3.5.5.2.2 Transformación

En el paso de transformación, el motor que se encargará de transformar los datos debe ser capaz de mover, limpiar, dividir, traducir, fusionar, ordenar y validar datos. Por ejemplo, datos estructurados como este, típicamente contenido en un registro médico ecléctico podría extraerse de los sistemas de información sanitaria y posteriormente convertido en un formato de datos estándar específico, ordenado por el criterio especificado (por ejemplo, nombre del paciente, ubicación o historial médico), y luego el registro. Esta capa es responsable de manejar los datos de los diversos datos.

3.5.5.2.3 Almacenamiento

Este elemento de la capa de agregación está constituido por los sistemas de ficheros y la base de datos. Destacando que hace poco la base de datos era el sistema de tratamiento de información principal pero debido a que los datos han asumidos otras formas diferentes en relación con la estructura se ha optado por los sistemas que utilizan Big Data ya que estos buscan una mayor variedad y las bases de datos son poco flexibles por lo que han sido sustituidas por los sistemas de ficheros.

3.5.5.3 Capa de procesamiento y análisis

Esta capa es la responsable de procesar todo tipo de datos y realizar el análisis apropiado. Este último se puede dividir en tres componentes principales: Hadoop Map / Reduce, informática de flujo y análisis en la base de datos, de acuerdo con el tipo de datos y el propósito del análisis.

MapReduce es el modelo de programación más utilizado en análisis de Big Data que proporciona la capacidad de procesar grandes volúmenes de datos en forma de lotes de manera rentable. Además, permite el análisis de datos no estructurados y estructurados en un entorno de procesamiento paralelo masivo (MPP).

La informática de transmisión admite el procesamiento de datos de flujo de alto rendimiento en tiempo casi real o tiempo real. Con un análisis en tiempo real, los usuarios pueden rastrear datos en movimiento, responder a eventos inesperados a medida que ocurren y determinar rápidamente próximas mejores acciones.

La informática de flujo es una herramienta analítica que ayuda a predecir la probabilidad de transacciones ilegales o uso deliberado de cuentas. El análisis en la base de datos se refiere a una minería de datos con un enfoque construido sobre una plataforma analítica que permite procesar datos dentro del almacén de datos. Este componente proporciona procesamiento paralelo de alta velocidad, escalabilidad y características de optimización orientadas a grandes análisis de datos, y ofrece un entorno seguro para la información empresarial confidencial. Sin embargo, los resultados proporcionados por el análisis en la base de datos no son actuales ni en tiempo real y, por lo tanto, es probable generar informes con una predicción estática.

3.5.5.4 Capa de visualización de los datos

Esta capa es la que está de cara al usuario, debido a que le permitirá conocer los resultados de todo el proceso de la analítica de Big Data. Esta encargada de preparar todos los elementos de los datos procesados y presentar los reportes, informes, gráficos, estadísticas y otros formatos de visualización, siendo este su objetivo principal. Esta debe garantizar que todos los formatos de presentación expresen de forma óptima el significado y conocimiento que aporta la información.

3.5.5.5 Capa de gobierno de datos

Esta capa está compuesta por la gestión de datos maestros (MDM), gestión del ciclo de vida y gestión de seguridad y privacidad de datos.

Esta capa enfatiza el "cómo" aprovechar los datos en la organización. El primer componente de la gobernanza de datos, la gestión de datos maestros, se considera los procesos, la gobernanza, las políticas, los estándares, y herramientas para gestionar datos. Los datos están debidamente estandarizados, eliminados, e incorporados con el fin de crear la inmediatez, integridad, precisión y disponibilidad de datos maestros para respaldar el análisis de datos y la toma de decisiones.

El segundo componente, la gestión del ciclo de vida de los datos, es el proceso de gestión de la información empresarial a lo largo de su ciclo de vida. Desde archivar datos, hasta mantener el almacén de datos, pruebas y entrega de diferentes sistemas de aplicación, para eliminar datos.

El tercer componente, la seguridad de los datos y la gestión de la privacidad, es la plataforma para proporcionar actividades de datos a nivel empresarial en términos de descubrimiento, evaluación de configuración, monitoreo, auditoría y protección (IBM, 2012).

3.5.6 Actores principales en una arquitectura de Big Data

En una arquitectura de analítica de Big Data se pueden encontrar distintos actores con sus respectivas responsabilidades como pueden ser gestión de operaciones, seguimiento, seguridad, como el procesamiento. Asimismo, existen otros que interactúan directamente con el sistema.

Tabla 3- Actores principales en una arquitectura de Big Data

Actor(es)	Responsabilidades
Clústeres In-house, Centros de Datos, Proveedores Cloud	Se encarga de proveer la infraestructura de Big Data
Usuarios finales	Son quienes se benefician y proporcionan información al sistema.
Consultores, especialistas de aplicaciones o plataformas.	Proveedor de aplicaciones Big Data
Líder de la Empresa, Consultores, Científicos de Datos, Arquitectos de la Información, Arquitectos de Software, Arquitectos de Seguridad, Arquitectos de Privacidad, Arquitectos de Redes de Datos	Sistema de Orquestación
Oficial / Jefe de seguridad, Especialista en seguridad	Estructura de Seguridad y Privacidad

Fuente: (Autores)

3.5.7 Infraestructura de Big Data

3.5.7.1 Plataformas

Hadoop

Es un Proyecto de Apache bajo la denominación de código libre, y ha sido influenciado por grandes empresas como son: Facebook, Microsoft, Yahoo! Su objetivo principal es comprender un conjunto de herramientas, aplicaciones y frameworks java para desarrollar sistemas y utilidades dentro de la computación con dos características fundamentales: escalabilidad y soporte a ambiente distribuidos.

Hadoop permite el procesamiento distribuido de grandes conjuntos de datos en grupos de computadoras utilizando modelos de programación simples. Está diseñado para escalar desde servidores individuales a miles de máquinas, cada una de las cuales ofrece computación y almacenamiento local. En lugar de confiar en el hardware para ofrecer alta disponibilidad, la biblioteca en sí está diseñada para detectar y manejar fallas en la capa de aplicación, por lo que ofrece un servicio de alta disponibilidad en la parte superior de un grupo de computadoras, cada una de las cuales puede ser propensa a fallas. (APACHE, 2016)

Hadoop está constituido por dos módulos:

Hadoop Distributed File System (HDFS): el sistema de ficheros sobre el que se ejecutan la mayoría de las herramientas que conforman el ecosistema Hadoop.

Hadoop MapReduce: el principal framework de programación para el desarrollo de aplicaciones y algoritmos.

Figura 18 Módulos de Hadoop

Fuentes: (Autores)

Spark

Spark es un motor de análisis unificado para el procesamiento de datos a gran escala (APACHE, 2018). Este fue desarrollado por la universidad de Berkely California con el fin de combinar un sistema de computación distribuida mediante un clúster, siendo considerado como el primer software de código abierto en realizar una programación distribuida y que es accesible a los científicos de datos.

La característica fundamental que representa esta tecnología es su capacidad de procesamiento de datos en tiempo real. Además, es fácil de utilizar y al ser de código abierto tiene un costo accesible, es compatible con varias tecnologías y brinda seguridad en el procesamiento de los datos.

3.5.7.2 Herramientas de recolección de datos

1- Chukwa

Es una herramienta de recolección de datos que tiene como propósito principal trabajar sobre los logs y realizar análisis. Está basado en HDFS y MapReduce por tanto es escalable y robusto, sirviendo de apoyo a Hadoop debido a que este según Serrat (2013) “No termina de trabajar bien con sistemas de logs ya que está más optimizado para trabajar con pocos ficheros de mayor tamaño; a contraposición de los sistemas de logs, que son directorios con un gran número de ficheros pequeños”. Una ventaja de esta herramienta es la flexibilidad que ofrece para la recolección de datos en forma distribuida y a esto se le suma la capacidad de adaptación a las nuevas tecnologías.

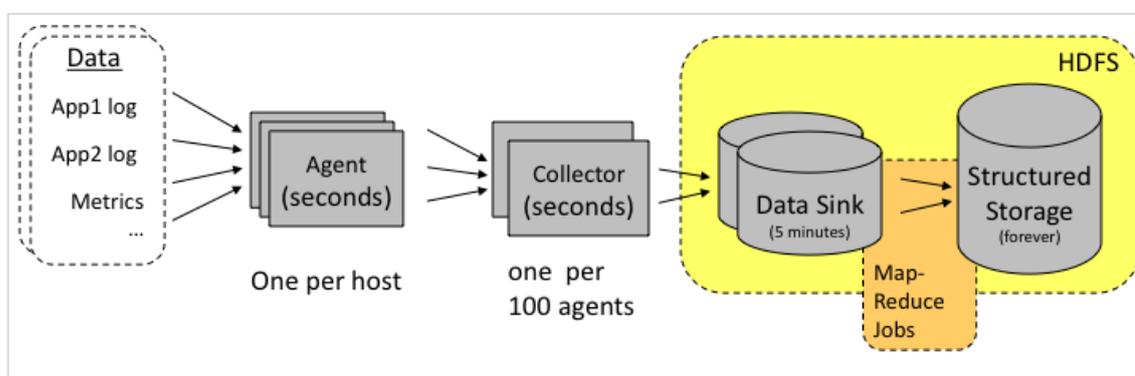


Figura 19- Arquitectura de Chukwa

Fuente (Apache,2007)

La arquitectura de Chukwa conta de 4 compontes:

Agentes: los procesos que se encargan de capturar datos.

Colectores: reciben los datos de los agentes y lo escriben en un almacenamiento permanente.

Trabajos MapReduce para trabajar con los datos.

HICC: es una interfaz web para visualizar datos.

Figura 20- La arquitectura de Chukwa

Fuente: (Autores)

2- Flume

Flume es una herramienta que trabaja bajo el concepto de la distribución y es utilizada para la recolección, agregación y transmisión de grandes volúmenes de datos. Su enfoque es la transmisión de datos en tiempo real destacándose por ser flexible y configurable. A raíz de sus características se puede adaptar a diferente tipo de situación como pueden ser:

monitorización de logs, descarga de información de redes sociales o mensajes de correo electrónico, entre muchas otras. A diferencia de Chukwa no va ligado exclusivamente con Hadoop.

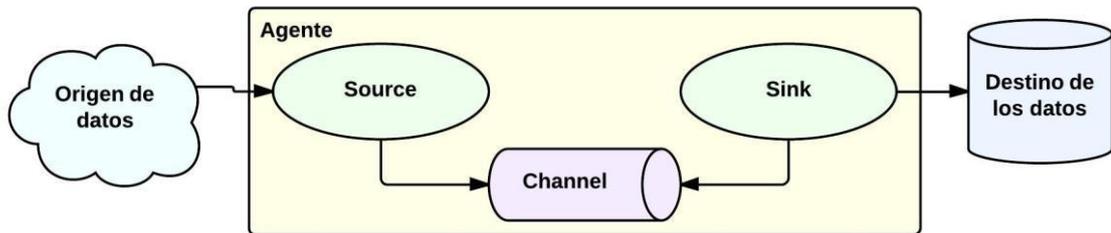


Figura 21- Arquitectura de Flume Fuente Big Data Análisis de herramientas y soluciones

Fuente (Serrat, 2013)

La arquitectura de esta herramienta está compuesta por agentes, procesos que tienen como tarea principal la recolección de datos y su envío al destino, quienes a su vez tienen tres componentes:

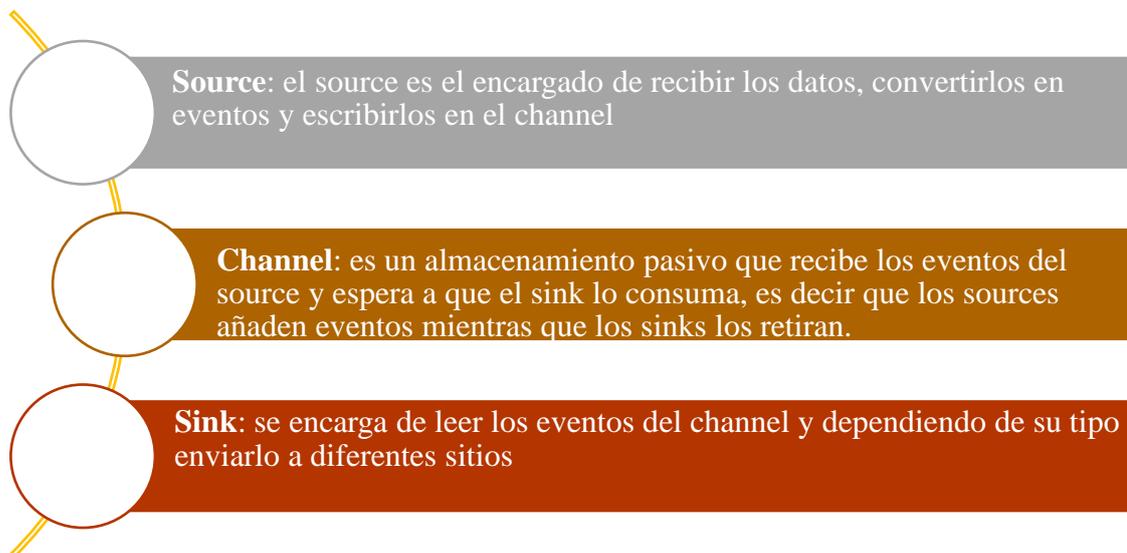


Figura 22- Componentes de un agente

Fuente: (Autores)

3- SQOOP

Apache Sqoop es una herramienta diseñada para transferir eficientemente datos masivos entre Apache Hadoop y almacenes de datos estructurados como bases de datos relacionales. Puede usar Sqoop para importar datos de almacenes de datos estructurados externos al sistema de archivos distribuidos de Hadoop o sistemas relacionados como

Hive y HBase. Por el contrario, Sqoop se puede utilizar para extraer datos de Hadoop y exportarlos a almacenes de datos estructurados externos, como bases de datos relacionales y almacenes de datos empresariales. (Apache, 2019)

3.5.7.3 Tecnologías para el almacenamiento

Cassandra

La base de datos Apache Cassandra es la elección correcta cuando necesita escalabilidad y alta disponibilidad sin comprometer el rendimiento. La escalabilidad lineal y la probada tolerancia a fallas en hardware básico o infraestructura de nube lo convierten en la plataforma perfecta para datos de misión crítica. (APACHE, 2016)

Cassandra es una base de datos distribuida, la cual cuenta con un diseño que puede ser ejecutada en una red de nodos de computadoras, simulando el comportamiento de un servidor. La topología típica de la red de Cassandra, la cual se compone por clúster de nodos, también llamado anillo de Cassandra, que se ejecuta en diferentes direcciones de red que están ubicadas en servidores físicos diferentes tal y como se muestra en la figura 24. (IBM, s.f)

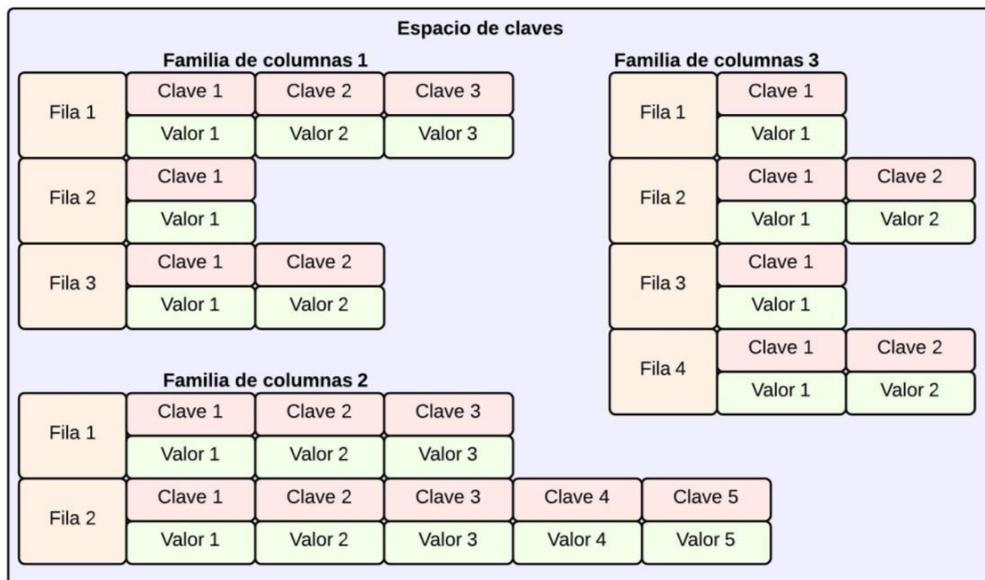


Figura 23- Ejemplo de un espacio de nombres de Cassandra. Fuente Big Data Análisis de Herramientas y Soluciones

Fuente (Serrat, 2013)

Hbase

Apache HBase es la base de datos de Hadoop, un almacén distribuido, escalable y de grandes cantidades de datos. Se utiliza para acceso de lectura /escritura aleatoria y en tiempo real a sus Big Data. El objetivo de este proyecto es el alojamiento de tablas muy

grandes (miles de millones de filas X millones de columnas) sobre grupos de hardware básico. Apache HBase es una base de datos de código abierto, distribuida, versionada y no relacional modelada a partir de Bigtable de Google. (Apache, 2016)

Esta herramienta aplicada a los procedimientos médicos se utiliza para la manipulación de los millones de registros que se almacenan y de los que se generan. La misma trabaja armónicamente bajo la estructura de Hadoop, software de código abierto. Es utilizada como apoyo a las decisiones clínicas y gestión de recursos implementando arboles de decisión que se encarga de verificar si se cumplen las condiciones de un atributo o un conjunto de ellos procedente con la verificación de la información que ya se encuentra en la base para proceder con la toma de decisiones.

Hive

Es un software de almacenamiento de datos que facilita la lectura, escritura y administración de grandes conjuntos de datos que residen en almacenamiento distribuido utilizando SQL. La estructura se puede proyectar en datos que ya están almacenados. Se proporciona una herramienta de línea de comandos y un controlador JDBC para conectar a los usuarios a Hive. (Apache, s.f).

Hive implementa una variante de SQL, llamada HQL (Hive QL). Este convierte la consulta HQL a un trabajo MapReduce que es ejecutado para obtener los datos. La finalidad de esto es permitir a usuarios que no cuenten con experiencia en programación de algoritmos MapReduce (que suelen ser bastante laboriosos), pero que sí cuenten con conocimientos de SQL, poder consultar datos. (Apache, s.f)

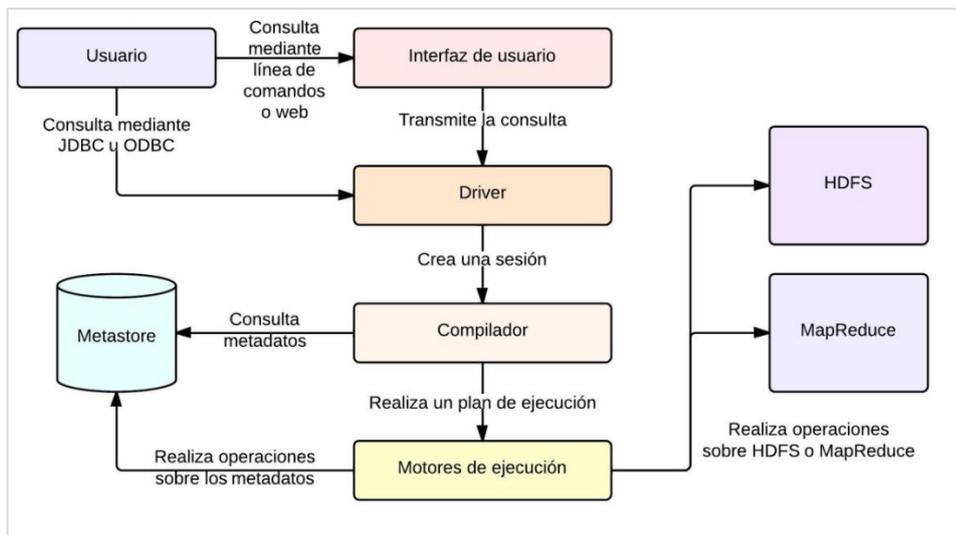


Figura 24- Arquitectura de Hive elaborada en fases o etapas

Fuente (Serrat, 2013)

3.5.7.4 Lenguaje de programación

Lenguaje Python

Es un lenguaje de programación de multipropósito considerado como una de las mejores opciones a la hora de trabajar no solo a nivel de entradas, sino que cuando hablamos de grandes cantidades de datos. Se dice esto debido a que gran parte del proceso de ETL (extracción, transformación y carga). En este sentido librerías como Tensorflow de Google hacen de Python un lenguaje totalmente apto para el aprendizaje automático (machine learning). (Morales, 2017)

En cuanto al área médica este lenguaje de programación es capaz de integrarse in extraer datos mediante herramientas HEDEA (Healthcare Data Extraction and Analysis) una herramienta que se ha diseñado de manera específica para extraer información de varios récords médicos usando el concepto de expresiones regulares. Esta herramienta también permite con varios conjuntos extensos de datos en varios formatos, como XML, por ejemplo, integrar esta capacidad para generar reportes de análisis y gráficos usando una base de datos centralizada. (Kumar, 2018)

Lenguaje R

R tiene un paquete para casi todas las aplicaciones cuantitativas y estadísticas imaginables. Esto incluye redes neuronales, regresión no lineal, filogenia, cartografía, mapas y muchos, etc. Maneja el álgebra de matriz particularmente bien (Morales, 2017).

En el campo de la medicina R es apto para el procesamiento datos e información. Por ejemplo, es usado en el campo de la genética para el descubrimiento de fármacos, la bioinformática, la epidemiología. Es importante decir que en lo relativo al análisis de datos R es utilizado de manera específica en este sector para analizar y predecir de diversas enfermedades, análisis de secuencias genéticas, análisis en la seguridad de medicamentos y también para analizar diversas permutaciones y combinación de medicamentos y productos químicos. Es una herramienta singular para los ensayos preclínicos. (Techvidvan, 2019)

Plataforma Pig

Apache Pig es una plataforma para analizar grandes conjuntos de datos que consta de un lenguaje de alto nivel para expresar programas de análisis de datos, junto con una infraestructura para evaluar estos programas. La propiedad sobresaliente de los programas Pig es que su estructura es susceptible de paralización sustancial, lo que a su vez les permite manejar conjuntos de datos muy grandes. (Allweil, 2015)

Según Sebastian Schneeweiss (2014):” *Pig como herramienta junto a tecnología de Big Data en Arboles de decisión se aplican en el Análisis de estudios clínicos*”.

3.5.7.5 Herramientas de visualización

Tableau

Desde la conexión hasta la colaboración, Tableau es la plataforma de análisis integral más eficaz, segura y flexible para sus datos. Aumente el potencial de las personas con el poder de los datos. Diseñado para un individuo y adaptado para toda una empresa. Tableau es la única plataforma de inteligencia de negocios que convierte sus datos en información útil (Tableau). Esta es una herramienta de visualización de datos que maneja grandes cantidades de datos, en volúmenes de “Big Data”, es muy potente y es muy utilizada en el área de la Inteligencia de Negocios ya que simplifica los datos en un formato muy fácil de entender. Este es una herramienta muy intuitiva con una interfaz gráfica amigable y fácil de comprender. También, cuenta con una gama de productos para adaptarse a las necesidades de quienes lo consumen. Tiene alta compatibilidad con las extensiones de archivos más usadas como son: Microsoft Excel, archivos de texto plano, Microsoft Access, archivos JSON, archivos pdf, datos geoespaciales, Amazon Athena, Amazon EMR Hadoop Hive, Amazon Redshift, Azure SQL Data Warehouse, Hadoop, Databricks, Dropbox, Google Analytics, Google BigQuery, Google Drive, Google Sheets, Hortonworks Hadoop Hive, IBM BigInsighths, MySQL, Salesforce, Sap HANA, Splunk, PostgreSQL, Teradata, entre otros.

Power BI

La herramienta de Microsoft Power BI es una solución de análisis empresarial que permite visualizar los datos y compartir información con toda la organización, o insertarla en su aplicación o sitio web. Conéctese a cientos de orígenes de datos y dé vida a sus datos con los paneles e informes dinámicos (Microsoft, 2019). Esta herramienta, destinada a la inteligencia empresarial, que permite unir diferentes fuentes de datos modelizar y analizar datos para después, presentarlos a través de paneles e informes que puedan ser consultarlos de una manera muy fácil, atractiva e intuitiva, también puede unificar todos los datos de su organización y externos, se encuentren a en la nube o localmente. (Makesoft, 2017)

CONCLUSIÓN

En este capítulo se han expuesto los conceptos de ciencia de datos, analítica de datos, Big Data y analítica de Big Data términos que a menudo son objeto de confusiones. También se hace diferencia de lo que es estadística y ciencia de datos.

Además, se hace hincapié en el desarrollo y definición de conceptos como la inteligencia de negocios y las tecnologías que más utiliza el sector de la medicina en lo referente a los historiales médicos.

En este mismo sentido, fueron descritos tres modelos de arquitecturas de analítica de Big Data. También se definen algunas de los recursos y herramientas tecnológicas que se toma como referencia en el campo, es decir, la historia clínica.

Por consiguiente, ya que se tiene total dominio del objeto de estudio de la presente investigación, es entonces, necesario posteriormente ejecutar la propuesta conociendo de manera especializada, todo el dominio de campo y las herramientas que se van a usar en la misma.

**CAPÍTULO IV: CARACTERIZACIÓN Y DIAGNÓSTICO
DE LA SITUACIÓN ACTUAL DEL SISTEMA DE GESTIÓN
DE INFORMACIÓN ADMINISTRATIVA Y DEL
HISTORIAL MÉDICO DE LOS PACIENTES DEL
HOSPITAL REGIONAL TAIWÁN 19 DE MARZO DE
AZUA, REPÚBLICA DOMINICANA**

INTRODUCCIÓN

El presente capítulo que se titula Caracterización y Diagnóstico de la Situación Actual del Sistema de Gestión de Información Administrativa y del Historial Médico de los Pacientes del Hospital Regional Taiwán 19 de Marzo de Azua, República Dominicana, expone la realidad de la entidad médica de cara a dos aspectos que desde inicios de este trabajo final se vienen exponiendo: historia médica de pacientes y el presupuesto del centro.

Este se fragmenta en tres segmentos de las cuales el primero explica aspectos como la misión, visión y valores de la entidad de salud y otros aspectos organizacionales. En segundo lugar, se habla de los sistemas tomando como principal objeto el sistema propio de gestión hospitalaria. Este apartado se desarrolla desde las perspectivas de hardware y software.

Finalmente, el tercero explica aspectos como la operación y manejo de pacientes en el área de emergencia y paciente que vienen por referencia de otro centro de salud. Se manejan varios datos estadísticos y también se presenta un panorama de cómo se encuentra el hospital mediante el uso de la herramienta FODA, y para concluir se expresan los resultados de las encuestas que fueron realizadas en función a al personal del centro de salud, es decir, especialistas de la salud, personal administrativo y pacientes.

4.1. Antecedentes históricos del Hospital Taiwán



Figura 25- Hospital Regional Taiwán 19 de Marzo

Fuente: (Hospital Taiwán, 2019)

El Hospital Regional Taiwán 19 de Marzo, es una institución de salud pública inaugurada el día 19 de Marzo del año 2005, que ofrece servicios de tercer nivel de atención, pertenece a la Regional VI, El Valle y se encuentra ubicado en la Ciudad de Azua de Compostela, calle Independencia, esquina 27 de Febrero, número 101, sector Simón Stridels, Azua, República Dominicana.

Este centro de salud surge a raíz de las relaciones diplomáticas entre la República de Taiwán y el gobierno dominicano como una iniciativa para que la provincia de Azua y otras provincias aleñas del sur de la República Dominicana pudieran obtener servicios de salud de calidad.

Este centro hospitalario posee una cartera de servicios, los mismos están anclados a los procesos de la entidad.

Tabla 4- Procesos y Servicios que ofrece el Hospital Taiwán

Procesos	Servicio
Consultas Externas	<ul style="list-style-type: none"> • Servicios de Consultas Externas
Apoyo Diagnóstico	<ul style="list-style-type: none"> • Servicios de Estudios e Imágenes Médicas • Servicios de Exámenes de Laboratorio clínico
Atención Quirúrgica	<ul style="list-style-type: none"> • Servicios de Procedimientos Quirúrgicos
Hospitalización	<ul style="list-style-type: none"> • Servicios de Hospitalización
Atención Emergencia	<ul style="list-style-type: none"> • Servicios de Emergencias
Programas de Atención y Prevención	<ul style="list-style-type: none"> • Servicios de Atención y • Prevención • Hemodiálisis
Servicios Sociales	<ul style="list-style-type: none"> • Servicios Sociales
Docencia e Investigación	<ul style="list-style-type: none"> • Servicio de Formación de Profesionales de la Salud

Fuente: (Hospital Taiwán, 2020)

4.1.1 Misión

Somos un Hospital Regional que garantiza servicios de salud especializados aplicando estándares de calidad para satisfacer las necesidades de los usuarios.

4.1.2 Visión

Ser reconocido como un Hospital Regional comprometido con la mejora continua de nuestros procesos para garantizar la calidad de los servicios de salud.

4.1.3 Valores

- Humanismo
- Equidad
- Compromiso
- Seguridad
- Respeto

4.1.4 Organigrama

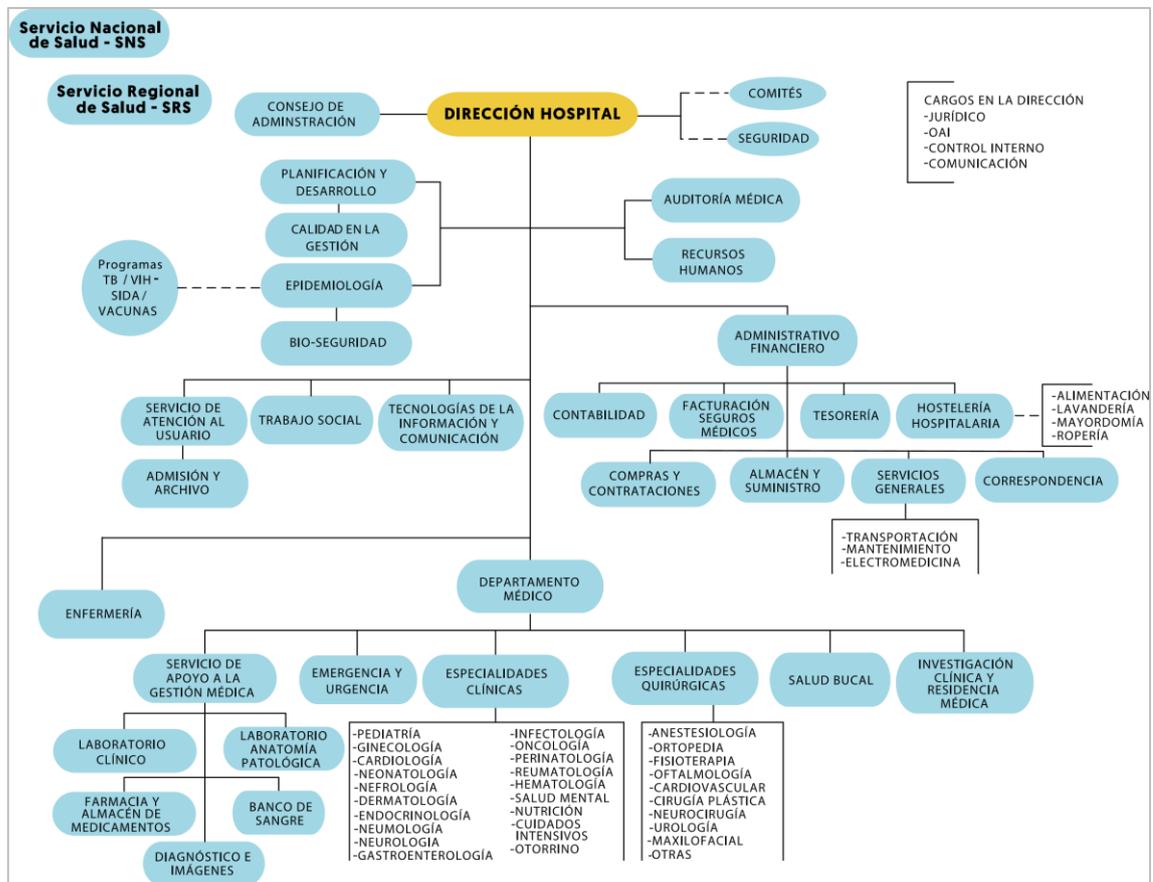


Figura 26- Organigrama del Hospital Taiwán

Fuente: (Hospital Taiwán, 2019)

4.2 Procesos para la Creación y/o Gestión de Historiales Médicos

Desde su inauguración en el 2005, los historiales médicos en el Hospital Regional Taiwán 19 de Marzo en Azua son trabajados de manera manual, lo que implica que cada doctor debe completar los formularios con los datos de cada paciente y los diagnósticos. Esta situación influye de forma negativa en la gestión médica-administrativa que se realiza en el hospital, debido a que retrasa procesos vitales como es el ingreso y el seguimiento de los pacientes. De igual manera, no es posible elaborar estadísticas en base a la información que se almacena en los historiales médicos, representado un activo importante para el centro.



Figura 27- Doctora completando los formularios de los pacientes

Fuente (Autores)

Cabe destacar que se han realizado esfuerzos por parte del centro de salud para contrarrestar esta situación, como es el desarrollo del Sistema de Gestión Hospitalario (SISGEHOSPI) del cual se estará hablando más adelante. Este posee un submódulo llamado “*Paciente*” y que una de sus funcionalidades es la de crear un formulario que contenga los siguientes datos de los pacientes:

1. Nombre
2. Apellido
3. Número de expediente
4. Edad
5. Sexo
6. Tipo de documento
7. No. Identificación
8. Nacionalidad
9. Dirección
10. Sector

- 11. Provincia
- 12. ARS
- 13. Padre
- 14. Madre

Este es el único proceso de la creación o/y gestión de los historiales médicos que esta digitalizado, el resto sigue realizándose manualmente. Este formulario es completado por un personal administrativo al momento que el paciente se presenta al centro luego de ser referido desde una Unidad de Atención Primaria (UNAP), el mismo es obligatorio para que un paciente pueda ser atendido por el personal de la salud del centro, destacando que hay casos en los cuales este procedimiento debe obviarse, como puede ser una persona que sufrió un accidente y no posea un documento de identidad.

Por otra parte, de acuerdo con las encuestas realizadas sobre el tiempo de espera de los pacientes en el Taiwán se observa que el 49% de los pacientes espera 30 minutos, el 25% permanece una (1) hora para ser atendido, mientras el 16% espera dos horas y un 10% tres (3) horas (ver gráfico No. 1). Lo cual demuestra la debilidad que existe para dar atención médica a los pacientes en el tiempo establecido por el Servicio Nacional de Salud, (Tabla No. 5) siendo una de las causas principales de esta situación la gestión de los historiales médicos realizada allí.

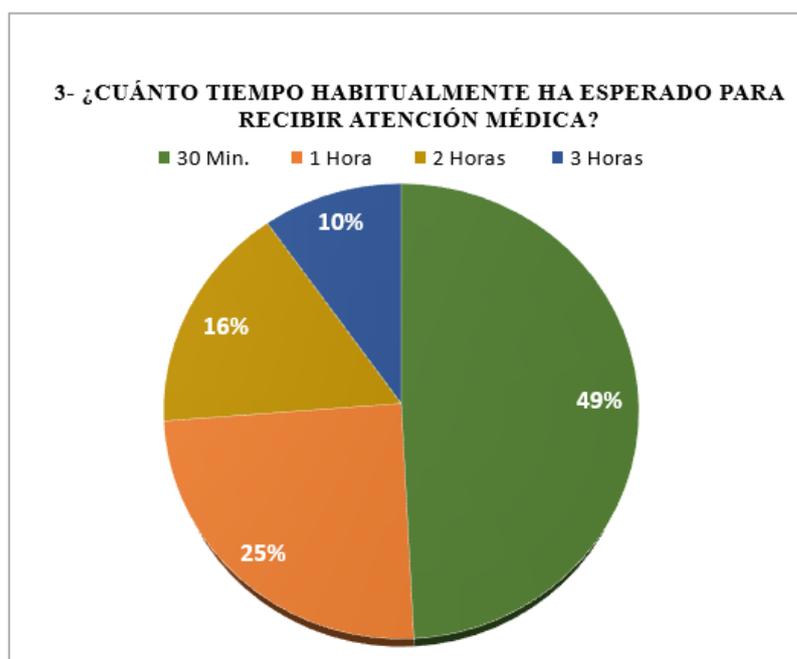


Gráfico 1- Tiempo para recibir atención médica

Fuente: (Autores)

Tabla 5- Prioridad según el estado del paciente

PRIORIDAD I	PRIORIDAD II	PRIORIDAD III	PRIORIDAD IV	NO PRIORIDAD
Inmediata	Muy Urgente	Urgente	Menos Urgente	No Urgente
Condición de salud extrema que amenaza la vida del paciente.	Condición de salud de alto riesgo que representa una amenaza al estado de salud del paciente.	Condición aguda, no amenazante de la vida.	Condición de salud que puede ser aguda, pero no compromete el estado general del paciente.	Condición clínica relacionada con problemas agudos o crónicos, sin evidencia de deterioro que comprometa el estado general del paciente.
Requiere una intervención médica inmediata.	Debe recibir atención médica inmediata.	Requiere consulta médica no inmediata.	No representa un riesgo evidente.	No representa un riesgo para la vida.
Atención inmediata con médico de la sala de emergencia.	Atención dentro de los siguientes 30 minutos con médico de la sala de emergencia.	Atención hasta 2 horas con médico y observación en sala de medicación o sala de emergencia (cuando hay necesidad de reposo en posición horizontal).	Atención de 2 hasta 4 horas y remisión a especialistas.	Atención, cita por consulta externa en ambulatorio o consulta médica agendada.

Fuente: (Servicio Nacional de Salud, 2017)

Tabla 6- Sistema de Triage Manchester

TABLA MODELO MANCHESTER DE TRIAJE			
Prioridad 1:	Atención	Inmediata	(Identificados con el color rojo).
Prioridad 2:	Atención	Muy urgente	Antes de 10 min, (color naranja)
Prioridad 3:	Atención	Urgente	Antes de 60 min, (color amarillo)
Prioridad 4:	Atención	Menos Urgente	Antes de 120 min, (color verde)
Prioridad 5:	Atención	No urgente	Antes de 240 min, (color azul)

Fuente: (Servicio Nacional de Salud, 2017)

En este mismo sentido, se destaca el proceso de almacenamiento de los historiales médicos. Después de completado el formulario de ingreso del paciente, se le asigna un número de expediente, y se adjuntan los documentos que forman parte de los procesos médicos que se le realizan al paciente, es decir, resultados de las analíticas, el formulario de ingreso al centro, las placas, sonografías y otros. Estos documentos son guardados en folders físicos con el número de expediente visible y es colocado en un gabinete (figura no. 29).



Figura 28- Gabinetes donde son guardados los historiales médicos.

Fuente: (Autores)

Durante la visita al centro, la encargada de archivos comentó lo siguiente: *“Existen tres áreas donde se guardan estos registros y que semanal se le proporciona una lista con las personas que tienen cita, con la finalidad de que ella y su equipo de 5 mujeres puedan ubicar los historiales médicos y entregarlos el día antes a cada doctor en su consultorio”*.

Con la finalidad de contrarrestar esta situación, se planea desarrollar por parte del departamento de tecnología un submódulo llamado “EXPEDIENTE” cuya finalidad es ofrecer a los especialistas de la salud un entorno digital donde se puedan consultar los historiales médicos de sus pacientes, evitando demoras por no recibir a tiempo toda la documentación.

Un caso que llama la atención es el seguimiento al historial de las embarazadas menores de edad. Estas deben retener todos los análisis, diagnósticos y documentación que forman parte del historial médico, y las mismas deben guardarlo y entregarlo al momento de ser ingresadas para realizarse la cesárea o parto normal.

A raíz de la gestión manual de los historiales médicos, el valor de la información que está contenida en ellos no es medible. Es evidente que la digitalización de estos ofrecería múltiples beneficios, desde la reducción de tiempo como la incorporación de nuevas tecnologías que permitirían generar estadísticas e informes que sustentarían la toma de decisiones y desarrollo de los procesos.

Por tanto, entre las situaciones que pueden contrarrestarse con la implementación de esta propuesta están:

- Predicción de brotes de virus: lo cual según los encuestados el 91% está a favor del desarrollo e implementación de un sistema que pueda predecir brotes de virus antes de que estos se propaguen y dejen mayores consecuencias.
- Reducción de tiempo en el ingreso de los pacientes.

- Predecir el tiempo estimado por consulta.
- Reducir la población de pacientes a través del seguimiento a distancia.
- Automatización de las consultas de los historiales médicos.



Gráfico 2- Interés en predecir enfermedades

Fuente: (Autores)

Como se observa en el gráfico no. 2 el 91% de los encuestados está a favor de un sistema para predecir enfermedades y apenas el 9 % no está de acuerdo.

4.3 Procesos para la planificación del presupuesto

El presupuesto del Hospital Regional Taiwán es financiado por dos fuentes, la primera de ellas por su naturalidad de organización pública proviene de los fondos del Estado Dominicano o como es comúnmente conocido “Fondos operativos”. La cantidad que reciben por parte del gobierno es de 4,356,060 millones de pesos dominicanos mensualmente de los cuales el 40% está destinado a la dirección de PROMESE/CAL que se encarga de suministrar los medicamentos esenciales a los centros de salud públicos de la República Dominicana. El 60% restante es empleado para cubrir los materiales que no son proporcionados por PROMESE/CAL, como son los materiales gastables en las salas de cirugías, en las reparaciones de las infraestructuras, combustible, cocina, higiene y otros gastos.

La otra fuente de financiamiento es por la venta de servicios que consiste en los ingresos que recibe el hospital de manera interna a través de las aseguradas de salud, y representa alrededor del 40% del presupuesto esto es equivalente a 2,200,000 millones de pesos dominicanos que ingresan mensualmente.

El proceso de elaboración del presupuesto se realiza de manera manual, salvo los reportes que son enviados al Servicio Nacional de Salud, que son los que se encargan de proporcionarle el presupuesto al centro de salud. El departamento de contabilidad cuenta con unas plantillas de Excel y un libro de banco (figura No. 30) donde se registran todos los gastos manualmente y que mensualmente son digitalizados en un documento de Excel estándar que es proporcionado a los centros de salud públicos.

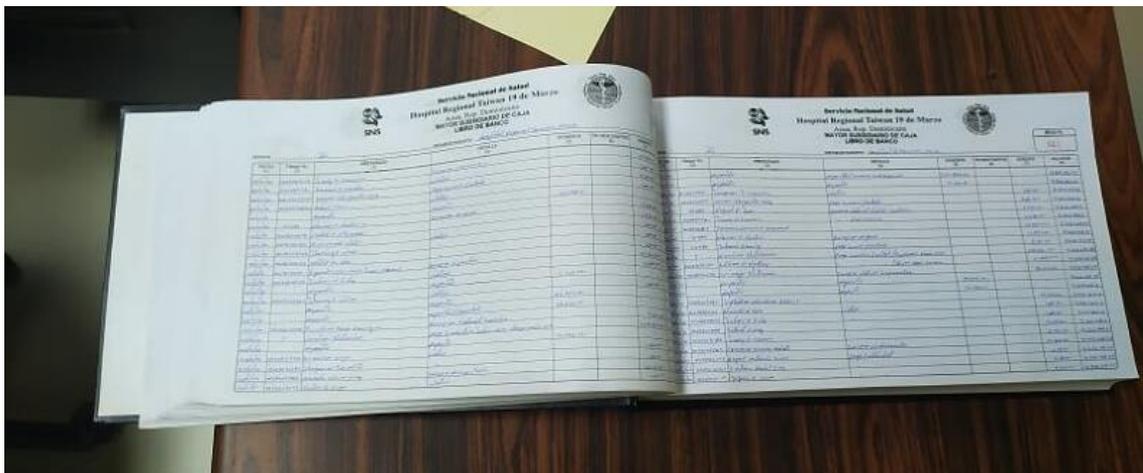


Figura 29- Libro de Banco

Fuente: (Autores)

Debido a la ineficiencia del proceso de elaboración del presupuesto, el 93% de los encuestados entiende que es favorable automatizar la elaboración del presupuesto y apenas un 7% no estaría de acuerdo (grafica No. 3). Mientras que el 80% de los encuestados considera altamente importante un sistema que pueda predecir el presupuesto (ver gráfico No. 4) debido a que en varias ocasiones según Nelson Gonzales administrador del centro de salud, este supera los ingresos que reciben de las dos fuentes de financiamiento, representando un déficit en la gestión administrativa por falta de recursos monetarios.

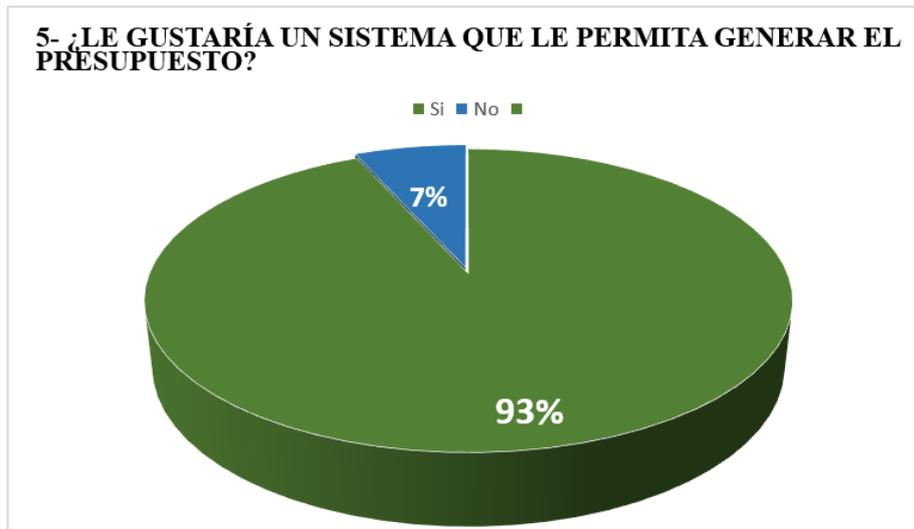


Gráfico 3- Necesidad de un sistema para generar el presupuesto

Fuente: (Autores)

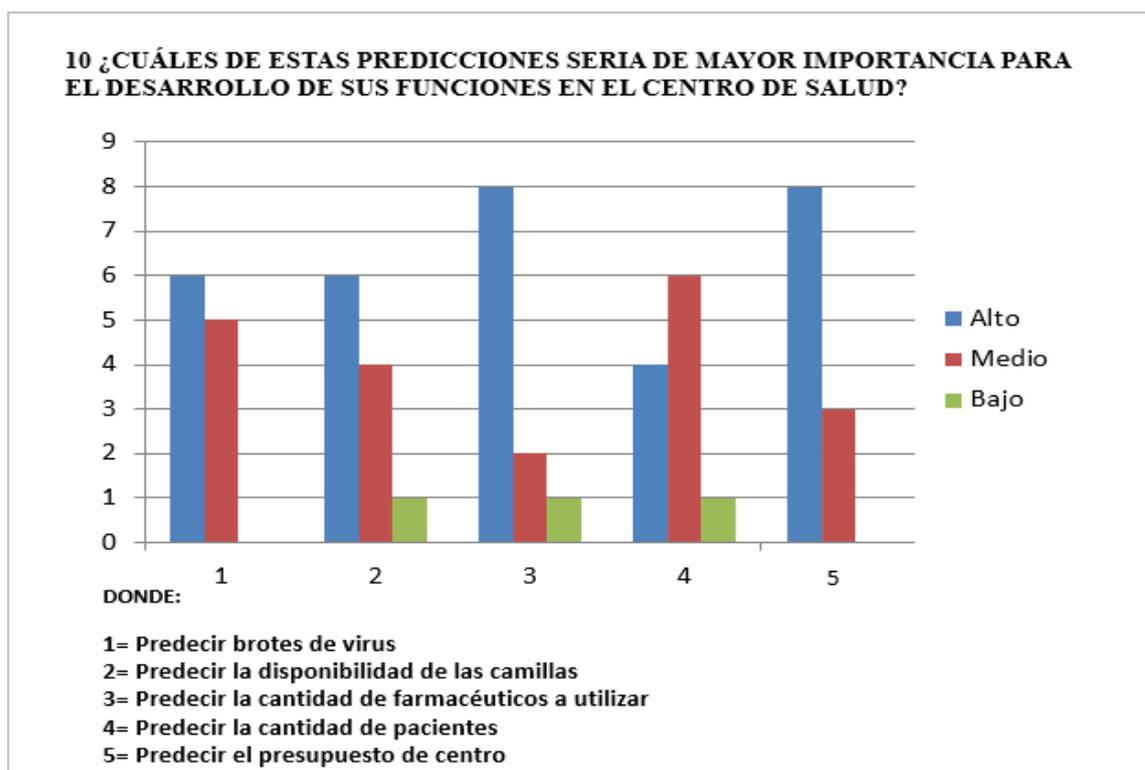


Gráfico 4- Predicción de eventos médicos/administrativos en el hospital

Fuente: (Autores)

Con la implementación de la propuesta que se plantea en el presente documento, existen diversas áreas que pueden ser beneficiadas, como es el departamento de finanzas donde urge la necesidad de mejorar los procesos que allí se realizan y queda demostrado cuando el administrador Lic. Gonzales expresaba que en muchas ocasiones deben buscar financiamiento en el sector privado, el cual posteriormente deberá ser saldado.

Por tanto, entre los procesos que impactarían una gestión eficiente de los recursos económicos del Hospital Regional Taiwán 19 de Marzo, se pueden destacar:

- **Gestión de las despensas de la cocina.**

Durante el levantamiento de información en el Hospital, el personal explicaba que prácticamente todo el presupuesto en palabras del encargado de informática “Se les iba por la cocina”. Por esta razón, la directora solicitó al departamento de informática la incorporación de una herramienta tecnológica y de ahí surge el módulo de la cocina cuya finalidad es el registro de los productos que son obligatorios para abastecer las necesidades de alimentación del centro. Este módulo, genera un inventario sobre lo que se ha comprado y consumido, permitiéndole a la directora consultar información relacionada a esto. Sin embargo, se observó que algunos productos consumidos no están siendo registrados por el personal correspondiente.

Por tanto, con la incorporación de este sistema de Analítica de Big Data, se automatizará de forma eficiente todo este proceso y tendrá como propuesta de valor un análisis predictivo de los productos de mayor consumo, evitar compras innecesarias, proveedores confiables, funcionalidad que de acuerdo con las encuestas el 100% los encuestados están a favor (gráfico No. 5). Asimismo, definir los gastos mensuales, debido que en ocasiones el flujo de pacientes puede aumentar como disminuir y así poder optimizar los gastos y aprovechar al máximo los insumos.



Gráfico 5- Definir proveedores confiables

Fuentes: (Autores)

- **Gestión de medicamentos**

Otra situación que afecta los ingresos financieros que recibe el hospital, es el alto porcentaje de recursos económicos que es destinado a la adquisición de medicamentos, tal y como se resaltó anteriormente el 40% del presupuesto recibido por el Servicio Nacional de Salud está destinado a la compra de medicamentos. Dicha situación, afecta directamente a la gestión administrativa y médica que son realizadas en el hospital.

Es, por tanto, que con la implementación de esta propuesta con la cual se realizará un análisis predictivo de diversos aspectos como es el surgimiento de un nuevo brote de un virus, indicaría cuales medicamentos serán más demandado en las diferentes etapas del año, agilizando así la gestión de estos y, por ende, se reflejara una optimización en el manejo del presupuesto.

4.4 Sistema de Gestión de Información del Taiwán

Como en la mayoría de los centros de salud públicos en la República Dominicana, no existen sistema que faciliten la gestión eficiente de los historiales médicos y todos los procedimientos que conlleva la gestión de un centro de salud, el Hospital Taiwán no es la excepción. Los historiales médicos actualmente son manejados en papel y archivados físicamente para su posterior búsqueda y consulta. Tampoco existe un sistema de citas y cada paciente tiene que esperar a ser atendido en sentido estricto, un proceso que puede tardar horas.

De igual manera, se presentan inconvenientes para elaborar el presupuesto del centro debido a que los procesos que complementan a este no están digitalizados y no se lleva un control absoluto.

En el 2013, el departamento de tecnología del hospital gracias a los avances tecnológicos propone un sistema para la gestión de los procesos administrativos y médicos que se realizan en el centro. Este lleva como nombre *Sistema de Gestión Hospitalario* (SISGEHOSPI) y el principal objetivo que se perseguía era poder generar un formulario único para cada paciente, como paso fundamental para estructurar el historial médico de estos.

A pesar de todos los esfuerzos por parte del departamento de tecnología, el proyecto aún sigue en fase de desarrollo a penas el 25% está en funcionamiento, ya que existen factores que no dependen de este, como es la gestión administrativa, la disposición de adaptarse al cambio por parte del personal, como en un momento nos expresaba el encargado de tecnología el Lic. Vázquez “*La mayoría de los doctores son tecnofóbicos, representando esto una barrera para la implementación del sistema*”.

En la actualidad el SISGEHOSPI está compuesto por 3 módulos y 14 submódulos:

Módulo de Gestión

Este módulo de gestión está encargado de la administración de los pacientes y de los servicios que son brindados en la institución. Está conformado por los siguientes submódulos de los cuales sólo los dos primeros están en funcionamiento:

- ✓ Pacientes
- ✓ Usuarios
 - Expedientes
 - Tipos de servicios
 - Servicios
 - Médicos
 - Citas

Módulo de Admisiones-prestación de servicios

Este módulo de admisiones-prestación de servicios, está encargado de los procesos que están involucrados con las ARS. Su objetivo principal es generar reportes relacionados a las productividades facturadas, médicas, servicios y reporte de ARS. En este último se listan las personas que poseen seguros de salud y que han recibido atención en el centro.

Está conformado por los siguientes submódulos:

- ✓ Prestación de servicios
- ✓ Consultas
- ✓ Reportes

Módulo de Gestión administrativa

Este módulo de gestión administrativa está encargado de los procesos administrativos como lo dice su nombre, se realizan los inventarios de la cocina exclusivamente, registro de usuarios y sus roles. Está conformado por los siguientes submódulos:

- ✓ Informática
- ✓ Administrativo financiero
- ✓ Cocina
- ✓ RRHH

A pesar de la existencia del sistema SISGEHOSPI desde el 2013, no todo el personal posee conocimiento de este (gráfico No. 6), debido a que sus funcionalidades son mínimas y por ende no ofrecen un beneficio cuantificable.

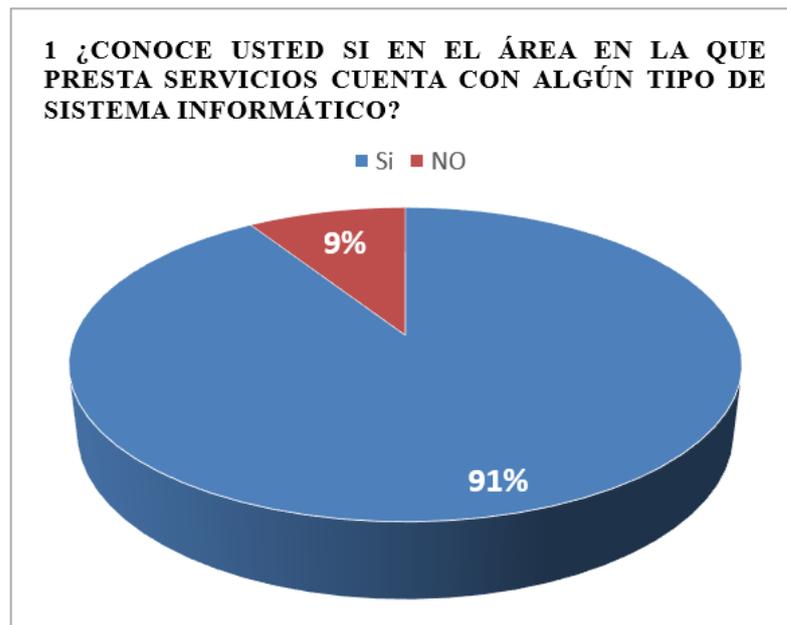


Gráfico 6- Conocimiento del SISGEHOSPI

Fuente (Autores)

Como se puede observar en el gráfico no. 6 el 91% de los encuestados está consciente de la existencia de un sistema informático en el centro, aunque no interactúe directamente con este, como es el caso de los pacientes porque no tiene ninguna funcionalidad con las que este pueda interactuar, mientras que el 9% desconoce la existe de uno.

4.4.1 Infraestructura de Hardware y Software del Centro de Salud

Actualmente este sistema hospitalario está habilitado para funcionar de manera local y se encuentra instalado en cada estación de trabajo de las áreas que la utilizan. A continuación, se mencionan las especificaciones del servidor del Hospital Regional Taiwán y las especificaciones de los ordenadores que tienen instalado el Sistema de Gestión Hospitalario.

Especificaciones de Servidor

- Windows Server 2012 64 bits
- .NET Framework 4.0

- SQL Server Setup Support Files
- Microsoft Windows Installer
- SQL Server Express y SQL Server 2008 R2
- Microsoft Common Language Runtime 4.5

Hardware

- Intel Processor Core I7
- 32 GB de Memoria RAM
- 2 TB de SSD Hard Drive

Las especificaciones básicas de ordenadores equipadas con SIGHOSPI

- 2 RAM o superiores
- Procesadores Intel i3
- Disco duro de 100 GB
- Sistema Operativo Windows 7

4.4.2 Entorno y plataforma de desarrollo

El sistema actual tiene una estructura básica desarrollada en el lenguaje de programación C # de la compañía Microsoft. Este software se compone de varios módulos como ya se ha planteado en esta investigación, los mismos han sido creados a medida que han ido surgiendo las necesidades en el centro médico. Para el manejo de los datos generados y guardado en el sistema SISGEHOSPI, se administran y manipulan mediante el software SQL Server Express que a su vez también proviene de compañía de Microsoft, el cual se realizan sólo funciones básicas como consultas, creación de nuevos campos y limpieza de los datos.

4.4.3 Versiones del sistema

El sistema desde su inicio ha experimentado 13 versiones, las cuales no se encuentran documentadas en su totalidad ya que los cambios han sido mínimos.

4.4.4 Seguridad del sistema

El SISGEHOSPI funciona por medio de la intranet del hospital, el mismo no cuenta con una estructura de seguridad muy sofisticada, sino más bien que utiliza los medios de control integrados en el mismo Windows Server, tales como su sistema de firewall. Al igual que la creación de usuarios, asignación de roles y demás características del Active Directory el cual asigna los diferentes permisos y accesos que tendrán los usuarios dentro del sistema, permitiendo así que cada uno tenga los permisos a las tareas referentes a su posición de trabajo, evitando cualquier fuga de información. En la misma dirección, el servidor está programado para que las computadoras cierren sección cada 5 minutos de inactividad para reducir la posibilidad de la infiltración de cualquier persona no autorizada.

4.4.5 Copias de Seguridad

Como un control correctivo rutinario para asegurar la calidad y la integridad de las informaciones, se programó una copia de seguridad automática todos los días a las 6 de la tarde en su servidor. Estos se copian mensualmente en un disco externo de 2 terabytes con el fin de salvaguardar los datos como método alternativo.

4.5 Operación y manejo para pacientes Emergencia

Para el caso de los pacientes admitidos en el área de emergencia actualmente en el Hospital Taiwán, se maneja haciendo una búsqueda en el sistema SISGEHOSPI para verificar si el individuo tiene consultas o ingresos previos. En caso de que el paciente no tenga ingresos o registros previos, entonces se procede a registrarlo en el sistema con el fin de generar su número de expediente e iniciar la creación de la historia médica exclusivamente en este centro.

Finalmente solventa su situación y se refiere con los médicos al área de necesidad para realizarle los procesos correspondientes, realizando y pautando su cita específica.

En caso del que el individuo posea un registro de historia médica, entonces se registra la visita realizada en el área de emergencia y este registro se anexa a esta. Así pues, a la hora de verificar y utilizar este historial entonces se cataloga como una unidad, es decir, que el paciente posee un solo número de historia médica, a los cuales se adjuntan varios procesos o visitas. En algunos casos si el paciente no aparece, el sistema permite crear un registro nuevo, lo cual puede generar duplicados, situación que el sistema permite solucionar unificando ambos registros con aquellas informaciones que no se repiten.

4.6 Operación y manejo de pacientes referidos

Para este tipo de pacientes se debe realizar una llamada a la dirección para verificar si el centro cumple con las condiciones de lugar para hacer la recepción del individuo/os. En este caso se verifican aspectos de lugar: si el centro maneja el trauma o proceso por el cual la persona será admitida, la disponibilidad de habitaciones, los mobiliarios y el equipamiento necesario para que el ingreso del paciente sea viable.

Si todas las condiciones previas están suplidas por el centro, entonces se admite como paciente y se maneja registrando en el sistema sólo como una consulta externa, guardando el proceso realizado, información del médico y las fechas del proceso.

En caso de que el centro no cumpla con las condiciones mencionadas entonces se procede a referirlo a otro centro.

4.7 Análisis FODA para la analítica de Big Data

Cuando se habla de recursos para evaluar la situación de las organizaciones en sus dos dimensiones, no se debe pasar por alto el FODA. Según Riquelme (2016): “El análisis FODA o DAFO es una herramienta de planificación estratégica, diseñada para realizar un análisis interno (Fortalezas y Debilidades) y externo (Oportunidades y Amenazas) en la empresa”.



Figura 30- Esquematización del Análisis FODA

Fuentes: (Autores)

A continuación, se describen aquellos factores que son parte del sistema actual que facilitarán la incorporación de la presente propuesta:

Fortalezas

Con el propósito de determinar su incidencia sobre la posibilidad de implementar la propuesta se utilizó una escala de valores siendo (1) el valor más bajo y (5) el más alto, iniciando con aquellos factores internos que facilitan la implementación de la propuesta, siendo estas las siguientes:

- Normas y disposiciones legales específicas que establecen la estructura y elementos integrales del historial médico y presupuesto como entidad del gobierno dominicano.
- Archivos físicos con el historial médico de los pacientes.
- Programas de modernización y remodelación de la planta física.
- Un departamento de tecnología comprometido y dedicado.
- Compromiso con el cumplimiento de los objetivos planteados por la Organización de las Naciones Unidas y la Organización Mundial de la Salud.

Oportunidades

Las oportunidades del proyecto se visualizan dentro de aquellas áreas a ser beneficiadas, generando un alto desarrollo en la efectividad de las actividades presentadas en la propuesta, donde actualmente este centro de salud carece de procesos totalmente eficientes para el manejo de estos.

- Promover la centralización de los hospitales de la República Dominicana.
- Fortalecimiento en las plataformas de las entidades de salud.
- Predecir pronósticos sobre aspectos de interés en el área de la salud.
- Experiencias y resultados obtenidos por centros de salud que funcionan bajo un modelo de autogestión.
- Desarrollo de planes de gobierno electrónico.

Debilidades

Luego de realizar el diagnóstico de la situación actual, se encontraron aspectos que fueron categorizados como debilidades, por lo que se incluyen:

- Deficiencia tanto en la gestión de los historiales médicos como la elaboración del presupuesto.
- Carencia de personal capacitado para el manejo de los recursos tecnológicos (Informáticos).

- Resistencia al cambio por parte del personal operativo.
- Personal médico con escasa afinidad con los historiales médicos y la aplicación de la tecnología en los procedimientos médicos.
- Personal administrativo con mínima relación a los procesos automatizados para la definición del presupuesto.
- Sistemas desactualizados que no brindan los mejores beneficios a la gestión en general del hospital.

Amenazas

Al igual que las oportunidades, las amenazas forman parte del análisis del entorno externo. Por lo que existe la posibilidad que situaciones externas sean capaces de afectar de manera negativa al proyecto si no se toman en cuenta.

- Baja de inversión en la infraestructura tecnológica tanto de hardware como de software.
- Deficiencia en el proceso de comunicación y seguimiento de las unidades de atención primarias (UNAP) con el hospital.
- Proceso de digitación de todos los historiales médicos físicos con todas sus informaciones.
- El presupuesto asignado al centro para la agilización de la implementación del nuevo sistema.
- Carencias de oportunidades para la capacitación del personal del hospital para el uso de nuevas herramientas.



Figura 31- Resumen del análisis FODA

Fuente: (Autores)

4.8 Presentación y Análisis de la información recopilada a través de encuestas

Durante el proceso de levantamiento y recopilación de información en este centro de salud relacionadas con la gestión de los procesos que influyen en los historiales médicos como en el presupuesto se definieron tres tipos de encuestas según el tipo de individuo, es decir, encuestas al personal de la profesión de medicina, personal administrativo y pacientes.

Se observó a través de las encuestas realizadas que el 82% del personal médico posee conocimiento de los historiales médicos electrónicos (HME) mientras sólo un 18% no conoce el concepto. Por tanto, es una fortaleza que puede aprovecharse en esta propuesta.

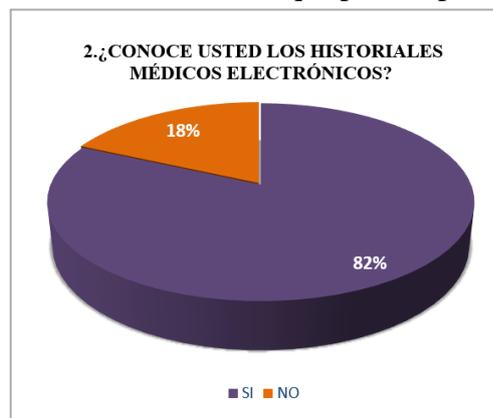


Gráfico 7- Conocimiento sobre los HME

Fuente: (Autores)

En comparación al conocimiento sobre lo HME, existe una gran diferencia al momento de cuestionar a los especialistas de la salud en lo referente a las tecnologías aplicadas a los HME como puede ser la analítica de Big Data y queda reflejado que sólo el 27% tienen conocimientos (gráfico No.8), sin embargo, el 73% expresa desconocimiento de esta herramienta y sus aplicaciones en el sector salud.

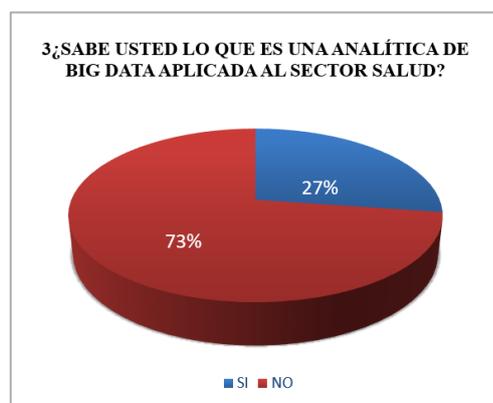


Gráfico 8- Conocimiento sobre la Analítica de Big Data

Fuente: (Autores)

A pesar de que el centro no cuenta con grandes avances tecnológicos que efficienten los procesos, todo el personal médico comprende la necesidad de adaptar los servicios del centro médico a las nuevas tecnologías, es por esto por lo que el 100% está de acuerdo.



Gráfico 9- Adopción de nuevas tecnologías al sector salud

Fuente:(Autores)

Una vez más, queda reflejada la necesidad de automatizar la gestión de los historiales médicos cuando el 91% de los encuestados (gráfico no.10) sustentaron que les gustaría obtener de manera inmediata a los datos del historial de cada paciente, cuando apenas el 9% no estuvo de acuerdo.



Gráfico 10- Acceso inmediato a los Historiales Médicos

Fuente: (Autores)

Ampliando lo antes mencionado, los médicos al momento de ser consultados sobre cómo impacta la adopción de nuevas tecnologías para la predicción de eventos en el sector salud y consigo mejora de la atención del paciente como la propagación de un virus como el COVID-19, se puede observar en el gráfico no. 11 que el 66% opinó que el uso de tecnologías en el sector salud podría considerarse en un 75%, ya que existen procesos que

son obligatorios que un especialista realice, el 34% entendía que en un 50% era oportuna la adopción de estas, mientras que nadie opinó que incidía en un 25%.

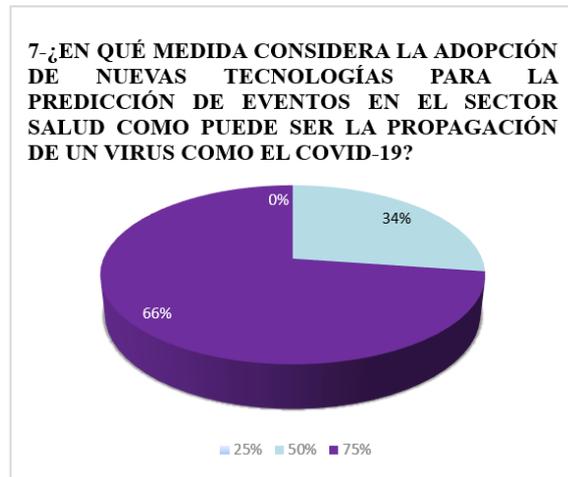


Gráfico 11- Adopción de tecnologías para predicciones

Fuente: (Autores)

Otro factor tomado en consideración durante el levantamiento de información en el centro de salud fue el uso de medios tecnológicos para el trabajar con los pacientes presentado en el gráfico no. 12, y se obtuvo como resultado que sólo el 36% de los doctores afirmaron esta expresión, mientras que 64% expresó que no utiliza medios tecnológicos para el desempeño de sus funciones.

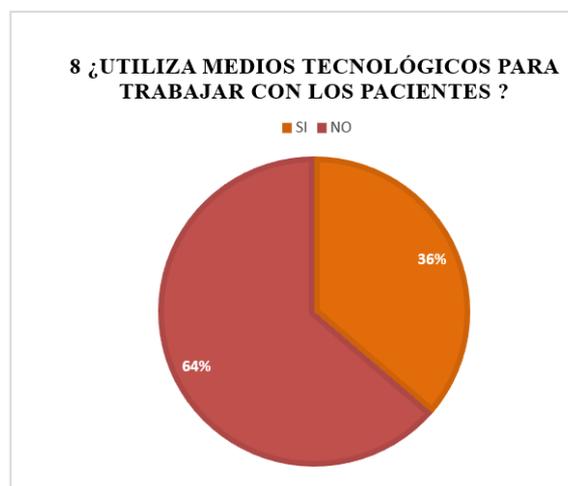


Gráfico 12- Uso de tecnologías para el trabajo

Fuente: (Autores)

Por otra parte, el 91% de los encuestados demostró estar de acuerdo con el desarrollo e implementación de un sistema que genere estadísticas que permitan a la directiva tomar decisiones que influyan en los procedimientos médicos, mientras que el otro 9% expresó no estar de acuerdo.

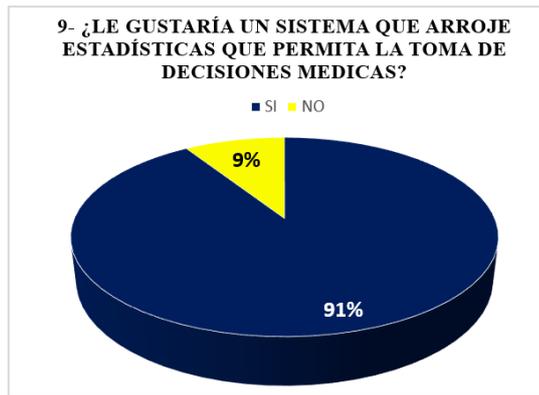


Gráfico 13- Sistema que genere estadísticas

Fuente: (Autores)

Uno de los factores a contrarrestar con la propuesta de la Analítica de Big Data es disminuir los impactos generados por las enfermedades que las personas que acuden con regularidad al hospital padecen, se indagó acerca de las enfermedades más frecuentes y como se puede observar en el gráfico no. 14 dentro de las 5 principales afecciones que se presentan a menudo en el centro médico son dengue, diabetes, hipertensión y finalmente malaria.

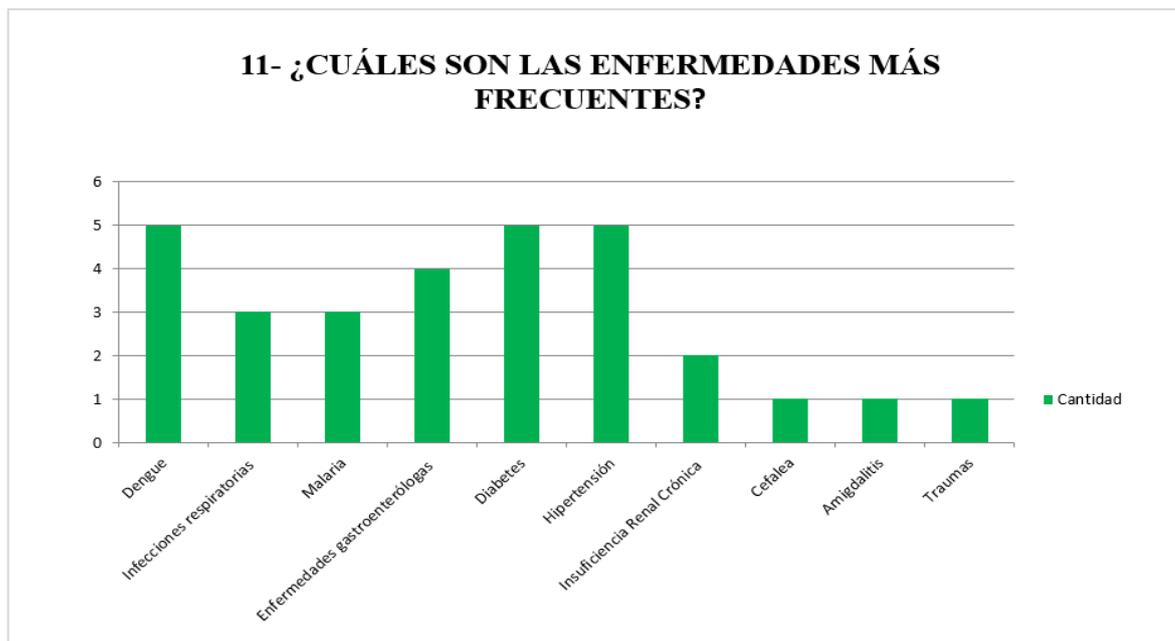


Gráfico 14- Enfermedades más frecuentes en el Hospital

Fuente: (Autores)

Se tomaron en consideración los principales beneficios que ofrecen los sistemas que aplican la Analítica de Big Data sobre los historiales médicos y de acuerdo con el gráfico no. 15 más del 60% los encuestados opinaron que el grado de importancia que radica en estos sistemas era alta.

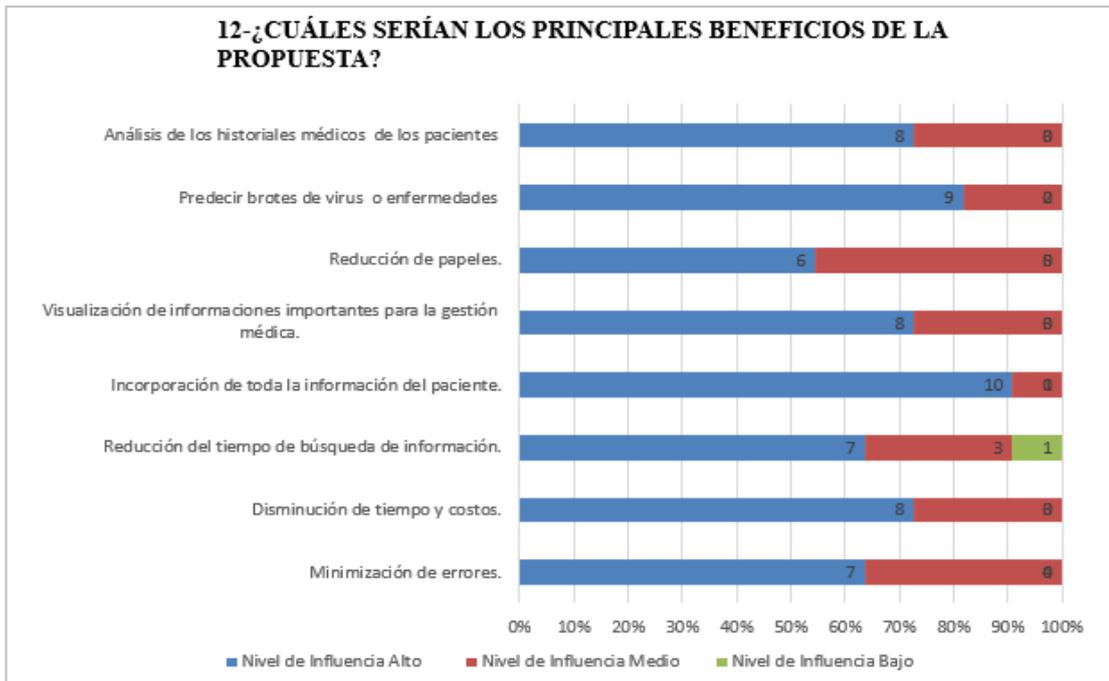


Gráfico 15- Beneficios de la propuesta

Fuente: (Autores)

Además de las encuestas realizadas a los especialistas de salud, fue tomado en cuenta el personal administrativo. Actualmente el Hospital Regional Taiwán presenta un tiempo de respuesta en el proceso de otorgar turnos variado, de manera que 7% del personal está de acuerdo que para otorgar turnos a los pacientes se toman de 0 a 4 minutos, el 29% opina que de 5 a 10 minutos y finalmente el 73 % de los encuetados concuerda que toma más de 15 minutos, según el gráfico no. 16.

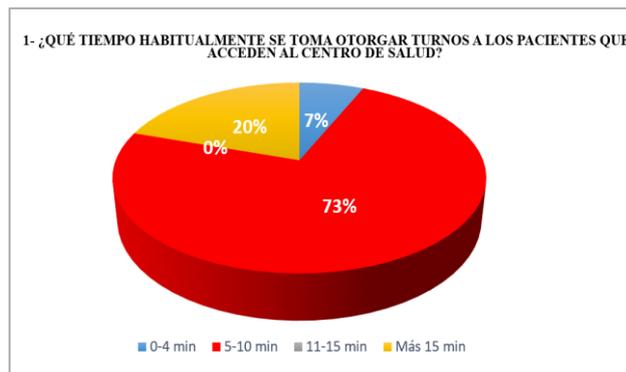


Gráfico 16- Tiempo en otorgar turnos

Fuente: (Autores)

Otra variable que se sometió a escrutinio estadístico fue la demora en canalizar y legalizar la admisión de un paciente, como se muestra en el gráfico no.17, el 53% estuvo de acuerdo en que toma de 5 a 10 minutos, el 40% de los encuestados que toma de 11 a 15 minutos y finalmente el 7% que toma más de 15 minutos.

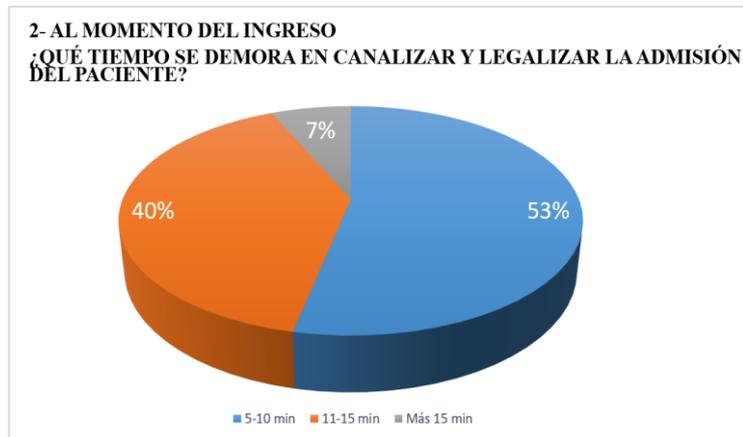


Gráfico 17- Demora en canalizar y legalizar admisión

Fuente: (Autores)

En este mismo orden, se consultó la cantidad de personas que ingresan por día al hospital y según los encuestados estos fueron los resultados: un 54% está de acuerdo que se realizan un promedio de 10 ingresos, un 33% concuerda que son más de 20 ingresos, el 13% restante dice que son 20 ingresos (gráfico no. 18).

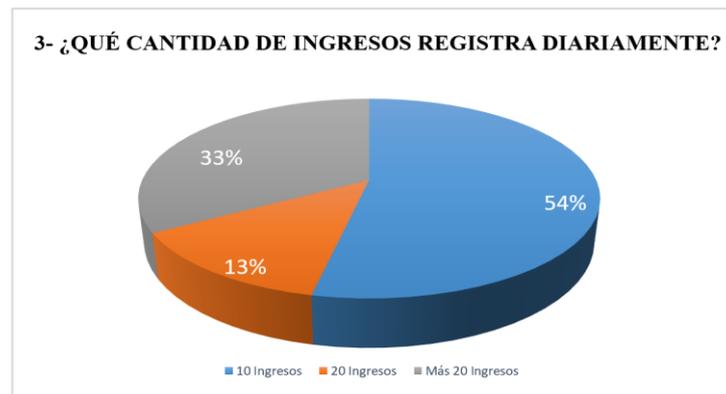


Gráfico 18- Ingresos por día

Fuente: (Autores)

También el personal administrativo en un 53% concuerda en que se utilizan controles estadísticos haciendo uso de Tics y medios convencionales como el papel, el 27% que sólo se utilizan controles automatizados (usando computadores o Tics), el 20% que esto se hace de manera manual o que no está automatizado. Véase gráfico no. 19



Gráfico 19- Proceso de control estadístico

Fuente: (Autores)

Al momento de preguntarle a los encuestados sobre las herramientas tecnológicas que ayuden al personal a evaluar el trabajo y las metas cumplidas a través de estadística que aporten y apoyen a la toma de decisiones administrativas se puede observar en el gráfico no. 20 que 100% estuvo de acuerdo.



Gráfico 20- Sistema para arrojar estadísticas

Fuente: (Autores)

Tomando en consideración que uno de los beneficios de la aplicación de Analítica de Big Data sobre los historiales médicos es aportar al desarrollo de nuevos productos farmacéuticos a través de los diagnósticos de los pacientes, el 94% del personal administrativo está de acuerdo con que en el Hospital Taiwán se construya un sistema que permita el desarrollo de medicamentos más eficientes, mientras que sólo el 6% no está de acuerdo según el gráfico no. 21. De manera que se nota que el personal está conforme con la incorporación de tecnologías que aporten a la mejorar la calidad de la entidad.

7- ¿LE GUSTARÍA UN SISTEMA QUE PERMITA EL DESARROLLO DE NUEVOS PRODUCTOS FARMACÉUTICOS MÁS EFICIENTES?

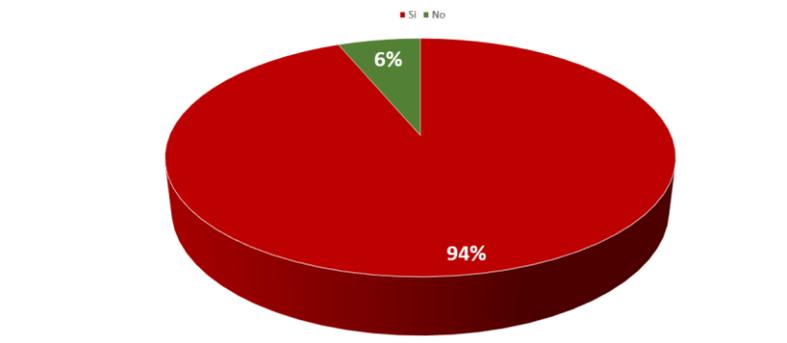


Gráfico 21- Desarrollo de nuevos productos

Fuente: (Autores)

Por último, se presentan los resultados de las encuestas realizadas a los pacientes del Hospital Regional Taiwán 19 de Marzo Azua. En lo referente a la satisfacción en el trato que recibieron en el servicio de asistencia médica según el gráfico no. 22 el 87% de los encuestados opinan que recibieron un trato amable, por el contrario, solo un 13% expresaba insatisfacción.

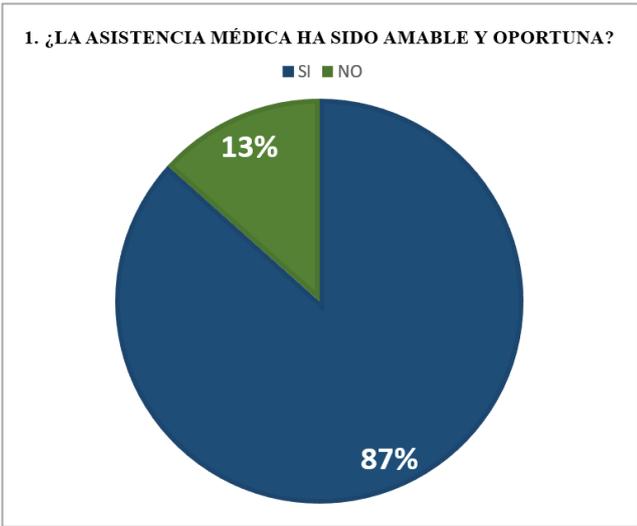


Gráfico 22- Asistencia amable y oportuna

Fuente: (Autores)

Es evidente que todo paciente que ingrese a un centro de salud espera ser atendido con prontitud. Existen varias razones, intolerancia al lugar, sentido de urgencia, cece de procedimientos, y queda demostrado cuando el 100% de los encuestados estuvo de acuerdo, ver gráfico no. 23. Por tanto, esto se convierte en una las prioridades que se deben tomar en cuenta al momento de hacer cambios en la gestion de procedimientos del hospital.

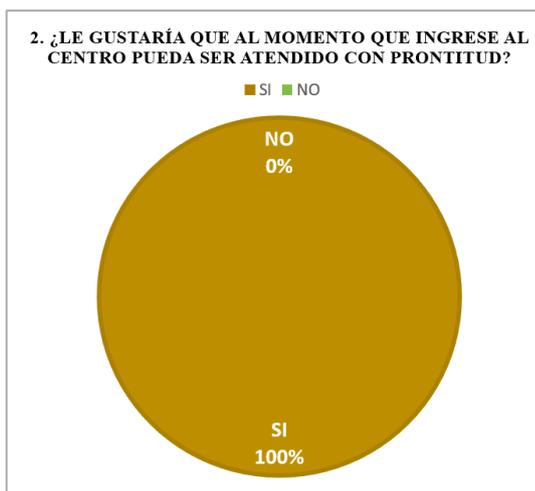


Gráfico 23- Atención en el centro de salud

Fuente: (Autores)

Para concluir el 100% de los encuestados como se muestra en los gráficos no. 24, 25 y 26 estuvieron a favor de un sistema que pueda ofrecer las siguientes funcionalidades:

- Predecir a tiempo virus como el covid-19 o enfermedades
- Desarrollar acciones preventivas para enfermedades
- Recibir un diagnóstico inmediato y veraz de su estado de salud

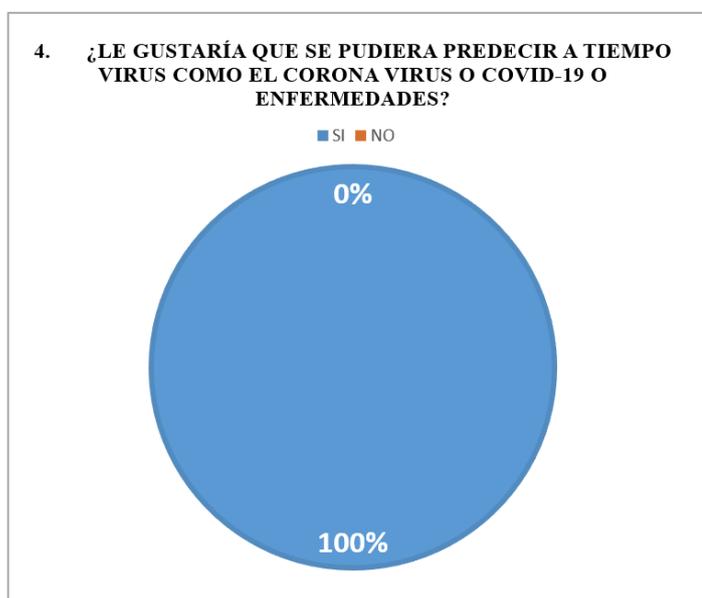


Gráfico 24- Predicción a tiempo de virus

Fuente: (Autores)

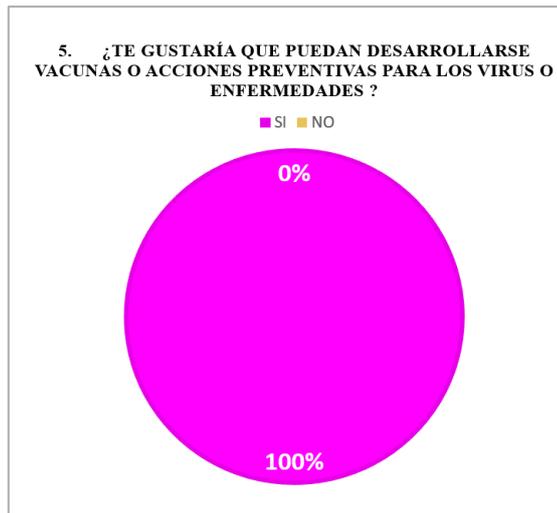


Gráfico 25- Desarrollo de vacunas o acciones preventivas

Fuente: (Autores)

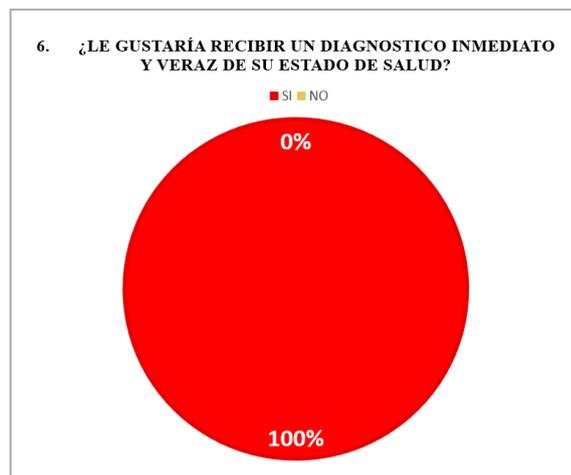


Gráfico 26 Recibimiento de un diagnóstico inmediato

Fuente: (Autores)

4.9 Resumen del analisis e interpretacion de los resultados de las encuestas

Luego de concluida la etapa de encuestas y entrevistas y una vez procesada la información relacionada a la gestión del Hospital Regional Taiwán 19 de Marzo, se identificaron una serie de problemas que trae consigo aspectos negativos que inciden directamente en la calidad de los servicios que se ofrecen.

Los principales resultados obtenidos a través de las herramientas de investigación utilizadas que evidencia la problemática actual por la cual atraviesa el hospital Taiwán, son:

- **Deficiente gestión de la documentación y archivos de los historiales médicos de los pacientes.**

En la actualidad los historiales médicos son trabajados de manera manual, lo que implica que los doctores deben completar toda la información de cada paciente a puño y letra. Esto afecta directamente en la calidad del servicio, porque conlleva que los doctores se demoren y los pacientes deban esperar hasta que se termine de completar este proceso. Además, existe la posibilidad de que no se entienda lo que se ha escrito, pero no sólo esto sino el riesgo que representa que los historiales médicos se almacenen en lugares físicos como se pudo observar anteriormente.

De igual manera, hay que destacar que el personal desconoce la importancia que radica en los historiales médicos y de las tecnologías que existen para la explotación de esto y queda demostrado que solo el 27% de los encuestados conoce de algunas herramientas.

También cabe mencionar de cara a los pacientes, que el tiempo de espera para ser atendidos suele prolongarse por largos periodos, cuestión que es poco favorable para los mismos. Este es uno de los motivos por el cual, estos están de acuerdo de que se implemente un sistema que, agilice el manejo de los historiales médicos y a su vez la asignación de turnos y citas de manera más efectiva para reducir estas tediosas esperas.

Por último, el hospital recibe diario más de 200 pacientes y la mayoría de estos con las siguientes condiciones de salud: dengue, malaria, diabetes, hipertensión y problemas estomacales. Por tanto, esto implica que deben estar preparados para dar la atención oportuna a estos pacientes, pero esta información sólo es manejada por los especialistas de la salud y no involucran a la alta directiva que son quienes toman las decisiones. Es por esto, que esta propuesta a fin de mejorar la condición de la gestión médica del Taiwán permitirá que todo el personal conozca sobre la información mencionada anteriormente y no sólo esto, sino que puedan ver la analítica que se realiza a los historiales médicos y conocer cuáles serían los futuros casos que pueden encontrarse durante la operatividad del día a día.

De tal modo, que con un sistema que realice predicciones de enfermedades permitirá que la directiva tome las mejores decisiones y el presupuesto no se vea afectando.

Otra de las problemáticas que afecta la gestión del hospital recae sobre la administración y es la definición del presupuesto, según cuenta la directora de este centro el 40% del presupuesto es destinado a la compra de fármacos. De igual manera, el abastecimiento de la cocina es otro de los grandes gastos que debe enfrentar, al punto que solicito asistencia del personal de tecnología.

Por lo que estas dos problemáticas que inciden en el buen manejo de los recursos económicos del centro de salud son variables fundamentales que pueden mejorar esta gestión.

CONCLUSIÓN

Al finalizar este capítulo se establecen puntualmente los aspectos institucionales del Hospital Regional Taiwán 19 de Marzo, se sabe por tanto que este proyecto vino como resultado de los esfuerzos del gobierno dominicano y las relaciones bilaterales con la República de Taiwán, de ahí su nombre.

En el mismo están descritos los problemas y deficiencias que presenta el centro, así como las fortaleza y oportunidades. Se detallan los sistemas usados tanto para la historia médica y la gestión del presupuesto. No se debe pasar por alto que se realizaron varias encuestas y entrevistas que incluyeron el personal administrativo, personal médico y en último lugar los pacientes.

En conclusión, se ha expuesto de manera resumida los pros y contras de los dos factores que impiden una buena gestión, es decir, el presupuesto para la toma de decisiones administrativas y la parte de registro de información personal de los pacientes.

CAPÍTULO V: PROPUESTA DE UN DISEÑO DE UN SISTEMA DE INFORMACIÓN PARA LA TOMA DE DECISIONES MÉDICAS Y ADMINISTRATIVAS EN EL HOSPITAL REGIONAL TAIWÁN 19 DE MARZO DE AZUA, R.D. APLICANDO ANALÍTICA DE BIG DATA SOBRE EL HISTORIAL MÉDICO DE LOS PACIENTES.

INTRODUCCIÓN

Hoy en día el sistema de salud dominicano tiene muchas carencias significativas, dentro de las cuales cabe mencionar el historial médico o clínico, que es el medio que se utiliza para guardar las informaciones del paciente y documentar los procesos por los cuales este que ha pasado. Este logra optimizar la relación médico-paciente, ya que hace posible saber la historia médica, es decir, la trayectoria e intervenciones que una persona ha tenido durante su vida. Este histórico da informaciones vitales para la toma de decisiones de cara a la medicina y futuros procesos a los que un paciente podría ser sometido.

A pesar de la gran importancia de estos, en el país no se aprovecha el potencial de las informaciones que estos guardan, así mismo carecen de un manejo eficiente de la misma, ya que suelen extraviarse, duplicarse, destruirse incluso no estar disponibles en situaciones de emergencia. En el mismo ámbito, el manejo del presupuesto es un tema importante dentro de estas instituciones sanitarias, ya que es susceptible a muchas variaciones, por diversas razones provocando inestabilidad económica. Por consiguiente, es vital poder tener un buen manejo de estas distribuciones para el uso responsable del mismo.

En este capítulo se propone desarrollar un novedoso sistema para el análisis los historiales médicos electrónicos (HME), implementando una Analítica de Big Data mediante tecnologías especializadas a través de un sistema en línea. Este sistema funciona en tiempo real con las informaciones más actualizadas del paciente, lo que permite que la toma de decisiones sea efectiva y certera, siendo el sistema capaz de utilizar todos los datos generados para producir nuevas informaciones insumo para la toma de decisiones. De igual manera la propuesta cuenta con un módulo para el manejo del presupuesto, con la finalidad de eficientizar la distribución de recursos.

5.1 Fundamentación de la propuesta

En las últimas décadas con la incorporación de la tecnología en diversas áreas como el sector salud, es decir, e-salud ha traído como efecto la generación de billones de datos sin importar la estructura de estos, ya que provienen de diferentes fuentes como son los historiales médicos electrónicos, redes sociales, diagnósticos médicos, resultado de analíticas, páginas web entre muchas más.

Asimismo, apegados a los objetivos de la agenda 2030 planteada por la ONU se destaca el objetivo número 3 “*Garantizar una vida sana y promover el bienestar para todos en todas las edades*”, donde radica la importancia de esta propuesta, ya que es uno de los elementos que componen las metas a trazarse por cada país miembro de la organización.

Según la Organización Mundial de la Salud (OMS), el avance de los países del continente americano en materia de e-salud es diverso. Si bien el 61% de ellos ya tiene una estrategia nacional de e-salud, muchos aún deben pasar de la fase de la formulación de políticas y estrategias de e-salud a la de implementación como es el caso de la República Dominicana.

Los datos disponibles de los países miembros de la OMS en la región, a través de su oficina regional y la Organización Panamericana de la Salud (OPS), muestran un panorama mixto de las tecnologías relacionadas con la e-salud, pues algunas de ellas muestran una tasa de adopción amplia y otras aún están en un estado incipiente. La tele-salud y los macrodatos, por ejemplo, se encuentran en una etapa inicial, siendo este último aspecto un factor decisivo para la aplicación de un sistema de Analítica de Big Data sobre los historiales médicos. Por tanto, con la implementación de este sistema tanto en el Hospital Taiwán como en el resto de los centros de salud en la República Dominicana representaría un avance de cara a las metas que como país se debe asumir.

Esta misma entidad sostiene que si bien un 52.6% de los países participantes tiene un sistema de información electrónico de salud (HIS), únicamente un 26.3% tiene legislación que apoya su uso en los sistemas nacionales, lamentablemente República Dominicana no posee un sistema de esta índole, lo que implica que no compite con el resto de los países americanos. Con este sistema de información para la toma de decisiones médicas y administrativas aplicando Analítica de Big Data sobre el historial médico de los pacientes, son varios los aspectos entorno a la e-salud que serán abordados, como es la implementación de historiales médicos electrónicos que de acuerdo con la OMS el 73.7% de los países miembros en esta región presentan factores que impiden este avance. De igual manera, dispositivos de monitorización a distancia, aplicaciones móviles, dispositivos personalizados para la salud, medicina a distancia y otras tecnologías aplicadas a la salud y la gestión de recursos económicos como sistemas de predicción de productos farmacéuticos como alimenticios.

En el plano nacional, el hospital Regional Taiwán es el centro de salud público principal de la región sur de la República Dominicana, sin embargo, se encuentra en condiciones

que le imposibilitan una gestión eficiente y eficaz. Lo cual radica en la ausencia de un sistema informático que permita una administración financiera adecuada como la gestión de los historiales médicos en formato electrónico. Asimismo, el flujo de pacientes que presenta este hospital es alto y la mayoría de estos con enfermedades como diabetes, dispersión alta, dengue, malaria y problemas estomacales, y al no poseer un sistema que pueda brindar datos precisos sobre los síntomas de las personas como su historia médica, la directiva del centro no recibe informaciones necesarias que permitan la toma de decisiones tanto de índole clínico como administrativas por ejemplo, determinar el presupuesto, más allá poder realizar estudios profundos sobre las distintas situaciones que se podrían presentar en el hospital.

Ante la situación, es evidente la necesidad que posee el hospital de la implementación de un sistema de información para la toma de decisiones médicas y administrativas aplicando analítica de Big Data sobre el historial médico de los pacientes. Dicha situación influye directamente en la manera cómo se maneja la información que se encuentra recopilada en los récords médicos de los pacientes como en los procesos financieros, dígame, la compra y venta de servicios.

De acuerdo con las capas o fases como pueden denominarse a las partes que componen una arquitectura de Analítica de Big Data, lo primero es definir las fuentes de datos. En este caso, la fuente principal son los historiales médicos y toda la documentación que es utilizada para la elaboración del presupuesto, estos pasaran hacer electrónicos para poder ser incorporados con mayor facilidad al sistema.

Otro aspecto que amerita automatización es la manera en que se consultan las historias médicas en el centro de salud, y es que, por encontrarse en formato físico, difícilmente se haga uso de ellos para dar lugar a un diagnóstico o el tratamiento de un paciente.

El aspecto más importante de la implementación de este sistema es la predicción de enfermedades con la finalidad de que se puedan tomar medidas preventivas como decisiones oportunas. Asimismo, la predicción de los factores que inciden en la determinación del presupuesto del centro de salud. Además, tendrá un valor agregado el cual es la creación de un entorno que se adapta perfectamente a las tecnologías que forman parte de la e-salud como la centralización de los datos en el sistema de salud dominicano. Esto último, permitirá que los pacientes puedan ser recibidos en los centros de salud y que su información se encuentre integra, confidencial y disponible para ofrecer un servicio de calidad.

El sistema de información para la toma de decisiones médicas y administrativas aplicando Analítica de Big Data sobre el historial médico de los pacientes colocaría al Hospital Regional Taiwán 19 de Marzo a la vanguardia en relación a los avances tecnológicos en el sector de salud y al mismo tiempo a la República Dominicana al nivel de países como Brasil, Cuba, Estados Unidos, México, Perú y demás países primermundistas, en el desarrollo de sistemas que forman parte de la Estrategia Nacional para la incorporación de la e-salud que propuso la Organización Mundial de la Salud a sus estados miembros. Al mismo tiempo, el sistema de salud dominicano podrá proporcionar herramientas para

favorecer la eficiencia en los procesos de asistencia médica, la gestión de recursos financieros, predicción de enfermedades, gestión de recursos farmacéuticos, recursos alimenticios, brindar un servicio de calidad a los pacientes, salvaguardar los datos y garantizar la interoperabilidad entre los centros de salud mediante el uso de estándares.

5.2 Presentación de la Propuesta

La siguiente propuesta tiene como objetivo principal diseñar una arquitectura de Analítica de Big Data para aplicarla a los historiales médicos de los pacientes del Hospital Regional Taiwán 19 de Marzo, con la finalidad de realizar análisis predictivo que permita la toma de decisiones orientadas a las gestiones médicas y administrativas. Al mismo tiempo, servir de modelo para la implementación de esta en otros centros de salud de la República Dominicana, desarrollar y orientar acciones que influyan en la realización de las metas trazadas por la ONU en el objetivo número 3 de la agenda 2030.

5.2.1 Documento Visión del Proyecto

5.2.1.1 Propósito

Con la presentación del Sistema de Gestión Médica y Administrativa SIGEMA propuesto en esta investigación se pretende demostrar el funcionamiento, alcance e implicaciones de este. Asimismo, busca que la República Dominicana pueda sentar las bases en materia de la salud predictiva y preventiva, a través del desarrollo de arquitecturas de Analítica de Big Data. En el mismo sentido, trabajar desde la perspectiva tanto médica como la administrativa mediante la incorporación de tecnologías que beneficien exponiendo y dando predicción a las enfermedades y afecciones que en este caso afectan la región Sur de la República Dominicana.

Por otra parte, se persigue organizar los recursos del presupuesto que el SNS destina para la región sur del país a través de la identificación de las mejores prácticas que provee el análisis predictivo en el Hospital Regional Taiwán 19 de Marzo.

Finalmente, que este proyecto sirva como modelo para organizar y centralizar un sistema de gestión de historiales médicos que pueda emplearse a nivel nacional y que la República Dominicana como miembro de la Organización Mundial de la Salud, sea un ejemplo en el desarrollo de sistemas orientados a la e-salud.

5.2.1.2 Alcance

El sistema propuesto es aplicable al Hospital Regional Taiwán 19 de Marzo, que es un centro de carácter público. Sin embargo, la idea es centralizar este proyecto a través de incorporación de las diferentes estancias de salud tanto públicas como privadas en la República Dominicana. Esto incluye hospitales, centros médicos especializados, laboratorios y centros de investigaciones.

5.2.1.3 Posicionamiento

Con la implementación total de este sistema en el Hospital Regional Taiwán 19 de Marzo, automáticamente se creará un modelo de referencia para todo el sector de salud de la República Dominicana. De tal modo que, se estandaricen los procesos médicos-administrativos y a raíz de ello, obtener la información necesaria para generar analíticas que permitan la toma de decisiones para la reducción del impacto económico producto de las secuelas ocasionadas por enfermedades.

Por otro lado, con la automatización de los historiales permitirá un entorno interoperable no solo en el sur del país sino en todo el territorio nacional. Esto por consiguiente genera credibilidad, calidad y mejora de los servicios de este sector. Todas estas ventajas aumentan la esperanza y la calidad de vida de los ciudadanos y se evitará la realización de procedimientos innecesarios representando un ahorro de dinero y tiempo tanto para el centro o laboratorio médico, como también para el paciente.

5.2.1.4 Descripción del personal involucrado

En esta sección se describen todas las personas que influyen en el sistema y cuáles son sus responsabilidades:



Figura 32- Personal Involucrado

Fuente: (Autores)

Persona que gestiona y asiste a los pacientes en aspectos de salud en las áreas de emergencia, consulta y procedimientos. Crea, consulta y modifica historiales clínicos electrónicos.

Se encarga de analizar las muestras médicas y ofrecer los resultados. Modifica y consulta los historiales clínicos electrónicos de acuerdo a lo que arrojan las pruebas.

Aquella persona que consume o utiliza los servicios de salud en las diferentes áreas de los centros de salud.

Aquel personal que asiste a los doctores con la atención a los pacientes en términos de urgencia de atención y según los recursos existentes y que necesitan los pacientes. Crea, consulta y modifica historiales clínicos electrónicos.

Aquel personal que gestiona el ingreso del paciente y de esta manera sirve como canalizador entre médico y paciente. Crea y consulta historiales clínicos.

Figura 33- Definición del personal involucrado

Fuente: (Autores)

5.2.1.5 Entorno de Usuario

El acceso al sistema es posible a través del uso de una conexión a internet, y un navegador web, mediante ordenadores de escritorio, portátiles, teléfonos y tabletas inteligentes. El sistema se diseña de manera tal que como se especifica en la descripción de personal involucrado, las funciones serán limitadas atendiendo a la naturaleza del usuario que ingrese para evitar confusiones y uso erróneo del sistema. Los informes generados por las consultas podrán ser guardados en formato PDF para su impresión o envío.

También, será posible la integración de dispositivos 'wearables' y la posibilidad de integrarlo mediante el uso de tecnología IoT. Otro aspecto por tomar en cuenta son los niveles de seguridad que serán definidos a través de políticas y controles.

5.2.2 Especificaciones del diseño del Sistema

La propuesta se centra en un conjunto de requisitos tanto funcionales como no funcionales los cuales son claves para el correcto desarrollo e implementación de esta.

5.2.2.2 Requisitos Funcionales

Los requisitos funcionales son aquellos que describen las funcionalidades del sistema que son visibles por el usuario. Estos requisitos representan los acontecimientos a suceder durante la interacción de algún agente externo con el sistema.

RF01. El sistema tendrá un mecanismo de autenticación de usuarios compuesto por nombre de usuario, contraseña y pregunta de seguridad para identificar doctores, personal de enfermería, analistas clínicos, personal administrativo y paciente. La contraseña debe ser una combinación de letras, números y caracteres especiales, con una longitud mínima de 8 caracteres, y máxima de 16.

RF02. El sistema permitirá al personal médico realizar diagnósticos basados en consultas y pruebas clínicas realizadas al paciente.

RF03. El sistema permitirá que el personal autorizado consulte estadísticas a través de un 'dashboard' de los pacientes con sus principales problemas y afecciones, ingresando el número de historial médico o cédula.

RF04. El sistema permitirá que el personal autorizado consulte estadísticas sobre la cantidad de fármacos usados por mes y año a través de un 'dashboard' ingresando el nombre o código del fármaco. Asimismo, ver un historial de los productos y los proveedores de dichos productos.

RF05. El sistema permitirá que el personal autorizado consulte estadísticas sobre la cantidad de productos de cocina usados por mes y año a través de un 'dashboard' ingresando el nombre o código del producto de cocina. Asimismo, ver un historial de los productos y los proveedores de dichos productos.

RF06. El sistema limitará las acciones de los usuarios dependiendo de las funciones asignadas previamente a cada rol.

RF07. El sistema deberá enviar una notificación sobre el origen de un nuevo brote de una enfermedad.

RF08. El sistema deberá enviar una notificación sobre el uso masivo de un producto.

RF09. El sistema deberá enviar una notificación sobre el uso masivo de un fármaco.

RF10. El sistema deberá generar análisis predictivo.

5.2.2.3 Requisitos no Funcionales

Estos requisitos describen todos los aspectos que, a pesar de ser visibles por el usuario, no tienen conexión directa con el funcionamiento funcional del sistema. Son conocidos como atributos de calidad debido a que son utilizados para juzgar las operaciones realizadas por un sistema en vez de sus comportamientos.

RNF01. El sistema encriptará las credenciales de autenticación para evitar que en el proceso de solicitud sea expuesto a personas malintencionadas.

RNF02. El sistema deberá tener políticas y controles de seguridad.

RNF03. El sistema deberá mantener una conexión permanente a internet para operar.

RNF04. El sistema deberá ser compatible con los navegadores web principales de computadoras personales, teléfonos inteligentes y tabletas.

RNF05. El tiempo de respuesta ante cada acción debe ser menor de 5 segundos para evitar ralentizar el proceso.

RNF06. La interfaz del sistema debe ser amigable e intuitiva al usuario.

RNF07. La interfaz debe guardar similitud tanto en ordenadores, como en dispositivos de mano.

RNF08 El sistema deberá ofrecer soporte de ayuda.

RNF09. El sistema actualizará siempre la información ordenada por fecha descendente, es decir, las últimas modificaciones se verán primero en la lista.

RNF10. El sistema deberá cerrar sesión después de 5 minutos de inactividad.

RNF11. El sistema siempre guardará información de todos los cambios realizados a los historiales médicos de los pacientes con los datos siguientes: número de historial médico electrónico, nombre y apellido del paciente, fecha de la modificación y usuario que realiza la modificación.

RNF12. El sistema deberá de ser 'responsive' y adaptar su visualización a los distintos tipos de dispositivos.

5.2.2.4 Diagramas y Especificaciones de Casos de Uso

En esta sección se presentan los casos de usos que representan la interacción que realiza el usuario con el sistema en cada caso aislado de uso. Estos casos de uso son una descripción de cada uno de los pasos que se llevan a cabo para la realización de algún proceso. Todo personaje o entidad que tiene participación en un caso de uso es un actor. A continuación, se presentan los distintos casos de uso del sistema.

- **Caso de uso General**

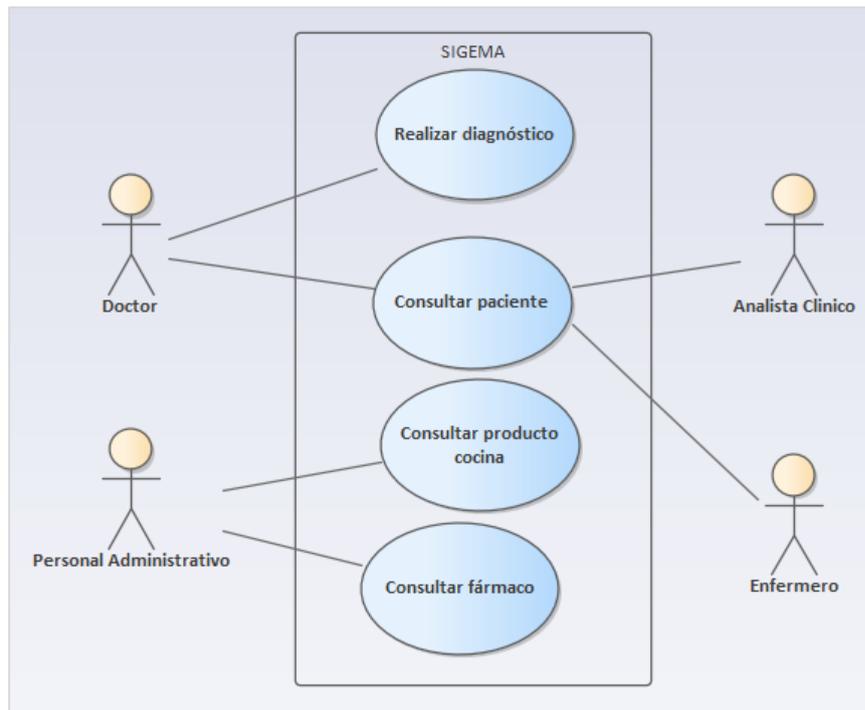


Figura 34- Diagrama de caso de uso general

Fuente (Autores)

- **Caso de uso para autenticarse**

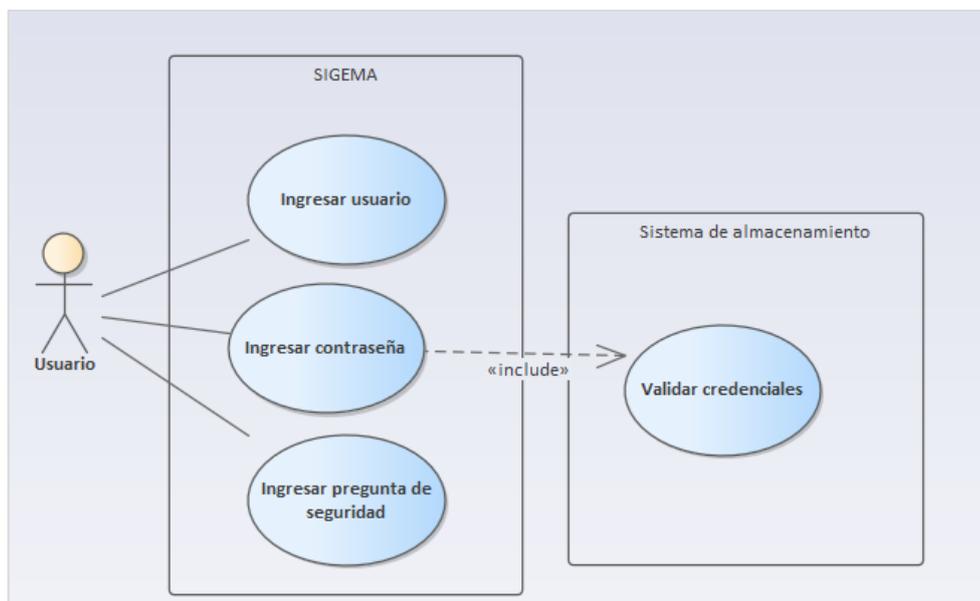


Figura 35- Diagrama de caso de uso autenticación de usuario

Fuente (Autores)

Tabla 7- Caso de uso para autenticarse

Caso de Uso	Autenticación de Usuario	<<CU-01>>
Escenario	Autenticación	
Fuente	RF01 RF06	
Actor(es)	Doctor, Enfermero/a, Analista Clínico, Personal Administrativo y Paciente	
Descripción	Proceso en el que el actor ingresa sus credenciales en el sistema según su responsabilidad en el sistema.	
Propósito	Autenticar el usuario que utilice el sistema para limitar e identificar sus acciones dependiendo de sus roles asignados.	
Precondiciones	Acceder a la pantalla de login.	
Postcondición	Iniciar sesión en el sistema.	
Flujo Básico		
Paso	Actor(es)	Sistema
FB1	El usuario introduce su nombre de usuario pulsa el botón “ Siguiente ”.	
FB2		El sistema recibe el nombre de usuario
FB3		El sistema valida la credencial
FB4	El usuario introduce su contraseña.	
FB5	El usuario pulsa el botón “ Acceder ”.	
FB6		El sistema recibe la contraseña.
FB7		El sistema valida la credencial.
FB8	El usuario selecciona la pregunta de seguridad.	
FB9	El usuario ingresa la respuesta.	
FB10	El usuario pulsa el botón “ Acceder ”.	
FB11		El sistema valida la información.
FB12		Si las credenciales la información es validad, el sistema crea una sesión para el usuario.
FB13		El sistema muestra la pantalla principal

Flujo Alterno		
FA1 a FB1		
FA1.1	El usuario introduce su nombre de usuario pulsa el botón “Siguiete” .	
FA1.2		El sistema recibe el nombre de usuario
FA1.3		Si el usuario no es válido, el sistema solicita nuevamente.
FA1.4	El usuario ingresa nuevamente el nombre de usuario y sigue al FB4	
FA2 a FB4		
FA2.1	El usuario introduce su contraseña.	
FA2.2	El usuario pulsa el botón “Acceder” .	
FA2.3		El sistema recibe la contraseña.
FA2.4		Si la contraseña no es válida el sistema solicita nuevamente.
FA2.5	El usuario ingresa nuevamente la contraseña y sigue FB8	
FA3 a FB8		
FA3.1	El usuario selecciona la pregunta de seguridad.	
FA3.2	El usuario ingresa la respuesta.	
FA3.3	El usuario pulsa el botón “Acceder” .	
FA3.4		Si la información no es válida, el sistema solicita nuevamente
FA3.5	El usuario ingresa nuevamente los datos.	
Flujo Error		
FE1 A FB1		
FE1.1	El usuario ingres al sistema	
FE1.2		El sistema muestra error de conexión.
FE1.3	El usuario valida su conexión.	

Fuente: (Autores)

- **Caso de uso para realizar diagnóstico**

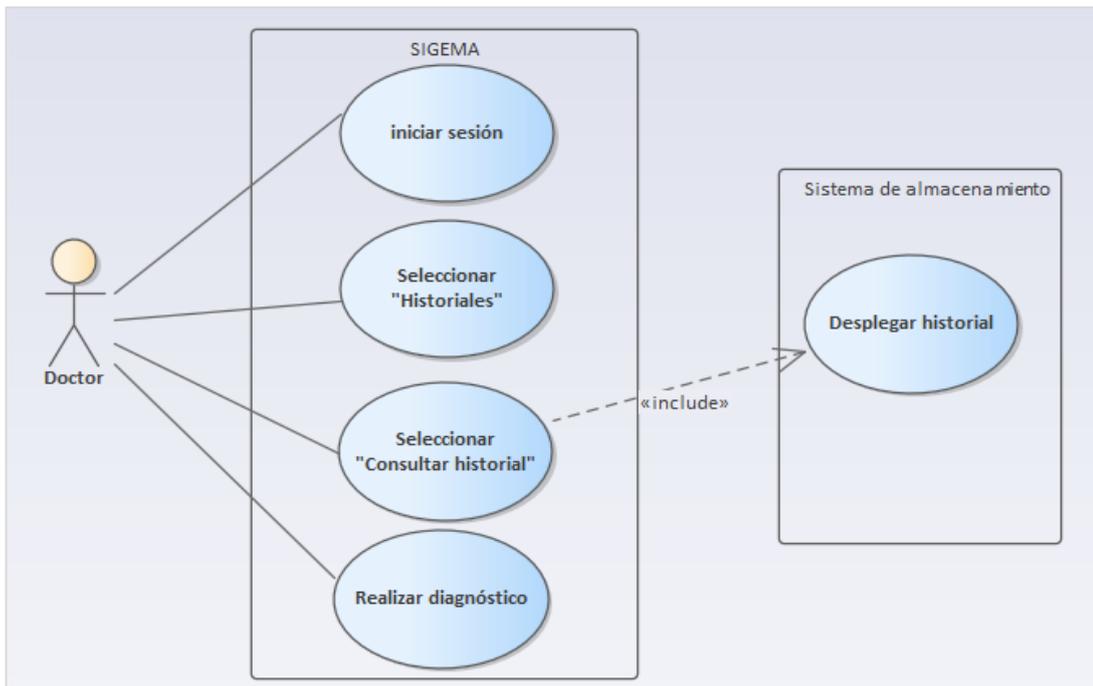


Figura 36- Diagrama de caso de uso realizar diagnóstico

Fuente (Autores)

Tabla 8- Caso de uso para realizar diagnóstico

Caso de Uso	Realizar diagnóstico	<<CU-02>>
Fuente	RF02 RF07	
Actor(es)	Doctores	
Descripción	El doctor autorizado realiza un diagnóstico del estado de salud del paciente de acuerdo con lo detallado por el sistema en el historial médico.	
Propósito	Dar un diagnóstico al paciente sobre su estado de salud.	
Precondiciones	Ser usuario del sistema y tener los permisos correspondientes. Acceder al historial del paciente.	
Postcondición	Diagnosticar el estado de salud del paciente.	
Flujo Básico		
Paso	Actor(es)	Sistema
FB1	El usuario ingresa al sistema	
FB2		El sistema despliega el menú principal
FB3	El usuario selecciona la opción "Historiales"	
FB4		El sistema despliega las opciones: Crear

		historial y Consultar historial
FB5	El usuario selecciona “Consultar” e ingresa el número de paciente.	
FB6		El sistema despliega los datos del paciente.
FB7	El usuario valida la información y procede.	
Flujo Alterno		
FA1 a FB5		
FB1.1	El usuario selecciona “Consultar” e ingresa cédula del paciente	
FB1.2		El sistema despliega los datos del paciente.
FB1.3	El usuario valida la información y procede.	
Flujo Error		
FE1 a FB1		
FE1.1	El usuario ingresa al sistema	
FE1.2		El sistema envía un mensaje de conexión
FE1.3	El usuario comprueba su conexión a internet	

Fuente (Autores)

- Caso de uso para consultar paciente

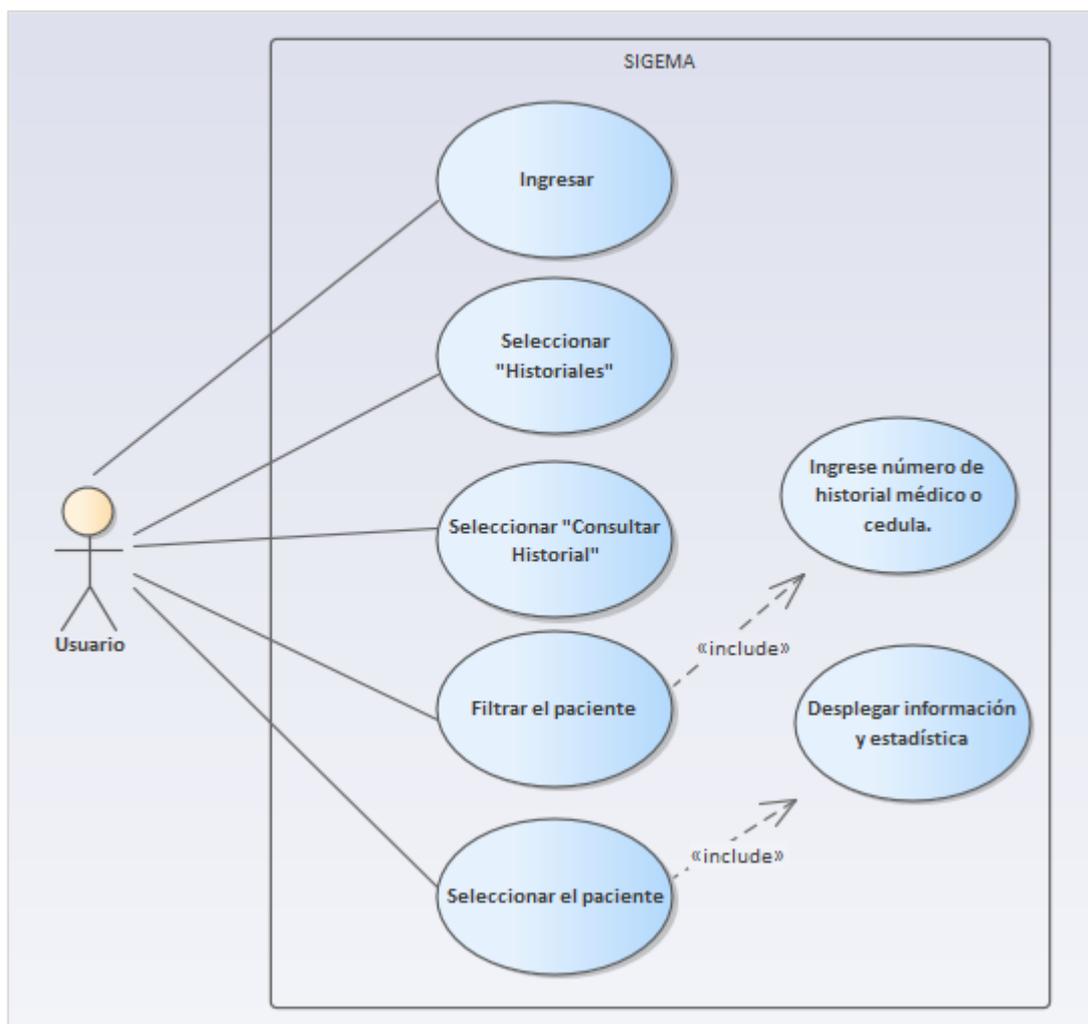


Figura 37- Diagrama de Caso de uso para consultar paciente

Fuente (Autores)

Tabla 9- Caso de uso para consultar paciente

Caso de Uso	Consultar paciente	<<CU-03>>
Fuente	RF03, RF07	
Actor(es)	Doctor, analista clínico, enfermero y personal administrativo.	
Descripción	El usuario consulta las estadísticas que son generadas al paciente con la información contenida en el historial médico.	
Propósito	Permitir que el usuario tome decisiones sobre el paciente en base a la analítica que se le ha realizado al historial médico de este.	
Precondiciones	Tener usuario y contraseña. Tener los privilegios de acceso.	

Postcondición	Indicaciones de tratamientos en caso de necesitarlo el paciente.	
Flujo Básico		
Paso	Actor(es)	Sistema
FB1	El usuario ingresa al sistema	
FB2		El sistema despliega los módulos
FB3	El usuario selecciona “Historiales”	
FB4		El sistema despliega las opciones.
FB5	El usuario selecciona “Consultar Historial”	
FB6		El sistema despliega un listado de pacientes
FB7	El usuario filtra al paciente por número de historial	
FB8		El sistema muestra al usuario.
FB9	El usuario selecciona al paciente y consulta la información.	
Flujo Alternativo		
FA1 a FB7		
FA1.1	El usuario filtra al paciente por cédula	
FA1.2		El sistema muestra al usuario.
FA1.3	El usuario selecciona al paciente y consulta la información.	
Flujo Error		
FE1 a FB1		
FE1.1	El usuario ingresa al sistema	
FE1.2		El sistema muestra error de conexión
FE1.3	El usuario comprueba la conexión	

Fuente (Autores)

- Caso de uso para consultar fármaco

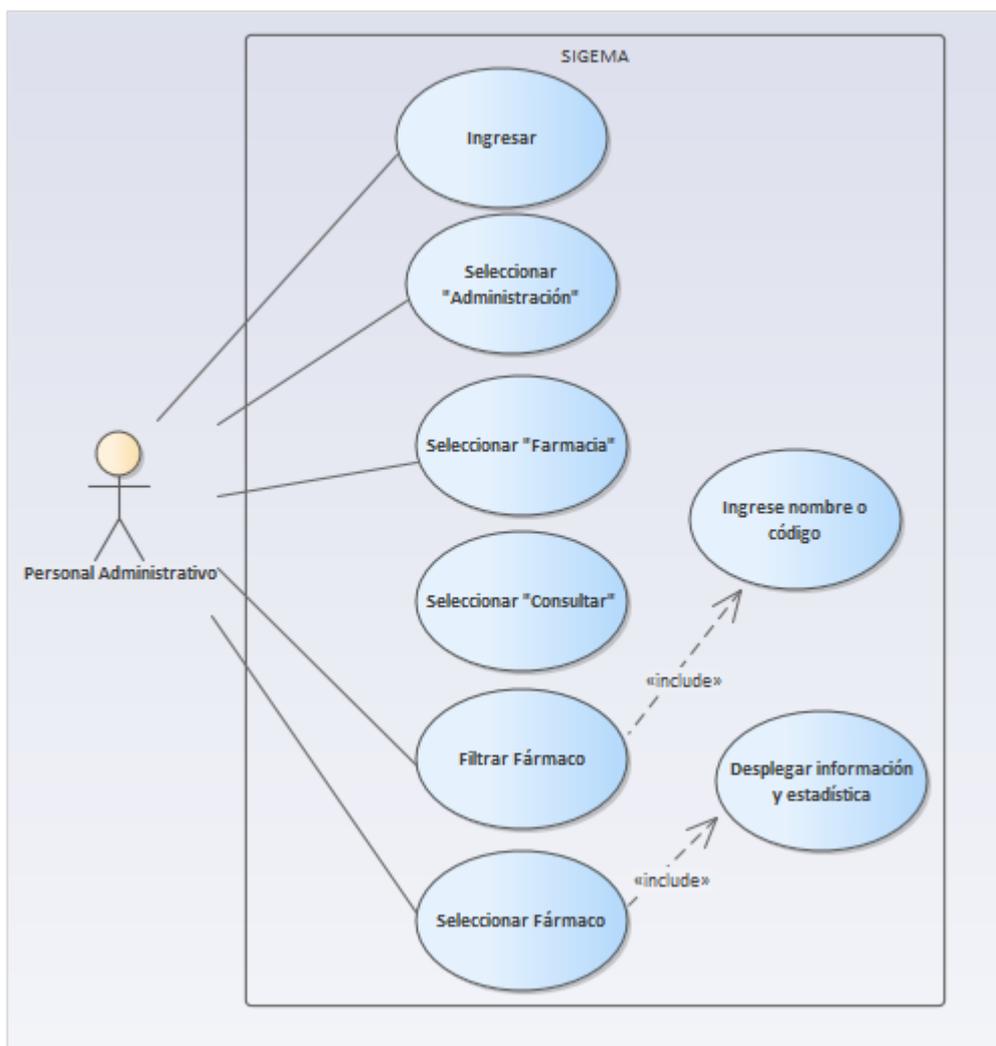


Figura 38- Diagrama de Caso de uso para consultar fármaco

Fuente (Autores)

Tabla 10- Caso de uso para consultar fármaco

Caso de Uso	Consultar fármaco	<<CU-04>>
Fuente	RF04, RF09	
Actor(es)	Personal administrativo	
Descripción	El personal administrativo consulta las estadísticas que son generadas sobre los fármacos	
Propósito	Permitir que el personal administrativo tome decisiones sobre los fármacos en base a la analítica que se le ha realizado a los procesos financieros.	
Precondiciones	Tener usuario y contraseña. Tener los privilegios de acceso.	
Postcondición	Toma de decisiones	

Flujo Básico		
Paso	Actor(es)	Sistema
FB1	El usuario ingresa al sistema	
FB2		El sistema despliega los módulos
FB3	El usuario selecciona “Administrativo” y luego “Farmacia.”	
FB4		El sistema despliega las opciones.
FB5	El usuario selecciona consultar	
FB6		El sistema despliega un listado de fármacos
FB7	El usuario filtra al fármaco por código	
FB8		El sistema muestra al usuario.
FB9	El usuario selecciona al fármaco y consulta la información.	
Flujo Alternativo		
FA1 a FB7		
FA1.1	El usuario filtra al paciente por nombre	
FA1.2		El sistema muestra al usuario.
FA1.3	El usuario selecciona al fármaco y consulta la información.	
Flujo Error		
FE1		
FE1.1	El usuario ingresa al sistema	
FE1.2		El sistema muestra error de conexión
FE1.3	El usuario comprueba la conexión	

Fuente (Autores)

- **Caso de uso para consultar producto de cocina**

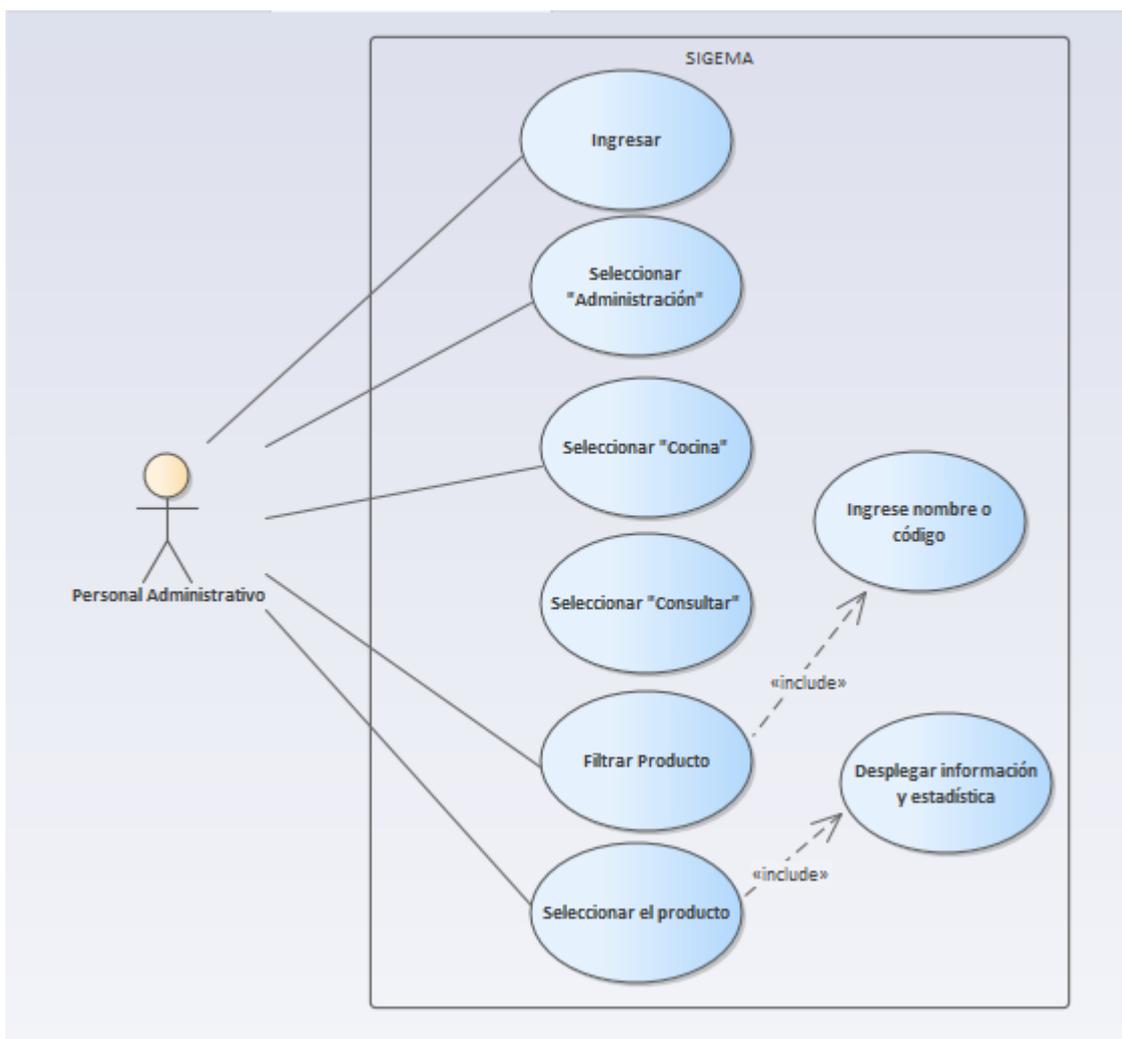


Figura 39- Diagrama de Caso de uso para consultar producto de cocina

Fuente (Autores)

Tabla 11- Caso de uso para consultar producto cocina

Caso de Uso	Consultar producto cocina	<<CU-05>>
Fuente	RF05, RF08	
Actor(es)	Personal administrativo	
Descripción	El personal administrativo consulta las estadísticas que son generadas sobre los productos de cocina.	
Propósito	Permitir que el personal administrativo tome decisiones sobre los productos de cocina en base a la analítica que se le ha realizado a los procesos financieros.	
Precondiciones	Tener usuario y contraseña. Tener los privilegios de acceso.	

Postcondición	Toma de decisiones	
Flujo Básico		
Paso	Actor(es)	Sistema
FB1	El usuario ingresa al sistema.	
FB2		El sistema despliega los módulos.
FB3	El usuario selecciona “Administrativo” y luego “Cocina.”	
FB4		El sistema despliega las opciones.
FB5	El usuario selecciona consultar.	
FB6		El sistema despliega un listado de productos de cocina.
FB7	El usuario filtra al producto de cocina por código.	
FB8		El sistema muestra al usuario.
FB9	El usuario selecciona al producto de cocina y consulta la información.	
Flujo Alternativo		
FA1 a FB7		
FA1.1	El usuario filtra al producto de cocina por nombre.	
FA1.2		El sistema muestra al usuario.
FA1.3	El usuario selecciona al producto de cocina y consulta la información.	
Flujo Error		
FE1 a FB1		
FE1.1	El usuario ingresa al sistema.	
FE1.2		El sistema muestra error de conexión.
FE1.3	El usuario comprueba la conexión.	

Fuente (Autores)

5.2.2.3 Diagrama de Modelado del Sistema

En esta sección mediante el lenguaje unificado de modelado (UML) se describirá de manera gráfica el comportamiento esperado de los diferentes procesos que integran del sistema propuesto. El objetivo del modelado del sistema es mostrar las partes esenciales del sistema mediante la notación y abstracción gráfica del mismo. Se presentan los diagramas de secuencias y el diagrama de clase del sistema:

5.2.2.3.1 Diagramas de Secuencia

- Diagrama de Secuencia Autenticación

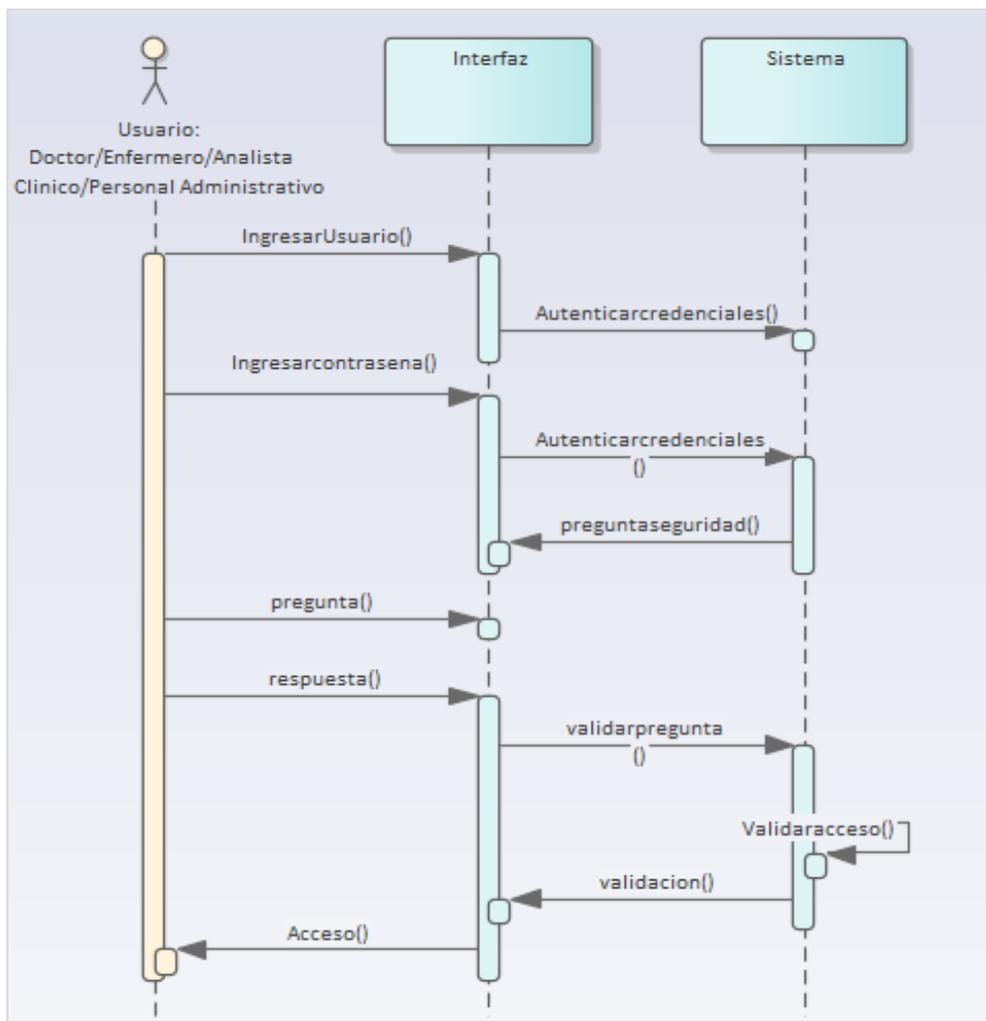


Figura 40- Diagrama de secuencia para autenticación.

Fuente (Autores)

- Diagrama de Secuencia para realizar diagnóstico

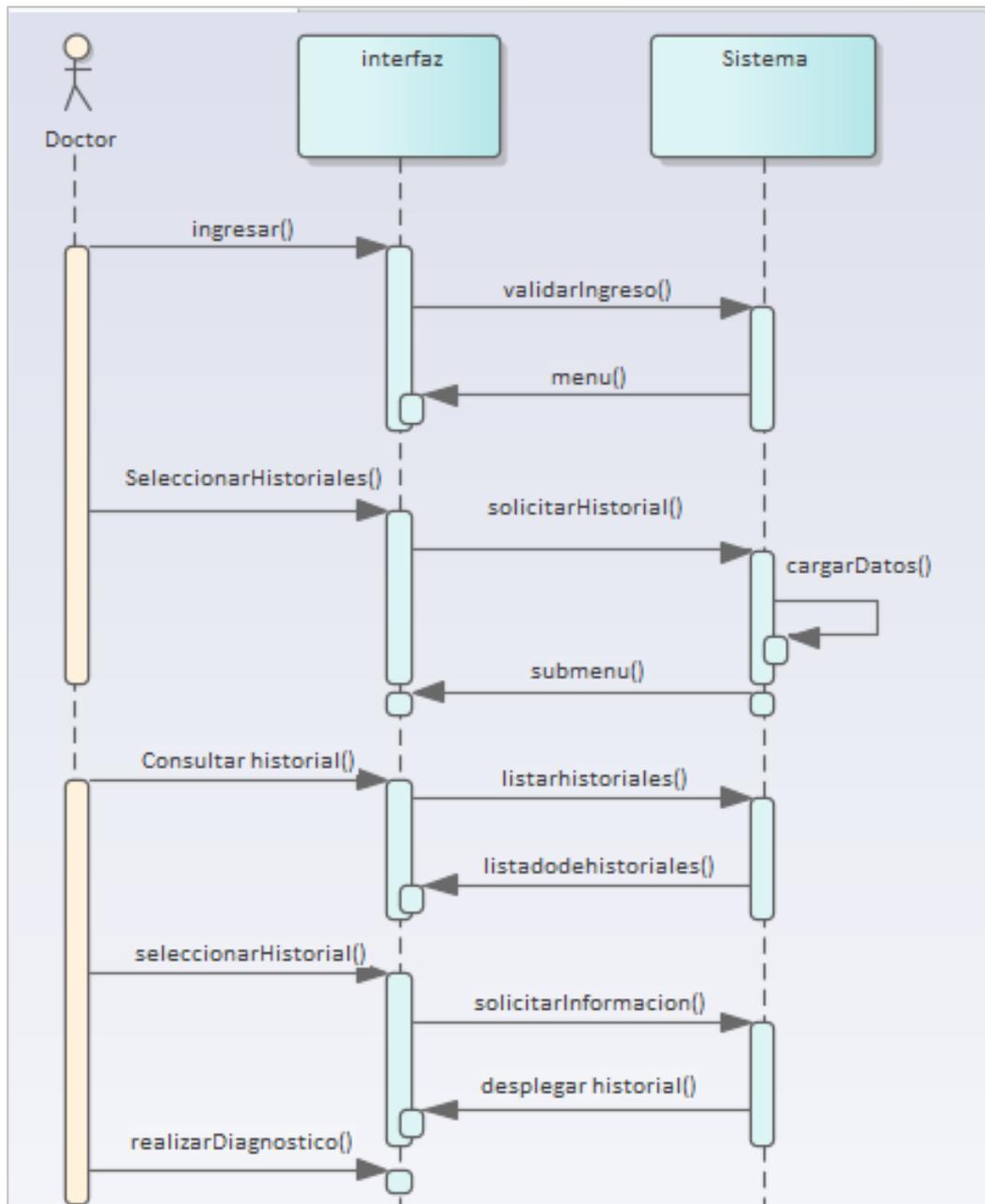


Figura 41- Diagrama de Secuencia Realizar Diagnostico

Fuente (Autores)

- **Diagrama de Secuencia para consultar paciente**

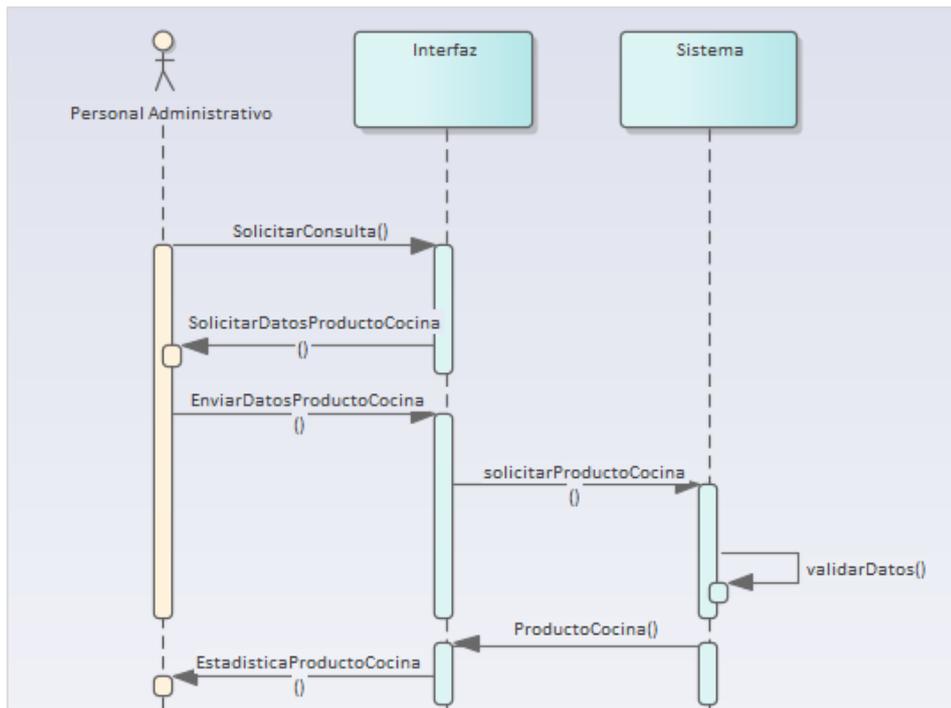


Figura 42- Diagrama de Secuencia para consultar paciente

Fuente (Autores)

- **Diagrama de Secuencia para consultar producto de cocina**

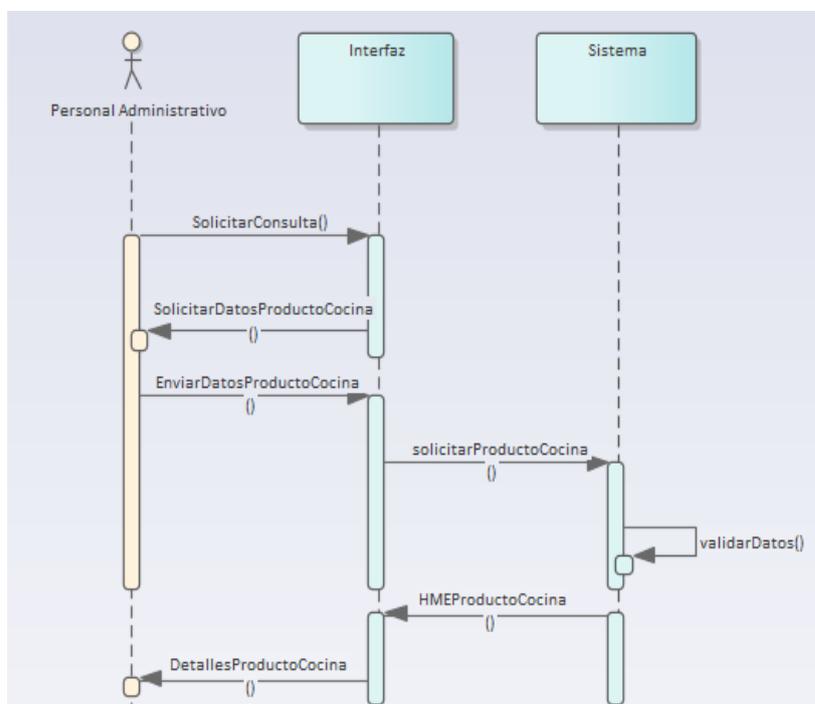


Figura 43- Diagrama de Secuencia para consultar producto cocina

Fuente (Autores)

- **Diagrama de Secuencia para consultar fármaco**

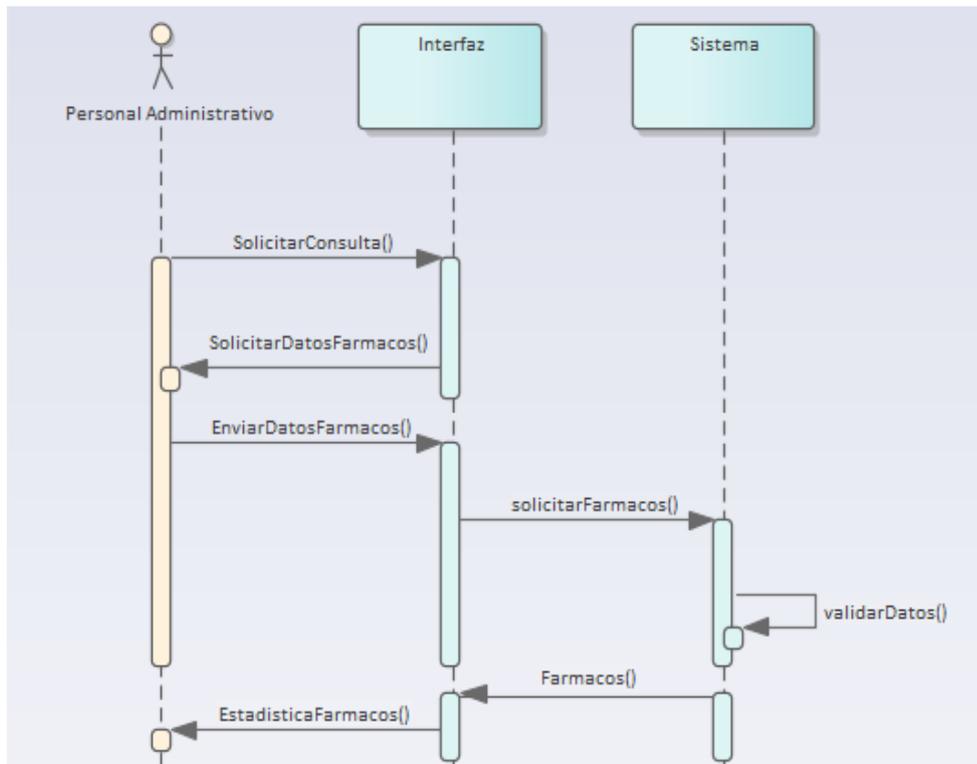


Figura 44- Diagrama de Secuencia para consultar fármaco

Fuente (Autores)

5.2.2.3.2 Diagrama de clases

Este diagrama describe como estará estructurado el sistema, mostrando sus clases, atributos, métodos y relaciones entre los objetos que lo componen.

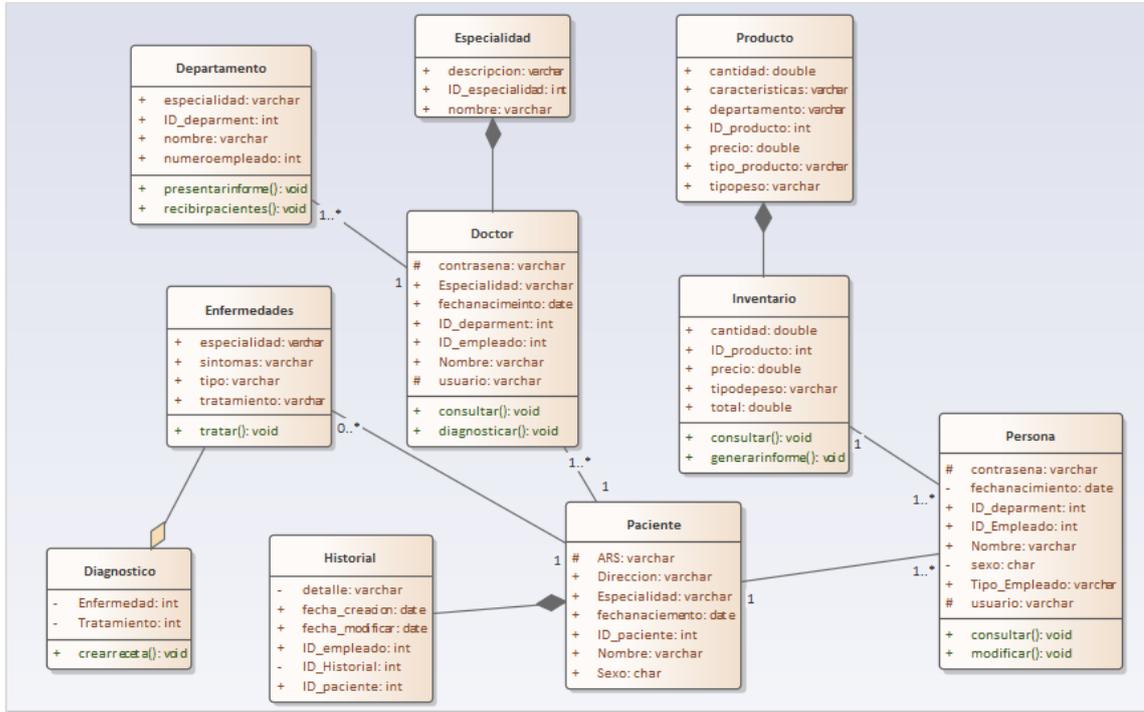


Figura 45- Diagrama de Clase

Fuente: (Autores)

5.2.3 Propuesta de solución Sistema de Gestión de Información de Analítica de Big Data

El objetivo principal de la propuesta como se ha mencionado anteriormente es proporcionar al Hospital Regional Taiwán 19 de Marzo Azua, un sistema de gestión de información para la toma de decisiones médicas y administrativas aplicando la Analítica de Big Data sobre los historiales médicos de los pacientes, cuyas funcionalidades principales serán la predicción de brotes de enfermedades y la definición del presupuesto.

Es importante destacar que dentro de una infraestructura de un sistema de información, como puede ser un sistema de gestión de datos, conviven distintas tecnologías y con respecto a esta propuesta, los autores de este trabajo de investigación para maximizar el diseño propuesto, no sólo se enfocaron en las principales componentes de una infraestructura de Big Data (Hadoop y NoSQL, etc.) sino que se plantea rediseñar el

sistema de base de datos relacional a los fines de integrar ambos sistemas y con ello tener un mejor funcionamiento y capacidad del sistema de gestión de analítica de Big Data

5.2.3.1 Diseño de la Arquitectura del Sistema de Información de Analítica de Big Data

El presente proyecto posee el enfoque de integrar las bases de datos tradicionales o relacionales existen en el Hospital Taiwán 19 de Marzo, Azua R.D, tales como SQL Server Express y SQL Server 2008 R2 con las de Big Data que serán descrita más adelante

A continuación, se presenta el diseño para la base de datos estructurada, la cual manejaría las fuentes de datos tradicional del hospital

5.2.3.1.1 Base de Datos Relacional

Se presenta la parte correspondiente a la base de datos estructurada el cual está integrado al sistema de gestión de Analítica de Big Data.

- **Diagrama Entidad Relación**

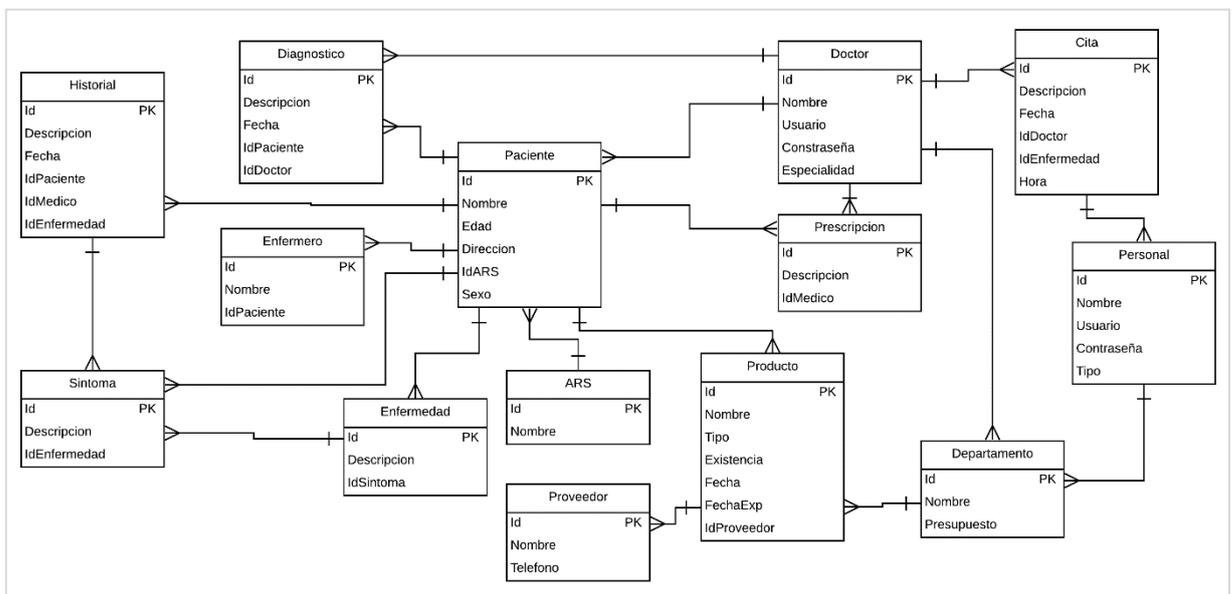


Figura 46- Diagrama Entidad-Relación

Fuente: (Autores)

Esquema de datos

1. - **Paciente:** Almacena las informaciones personales de los pacientes, como es su nombre, apellidos, fecha de nacimiento, sexo entre otras.
2. - **Doctor:** Almacena todas las informaciones relacionadas con los doctores tales como: especialidad, nombre y otros.
3. - **Personal:** Almacena informaciones personales de los colaboradores de diversas áreas en el hospital.
4. – **Enfermedad:** Almacena todas las informaciones sobre el padecimiento de un paciente.
5. – **Enfermero:** Almacena información relacionada a los datos clínicos de una persona.
6. – **Departamento:** Almacena todas las informaciones sobre los departamentos que componen el hospital y sus funciones.
7. – **Cita:** Almacena las informaciones sobre el próximo encuentro que tendrá un doctor con su paciente.
8. – **Prescripción:** Almacena toda la información sobre la receta que elabora un médico para una determinada medicación o indicar algún tratamiento.
9. -**Producto:** Almacena las informaciones sobre las características de los productos que se utilizan en el hospital como puede ser: material gastable, fármacos o alimentos
10. -**Proveedor:** Almacena las informaciones concernientes a los productos que son adquiridos por el centro de salud.
11. -**Diagnostico:** Almacena todas las informaciones sobre el diagnostico que recibe un paciente.
12. -**Historial:** Almacena toda la historia clínica de un paciente desde el primer momento que llego al hospital hasta la actualidad.
13. -**Síntoma:** Almacena todas informaciones vinculadas a los síntomas que presenta un paciente.
- 14.- **ARS:** Almacena las informaciones de las Asegurados de salud.

Diccionario de datos de la base de datos (Tablas)

Tabla 12- Paciente

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Paciente						
1	ID	PK				
2	Nombre	Nombre de paciente	Varchar	50	Si	No
3	Edad	Edad de paciente	Int	3	Si	No
4	Dirección	Dirección	Varchar	100	Si	No
5	Sexo	Género	Varchar	1	Si	No
6	IdARS	FK	Int	-	-	-

Fuente: (Autores)

Tabla 13- Doctor

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Doctor						
1	ID	PK				
2	Nombre	Nombre de doctor	Varchar	50	Si	No
3	Usuario	Nombre de usuario	Varchar	20	Si	No
4	Contraseña	Contraseña	Varchar	16	Si	No
5	Especialidad	Area Médica	Varchar	50	Si	No

Fuente: (Autores)

Tabla 14- Personal

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Personal						
1	ID	PK				
2	Nombre	Nombre de personal	Varchar	50	Si	No
3	Usuario	Nombre de usuario	Varchar	20	Si	No
4	Contraseña	Contraseña	Varchar	16	Si	No
5	Tipo	Tipo de usuario	Varchar	30	Si	No

Fuente: (Autores)

Tabla 15- Enfermedad

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Enfermedad						
1	ID	PK				
2	Descripción	Detalles de la tabla	varchar	50	Si	No
3	IdSintoma	FK	Int	-	-	-

Fuente: (Autores)

Tabla 16- Enfermero

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Enfermero						
1	ID	PK				
2	Nombre	Nombre de enfermero	Varchar	50	Si	No
3	IdPaciente	FK	Int	-	-	-

Fuente: (Autores)

Tabla 17- Departamento

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Departamento						
1	ID	PK				
2	Nombre	Nombre de Departamento	Varchar	50	Si	No
3	Presupuesto	Asignación Presupuestal	Int	10	Si	No

Fuente: (Autores)

Tabla 18- Cita

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Cita						
1	ID	PK				
2	Descripcion	Detalles de la tabla	Varchar	50	Si	No
3	Fecha	Fecha de creación	Date	Default	Si	No
4	IdDoctor	FK	Int	-	-	-
5	IdEnfermedad	FK	Int	-	-	-
6	Hora	Hora de cita	Time	Default	Si	No

Fuente: (Autores)

Tabla 19- Prescripción

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Prescripción						
1	ID	PK				
2	Descripción	Detalles de la tabla	varchar	50	Si	No
3	IdMedico	FK	Int	-	-	-

Fuente: (Autores)

Tabla 20- Producto

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Producto						
1	ID	PK				
2	Nombre	Nombre producto	Varchar	50	Si	No
3	Tipo	Tipo de producto	Varchar	20	Si	No
4	Fecha	Fecha	Data	Default	Si	No
5	FechaExp	Fecha de caducidad	Date	Default	Si	No
6	IdProveedor	FK	int	-	-	-

Fuente: (Autores)

Tabla 21- Proveedor

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Proveedor						
1	ID	PK				
2	Nombre	Nombre proveedor	varchar	50	Si	No
3	Teléfono	Contacto	Int	15	Si	No

Fuente: (Autores)

Tabla 22- Diagnostico

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Enfermero						
1	ID	PK				
2	Nombre	Nombre de enfermero	Varchar	50	Si	No
3	IdPaciente	FK	Int	-	-	-

Fuente: (Autores)

Tabla 23- Historial

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Historial						
1	ID	PK				
2	Descripción	Detalles de la tabla	Varchar	50	Si	No
3	Fecha	Fecha de creación	Date	Default	Si	No
4	IdPaciente	FK	Int	-	-	-
5	IdMedico	FK	Int	-	-	-
6	IdEnfermedad	FK	Int	-	-	-

Fuente: (Autores)

Tabla 24- Síntoma

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
Síntoma						
1	ID	PK				
2	Descripción	Detalles de la tabla	varchar	50	Si	No
3	IdEnfermedad	FK	Int	-	-	-

Fuente: (Autores)

Tabla 25- ARS

Tabla	Elemento	Descripción	Tipo de dato	Longitud	Requerimiento	Acepta valor nulo?
ARS						
1	ID	PK				
2	Nombre	Nombre de ARS	Varchar	50	Si	No
3	Plan	Tipo de Plan	Varchar	30	Si	No
4	Cobertura	Cobertura	Varchar	20	Si	No

Fuente: (Autores)

5.2.3.1.2 Arquitectura del sistema de información de Analítica de Big Data

La solución de la arquitectura de analítica de Big Data propuesta para el sistema de información para la toma de decisiones médicas y administrativas en el Hospital Taiwán de Azua como se observa en la figura No. 47, los autores de este trabajo de investigación consideran que es la más apropiada tal como se ha explicado más adelante en el análisis y elecciones de todos los componentes que la constituyen.

La presente arquitectura muestra toda la infraestructura necesaria para el sistema de gestión de analítica de Big Data integrada por las bases de datos relacional y NoSQL. Además, se presentan cada uno de los componentes de la arquitectura de Big data (Recolección, almacenamiento, procesamiento, análisis y visualización).

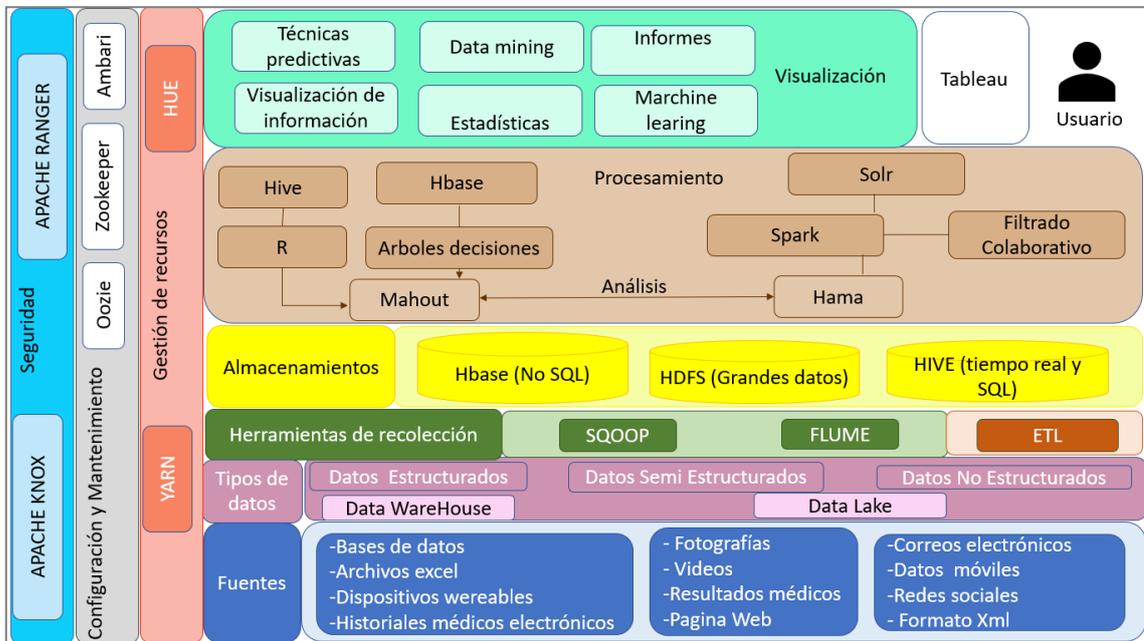


Figura 47 Arquitectura de Analítica de Big Data

Fuente: (Autores)

Esta arquitectura es escalable, tolerante a fallos, ofrece tecnologías de renombre, que son utilizadas actualmente por países desarrollados como España y que son eficientes para los procesos para los cuales se han escogido. Asimismo, por su condición de código abierto su costo es relativamente bajo, lo que representa un beneficio económico para el centro y el país. Además, el sistema propuesto es robusto, cualidad que transmite confianza para su implementación. De igual manera, esta se adapta a las necesidades presentes y futuras de la institución, por los diversos análisis que se pueden realizar con las informaciones que se encuentran almacenadas en los historiales médicos. Por lo tanto, la gestión de búsqueda y consulta será beneficiadas.

Por último, procesos como el ingreso de pacientes, ya sea para consultas como emergencias y no se puede dejar a un lado la gestión de los procesos que influyen en la elaboración del presupuesto. Todo esto se logrará con la integración de los diversos sistemas gestores de datos tradicionales como no tradicionales que son los que manejan datos no estructurados como semiestructurados.

5.2.3.1.2.1 Tecnologías de Desarrollo Sistema de Analítica de Big Data

5.2.3.1.2.1.1 Componentes de las bases de datos relacional

La utilización de un Data Warehouse en la propuesta es por su capacidad para gestionar datos estructurados como se explicó anteriormente y este está compuesto por siguientes componentes:

Fuentes de datos

Este es el componente en cual se realiza la captura de los datos estructurados, es decir, áreas que generan mayor volumen de datos diariamente. Entre las que se pueden destacar: documentos en Excel y las tablas de las bases de datos relacional. Con el propósito de las consultas datos sobre los pacientes, disponibilidad de productos en la cocina, citas, ingresos, farmacia y así sucesivamente en esa índole.

Extracción y transformación

En esta parte es donde los datos son transformados según sea necesario, desde su origen, hasta su respectivo lugar en el data warehouse.

Servidor de datos

Este componente de gestión es donde se determina el lugar donde estarán alojadas los datos, ya sea de manera centralizada o descentralizada y se ocupa de su correcto funcionamiento, disponibilidad, seguridad y mantenimiento de los datos.

Herramientas de acceso

En este punto es donde entran las diferentes herramientas mediante de las cuales se estarán gestionando este tipo de datos en sus diferentes dimensiones, en este caso se ha seleccionado Tableau para el sistema en general.

Repositorio/Metadatos

Los metadatos con informaciones referentes a los mismos datos que se almacenan en el data warehouse y hacen una descripción de estos. Permitiendo así que el usuario especializado pueda tener un mayor control sobre a la hora de realizar tareas especializadas con los mismos. También tiene otras funciones como:

- Especifica que usuario genero la información y con qué propósito
- Detallar el tipo de relación de los distintos datos y limitar la veracidad y validez de los mismos.
- Especificar las reglas de transformación de los datos

5.2.3.1.2.1.2 Componentes del Sistema de Información de Analítica de Big Data

Para la implementación del sistema se ha seleccionado el ecosistema Hadoop, debido a que es una plataforma de código abierto, escalable y tolerante a fallos. Esta permite el almacenamiento y procesamiento de grandes conjuntos de datos. Gracias a las características que ofrece, la hace ideal para trabajar en una implementación de Big Data.

Según (Anchiraico, 2017) estas son las características que hacen ideal la plataforma Hadoop:

Escalabilidad

Un clúster Hadoop puede estar compuesto desde una sola máquina hasta, literalmente, miles de máquinas.

La tolerancia a fallos

Los servicios de Hadoop llegan a ser tolerantes a fallos gracias a la redundancia. Por ejemplo, en el caso del sistema de archivos distribuido de Hadoop (Hadoop Distributed File System - HDFS), considerando que el clúster tiene al menos tres nodos (por defecto), se replican automáticamente los bloques de datos a esos tres nodos separados.

Código abierto

El desarrollo de Hadoop está a cargo de la comunidad gobernada bajo la licencia de la Apache Software Foundation. Aquí se puede aportar en la mejora de Hadoop mediante la adición de características, corrección de fallos de software, mejora del rendimiento o en la escalabilidad.

Almacenamiento y procesamiento distribuido

Los grandes conjuntos de datos se dividen automáticamente en trozos más pequeños, llamados bloques, que se distribuyen a través de los nodos del clúster. Cada máquina procesa su bloque local de datos, lo que significa que el procesamiento se distribuye, a la vez, mediante cientos de CPU y gigabytes de memoria.

Hardware genérico

Todos los servicios de Hadoop se pueden ejecutar en hardware genérico (commodity hardware). Con esto no solo reduce los costos de implementación, sino también, los costos de soporte y mantenimiento.

De manera más específica, diríamos que Hadoop es un motor que provee almacenamiento, vía HDFS; y cómputo a través de las capacidades de YARN (Yet Another Resource Negotiator).

A continuación, se expone y un esquema general de cuáles son los componentes del sistema y su descripción.

1. Fuentes de datos

La primera acción por ejecutar es un análisis del sistema actual con el propósito de identificar aquellos elementos que serán utilizados en la propuesta. En primer lugar, serán definidas las fuentes de los datos que utiliza el sistema SISGEHOSPI del Hospital Taiwán, con la finalidad de reutilizarlas y extenderlas. Estas son la página principal del historial médico, los resultados de análisis, las solicitudes de compras en general, las plantillas de Excel que son usadas en finanzas y la base de datos relacional. Además, de las fuentes tradicionales de información se debe implementar un sistema de recolección de datos provenientes de aparatos como tomógrafos, dispositivos móviles, aplicaciones, redes sociales médica.

Fuentes Taiwan	Nuevos Datos	Otros Datos
<ul style="list-style-type: none">• Archivos en papel• Documentos excel• Historial medico tradicional• Pruebas clinicas en fisicas• Correos electronicos• servidores	<ul style="list-style-type: none">• Historiales medicos electronicos• Pruebas clinicas electronicas• Multimedia (Audios, videos)• Sensores• Procesos financieros digitalizados• Estadisticas en informe de epidemiologia• Pagina Web• Redes sociales	<ul style="list-style-type: none">• Dispositivos moviles• Estadisticas e informes del SNS• Monitorizacion• estudios de investigacion• Estadisticas e infomes ONE• Estadisticas e informes de pro-consumidor• dispositivos IoT

Figura 48- Fuente de Datos

Fuente: (Autores)

Gracias a la introducción de nuevas fuentes de datos como se observa en la figura no. 48 se puede analizar una gran cantidad de nuevas informaciones y por ende mejorar otros procesos que son vitales para el funcionamiento de este centro de salud.

Las fuentes de los datos serán:

- Estructurado: fuente tradicional (CRM, ERP, redes, servidores, web log data, etc.)
- Semiestructurado: XML
- No estructurado: e-mail, mensajes instantáneos- SMS, WhatsApp, fotos, audios, videos, etc.

Los dos últimos son el foco cuando se trabaja con Big Data, debido a que son datos que no pueden almacenarse en los sistemas tradicionales.

2. Recolección de Datos

En la actualidad la información que posee el Hospital Taiwán, en su mayoría se encuentra físicamente, mientras que una pequeña porción está en digital como es el formulario de recepción del paciente, las tablas en Excel y los documentos en Word que son datos estructurados. Las nuevas fuentes de datos que incluirá el sistema propuesto son datos no estructurados, por tanto, se necesitaran herramientas específicas orientadas al ecosistema Hadoop, que más adelante veremos, con la finalidad de que estos datos sean capturados, almacenados y procesados para generar información de valor para la toma de decisiones.

A continuación, se describirán como se configurará este ecosistema con las herramientas seleccionadas donde el 90% de estas son open source.

Para iniciar, tendremos que recolectar los datos de todas aquellas fuentes que no generan datos estructurados, para lo cual se utilizará Flume, un servicio que se encarga de la recolección, agregación y movimiento de los flujos de datos específicamente de aquellos que se originan de la e-salud lo cuales en su mayoría son datos no estructurados. Flume trabaja bajo el concepto de distribuido, lo que facilita su implementación en diversos sistemas. Este ofrece dos grandes ventajas de cara a este sistema: trabaja con datos no estructurados como se mencionó anteriormente y es un altamente disponible, esto permite la recolección en tiempo real. Sin embargo, entrega los datos sin procesar, por esta razón se ha seleccionado la herramienta Sqoop, que permite procesar y extraer grandes cantidades de datos de sistemas de gestión de bases de datos relacionales, u otras aplicaciones open source, es decir, realizar el proceso de ETL, véase fig. 49. Además, se utiliza Sqoop para importar datos estructurados externos en el sistema de archivos de Hadoop (HDFS) o a sistemas relacionados como Hive y Hbase los cuales formaran parte de la siguiente fase de la arquitectura. Del mismo modo, se puede utilizar para extraer datos de Hadoop y exportarlos a los almacenes de datos estructurados externos tales como las bases de datos relacionales.

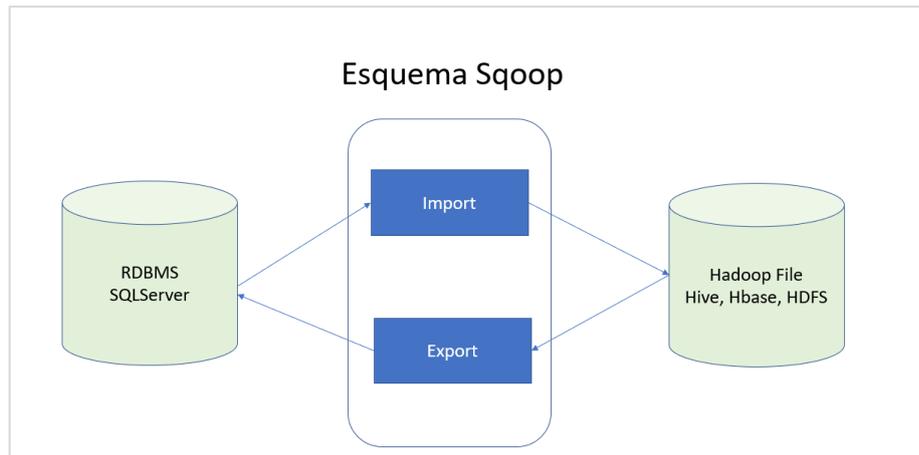


Figura 49- Estructura de la herramienta Sqoop

Fuente (Autores)

Después de que los datos son recolectados y aplicado el proceso de ETL, el siguiente paso es el almacenamiento de esos datos y la gestión de los recursos. Se implementará un Data Warehouse para los datos de origen estructurado y que solo se deba aplicar el proceso de extracción, transformación y carga (ETL). Por otra parte, se integrará un Data Lake que trabajará los datos no estructurados y que será gobernado por el framework Hadoop como se muestra en la figura no. 50.

Como sistema de gestor de datos se implementará un Data Lake por su capacidad para gestionar y administrar datos no estructurados y semiestructurados, es decir, almacena información que no está preparada y lista para el consumo, sino que se recoge en estado natural. Esto permite que los usuarios puedan dar a los datos un uso más creativo, que no queda marcado por el fin para el que se han definido al momento de su carga, como sucedía en el Data Warehouse. El Data Lake conserva todos los datos en su estado natural, también incluyendo a esos perfiles técnicos con exigencias de análisis más avanzadas, que son quienes recurren a capacidades como análisis estadístico y modelado predictivo. Mientras que el Data Warehouse es más selectivo y guarda solo los que considere útiles, y basados en esto el data Lake es mejor herramienta para trabajar con esta propuesta ya que permite un mayor margen de almacenamiento de información en su estado de creación de manera tal que puede ser utilizada para hacer diferentes tipos de análisis desde diferentes puntos de vista y seguir conservando los archivos en su estado bruto.

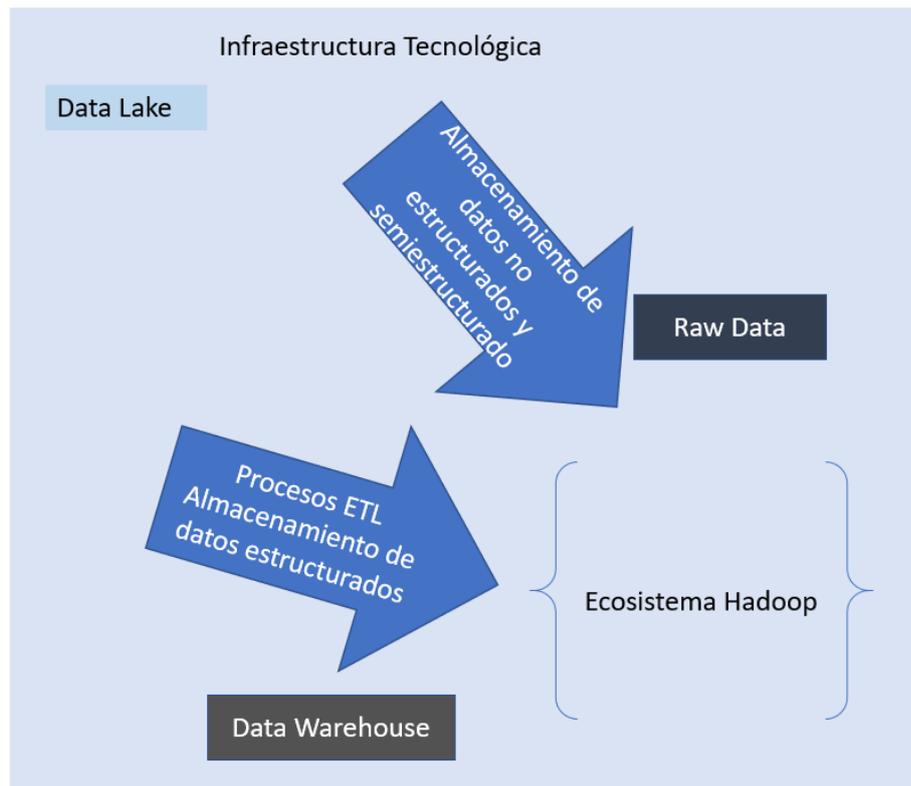


Figura 50- Infraestructura tecnológica

Fuente: (Autores)

3. Sistema de gestión de almacenamiento de datos

En este sistema se utiliza el sistema de almacenamiento principal del ecosistema de Hadoop, conocido por sus siglas como HDFS, es decir, Hadoop Distributed File System. Este adquiere las mismas características de Hadoop en cuanto a escalabilidad y tolerancia de fallos permitiendo así el uso de clúster de bajas prestaciones. Como complemento a esta fase se ha seleccionado la base de datos no relacional y que usa NoSQL para extraer datos Hbase que se ejecuta en HDFS y que ofrece acceso aleatorio en tiempo real a los datos. Además, añade algunas capacidades transaccionales a Hadoop, lo que permite a los usuarios realizar inserciones, actualizaciones, búsquedas y eliminación. Por último, se utilizará Hive una infraestructura de data warehouse que facilita la administración de grandes conjuntos de datos que se encuentran almacenados en un ambiente distribuido y es orientada a datos SQL a diferencia de Hbase.

4. Gestión de Recursos

En lo relativo a las herramientas de gestión de los recursos del ecosistema Hadoop se ha seleccionado Yet Another Resource Negotiator (YARN), ya que es el responsable de administrar el acceso a los recursos críticos de Hadoop. Este soporta múltiples motores de acceso a datos heterogéneos que son analizados con el procesamiento por lotes, consultas interactivas, búsquedas o técnicas de Machine Learning. Estos se pueden ejecutar simultáneamente y son gestionados de forma centralizada por YARN. La arquitectura de YARN desacopla el modelo de programación de la infraestructura de gestión de los recursos y delega muchas funciones programadas a los componentes de cada aplicación.

Hue

Otra herramienta que se utilizará para la gestión de las demás herramientas que trabajan bajo el ecosistema de Hadoop es Hue. Esta herramienta proporciona un entorno de interfaz gráfica con el objetivo de administrar y gestionar todas las herramientas que se sitúan bajo la sombrilla de Hadoop, ofreciendo al usuario la oportunidad utilizar y trabajar con gráficos, visualización de datos y navegadores.

La razón por la que se escoge esta poderosa y útil herramienta es para facilitar todas las herramientas de Hadoop en un solo ambiente y hacer todas las operaciones de configuración y gestión de estas herramientas para usuarios no expertos.

5. Configuración y Mantenimiento

En lo concerniente a la configuración y mantenimiento del sistema se proponen las siguientes tres herramientas que pertenecen al ecosistema Hadoop y se complementan. Oozie funciona como un motor de flujo de trabajo utilizado para ejecutar trabajos de Hadoop de forma lógica, permite a los usuarios crear y programar transformaciones de datos complejas mediante la combinación de Hive y Sqoop en una sola unidad lógica de trabajo. Además, se encargará de coordinar junto a YARN diversas plataformas orientadas al procesamiento de los datos. ZooKeeper tendrá la responsabilidad de coordinar y dar soporte a las aplicaciones y los servicios distribuidos. Por último, Ambari una plataforma de gestión y seguimiento de los clústeres, asimismo se encarga de monitorizarlos, mantenerlos y asegurarlos. El conjunto de estas plataformas nos garantizara el buen funcionamiento del sistema a través del mantenimiento del correcto funcionamiento de los clústeres.

6. Procesamiento de los datos

Una de las frases cruciales en una arquitectura de Analítica de Big Data son las herramientas de procesamiento, cuya finalidad es hacer y gestionar cambios en los datos para aumentar la fiabilidad y calidad de estos, así como herramientas más avanzadas para el análisis de los datos. Se ha seleccionado Spark como el motor de manejo de datos que usa Hadoop, debido a que cuenta con soporte para aplicaciones de aprendizaje máquina, procesamiento de flujo de datos y visualización. Además, por su capacidad para mantener en memoria grandes conjuntos de datos lo hace ideal cuando se habla de procesos del sector salud. Este es compatible con lenguajes como Python y lenguaje R. También combina SQL, analítica compleja y en conjunto con el filtrado colaborativo se convierten en las tecnologías de preferencia para el manejo de los historiales médicos.

Solr es una plataforma de búsqueda e indexación de datos almacenados en HDFS, proporciona búsquedas indexando los documentos a través de XML, JSON, CSV o HTTP para ayudar a encontrar los patrones de datos, relaciones y correlaciones.

Durante la etapa de procesamiento se reutilizará Hive en combinación con el lenguaje R como alternativa para scripting y sobre el cual se pueden modelar algoritmos y técnicas de análisis, y transformar esas técnicas en conocimiento útil en el área de salud. Este lenguaje prevalece porque es más versátil y aunque la curva de aprendizaje es más complicada que Python, tiene soporte con la arquitectura Hadoop y SQL Server posee una integración para trabajarlo. Por lo tanto, se escogió también frente a la herramienta Pig y Scala.

Desde el punto de vista del apoyo a decisiones clínicas y gestión de recursos se ha decidido la incorporación de los árboles de decisiones que trabajan directamente con Hbase que como se expresó anteriormente se encargaran de verificar si se cumplen las condiciones de un atributo o un conjunto que procede de ellos, con la verificación de la información que ya se encuentra en la base para proceder con la toma de decisiones.

7. Análisis de información

Por último, se presentan dos potentes tecnologías para el análisis avanzando de los datos, Mahout que es una plataforma de álgebra lineal distribuida y de expresiones matemáticas para el mundo de la ciencia de datos, permitiendo la propia implementación de algoritmos y con una alta presencia del machine learning. Su uso junto Spark nos proporcionará una herramienta consistente para la clasificación de grupos de pacientes, búsqueda de tendencias y patrones, recomendación de posibles tratamientos que puedan ajustarse a las necesidades y condiciones del paciente. Hama es una herramienta para el procesamiento de datos de forma paralela, a través de técnicas de grafos, algoritmos de redes y Deep learning, muy útil para el campo de la epidemiología.

Mahout debido a que es tan sencillo como ejecutar un proceso Java Mahout indicándole que algoritmo usar, la entrada de datos y un directorio donde escribir la salida. Por esto mismo, el usuario no tiene que implementar una sola línea de código para hacer uso de esta herramienta. Por lo tanto, se destaca frente a opciones como Scikit-learn ya que tiene soporte muy limitado para redes neuronales y no se puede usar para problemas de Deep learning.

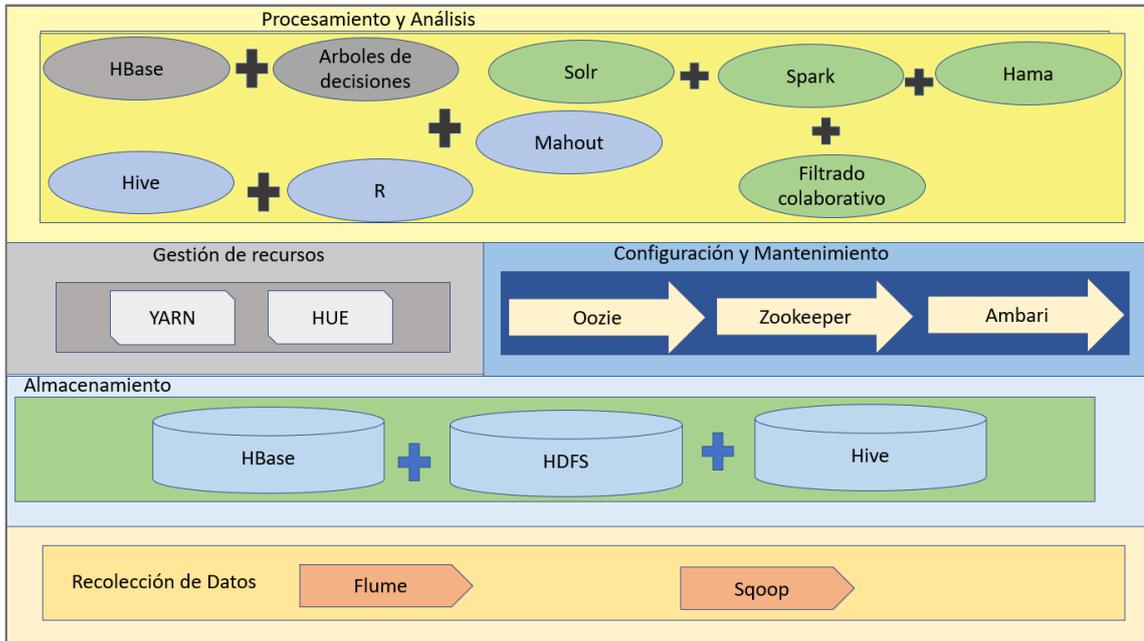


Figura 51- Ecosistema Hadoop

Fuente: (Autores)

La arquitectura descrita anteriormente nos permitirá obtener datos sobre la salud de la población gracias a las tecnologías y las técnicas de análisis de datos masivos que permitirán un análisis de datos a una gran velocidad, lo que otorgaría una posición ventajosa ante la prevención de enfermedades detectadas en alguna parte de la población.

Por otra parte, se podrán realizar análisis a partir de técnicas predictivas como es el caso planteado en la propuesta en conjunto con la visualización de información. De igual manera, el procesamiento distribuido que permite obtener una medicina preventiva y personalizada mediante el análisis del genoma humano, y también realizar un gran avance en cuanto a epidemiología se refiere. Asimismo, cabe destacar que las técnicas predictivas y el data mining son de gran ayuda en el campo de la farmacología siendo este un campo de acción de la propuesta, pudiendo avanzar en estudios predictivos.

En este mismo sentido, la realización de informes proporciona una visión holística de los pacientes, con la finalidad de prevenir enfermedades o minimizar sus efectos y en gran medida la prevención de efectos adversos en el campo de la farmacología, planteándose, como uno de los nuevos enfoques de la medicina predictiva y personalizada.

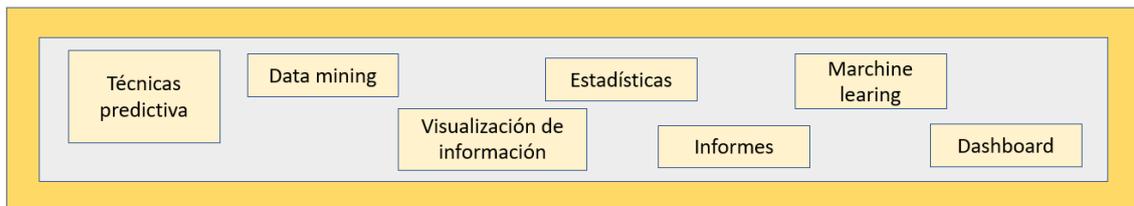


Figura 52- Análisis

Fuente (Autores)

8. Herramienta de Visualización

En lo relativo a la herramienta de visualización se ha decidido utilizar Tableau.

Se elige esta por su versatilidad con el ecosistema Hadoop y como una herramienta que soporta escalabilidad por la capacidad de conectarse a fuentes de datos remotas como Splunk o Hadoop, OData y Windows Azure Marketplaces y funciona con hasta un millón de filas de datos.

Aunque se trata de una herramienta cuyas capacidades se pueden aprovechar a través de la versión de pago. La capacidades y funciones que se pueden realizar con esta herramienta validan su calidad.

Otras capacidades que necesita el proyecto que se realiza en este trabajo requieren la posibilidad de conexión a nube y también habilidad de conexión con servicios web. Tableau es hábil para realizar análisis estadísticos, ver tendencias y pronósticos.

Un dato de suma importancia es el hecho de que la consultora Gartner sigue clasificándola como una de las herramientas líderes en el entorno de Business Intelligence y analítica de datos; basándose no solo en su fácil uso, si no en los comentarios positivos y en la satisfacción por parte de sus usuarios.

9. Seguridad de la Información

Todos los sistemas informáticos deben garantizar la seguridad de los datos, es decir, salvaguardar los tres pilares básicos que son la integridad, confidencialidad y disponibilidad de los datos, por ellos se ha seleccionado la plataforma de seguridad Apache Ranger que tiene como característica principal su capacidad para centralizar sus funciones. Esta ofrece controles de políticas para HDFS, Hive, HBase, Solr otros más. Dentro de las funcionalidades están:

- Permite el manejo de políticas para el acceso a ficheros, directorios, bases de datos, tablas y columnas. Estas políticas se pueden establecer para usuarios individuales o grupos.
- Administración de seguridad centralizada para administrar todas las tareas relacionadas con la seguridad en una interfaz de usuario central o mediante API REST.
- Autorización detallada para realizar una acción y / u operación específica con el componente / herramienta Hadoop y administrarse a través de una herramienta de administración central.
- Estandariza el método de autorización en todos los componentes de Hadoop.
- Soporte mejorado para diferentes métodos de autorización: control de acceso basado en roles, control de acceso basado en atributos, etc.
- Centraliza la auditoría del acceso del usuario y las acciones administrativas (relacionadas con la seguridad) dentro de todos los componentes de Hadoop.



Figura 53- Herramientas de seguridad

Fuente (Autores)

Como complemento a Apache Ranger, se implementará Apache Knox cuya función es ser una puerta de enlace perimetral que protege al clúster Hadoop. Permite un único punto de autenticación en un clúster Hadoop. Knox es capaz de proporcionar ayuda en el control, integración, supervisión y automatización de las necesidades administrativas y de análisis.

Las políticas de seguridad que soporta Knox son:

- Autenticación (LDAP y el proveedor de autenticación de *Active Directory*)
- Federación / SSO (encabezado HTTP basado en la federación de identidades)
- Autorización (*Service Level Authorization*)
- Auditoría

5.2.4 Prototipo de Interfaz de Usuario

A continuación, se muestra una serie de prototipos que deberán contemplar para el sistema del cual se alimentara la arquitectura de la analítica de Big Data para lograr el alcance de la propuesta. En estas se podrán observar de manera gráfica las opciones con las que contarán al desarrollar el sistema.

- **Prototipo de Interfaz de Usuario para Inicio**

Esta es la primera interfaz que aparece ante el usuario cuando ingresa al sistema. En esta se ofrece la opción de ingresar para utilizarlo. Cabe resaltar que mientras el usuario no se autentica, absolutamente todas las características del sistema se encontrarán ocultas. Esto aporta a lo que es la seguridad ya que no hay posibilidad de navegar sin previa autorización.

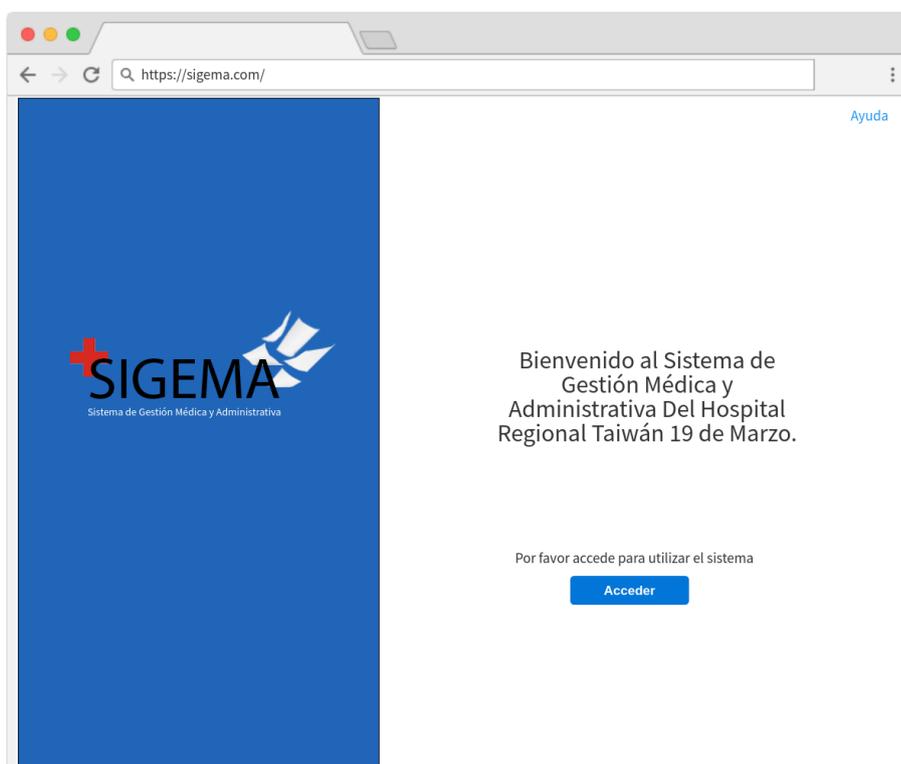


Figura 54- Pantalla inicial del sistema

Fuentes (Autores)

- **Prototipo de Interfaz de Usuario para Autenticación**

Por temas de seguridad, a diferencia de otras formas de ingresar a un sistema se han diseñado tres pantallas: la primera de ella solicita el nombre de usuario, la segunda la contraseña y por ultima la pregunta de seguridad.

La figura no. 55 es el prototipo de la primera pantalla para el proceso de autenticación en el sistema, la misma posee un campo donde debe ser ingresado el nombre de usuario de la persona que iniciará sesión. Una vez ingresado el nombre de usuario, se pulsa el botón color azul que dice “Siguiete”, lo que permite acceder a la siguiente pantalla.



Figura 55- Prototipo de Interfaz Gráfica (Autenticación 1).

Fuentes (Autores)

La figura no. 56 es el prototipo de la segunda pantalla para el proceso de autenticación en el sistema, la misma posee un campo donde debe ser ingresado la contraseña de la persona que iniciará sesión. Una vez ingresada la contraseña, se pulsa el botón color azul que dice “Acceder”, lo que permite acceder a la siguiente pantalla.



Figura 56- Prototipo de Interfaz Gráfica (Autenticación 2).

Fuentes (Autores)

Por último, la figura no. 57 es el prototipo de la tercera y última pantalla para el proceso de autenticación en el sistema, la misma posee dos campos donde debe seleccionar la pregunta de seguridad e ingresar su respuesta. Una vez realizado lo anterior, se pulsa el botón color azul que dice “Acceder”, lo que permite ingresar al sistema.



Figura 57- Prototipo de Interfaz Gráfica (Autenticación 3).

Fuente: (Autores)

- **Prototipo de Interfaz de Usuario para el menú**

A continuación, se muestra la pantalla principal que contiene los módulos del sistema los cuales serán claves para ingresar toda la información que el sistema necesitara para hacer todos los análisis correspondientes. Para acceder a cada uno de ellos, el usuario solo debe dar clic sobre el módulo que desee navegar.



Figura 58- Pantalla que muestra el menú del sistema

Fuente: (Autores)

- **Prototipo de Interfaz de Usuario para Creación de Historial Médico Electrónico**

Al dar clic al módulo “Historiales”, se desplegará la siguiente pantalla que contiene los datos fundamentales para la creación de los historiales médicos que es uno de los pilares sobre los cuales se soporta la propuesta. Se puede observar los campos que serán necesarios conocer para realizar análisis predictivos sobre la condición del paciente individual como la población.

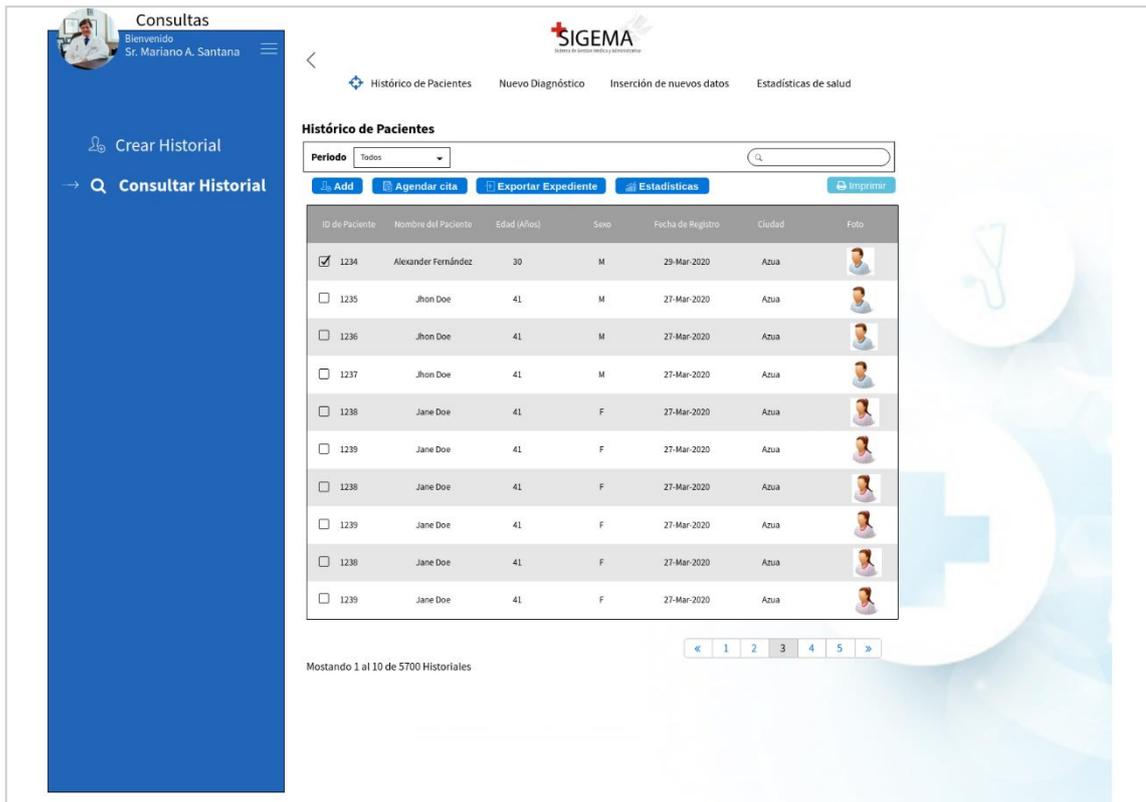


Figura 60- Listado de Pacientes

Fuente: (Autores)



Figura 61- Pantalla principal del Historial Médico

Fuente: (Autores)

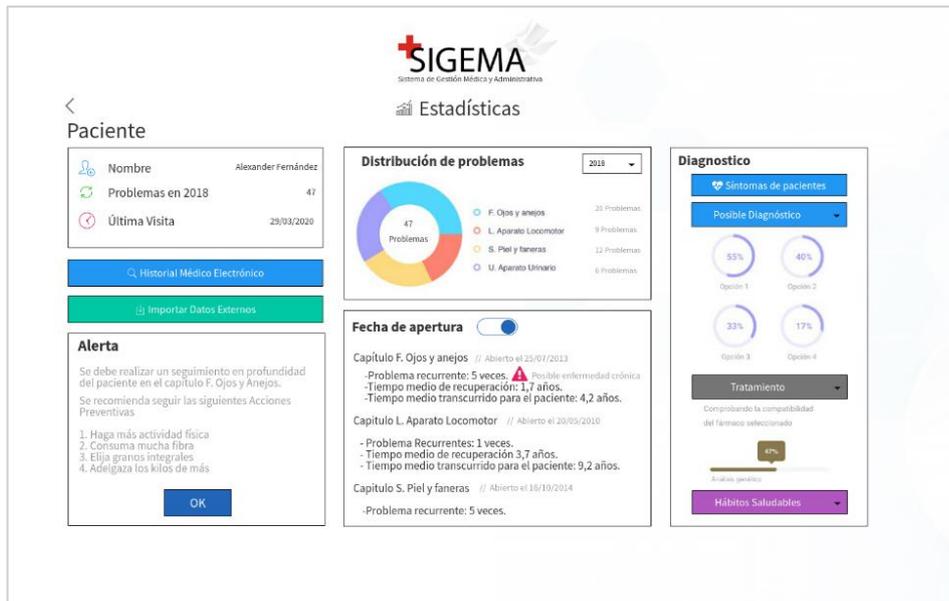


Figura 62- Análisis del historial médico

Fuente: (Autores)

- **Prototipo de Interfaz de Usuario para crear producto de cocina**

Al dar clic al módulo “Administración” y luego “Cocina”, se desplegará la siguiente pantalla que contiene los datos fundamentales para la creación de los productos de la cocina que son parte de los datos que son fundamentales para la propuesta. Se puede observar los campos que serán necesarios conocer para realizar análisis predictivos.

SIGEMA
Sistema de Gestión Médica y Administrativa

RRHH | **Cocina** | Farmacia | Finanzas

Administración
Bienvenido Sr. Mariano A. Serrata

→ **Crear Producto**
🔍 Buscar Producto

Cocina Registro de Nuevo Producto

ID de Artículo:
Nombre de Artículo:
Unidad de Medida:
Precio/U.M:
Cantidad:
Proveedor:
Categoría:
Ubicación Física:
Fecha:

Observación o comentario:

Cancelar Aceptar

Figura 63- Crear un producto de cocina

Fuente: (Autores)

- **Prototipo de Interfaz de Usuario para consultar productos de la cocina**

Así como se puede crea un producto de cocina, se puede consultar como se observa en la figura no. 64. Primero se selecciona la opción “Consultar”, y el sistema despliega el listado de productos, el usuario puede filtrar por nombre o código. En la figura no. 65 se puede observar la pantalla que despliega el análisis de ese producto.

The screenshot displays the SIGEMA (Sistema de Gestión Médica y Administrativa) interface. On the left, a blue sidebar contains the user's name 'Sr. Mariano A. Serrata' and options to 'Crear Producto' and 'Buscar Producto'. The main header shows the SIGEMA logo and navigation tabs for 'RRHH', 'Cocina', 'Farmacia', and 'Finanzas'. The 'Cocina' section is active, showing a search bar for 'Categoría' set to 'Todos'. Below the search bar are buttons for 'Agregar', 'Exportar Inventario', 'Estadísticas', and 'Imprimir'. A table lists kitchen products with the following data:

ID de Artículo	Nombre	U. Medida	Cantidad	Precio	Categoría	Ubicación	Total	
<input checked="" type="checkbox"/>	1235	Habichuela Roja	Libra	60	20	Cereal	Almacen1-05	\$1200
<input type="checkbox"/>	1236	Lechuga	Unidad	30	10	Cereal	Almacen1-04	\$1500
<input type="checkbox"/>	1237	Plátano Verde	Unidad	100	14	Viveres	Almacen1-01	\$1400
<input type="checkbox"/>	1238	Plátano Maduro	Unidad	100	14	Viveres	Almacen1-01	1600
<input type="checkbox"/>	1239	Guineo Maduro	Unidad	300	5	Fruta	Alacen1-04	1500
<input type="checkbox"/>	1240	Botellones de Agua	Unidad	300	200	Líquido	Almacen1-06	60000
<input type="checkbox"/>	1241	Aroz Integral	libra	50	50	Cereal	Almacen1-05	2500
<input type="checkbox"/>	1242	Pollo	Libra	100	60	Carnes	Almacen2-02	6000

At the bottom of the table, it indicates 'Mostando 1 al 8 de 30 Páginas' and a pagination control showing pages 1, 2, 3, 4, 5.

Figura 64- Listado de productos de la cocina

Fuente: (Autores)



Figura 65- Análisis del producto

Fuente: (Autores)

- **Prototipo de Interfaz de Usuario con la información de un fármaco**

Al dar clic al módulo “Administración” y luego “Farmacia”, se desplegará la siguiente pantalla que contiene los datos fundamentales para la creación de los fármacos que son parte de los datos significativos para la propuesta. Se puede observar los campos que serán necesarios conocer para realizar análisis predictivos sobre ese fármaco individual.

The screenshot shows the 'SIGEMA' web application interface. At the top, there is a logo and the text 'SIGEMA Sistema de Gestión Médica y Administrativa'. Below this, there are navigation tabs for 'RRHH', 'Cocina', 'Farmacia', and 'Finanzas'. The 'Farmacia' tab is selected, and the page title is 'Registrar Nuevo Producto'. On the left, there is a blue sidebar with a user profile 'Administración' and 'Bienvenido Sr. Mariano A. Serrata'. The sidebar contains two options: 'Crear Producto' and 'Buscar Producto'. The main form area contains the following fields:

- ID de Artículo:
- Nombre de Artículo:
- Unidad de Medida:
- Precio/U.M:
- Cantidad:
- Laboratorio:
- Proveedor:
- Categoría:
- Ubicación Física:
- Fecha:

There is also a large empty text area labeled 'Observación o comentario' and a 'Select' dropdown menu. At the bottom right, there are two buttons: 'Cancelar' (red) and 'Aceptar' (blue).

Figura 66- Plantilla para registrar un fármaco

Fuente: (Autores)

- **Prototipo de Interfaz de Usuario para consultar fármaco**

Así como se puede crea un fármaco, se puede consultar como se observa en la figura no. 67. Primero se selecciona la opción “Consultar”, y el sistema despliega el listado de fármacos, el usuario puede filtrar por nombre o código. En la figura no. 68 se puede observar la pantalla que despliega el análisis de ese fármaco.

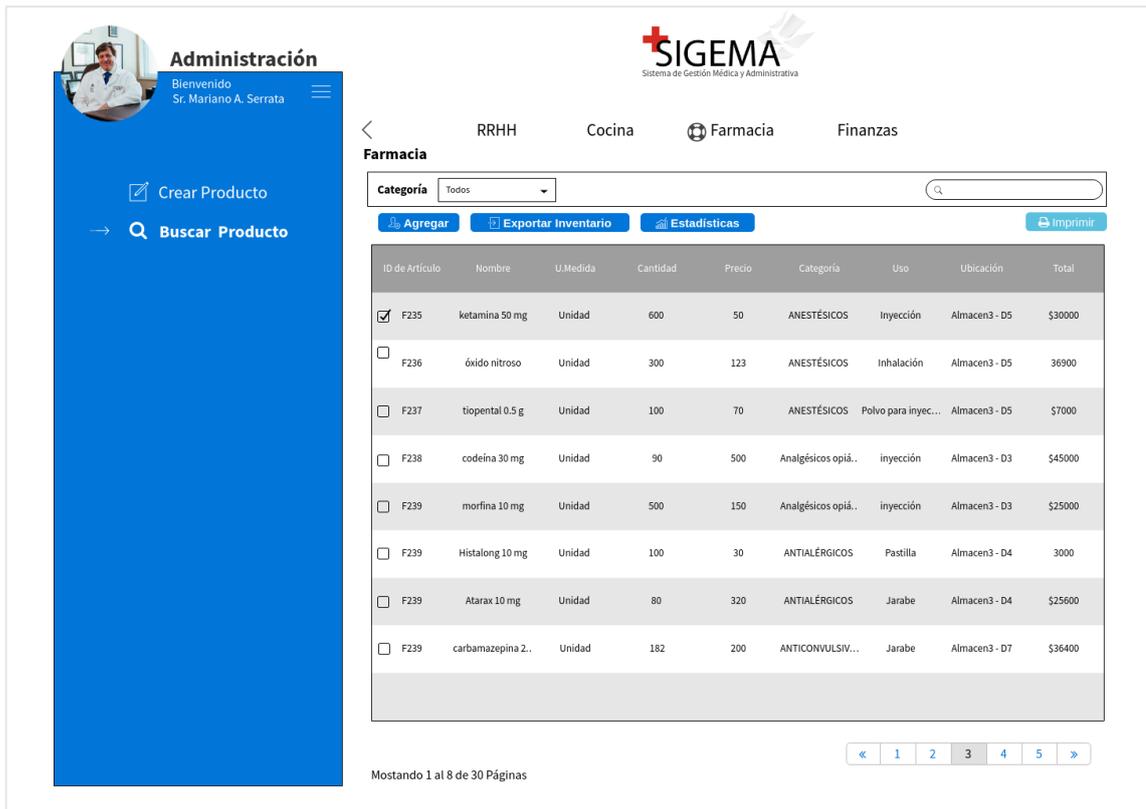


Figura 67- Listado de fármacos

Fuente: (Autores)



Figura 68- Análisis del fármaco

Fuente: (Autores)

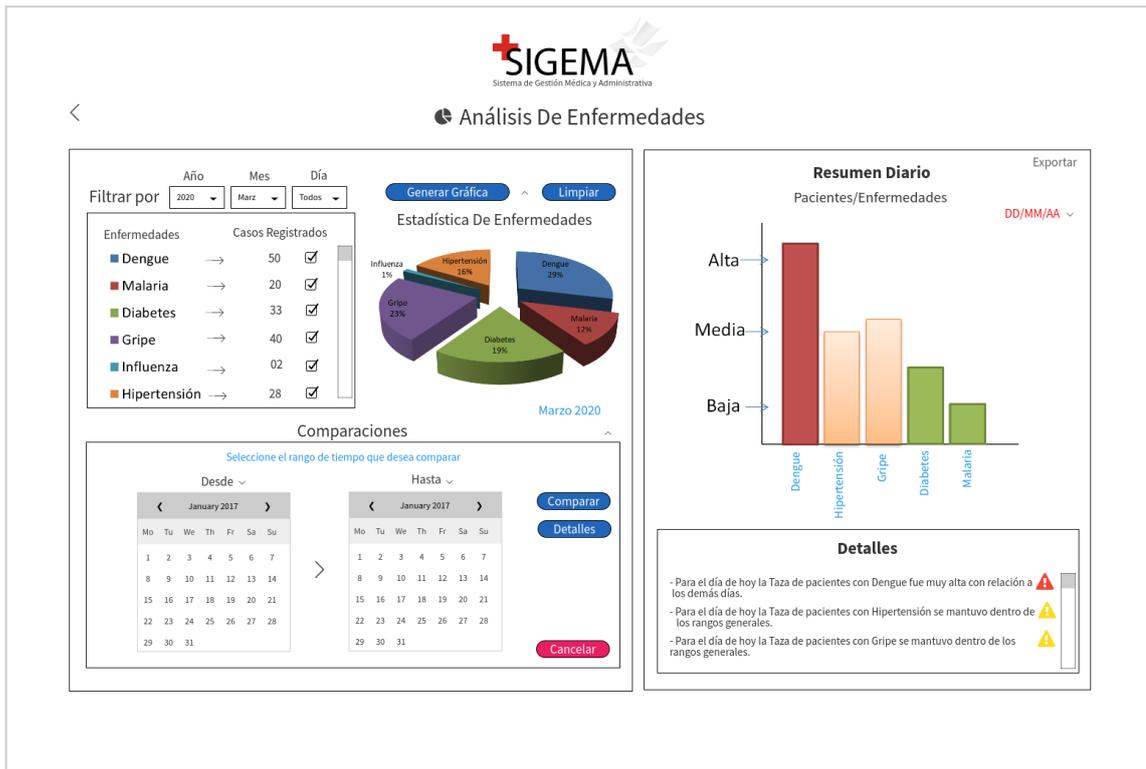


Figura 69- Predicción enfermedad

Fuente: (Autores)

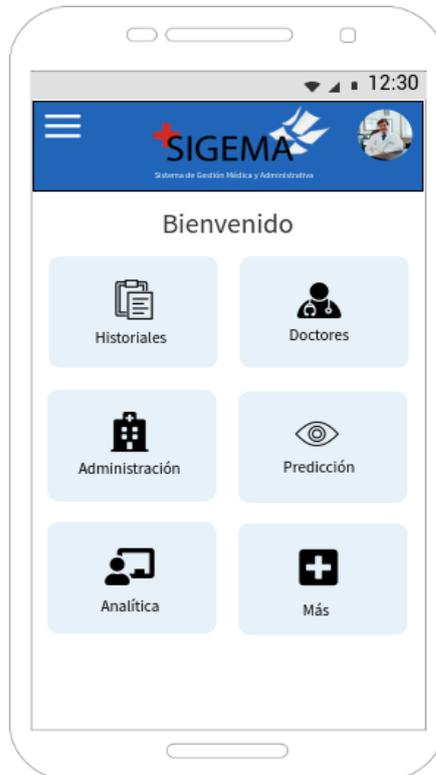


Figura 70- Prototipo Móvil

Fuente: (Autores)

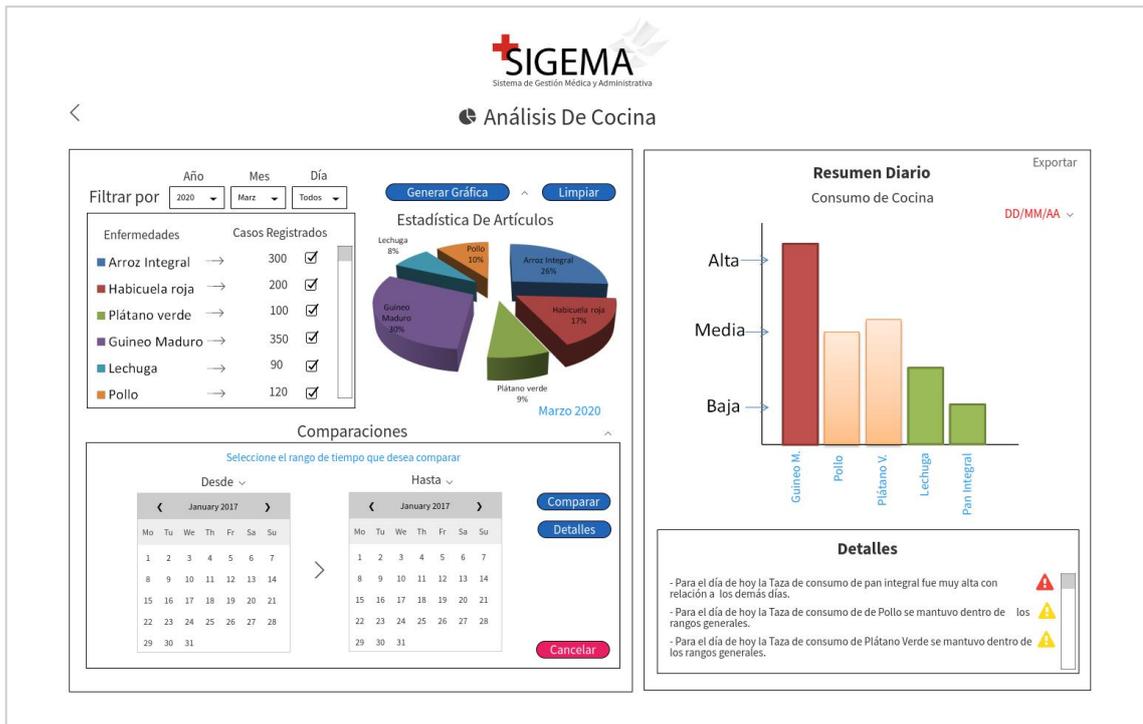


Figura 71- Predicción de productos de cocina

Fuente: (Autores)

5.2.5 Estudio de Factibilidad de la Propuesta

Cuando se habla de construcción y realización de proyectos, y no menos importante sobre el impacto que estos pudieran tener en un tiempo determinado, en este sentido es necesario hablar de la factibilidad o viabilidad de este. Según Kendall en su libro, la viabilidad de un proyecto debe ser medida bajo tres aspectos, el técnico que representa las tecnologías que soportan la propuesta, el operativo que se basa en la capacidad del personal existente para llevar a cabo el proyecto, y por último el económico en el que se determina si la inversión realizada tendrá retorno una vez se completado el proyecto.

De acuerdo con lo anterior se deben considerar los tres tipos de viabilidad y de esta manera, evaluar si un proyecto o idea resulta beneficioso, por el contrario, representa pérdidas. Es por esto por lo que no se debe pasar por alto la aplicación de esta en este proyecto de Big Data.

5.2.5.1 Factibilidad Técnica

Gracias a los avances tecnológicos obtenidos en la última década, con relación a las tecnologías que trabajan con ciencia de datos, Big Data y analítica de Big Data, en la factibilidad técnica se requiere destacar tres aspectos: Arquitectura de hardware, arquitectura de software y herramientas de ciencia de datos.

En primer lugar, la arquitectura de hardware que demanda esta solución puede ser provista por el Estado Dominicano a través de la Oficina Presidencial de Tecnología (OPTIC). En este sentido hablamos de los servidores y sus recursos.

- Colocación de Servidores.
- Servicio de espacio de almacenamiento en la nube.
- Servicio de Redundancia de Procesamiento (DCP).
- Espacio y servicios para alojamiento de Aplicaciones.

En caso de que no quiera ceder, la gestión de hardware y los demás aspectos pasarían a manos de un tercero u empresa privada, bien pudiera contratarse los servicios de AWS de Amazon o bien Google Cloud Services. De manera que el almacenamiento, gestión y distribución de datos estaría garantizada, además la seguridad de los datos que sería tomada en cuenta. Con la salvedad que esto implicaría en la factibilidad económica.

El proyecto es viable en cuanto a software se refiere, debido a que son herramientas intuitivas y fácil adquisición. Como una segunda opción está adquirir servicios de terceros, como podría ser Amazon que provee a empresas, corporaciones, un servicio llamado Software como Servicio en que permite una arquitectura que provee todo el software requerido, pero el factor económico incide directamente.

5.2.5.2 Factibilidad Operacional

La propuesta trae consigo una adecuación y diseño en la captura de los datos para el historial médico electrónico de los pacientes, es propicio entrenar correctamente a los usuarios en el uso de las aplicaciones, pues es necesario que los datos introducidos ingresen de manera correcta y limpia.

Por otro lado, para implementar, mantener la arquitectura y el sistema se quiere la incorporación de profesionales competentes en el manejo de ciencias de datos como son los analistas y especialistas de datos con conocimiento de analítica de Big Data. De igual manera, capacitar a los altos directivos para que comprendan las analíticas producto del sistema y tomen las decisiones oportunas.

5.2.5.3 Factibilidad Económica

Al aplicar este sistema y arquitectura de analítica de Big Data en el Hospital Taiwán 19 de Marzo sería el primer proyecto de esta índole que incorporaría el sector de la salud dominicano, y la República Dominicana avanzaría en cuanto a medicina predictiva y preventiva se refiere. De igual manera, trae consigo una serie de beneficios económicos a raíz de la eliminación de procesos manuales con la automatización de esto.

El sistema propuesto, no solo ofrece beneficios económicos a este centro de salud, sino también a los pacientes, aseguradora y al propio estado dominicano. De manera desglosada, estos serían algunos de los principales factores económicos que apoyan la propuesta:

- Reducción de los costos de operación en hospital Taiwán, se dice esto porque contar con el apoyo de un sistema que te ayude a descubrir con antelación enfermedades y patrones que ayuden a predecir, y posteriormente ofrecer acciones para erradicar y minimizar el impacto de estas.
- Disminución de gastos de operación del presupuesto a partir de toma de mejores decisiones en las áreas involucradas en el presupuesto del centro, esto mediante la incorporación de algoritmos y herramientas analíticas que apoyen la toma de decisión en las áreas a fines como es el caso de la selección de los alimentos, su tipo, cantidad y duración.
- Minimización de los gastos con relación al 40% del presupuesto que se extrae para los medicamentos, esto se logra por medio de que la solución es capaz de recomendar mejores proveedores y medicamentos de calidad superior, a más bajo costo.
- Finalmente, aunque es un aspecto que se mencionó en la viabilidad técnica, el impacto de la inversión en tecnología sería mínimo ya que las herramientas utilizadas en esta solución son código abierto. Además de que los gastos de la arquitectura de hardware se estarían solicitando los recursos provistos por la OPTIC debido a que el centro es de carácter público y trabaja con recursos del estado.

5.2.6 Escalabilidad

La implementación de SIGEMA contempla la idea de un sistema que no solo serviría únicamente en el Hospital Regional Taiwán, sino que es hábil para aplicarse en todo el territorio nacional, con esta iniciativa lo que se persigue es incluir todo el sistema de salud de la República Dominicana a funcionar bajo un esquema estándar y de esta manera reunir tanto a instituciones como a profesionales de la salud, incluidos centros de investigación.

Centros de Salud

Luego del lanzamiento e incorporación del SIGEMA se realizaría un programa para presentar la idea de negocios al Estado Dominicano como miras a la centralización de esta solución en el marco de la salud predictiva y preventiva, como la e-salud. De manera que se incluirían todos los centros públicos y privados.

Centros de investigación

Se incorporarían los centros de investigación, esto incluye a universidades, institutos especializados, y otros centros de investigación. Esto debido a que se tendrían fuentes confiables para futuras evaluaciones y capacidades de predicciones más atinadas.

Esta acción incluye la utilización de conjuntos de datos de brotes que poseen instituciones de renombre tanto nacionales como internacionales.

Laboratorios

Se incorporarían también los laboratorios que se constituyen como fuentes de datos fiables. Esto ayudaría a tomar en cuenta varios factores con relación a los procesos que se realizan en estos lugares y también habilitarían información de proveedores.

5.2.7 Cronograma de Implementación

El cronograma de implantación de esta prepuesta se ha segmentado en seis etapas ya que por la complejidad de este tipo de proyectos se requiere personal experto en el campo de la ciencia de datos y Big data, por lo cual las siguientes etapas serán aplicadas en una duración de cuatro meses:

1. Definición y puesta en marcha de la infraestructura de Hardware y Software

- Configuración e instalación de Servidor y su arquitectura
- Instalación y Definición de servidor de base de datos SQL Server
- Instalación de servicios y Servidor Web

- Configuración de ambiente para herramientas de Hadoop
- Publicación de Pagina Nuevo Portal (SIGEMA)
- 2. Definición de las fuentes de datos y sus tipos**
 - Definición de Fuentes de datos estructurados
 - Definición de fuentes de datos no estructurados
 - Definición de fuentes de datos semiestructurados
 - Adaptación y adecuación de base de datos de historiales médicos del centro
 - Diseño y construcción de Data WareHouse
- 3. Recolección de datos**
 - Instalación y configuración de Flume
 - Instalación y configuración de Sqoop
 - Diseños y Configuración de herramientas y Framework para ETL
 - Recolección de datos provenientes de fuentes estructuradas
 - Recolección de datos provenientes de fuentes no estructuradas
 - Recolección de datos provenientes de fuentes semiestructuradas
- 4. Almacenamiento de datos y configuración de data lake**
 - Contratación de servicios de AWS y definición data lake
 - Migración de data estructurada a Data lake
 - Migración de data no estructurada a Data lake
 - Instalación y configuración de Hbase
 - Instalación y configuración HDFS
 - Instalación y configuración de Hive
- 5. Incorporación y configuración de herramientas de procesamiento y análisis**
 - Instalación y configuración de Solr
 - Instalación y configuración de Spark
 - Instalación y configuración de Hama

- Instalación y configuración de Mahout

- Instalación y configuración de R

6. Definición y puesta en marcha de las herramientas de visualización

- Instalación y configuración de Tableau
- Diseño y desarrollo de repodes e informes
- Diseño y desarrollo de Dashboard

7. Implementación de herramientas de configuración para manejo arquitectura

- Configuración de Hue
- Instalación y configuración de Oozie
- Instalación y configuración Zookeeper
- Instalación y configuración de Amari
- Instalación y configuración de YARN

8. Desarrollo de Pruebas

- Pruebas de Ambiente Hadoop
- Pruebas de Usuario final

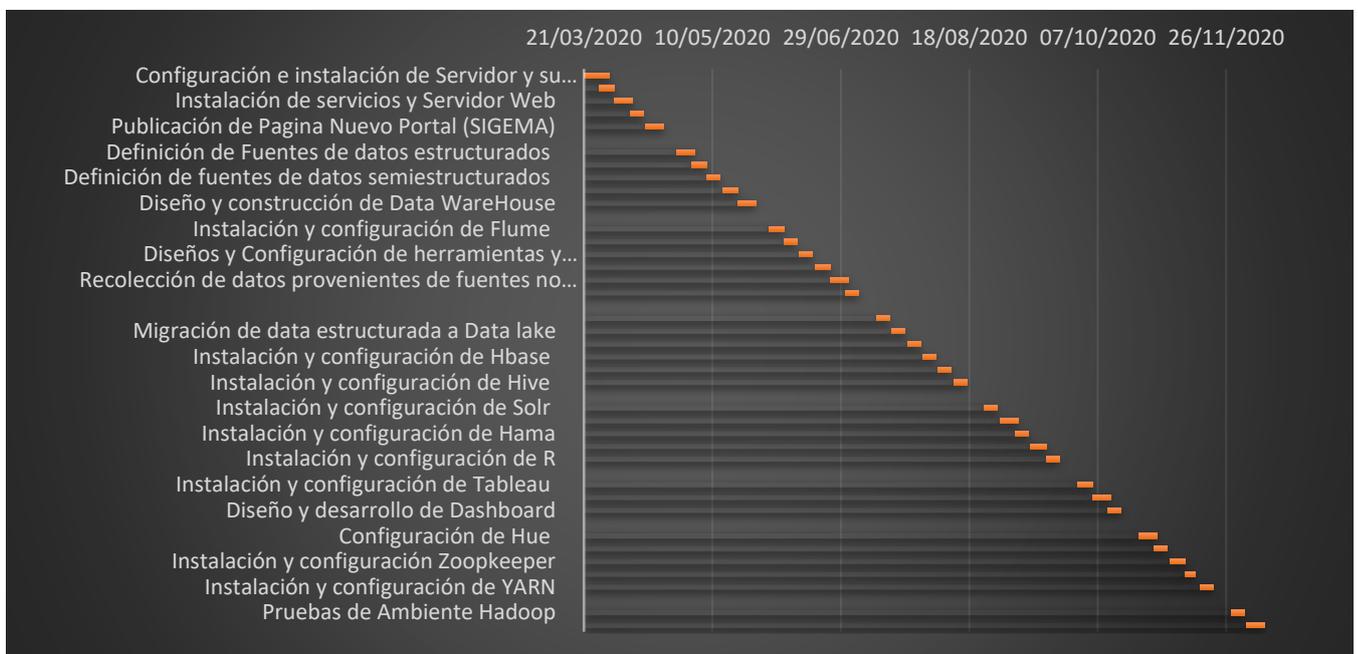


Figura 72- Diagrama de Gantt Cronograma de Implementación

Fuente (Autores)

CONCLUSIÓN

Un porcentaje significativo de los historiales médicos en las clínicas y hospitales de la República Dominicana aún son manejados de manera manual, utilizando archiveros, anaqueles y cuartos de documentos. Algunos centros médicos, han optado por automatizarlos por medio de un sistema, lo que les permite saber de manera eficiente la ubicación exacta de estos, pero no la mejor forma de gestionarlos para su posterior uso.

El sistema propuesto busca digitalizar estas informaciones que se encuentran guardadas en estos documentos localizados en archiveros, de forma que sea posible acceder a estos y utilizarlos de la mejor manera posible. Con la implementación de este se reducen los tiempos entre procesos, se agiliza la gestión y la precisión de los datos de los pacientes, aumentando así la comodidad y efectividad del trabajo del médico.

El foco de esta propuesta no se limita solamente a esto, ya que tiene como finalidad la explotación masiva de estos archivos, implementando nuevas tecnologías que van más allá de los métodos tradicionales, siendo este nuevo sistema capaz de procesar todo tipo de datos para la predicción de posibles problemas relativos a la salud como brotes, enfermedades, epidemias, así como también proyecciones para la toma de decisiones y mediante herramientas de Big Data.

Por tanto, con la implementación esta propuesta, es posible predecir factores que pueden afectar el presupuesto en las distintas áreas clínicas. Este sistema es capaz de analizar las informaciones de los gastos que se generan a partir de las necesidades y exigencias del hospital Taiwán.

CONCLUSIONES

Ante la definición de la agenda 2030 por la Organización De Las Naciones Unidas (ONU) de los 17 objetivos de Desarrollo Sostenible (ODS) todos los países miembros incluyendo República Dominicana, tienen la obligación de cumplir todas las metas que se han trazado para cada objetivo. Este proyecto está orientado a hacer frente al objetivo no. 3 “*Garantizar una vida sana y promover el bienestar para todos en todas las edades*”, a través de la incorporación de la tecnología al proceso de análisis de la información que reside en los históricos médicos de los pacientes que acuden al Hospital Regional Taiwán 19 de Marzo Azua. Este último es el centro principal de la región sur de R.D.

Con el avance de la tecnología y la forma en la que esta se ha estado introduciendo en todas las áreas, se está volviendo un elemento indispensable en nuestras vidas. Actualmente, procesos que solían ser tediosos, largos, repetitivos y complicados ha sido posible automatizarlos y hacer que los mismos sean muchos más fáciles y sencillos de realizar de una manera más rápida y eficaz reduciendo incluso costos.

Por tanto, con la aplicación de diversas técnicas de Big Data en el sector salud y la observación de la carencia de un modelo que englobe el funcionamiento de un sistema y el hospital Taiwán maneja un alto volumen de datos que los sistemas tradicionales no podrían manejar se ha decidido la realización de esta propuesta para cubrir nicho aun no cubierto no solo en este centro sino en todo el sector de salud dominicano basada en el diseño de una arquitectura de analítica de Big Data, debido a que es necesario contar con un sistema que pueda recibir diversos tipos de datos sin importar su estructura, su tamaño, su origen como su estado y el concepto de Big Data que puede denominarse una tecnología nueva es la que encaja perfectamente para el diseño de la arquitectura.

Se ha logrado con los objetivos definidos de resaltar la importancia de un sistema de información con las cualidades explicadas en el modelo para la mejora de la calidad sanitaria, nuevos enfoques médicos y la eficiencia económica. Esto mediante el diseño de una arquitectura que abarca el funcionamiento del sistema de información y el flujo de los datos desde las fuentes hasta la toma de decisiones.

Cabe mencionar que la colaboración de la directiva del Hospital Regional Taiwán 19 de Marzo, ha sido clave para la obtención de datos que permitieron conocer a fondo la situación en la cual se encuentra sumergido este centro.

Durante el proceso investigativo se puede observar la deficiencia que existe no solo en este de salud sino en el sector completo, sobre las tecnologías que pueden aplicarse en diversas áreas de la salud con la finalidad de mejorar su funcionalidad, facilitar la toma de decisiones y sobre todo ofrecer un servicio de calidad. Hasta el momento ningún centro de la República Dominicana cuenta con un tipo de sistema como el que se ha propuesto.

Este trabajo de investigación tiene un enfoque sobre el método actual utilizado para la gestión y el manejo de los historiales médicos en el Hospital Taiwán, debido que este es la fuente primordial para todo tipo de decisiones, el cual se realiza de forma prácticamente manual, lo que hace que este proceso sea más riesgoso, tedioso, lento y a su vez al no estar digitalizados dificulta que todas estas informaciones puedan ser utilizadas para beneficio de la comunidad. No obstante, con las implementaciones de nuevas tecnologías es posible subsanar y optimizar la distribución más efectiva del presupuesto general de las diferentes áreas para así lograr cumplir con las necesidades que demanda el centro y a su vez evitando que al fin de cada periodo no hayan podido cubrir todas las necesidades.

De acuerdo con este trabajo, en base a estas necesidades que se han mencionado, de las cuales los métodos tradicionales ya se están viendo desfasados por los sistemas tecnológicos que son capaces de realizar la misma tarea, de manera más efectiva y a su vez sacando mayor provecho de los mismos insumos existente, mediante la utilización de un sistema de analítica de Big Data. Esta tecnología es relativamente nueva en el país, pero desde sus inicios ha demostrado ser robusta, adaptable, eficiente y con grandes resultados en las diferentes áreas del saber. Por lo cual, se ha construido una arquitectura basada en Hadoop para hacer capaz de analizar los historiales médicos de los pacientes y poder relacionar esa información con factores externos tales como el entorno, las enfermedades más comunes, las opiniones generales, entre otros, logrando hacer predicciones de sucesos importante. De igual manera, realiza un análisis de las necesidades del centro de cara a su historial de consumo y otros factores para sugerir formas óptimas de distribución del presupuesto.

Hemos elegido esta tecnología por encima de otras más tradicionales, porque la misma es más robusta y está en tendencia de desarrollo y crecimiento al igual que este proyecto, el cual tenemos como visión que pueda ser utilizado en todos los centros médicos del país. Dicha tecnología es perfecta para este tipo de situaciones, por capacidad de procesar cantidades masivas de datos de diferentes tipos y fuentes y se adapta a casi todas las nuevas tecnologías y por ende es capaz solucionar grandes problemáticas de esta índole y no obstante unificar los historiales médicos de pacientes a nivel nacional.

Esta propuesto es la antesala a un sistema centralizado en todo el país con la unificación de los historiales médicos de los pacientes, de manera tal que se podría consultar desde cualquier lugar. Por consiguiente, si un paciente tuviera la necesidad de pasar por un centro médico diferente al de costumbre, se podría verificar su récord agilizando así cualquier proceso y que al mismo tiempo brinda facilidad de costo porque las herramientas utilizadas son de código abierto exceptuando la herramienta de visualización.

RECOMENDACIONES

- Invertir en la implementación e incorporación de políticas que apoyen el desarrollo de tecnologías para aseguramiento y sostenimiento de la integridad de datos, debido a la importancia que representan los datos en esta era de la información, no menos importante el hecho de que este proyecto maneja datos confidenciales de las pacientes.
- Invertir en capacitación de los recursos humanos en manejo de tecnologías de la información y comunicación, prestando especial atención a manejo de herramientas de datos e inteligencia de negocios.
- Establecimiento de metas concretas y un modelo de evaluación basado en la mejora continua, para asegurar el compromiso con la calidad, estándares y mejores prácticas.
- Designación de un departamento de ciencia de datos que vele el buen funcionamiento y el mantenimiento de las estructuras de ciencias de datos, analítica de Big Data y que apoye al personal técnico y administrativo. Esto supone y asegura el mejor uso y aprovechamiento de las capacidades de esas herramientas a todo el personal operativo.
- Uso de fuentes de datos de confianza, certificadas y acreditadas, esto porque se está manejando datos relacionados con la salud, por lo tanto, con la vida de las personas. Aquí se incluyen revistas científicas e instituciones que cumplan con los más altos estándares de validación de información en términos científicos.
- Acorde con la recomendación anterior se recomienda que este proyecto tenga acceso a los DataSets relacionados con la familia de Virus Coronavirus, esto nos da acceso a todas las fuentes científicas de instituciones como la Casa Blanca en los Estados Unidos que ha provisto sus conjuntos de datos sin costo alguno.

REFERENCIAS

- Ruiz, B. (2018). *Algunas aplicaciones de la ciencia de datos y su futura importancia*. Press Noticias y Negocios. Recuperado de My: <https://www.mypress.mx/tecnologia/aplicaciones-ciencia-dato>
- Ortiz, E. (11 de enero de 2019). Historia del data science, la ciencia de datos. Pweb. Recuperado de <https://pweb.info/historia-del-data-science-la-ciencia-de-datos/>
- Kapow Software(s.f.). Infographic. Columnfivemedia. Recuperado de <https://www.columnfivemedia.com/work-items/infographic-intelligence-by-variety>
- Diebold, Francis X. (21 de septiembre de 2012)., On the Origin (s) and Development of the Term 'Big', 2012. Papers.ssrn. Recuperado de http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2152421
- Anónimo (24 de Feb de 2010). "Elementos de un Sistema de Información". Econlink. Recuperado de <https://www.econlink.com.ar/sistemas-informacion/elementos>
- IBM (s.f). IBM Knowledge Center. IBM. Recuperado de <https://www.ibm.com/support/knowledgecenter?origURL=api/redirect/dzichelp/v2r2/index.jsp>
- Soares, S. (22 de Julio de 2012). Big Data Reference Architecture.sunilsoares. Recuperado de <http://sunilsoares.wordpress.com/2012/07/22/big-data-reference-architecture-2/>
- Allweil, E. (6 de julio del 2015). Apache Org. Cwikiapache. Recuperado de <https://cwiki.apache.org/confluence/display/PIG/PigTools>
- Kumar, A. (24 de enero 2018). Health Care Informatics Research. researchgate. Recuperado de https://www.researchgate.net/publication/325039479_HEDEA_A_Python_Tool_for_Extracting_and_Analysing_Semi-structured_Information_from_Médical_Records
- Morales, A. (31 de agosto del 2017). Lenguajes de programación para realizar ciencia de datos. mappinggis. Recuperado de <https://mappinggis.com/2019/07/lenguajes-de-programacion-para-realizar-ciencia-de-datos/>
- TECHVIDVAN, T. (20 de noviembre de 2019). R Applications – 9 Real-world Use Cases of R programming. techvidvan. Recuperado de <https://techvidvan.com/tutorials/r-applications/>

- hadoop.apache (2016). General format. Apache. Recuperado de <https://hadoop.apache.org/>
- Universidad de Atlanca (24 de diciembre de 2019). Características de la data analytics: 4 formas de analizar datos. master-data-scientist. Recuperado de <https://www.master-data-scientist.com/caracteristicas-data-analytics/>
- Gil Press (28 de mayo del 2013). A Very Short History of Data Science. Forbes. Recuperado de <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/no.748ab44f55cf>
- Wang, Y., Kung, L., & Byrd, T. A. (2018). Big Data analytics: Understanding its capabilities and potential benefits. *Technological Forecasting & Social Change*, 3–13.
- Camargo, J. J. Camargo, J. F. y Aguilar, L. (2015). Arquitectura tecnológica para Big Data. *Revista Científica*, 21, 7-18. Doi: 10.14483/udistrital.jour.RC.2015.21.a1
- Seller, C. A. (2019). Diseño de un modelo para la creación de conocimiento y toma de decisiones en el área sanitaria (trabajo de grado). Universitat Politècnica de València.
- Morros, R. S., & Picaño, J. S. (2013). Big Data - Análisis de herramientas y soluciones. Universidad de Barcelona.
- Oficina Regional de la UIT para las Américas. (2018). Estudio sobre TIC y salud pública en América Latina la perspectiva de e-salud y m- salud. Suiza, Ginebra: ITU. Recuperado de https://www.itu.int/pub/D-STR-E_HEALTH.13-2018
- Salud, O. P. (2016). *Electronic Médica l Records in Latin America and the Caribbean: An Analysis of the current situation*. Washington, D.C.
- Reyes, G. V., Ernesto Bautista Thompson, J. A., & Penna, A. F. (2018). *MODELOS DE TECNOLOGÍAS DEL BIG DATA*. México: Universidad Popular Autónoma del Estado de Puebla.
- Secretaría de Salud de México. (2011). *Manual del Expediente Clínico Electrónico*. México.
- Carnicero, J., & Fernández, A. (2012). *Manual de Salud Electrónica*. Santiago de Chile.
- Bertolucci, J. (2013, Diciembre 31). *informationWeek*. Retrieved from Big Data Analytics: Descriptive Vs. Predictive Vs. Prescriptive: <https://www.informationweek.com/big-data/big-data-analytics/big-data-analytics-descriptive-vs-predictive-vs-prescriptive/d/d-id/1113279>
- Wu, M. (2013, Marzo 14). *Kroros ATLAS*. Retrieved from Big Data Reduction 1: Descriptive Analytics: <https://community.khoros.com/t5/Science-of-Social-Blog/Big-Data-Reduction-1-Descriptive-Analytics/ba-p/77766>
- Mariscal, J; Rentería, C & Arteaga, R. (2014). Lecciones de gestión pública en e-salud. Revista Digital del Departamento de Telemedicina de la Secretaría de Salud de Nuevo León y CENETEC- SALUD. Recuperado de:

<http://teleiberoamerica.com/publicaciones/Revista-eSalud-Mexico-2014.pdf>

- Anchiraico, J. (2017). Diseño De Una Arquitectura Big Data Para La Predicción De Crisis en El Trastorno Bipolar (tesis de posgrado). Universidad Complutense de Madrid.
- Andreu Pérez, J., Poon, C. C., Merrifield, R. D., Wong, S. T., & Yang, G.-Z., Big data for health. *IEEE Journal of Biomédica l and Health Informatics*, 19(4), pp.1193–1208, 2015.
- Hiltbrand, T., Behavior-Based Budget Management Using Predictive Analytics. *The Business Intelligence Journal*, 18(INL/JOU-12-26713), 2013: <http://www.osti.gov/scitech/biblio/1072389>.
- Zorrilla, M., García, D. (2017). Arquitecturas y tecnologías para el big data. Universidad de Cantabria.
- Harrison et al. *Next Generation Databases: NoSQL, NewSQL, and Big Data*. 2015. Apress.
- Gui1, H., Zheng1, R., Chao, M., Fan, H., y Xu, L. (2016). An Architecture for Healthcare Big Data Management and Analysis. *Springer International Publishing*, 100(38), 154–160 páginas.
- Raghupathi, W., Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2, 3. doi:10.1186/2047-2501-2-3.
- López D. (2016). Analysis of the possibilities of use of Big Data in organizations. (tesis de posgrado). Universidad de Cantabria, España.
- Instituto Gallego de Promoción Económica, Alianza Tecnológica Intersectorial de Galicia. (2019). Estado Del Arte de big data. Junta De Galicia, España.
- Reina A. (2018). San Francisco Open Data. (trabajo de grado). Universidad de Alicante, España.
- Linares G. (2016). La analítica de Big Data en el sector sanitario. (trabajo de grado). Universidad Oberta de Catalunya, España.
- Almela C. (2019). Diseño de un modelo para la creación de conocimiento y toma de decisiones en el área sanitaria. (trabajo de grado). Universitat Politècnica de València, Valencia.
- Galeano L., & Domínguez A. (2017). Prototipo de Laboratorio Hadoop Para Análisis Big Data en la Institución Universitaria Politécnico Grancolombiano. (trabajo de grado). Universidad Politécnico Grancolombiano, Colombia.
- Raghupathi, W., Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2, 3. doi:10.1186/2047-2501-2-3.
- ACCIONA (Dirección). (2017). *¿Qué es el Big Data Analytics?* [Video].
- Apache. (18 de 01 de 2019). Apache. Recuperado de <https://sqoop.apache.org/>

ANEXOS

ANEXO 1- ENCUESTA MÉDICOS

Estimado/a Encuestado/a, la información recolectada a partir de esta encuesta será utilizada para evaluar la adecuada mitigación de las barreras existentes que permitirá culminar con éxito la realización del trabajo de grado de UNAPEC en el que se propone el de diseño de un sistema de información para la toma de decisiones médicas y administrativas aplicando analítica de Big Data sobre el historial médico de los pacientes, razón por la cual será necesario conocer los aspectos siguientes.

- 1. ¿Conoce usted si en el área en la que presta servicios cuenta con algún tipo de sistema informático?**
 - SI
 - NO
- 2. ¿Conoce usted los historiales médicos electrónicos?**
 - SI
 - NO
- 3. ¿Sabe usted lo que es una analítica de Big Data aplicada al sector salud?**
 - SI
 - NO
- 4. ¿Considera usted que es necesario adaptar los servicios del centro de salud a las nuevas tecnologías de información y comunicación?**
 - SI
 - NO
- 5. ¿Le gustaría tener un sistema que pueda predecir brotes de virus antes de que estos se propaguen y dejen mayores consecuencias?**
 - SI
 - NO
- 6. ¿Le gustaría que al momento de recibir un paciente pueda obtener de manera inmediata su historial médico?**
 - SI
 - NO

7. ¿En qué medida considera la adopción de nuevas tecnologías para la predicción de eventos en el sector salud como puede ser la propagación de un virus como el covid-19?

- 25%
- 50%
- 75%

8. ¿Utiliza medios tecnológicos para el trato con los pacientes?

- SI
- NO

9. ¿Le gustaría un sistema que arroje estadísticas que permita la toma de decisiones médicas?

- SI
- NO

10. ¿Cuáles de estas predicciones sería de mayor importancia para el desarrollo de sus funciones en el centro de salud?

	Nivel de importancia		
	Alto	Medio	Bajo
Predecir brotes de virus			
Predecir la disponibilidad de las camillas			
Predecir la cantidad de farmacéuticos a utilizar			
Predecir la cantidad de pacientes			
Predecir el presupuesto de centro			

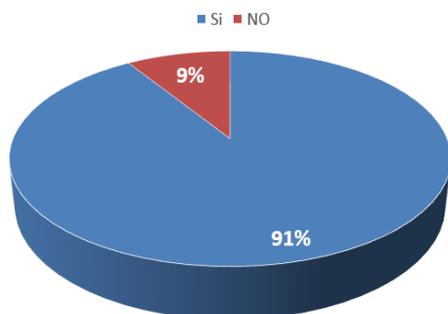
11. ¿Cuáles son las enfermedades más frecuentes?

12. ¿Cuáles serían los principales beneficios de la propuesta?

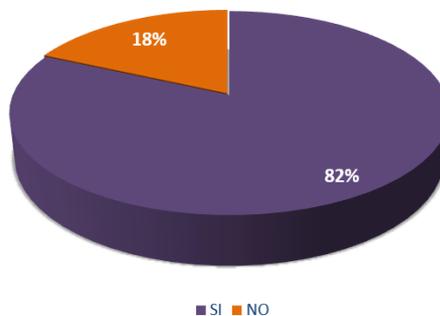
Beneficios	Nivel de Influencia		
	Alto	Medio	Bajo
Minimización de errores.			
Disminución de tiempo y costos.			
Reducción del tiempo de búsqueda de información.			
Incorporación de toda la información del paciente.			
Visualización de informaciones importantes para la gestión médica.			
Reducción de papeles.			
Predecir brotes de virus o enfermedades			
Análisis de los historiales médicos de los pacientes			

Anexo 1.1- TABULACION ENCUESTA MÉDICOS

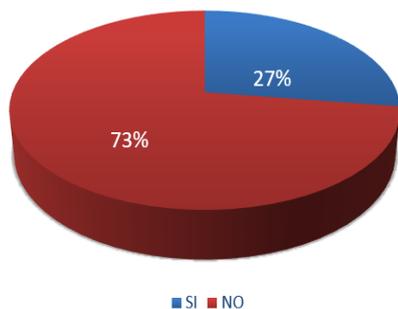
1 ¿CONOCE USTED SI EN EL ÁREA EN LA QUE PRESTA SERVICIOS CUENTA CON ALGÚN TIPO DE SISTEMA INFORMÁTICO?



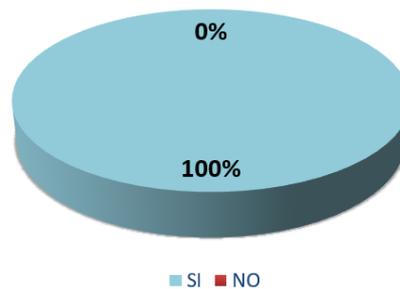
2.¿CONOCE USTED LOS HISTORIALES MÉDICOS ELECTRÓNICOS?



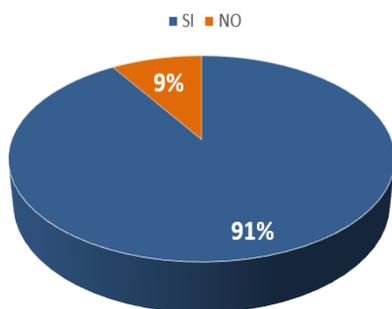
3.¿SABE USTED LO QUE ES UNA ANALÍTICA DE BIG DATA APLICADA AL SECTOR SALUD?



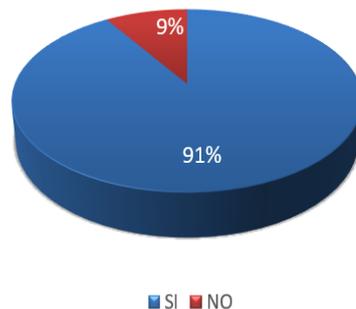
4- ¿CONSIDERA USTED QUE ES NECESARIO ADAPTAR LOS SERVICIOS DEL CENTRO DE SALUD A LAS NUEVAS TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN?



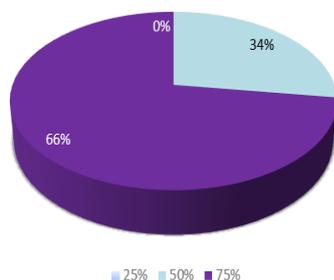
5 ¿LE GUSTARÍA TENER UN SISTEMA QUE PUEDA PREDECIR ENFERMEDADES ANTES DE QUE ESTOS SE PROPAGUEN Y DEJEN MAYORES CONSECUENCIAS?



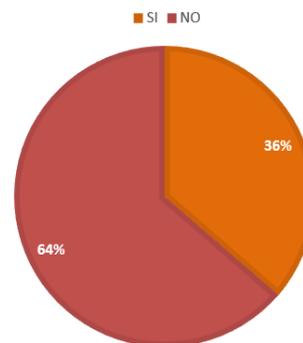
6- ¿LE GUSTARÍA QUE AL MOMENTO DE RECIBIR UN PACIENTE PUEDA OBTENER DE MANERA INMEDIATA SU HISTORIAL MÉDICO?



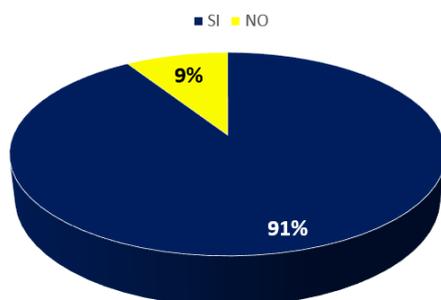
7-¿EN QUÉ MEDIDA CONSIDERA LA ADOPCIÓN DE NUEVAS TECNOLOGÍAS PARA LA PREDICCIÓN DE EVENTOS EN EL SECTOR SALUD COMO PUEDE SER LA PROPAGACIÓN DE UN VIRUS COMO EL COVID-19?



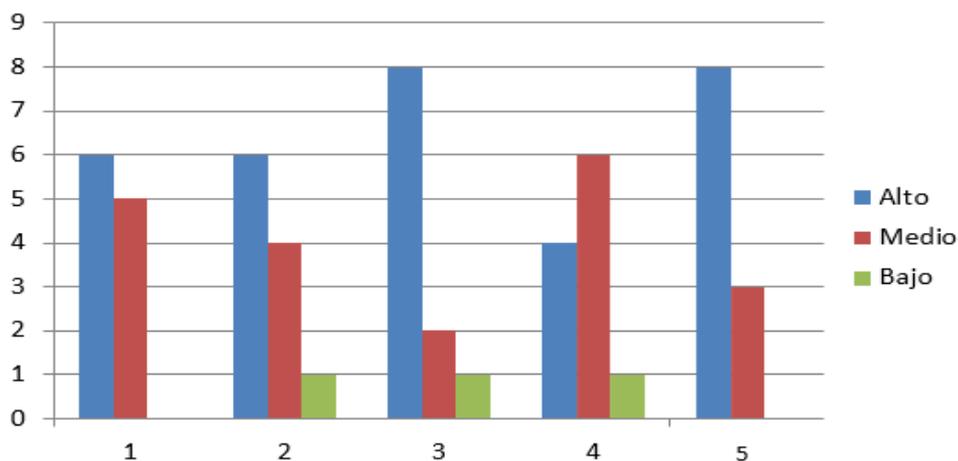
8 ¿UTILIZA MEDIOS TECNOLÓGICOS PARA TRABAJAR CON LOS PACIENTES ?



9- ¿LE GUSTARÍA UN SISTEMA QUE ARROJE ESTADÍSTICAS QUE PERMITA LA TOMA DE DECISIONES MEDICAS?



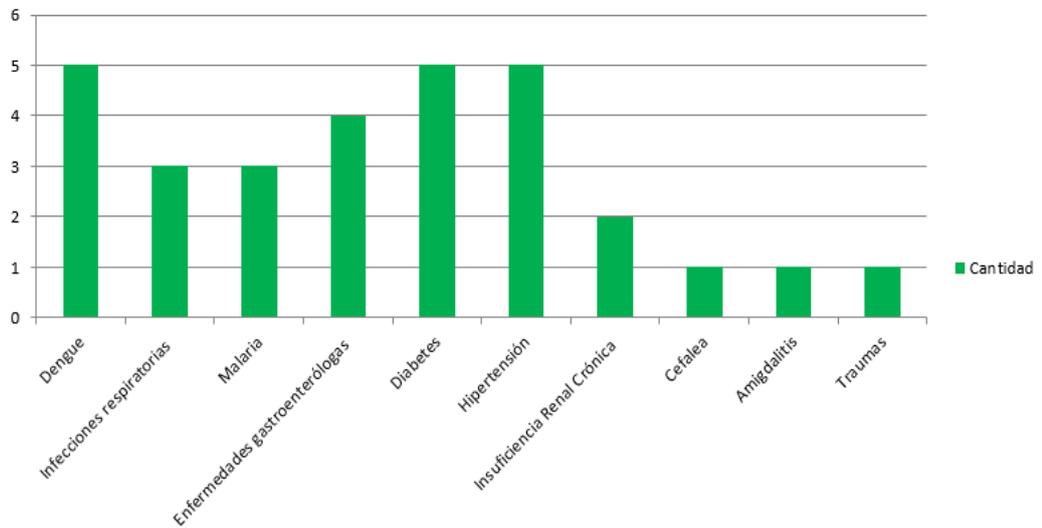
10 ¿CUÁLES DE ESTAS PREDICCIONES SERIA DE MAYOR IMPORTANCIA PARA EL DESARROLLO DE SUS FUNCIONES EN EL CENTRO DE SALUD?



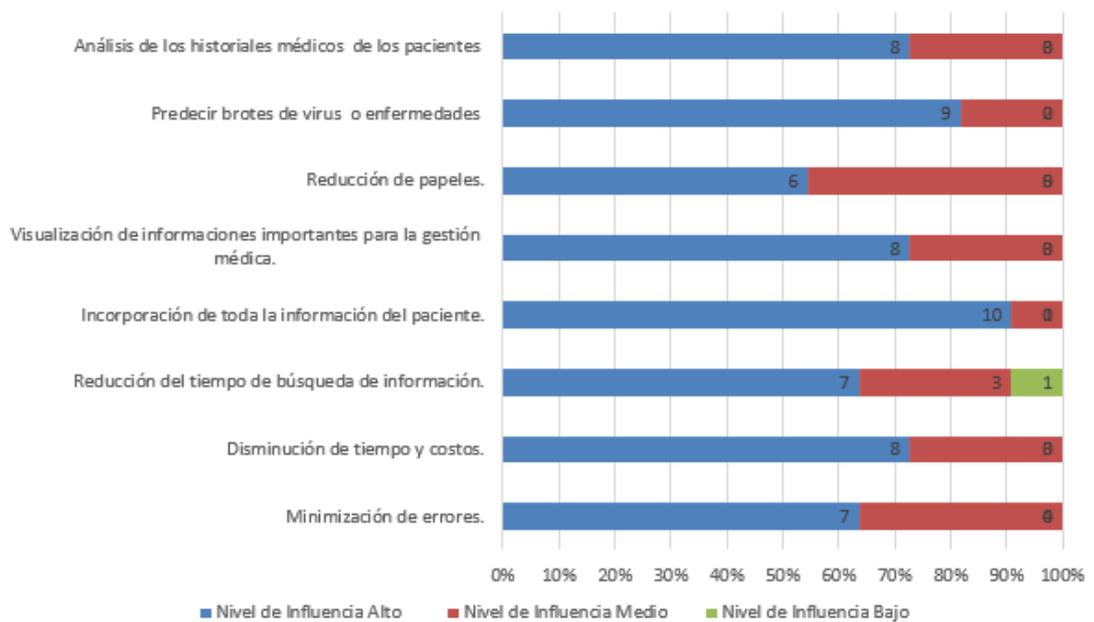
DONDE:

- 1= Predecir brotes de virus
- 2= Predecir la disponibilidad de las camillas
- 3= Predecir la cantidad de farmacéuticos a utilizar
- 4= Predecir la cantidad de pacientes
- 5= Predecir el presupuesto de centro

11- ¿CUÁLES SON LAS ENFERMEDADES MÁS FRECUENTES?



12-¿CUÁLES SERÍAN LOS PRINCIPALES BENEFICIOS DE LA PROPUESTA?



ANEXO 2- ENCUESTA PACIENTE

Estimado/a Encuestado/a, la información recolectada a partir de esta encuesta será utilizada para evaluar la adecuada mitigación de las barreras existentes que permitirá culminar con éxito la realización del trabajo de grado de UNAPEC en el que se propone el de diseño de un sistema de información para la toma de decisiones médicas y administrativas aplicando analítica de Big Data sobre el historial médico de los pacientes, razón por la cual será necesario conocer los aspectos siguientes.

1. ¿La asistencia médica ha sido amable y oportuna?

- SI
- NO

2. ¿Le gustaría que al momento que ingrese al centro pueda ser atendido con prontitud?

- SI
- NO

3. ¿Cuánto tiempo habitualmente ha esperado para recibir atención médica?

- 30 Min.
- 1 Hora
- 2 Horas
- 3 Horas

4. ¿Le gustaría que se pudiera predecir a tiempo virus como el covid-19 o enfermedades?

- SI
- NO

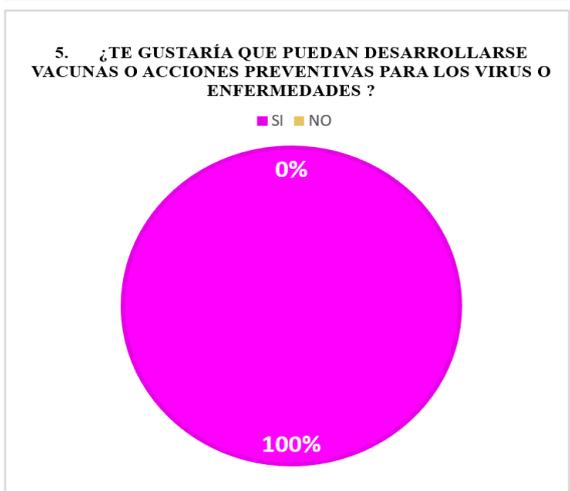
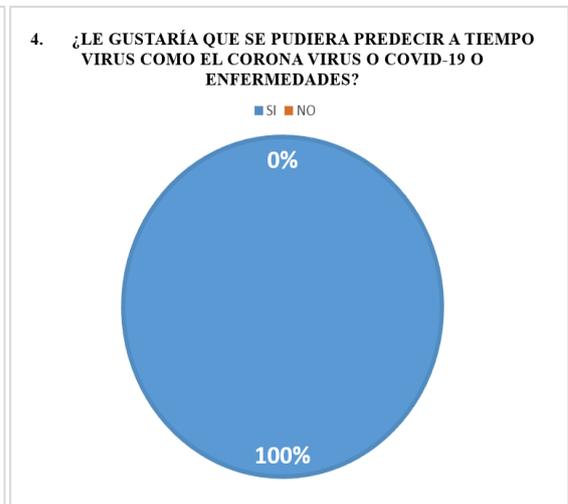
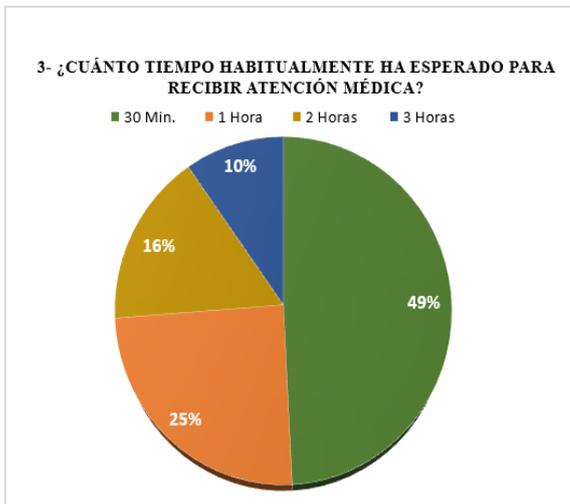
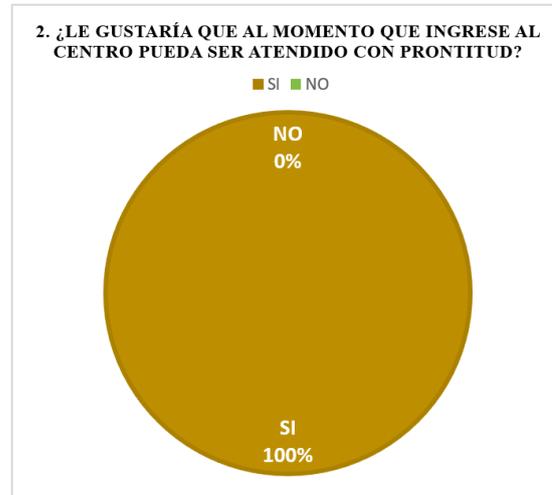
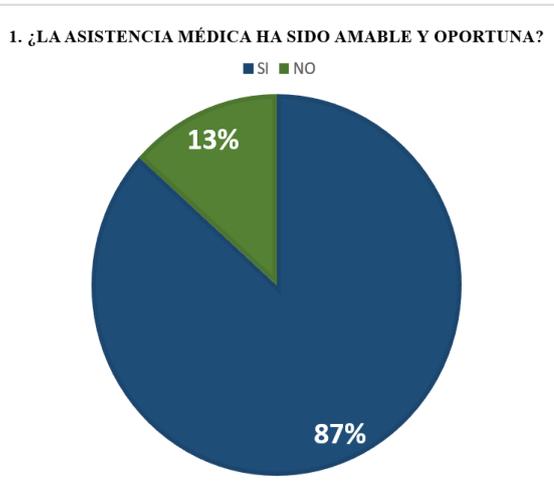
5. ¿Te gustaría que puedan desarrollarse vacunas o acciones preventivas para los virus o enfermedades?

- SI
- NO

6. ¿Le gustaría recibir un diagnóstico inmediato y veraz de su estado de salud?

- SI
- NO

ANEXO 2.1- TABULACION ENCUESTA PACIENTE



ANEXO 3. ENCUESTA PERSONAL ADMINISTRATIVO

Estimado/a Encuestado/a, la información recolectada a partir de esta encuesta será utilizada para evaluar la adecuada mitigación de las barreras existentes que permitirá culminar con éxito la realización del trabajo de grado de UNAPEC en el que se propone el diseño de un sistema de gestión de información para historiales médicos electrónicos, razón por la cual será necesario conocer los aspectos siguientes.

1. ¿Qué tiempo habitualmente se toma otorgar turnos a los pacientes que acceden al centro de salud?

- 1- 0-4 Min.
- 2- 10 Min.
- 3- 15 Min.
- 4- Más de 15 Min.

2. Al momento del ingreso ¿Qué tiempo se demora en canalizar y legalizar la admisión del paciente?

- 1- 5-10 Min.
- 2- 11-15 Min.
- 3- Más de 15 Min.

3. ¿Qué cantidad de ingresos registra diariamente?

- 1- 10 Ingresos
- 2- 20 Ingresos
- 3- Más de 20 Ingresos

4. ¿Qué tipo de proceso para el control estadístico utiliza con frecuencia?

- Manual
- Mixto
- Automatizado

5. ¿Le gustaría un sistema que le permita generar el presupuesto?

- SI
- NO

6. ¿Le gustaría un sistema que arroje estadísticas que permita la toma de decisiones administrativas?

- SI
- NO

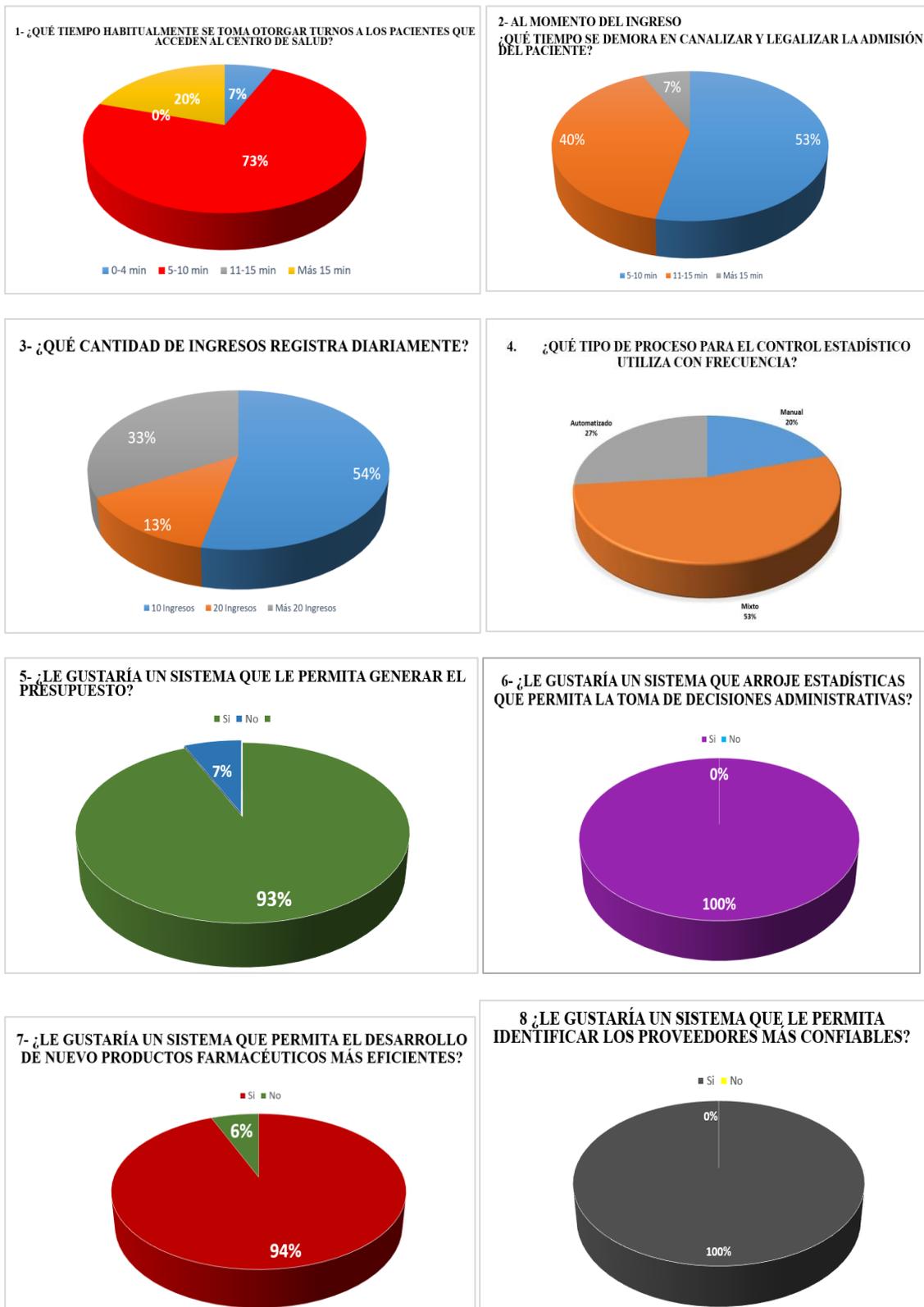
7. ¿Le gustaría un sistema que permita el desarrollo de nuevo productos farmacéuticos más eficientes?

- SI
- NO

8. ¿Le gustaría un sistema que le permita identificar los proveedores más confiables?

- SI
- NO

ANEXO 3.1- TABULACION ENCUESTA PERSONAL ADMINISTRATIVO



ANEXO 4 ENTREVISTA NO. 1

Estimado/a encuestado/a, la información recolectada a partir de esta entrevista será utilizada para la realización del trabajo de grado de UNAPEC en el que se propone el diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de Marzo en Azua, República Dominicana, aplicando analítica de Big Data sobre el historial médico de los pacientes.

Nombre del entrevistado/a: *Dr. Mercedes Rosso*

Centro de Salud: **Hospital Regional Taiwán 19 de Marzo.**

Cargo: **Directora General**

1. ¿De qué manera se manejan los historiales médicos en esta institución?

R: Se podría decir que mixto, porque la primera página de esto se hace en sistema, pero el resto, es decir, los resultados clínicos, las citas y demás.

2. ¿Cómo es el proceso de recibimiento por primera vez de un paciente?

R: El paciente debe ser referido obligatoriamente desde una UNAP

3. ¿Cómo es el proceso de recibimiento después de la primera vez?

R: La persona ya viene con una cita hecha en ocasiones y si es por emergencia se consulta en el sistema y luego se envía a buscar el historial a los archivos.

4. ¿Qué ventajas le ofrece la forma en que se maneja en que gestiona los historiales médicos en esta institución?

R: No existe una ventaja como tal, porque el proceso es manual, pero al menos podemos tener datos básicos de los pacientes.

5. ¿Ha manejado de otra manera los historiales médicos en esta u otra institución?

R: No.

6. ¿Qué le gustaría que sea diferente en la manera en que se manejan actualmente los historiales médicos en el centro?

R: Si

7. ¿Qué piensa de la creación de una plataforma que permita intercambiar la información generada en el centro médico con otros?

R: Seria muy novedoso porque mejoraría la gestión que realizamos aquí.

8. ¿De qué manera se organizan los historiales médicos?

R: Se organizan en gabinetes por numeración

9. ¿Cuál es el proceso para verificar y/o modificar los historiales médicos?

R: *Al momento de la cita, el doctor adjunta al folder toda la información*

10. ¿Cuál es el procedimiento para buscar un historial médico?

R: *Todas las semanas se listan las citas que se realizaran la semana siguiente y se le entrega a la encargada de archiveros quien junto a su equipo los buscan en los gabinetes y luego entregan a los doctores un día antes.*

11. ¿Cuánto tiempo toma la creación de un nuevo historial médico?

R: *Todo depende, porque al ser manual casi el 99% del proceso se puede demorar.*

12. ¿Puede un paciente solicitar su historial médico?

R: *No realmente, porque estos se encuentran en los archiveros.*

13. ¿Ese sistema analiza los historiales médicos?

R: *No.*

14. ¿Ese sistema me realiza algún tipo de análisis, el cual me pueda informar sobre alguna anomalía, como inicios de un brote de un virus o una enfermedad?

R: *No, solo se reportan los casos al Centro de Epidemiología de la República Dominicana y ellos emiten un comunicado.*

15. ¿Le gustaría tener un sistema que predijera el inicio de un brote de un virus o enfermedad?

R: *Si.*

ANEXO 5 ENTREVISTA NO. 2

Estimado/a encuestado/a, la información recolectada a partir de esta entrevista será utilizada para la realización del trabajo de grado de UNAPEC en el que se propone el diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de Marzo en Azua, República Dominicana, aplicando analítica de Big Data sobre el historial médico de los pacientes.

Nombre del entrevistado/a: Paoli Vázquez
Centro de Salud: *Hospital Regional Taiwán 19 de Marzo*
Cargo: Encargado de Informática

1. ¿Tienen un sistema informático?

R: *Si, desde el 2013 se viene desarrollando el Sistema de Gestión Hospitalario (SISGEHOSPI)*

2. ¿Tienen digitalizado los historiales pacientes?

R: *No, todo se realiza manualmente, es decir, papel y lápiz salvo la página principal de estos que se genera desde el sistema y se guarda en físico en los archiveros.*

3. ¿El ciclo del sistema, dígame se introducen los datos y que procede, ¿se quedan en las bases de datos y ya?

R: *Usamos SQL Server solo para consultas. Cuando el paciente llega al centro desde una UNAP (Unidad de Atención Primaria) se crea un registro con los datos básicos del paciente y esto se realiza y se guarda un historial, pero solo de las visitas que ha realizado al centro.*

4. ¿Tecnologías que usan?

R: *Para el desarrollo usamos Visual Studio y como lenguaje C Sharp, la aplicación es a nivel desktop y las computadoras tienen Windows desde XP hasta la versión 10. Se usa SQL server.*

5. ¿Cuáles mecanismos de seguridad utilizan?

R: *Lo primero que tenemos es la definición de roles, así como un usuario y contraseña y a los 5 minutos de inactividad se cierra la sesión.*

6. ¿Estructura del sistema?

R: *En la actualidad el SISGEHOSPI está compuesto por 3 módulos y 14 submódulos:*

Módulo de Gestión

Este módulo de gestión está encargado de la administración de los pacientes y de los servicios que son brindados en la institución. Está conformado por los siguientes submódulos de los cuales solo los dos primeros son los únicos que están en funcionamiento: pacientes, usuarios, expedientes, tipos de servicios, servicios médicos y citas.

Módulo de Admisiones/prestación de servicios

Este módulo de admisiones/ prestación de servicios, está encargado de los procesos que están involucrados con las ARS. Asimismo, genera reportes relacionados a las productividades facturadas, médicas, servicios y reporte de ARS. En este último se listan las personas que poseen seguros de salud y que han recibido atención en el centro. Está conformado por los siguientes submódulos prestación de servicios, consulta y reportes.

Módulo de Gestión administrativa

Este módulo de gestión administrativa está encargado de los procesos administrativos como lo dice su nombre, se realizan los inventarios de la cocina exclusivamente, registro de usuarios y sus roles. Está conformado por los siguientes submódulos: informática, administrativo financiero, cocina y RRHH.

7. ¿El sistema genera algún tipo de reporte?

R: *Se existe un módulo de reporte, y me genera los siguientes reportes: productividades facturadas, productividad médica, servicios y reporte de ARS (me dice las personas que usan esa ARS que estuvieron en el lugar)*

8. ¿Cuántas versiones ha sufrido?

R: *13 versiones*

9. ¿Documentan el sistema y sus procesos?

R: *no*

10. ¿Utilizan backup?

R: *Se hace automático backup diario, a las 6 de la tarde a todos son servidores locales que luego se pasan a disco duro mensual. Disco de 2 TERABYTE.*

ANEXO 6 ENTREVISTA NO. 2 ADMINISTRATIVA

Estimado/a encuestado/a, la información recolectada a partir de esta entrevista será utilizada para la realización del trabajo de grado de UNAPEC en el que se propone el diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de Marzo en Azua, República Dominicana, aplicando analítica de Big Data sobre el historial médico de los pacientes.

Nombre Nelson Gonzales
Centro de Salud: *Hospital Regional Taiwán 19 de Marzo*
Cargo Administrador

1. ¿Como definen el presupuesto del hospital?

R: *El Hospital Regional Taiwán tiene dos fuentes de financiamiento los fondos operativos que no es más que el presupuesto que recibe por parte del gobierno mensual 4,356,060 pesos durante 10 meses en el año y el 40% va la dirección de promesa que suministra los medicamentos y la otra fuente es la venta de servicios eso no es más que los ingresos que recibe el hospital e manera interna a través de las aseguradas 7 ARS que son: 2,200,000 millones mensual a doce meses.*

2. ¿Qué hacen cuando el presupuesto no les alcanza?

R: *Se recurre al sector salud*

3. ¿Existen un sistema para generar el presupuesto, ustedes introduzcan los datos y este se genere?

R: *Todo está manual*

4. ¿Cómo se realice el proceso de orden de compra?

R: *La liquidación u orden de compra cada 6 días se envía al servicio nacional de salud y se realiza así:*

- 1. Se hace una requisición de productos*
- 2. El encargado de compra cotización*
- 3. Luego se hace una orden de compra*
- 4. Se va al proveedor*
- 5. Luego se decepciona en el almacenamiento*

5. ¿Cuáles son los principales inconvenientes al momento de definir el presupuesto?

R: *Al no tener un sistema informático y que todo se realice manual no hay manera de validar si el proveedor está al día en impuesto internos.*

6. ¿Cómo llevan un control de los medicamentos y materiales?

R: *El departamento de activo fijo que lleva un control, se rotula todo y se lleva un control manual*

7. ¿Le gustaría un sistema que pueda predecir el presupuesto?

R: *Si*

8. ¿Le gustaría un sistema que pueda decirle la disponibilidad de recursos?

R: *Por supuesto, eso agilizaría la gestión que se realiza aquí en el hospital.*



DECANATO DE INGENIERÍAS E INFORMÁTICA

ESCUELA DE INFORMÁTICA

Anteproyecto de Trabajo de Grado

Título:

Propuesta de diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de marzo en Azua, República Dominicana, aplicando analítica de Big Data sobre el historial médico de los pacientes.

Sustentado por:

Manuel Ariel Galva Fernández 2012-1458

María Jose Gutierrez Perez 2016-1918

Wilkin A. Santana Gonzales 2014-0320

Asesor:

Eddy Alcantara

Distrito Nacional, República Dominicana

2019.

Contenido

1. TÍTULO DEL TEMA.....	xl
2. INTRODUCCIÓN.....	xli
3. JUSTIFICACIÓN.....	xliii
4. DELIMITACIÓN DEL TEMA Y PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN.....	xliv
4.1 Delimitación del Tema.	xliv
4.2 Planteamiento del problema de investigación.....	xliv
5. OBJETIVOS GENERALES Y ESPECÍFICOS.....	xlvi
5.1 Objetivo General.....	xlvi
5.2 Objetivos Específicos.....	xlvi
6. MARCO TEÓRICO REFERENCIAL.....	xlvi
6.1 Marco Teórico.....	xlvi
6.2 Marco conceptual.....	lii
7. DISEÑO METODOLÓGICO.....	lv
7.1 Métodos de Investigación.....	lvi
7.2 Técnicas de investigación.....	lvi
7.2.1 Técnica para recolección de la información.....	lvi
8. FUENTES DE DOCUMENTACIÓN.....	lviii
9. ESQUEMA PRELIMINAR DE CONTENIDO DEL TRABAJO DE GRADO.....	lx

1. TÍTULO DEL TEMA.

Propuesta de diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de marzo en Azua, República Dominicana, aplicando analítica de Big Data sobre el historial médico de los pacientes.

2. INTRODUCCIÓN.

En la presente investigación se presenta el anteproyecto de investigación de la propuesta de diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de marzo en Azua, República Dominicana, aplicando analítica de Big Data sobre el historial médico de los pacientes. El mismo contiene la justificación del tema investigado, la delimitación espaciotemporal y el planteamiento del problema de esta.

De igual manera, contiene el objetivo general que se pretende alcanzar en el proceso de investigación. Asimismo, el conjunto de objetivos específicos que complementan al general, abarcando de forma individual los diversos subtemas que sustentan a este.

De igual manera, se describe el marco referencial que es el conjunto de autores que servirán de base a la presente investigación y se definen los conceptos fundamentales de la investigación.

Por otra parte, presentamos el diseño metodológica donde se detallan los diferentes tipos de investigaciones a utilizar, los tipos de métodos y las técnicas que se emplearan para la recolección de la información. Además, se nombrarán las fuentes que servirán de apoyo a la elaboración de este anteproyecto, dividiéndolas en primarias y secundarias.

Por último, se presenta el esquema preliminar del contenido del trabajo de grado. El mismo está compuesto por cinco capítulos, el primero consiste en analizar teóricamente sistemas de información aplicados a la toma decisiones medicas mediante la implementación de Big Data y el uso de la analítica de datos. Seguido del capítulo dos en el cual se analiza teóricamente los Historiales Médicos Electrónicos y su impacto en la gestión de información.

En el capítulo tres se diagnostica el sistema de gestión del historial médico de los pacientes del Hospital Regional Taiwán 19 de marzo de Azua. Mientras que el capítulo 4 se enfoca en el análisis y reingeniería del sistema de gestión de información actual. Para finalizar el capítulo 5 trata de elaborar una propuesta

de diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de marzo de Azua, R.D. aplicando analítica de Big Data sobre el historial médico de los pacientes.

3. JUSTIFICACIÓN.

La presente investigación tiene como objetivo principal elaborar una propuesta de diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de marzo en Azua, R.D. aplicando analítica de Big Data sobre el historial médico de los pacientes.

El Hospital Regional Taiwán 19 de marzo de Azua es el principal centro de salud público de la zona sur de la República Dominicana, el cual a pesar de ser un centro que no tiene tantos años de funcionamiento como otros del país y haber surgido como fruto de las relaciones diplomáticas que hasta el año pasado mantuvo la República Dominicana con Taiwán, este no ha logrado cumplir a cabalidad la misión para la que fue concebido.

De tal modo, que se ha visto reducir progresivamente su cartera de servicios y a la fecha no ha podido poner en funcionamiento áreas tan vitales como la Unidad de Cuidados Intensivos ni el Banco de Sangre, los cuales están pendientes de equipamiento por el Servicio Nacional de Salud (SNS). Tampoco ha logrado mantener en funcionamiento el sistema de climatización constante, al punto que las áreas de internamiento no tienen aire acondicionado, como tampoco las salas de pediatría y ginecoobstetricia.

Como resultado de las situaciones planteadas, todos los pacientes con traumas craneoencefálicos deben ser trasladados a los hospitales del Gran Santo Domingo, tal y como están procediendo el resto de los hospitales de la región sur.

Por tanto, consideramos que con el diseño de este sistema de información podrán tomarse decisiones tanto medicas como administrativas, que mejorarían considerablemente la situación en la que se encuentre sumergida esta institución y por ende enfrentar los desafíos de salud por los que atraviesa la región sur del país. De igual manera, este sistema permitirá cuantificar los recursos económicos referentes a las áreas operativas y fármacos, es decir, en base a esta información preparar un presupuesto que pueda mantener el funcionamiento esperado.

Por otra parte, este sistema de información ofrecerá proyecciones de las enfermedades que frecuentan en la región. De igual manera, las temporadas en las cuales éstas se manifiestan y los lugares de mayor propagación y áreas de brote.

En este mismo sentido, cabe destacar que los usuarios serán beneficiados con un sistema de calidad y que garantice que los servicios ofrecidos por este centro de salud estén de acuerdo con todos los estándares de salud.

4. DELIMITACIÓN DEL TEMA Y PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN.

4.1 Delimitación del Tema.

El diseño del sistema propuesto se estará realizando en el periodo enero-abril del 2020 en el Hospital Regional Taiwán 19 de marzo localizado en la provincia de Azua, República Dominicana.

4.2 Planteamiento del problema de investigación.

EL Hospital Regional Taiwán 19 de marzo inaugurado en el 2005, con los mejores equipamientos con tecnología de punta, incluidos un tomógrafo y un mamógrafo. Este fue construido con el propósito de ser el centro principal de salud de la región sur de la Republica Dominicana. Lamentablemente este no ha cumplido su objetivo.

Desde el año 2009 Taiwán localizado en la provincia de Azua de nuestro país, ha presentado un sin número de inconvenientes tanto con los recursos médicos como los recursos económicos, lo que implica que no pueden ofrecer servicios de calidad a sus pacientes. En este centro, los doctores deben de completar los formularios de los pacientes con papel y lápiz, lo cual representa una situación delicada porque podría ser que estos sufran algún accidente, dígase que se moje el papel o en caso de haber un incendio.

Por otra parte, dicha situación afecta el proceso de atención al paciente, ya que no se conocen los datos actuales de los pacientes, es decir, el historial médico de manera rápida. Se empleado mucho tiempo en poder recibir a los pacientes y procesar con su evaluación. Además, destacamos que carecen de un sistema de información que les permita tomar decisiones en base a los registros médicos de los pacientes.

Otro de los inconvenientes que afecta a esta institución es el bajo presupuesto que recibe por parte del Gobierno Dominicano, el cual según María Teresa Morel (2019) es de 3.4 millones que se reduce a la mitad de los descuentos en medicamentos e insumos y casi RD\$ 3 millones por la venta de servicios al SENASA subsidiado para un centro que en este primer trimestre atendió 15,931 consultas; 695 partos y cesáreas y 7,586 emergencias, sin contar con las cirugías.

Por tanto, con un sistema que me permita conocer informaciones como cantidad de usuarios recibidos, enfermedades, brotes, virus recursos fármacos demandas y otros datos, se podrá preparar el presupuesto que satisfaga las necesidades que existen.

Por todo lo anteriormente planteado, surge la siguiente pregunta en la investigación:

1.- ¿Cuáles son las características y componentes de un sistema de información para digitalizar los historiales médicos de los pacientes que permita tomar decisiones medicas y administrativas aplicando analítica de big data en el hospital regional Taiwán 19 de marzo en azua, R.D.?

Para resolver esta pregunta se requiere responder las siguientes interrogantes:

- a.- ¿Cuál es la situación actual del sistema de gestión de información utilizado para la toma de decisiones en el hospital ?
- b. ¿ Cuáles son las herramientas y técnicas de Big Data usando analítica de datos propicias para el diseño de un sistema de información que automatice los historiales médicos de los pacientes para la toma de decisiones ?
- c. ¿ Que tipo de sistema de gestión de información es necesario para automatizar los historiales médicos de los pacientes y tomar decisiones?
- d. ¿ Es factible desarrollar un sistema de información para la toma de decisiones en el hospital regional Taiwán19 de marzo ?

5. OBJETIVOS GENERALES Y ESPECÍFICOS.

5.1 Objetivo General.

Elaborar una propuesta de diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de marzo de Azua, R.D. aplicando analítica de Big Data sobre el historial médico de los pacientes.

5.2 Objetivos Específicos.

- Analizar teóricamente sistemas de información aplicados a la toma de decisiones médicas mediante la implementación de Big Data y el uso de la analítica de datos.
- Analizar teóricamente los Historiales Médicos Electrónicos y su impacto en la gestión de información.
- Diagnosticar el sistema de gestión del historial médico de los pacientes del Hospital Regional Taiwán 19 de marzo de Azua.
- Análisis y reingeniería del sistema de gestión de información actual.
- Elaborar un diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de marzo de Azua, R.D. aplicando analítica de Big Data sobre el historial médico de los pacientes.

6. MARCO TEÓRICO REFERENCIAL.

61. Marco Teórico.

Big data (grandes volúmenes de datos) se refiere al conjunto de datos e información tan grandes y complejos que hace muy difícil su procesamiento utilizando herramientas de gestión de bases de datos convencionales. La cuestión es cómo acceder, distribuir y utilizar esta vasta cantidad de datos “no estructurados”. Los pacientes, las clínicas, los hospitales tienen cantidades masivas de datos clínicos, en formatos escritos en papel o electrónicos pero que permanecen sin utilizar por la dificultad e imposibilidad material de “digerirlos” de forma efectiva, por muy buenos deseos que pueda tener el equipo sanitario. Lo cierto es que esta dificultad puede tener consecuencias tanto en el control de los gastos médicos como en la mejora de las tasas de mortalidad. (Riskin D., 2012)

Este es el futuro de la salud, según ha publicado la prestigiosa revista Forbes: los Big Data representan una oportunidad para los innovadores y todos los que se preocupan por la salud, aumentan substancialmente la posibilidad de obtener información más efectiva de los datos y menores tasas de mortalidad de los pacientes. (Riskin D., 2012)

El Hospital la Fe en Valencia España ha puesto en marcha la plataforma Big Data la cual está permitiendo impulsar una gestión clínica más moderna, centrada en el paciente a través de la creación de unidades de gestión clínica y la medición de los resultados clínicos que permitan medir el valor de nuestras intervenciones. Lo cual se ha facilitado con la implementación de la Historia Clínica Electrónica (HCE), ya que el volumen de información generada se ha incrementado enormemente. Asimismo, destacan que el principal objetivo es desarrollar herramientas de apoyo a la toma de decisiones clínicas para facilitar la transformación del modelo asistencial actual y orientarlo hacia un modelo más participativo. (Hospital La Fe, 2018)

La plataforma ofrece la posibilidad de organizar la información de manera efectiva, integrando datos estructurados y no estructurados, junto con aquellos que provienen del uso habitual de las tecnologías de la información y la comunicación (TIC) por parte de los pacientes, profesionales y población en general. Además, permite analizar y visualizar los datos, a la vez que generar modelos predictivos y sistemas de ayuda a la decisión. (Hospital La Fe (2018))

Según el doctor Bernardo Valdivieso, director del Área de Planificación del Hospital la Fe y director del Área de Atención Domiciliaria y Telemedicina del Departament de Salut La Fe, responsable de la nueva Plataforma “todo esto nos permite generar más y mejor información para caracterizar de manera más eficiente tanto en las enfermedades como la provisión de servicios en salud lo que nos permite conocer el “valor de nuestras intervenciones” – resultados clínicos- e impulsar la medicina de precisión”. (Valdivieso B.,2018)

La incorporación de las tecnologías de la información y las comunicaciones por parte de los Servicios de Salud de las Comunidades Autónomas ha dotado en los últimos años a usuarios y profesionales de sistemas y aplicaciones que han facilitado a ambos colectivos el acceso a una información de calidad sobre la salud individual, al servicio de una atención sanitaria de calidad creciente. Hoy, todas las Comunidades Autónomas sin excepción tienen sistemas de Historia Clínica (o de Salud) Electrónica (o Digital) (HCE), en fase de implantación casi completa en Atención Primaria. Este nivel de implantación es sensiblemente inferior en Atención Especializada, aunque no está por debajo del que corresponde a otros países desarrollados.¹ Estos sistemas incorporan magníficas funcionalidades que aportan efectividad, eficiencia y seguridad al proceso de atención, pero no están concebidos para ofrecer información cuando el paciente debe ser atendido fuera del ámbito geográfico donde su información se ha generado. (SNS España)

En un aspecto económico y de tiempo, los EHRs a diferencia de otros métodos de investigación retrospectivos no requiere del reclutamiento del paciente o la colección de datos, procesos que son realmente costosos y que consumen mucho tiempo. Se estima que el uso de EHRs se puede disminuir el costo de las investigaciones, incrementar las investigaciones centradas en pacientes únicos, y puede acelerar en gran medida la frecuencia de descubrimientos clínicos. (Weiskopf y Weng, 2013)

Desde la introducción del concepto Big Data se han asociado una serie de características clave que lo definen, llamadas las «V» de Big Data. Inicialmente, estas características eran tres: volumen, velocidad y variedad. Posteriormente se han añadido otras características, como la veracidad de los datos, añadiendo así cuatro dimensiones y marcando las características que tradicionalmente se están asociando a Big Data, y otra quinta dimensión, el valor de los datos, que se ha añadido recientemente, generando así las denominadas cinco V del Big Data. Esta definición es válida en todos los sectores en los que se dé la confluencia de esta serie de características en mayor o menor medida, y así ocurre en el dominio de la salud, por lo que estos términos también son relevantes en el sector. (E. Menasalvas / C. Gonzalo / A. Rodríguez-González, 2014)

Se estima que los ahorros potenciales y los costos de la adopción generalizada de los sistemas de registros médicos electrónicos (EMR), modelos de importantes beneficios de salud y seguridad, y concluye que la aplicación y creación de redes eficaces EMR podrían finalmente ahorrar más de \$ 81 mil millones al año, mediante la mejora de la eficiencia de la asistencia sanitaria y Seguridad- y que la prevención y la gestión de HIT habilitado de enfermedades crónicas podrían finalmente duplicar esos ahorros al tiempo que aumenta la salud y otros beneficios sociales. (Hillestadetal., 2005)

Como consecuencia del almacenamiento de estos datos, los hospitales cuentan con una ingente cantidad de información no estructurada en forma textual o visual, que puede ser explotada con tecnologías de Big Data. Y es que Big Data representa el paradigma perfecto para el desarrollo de lo que se conoce como medicina basada en la evidencia. La Medicina Basada en la Evidencia es una expresión que se ha generalizado en castellano como equivalente a la expresión inglesa Evidence-Based Medicine (EBM). (Menasalvas E., Gonzalo C., Rodríguez A., 2013).

Los retos a los que nos enfrentamos y que hemos planteado no pueden demorar más una respuesta ordenada en el Sistema Nacional de Salud que potencie los efectos beneficiosos de la aplicación de *big data* en sanidad y en biomedicina en

España, reduciendo riesgos como pueden ser la pérdida de economías de escala en las inversiones tecnológicas requeridas o la dificultad de escalabilidad y de explotación unificada si las iniciativas no están alineadas. La tecnología está disponible y la industria que la ofrece muestra una agresividad comercial muy alta para introducirla (a cualquier precio), influenciada por la travesía del desierto en la que se encuentra desde que se iniciaron la crisis económica y los consiguientes dramáticos recortes en la inversión en tecnologías de la información y la comunicación. Estamos frente a una oportunidad histórica para aunar voluntades, políticas y tecnologías en una estrategia nacional. En este sentido, es procedente desarrollar una estrategia inicial en el ámbito de la investigación biomédica, donde los retos son tremendos pero los posibles beneficios de una explotación masiva de la información digital disponible en el Sistema Nacional de Salud son evidentes, alineando esfuerzos de las comunidades autónomas. Destaca como referencia en este país y primera experiencia importante de *big data* sobre información de pacientes el proyecto *Visc+* de Cataluña, que está comenzando su recorrido para uso científico. (Parra C., 2016)

Ministro de salud de Chile: Se trata de un nuevo modelo de atención en salud, centrado en el paciente, y que aprovecha el potencial de las tecnologías para acercar la atención a las personas, instalando una alternativa al modelo tradicional. Dicho proyecto, es la estrategia principal del Ministerio de Salud para informatizar el proceso asistencial. Esta innovación fue presentada través de reuniones convocadas por el Ministerio de Salud Emilio Santelices a expertos del área y fue lanzado oficialmente en el Simposio Salud País Digital el 14 de junio de 2018. (Santelices E., 2019)

Este sistema se subdivide en las áreas de pacientes, Profesionales de la salud y Para el área de inteligencia sanitaria para clasificar las funciones que ofrece, tales como, (videconferencia, chat, intercambio de archivos, lista de espera, registro inmunizaciones, receta electrónica, imágenes, agendamiento, información de farmacias, datos de contacto con el hospital digital...) Intra web del Hospital Digital, Registro Clínico Electrónico del Hospital Digital , Teleproceso (solución de telemedicina para consulta asincrónica), Visualizador de la Historia

Clínica Compartida (HCC), Datawarehouse para análisis predictivo (Big Data), Historia Clínica Compartida. (Santelices E. ,2019)

Una tercera área donde existe un claro potencial significativo para las grandes tecnologías de Big Data, desde un prisma más global, es la de gestión de enfermedades infecciosas. Las epidemias se extienden a través de personas y de los contactos persona-persona o por el entorno, y ello puede suponer un gran peligro para la sociedad, sobre todo en los países en vías de desarrollo. Sin embargo, debido a la gran movilidad de personas que se produce actualmente, esta idea ha de extenderse a los países con sistemas de salud más evolucionados. (Martínez M. ,2015)

Se pueden obtener datos sobre salud de historias clínicas electrónicas, dispositivos de telemedicina, pruebas clínicas, e incluso de wearables. Asimismo, aportan un valor añadido los datos epidemiológicos, los nutricionales y los genómicos, más relacionados con lo que se conoce como Real World Data (RWD) y con la medicina personalizada. Analizar esa información puede ayudar a tomar decisiones tanto a los médicos como a los gestores de los centros sanitarios, lo que repercute en un mejor servicio de salud para los pacientes. (Instituto de Ingeniería de Conocimiento, 2019)

Cada vez más, los médicos y otros profesionales de la salud se unen a expertos de otras disciplinas, especialmente tecnológicas, para mejorar su trabajo. El Big Data en salud posee gran volumen de información relacionada con los pacientes, las enfermedades y los centros sanitarios que, bien analizada con la tecnología adecuada, resulta de gran utilidad para atender con más rapidez y precisión a futuros pacientes. Un caso de éxito en este terreno fue el trabajo conjunto realizado por el Hospital Universitario Son Llàtzer y el Instituto de Ingeniería del Conocimiento (IIC) para desarrollar un sistema de detección precoz de sepsis, en el marco del proyecto. (Instituto de Ingeniería de Conocimiento, 2019)

6.2 Marco conceptual.

Las Tecnologías de Información y comunicación (TIC): son aquellos dispositivos, herramientas, equipos y componentes electrónicos, capaces de manipular información que soportan el desarrollo y crecimiento económico de cualquier organización. Cabe destacar que en ambientes tan complejos como los que deben enfrentar hoy en día las organizaciones, sólo aquellos que utilicen todos los medios a su alcance, y aprendan a aprovechar las oportunidades del mercado visualizando siempre las amenazas, podrán lograr el objetivo de ser exitosas. (Thompson y Strickland, 2004)

Tecnología de la Información (TI): se conoce como tecnología de información (TI) a la utilización de tecnología específicamente computadoras y ordenadores electrónicos para el manejo y procesamiento de información específicamente la captura, transformación, almacenamiento, protección, y recuperación de datos e información.

Sistema de información: es el conjunto de componentes que interaccionan entre sí con el propósito de alcanzar un objetivo determinado, el cual debe satisfacer las necesidades de información. Estos componentes pueden ser actividades, recursos materiales, personas o datos, que deben además procesar la información y distribuirla de la manera más correcta para satisfacer necesidades.

Ciencia de los datos: es el estudio de dónde proviene la información, qué representa y cómo se puede convertir en un recurso valioso para la creación de estrategias empresariales y de TI. La extracción de grandes cantidades de datos estructurados y no estructurados para identificar patrones puede ayudar a una organización a controlar los costos, aumentar la eficiencia, reconocer nuevas oportunidades de mercado y aumentar la ventaja competitiva de la organización. (Rouse M. ,2018)

Business Intelligence: es un proceso impulsado por la tecnología para analizar datos y presentar información procesable que ayuda a los ejecutivos, gerentes y otros usuarios finales corporativos a tomar decisiones comerciales informadas.

Base de datos: Es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.

Data mart: es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento.

Data Mining: es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

Data warehouses: es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta.

Big Data: es un término que describe un gran volumen de datos estructurados y no estructurados que inundan una empresa en el día a día. Pero no es la cantidad de datos lo que importa sino lo que las organizaciones hacen con los datos es lo que importa. Los grandes datos se pueden analizar para obtener información que conduzca a mejores decisiones y estrategias.

Petabytes: es una unidad de medida de memoria (2^50) que es igual a 1.024 Terabytes (en realidad 1.125.899.906.842.624 bytes). Se utiliza sobre todo en soluciones distribuidas de almacenaje y dentro de soluciones de empresa importantes.

Hadoop: es un marco de procesamiento distribuido de código abierto que gestiona el procesamiento y el almacenamiento de datos para aplicaciones de big data en grupos escalables de servidores informáticos.

Hadoop Distributed File System (HDFS): es un sistema de archivos distribuido diseñado para ejecutarse en hardware básico. Tiene muchas similitudes con los sistemas de archivos distribuidos existentes, Sin embargo, las diferencias con otros sistemas de archivos distribuidos son significativas.

Graphical User Interface (GUI): Es un sistema de componentes interactivos como iconos y otros objetos gráficos que ayudan al usuario a interactuar con el software de la computadora, como un sistema operativo.

Datos estructurados: información que se suele encontrar en la mayoría de las bases de datos son archivos de tipo texto que se suelen mostrar en filas y columnas con títulos o datos que pueden ser ordenados y procesados fácilmente por todas las herramientas de minería de datos.

Datos semiestructurados: son una mezcla de los datos estructurados y los no estructurados, no presenta una estructura perfectamente definida como los datos estructurados, pero si presentan una organización definida en sus metadatos donde describen los objetos y sus relaciones, y que en algunos casos están aceptados por convención.

Datos no estructurados: son datos binarios que no tienen estructura interna identificable. Es un conglomerado masivo y desorganizado de varios objetos que no tienen valor hasta que se identifican y almacenan de manera organizada.

Historial médico electrónico: es una recopilación de manera digital de la información acerca de un paciente que se puede utilizar para comprender mejor cualquier problema de salud que pueda sufrir.

Historial médico: es una recopilación de información acerca de un paciente que se puede utilizar para comprender mejor cualquier problema de salud que pueda sufrir.

7. DISEÑO METODOLÓGICO.

7.1 Métodos de Investigación.

El método de investigación que implementaremos será el analítico y deductivo, el primero porque este nos permitirá realizar un análisis profundo de la problemática que concierne. De tal manera, que se puedan identificar las posibles causas que generan que el Hospital Taiwán no funcione como se espera. Por su parte, el método deductivo nos permitirá llevar la situación desde lo general hasta lo particular.

7.2 Técnicas de investigación.

En nuestro caso los tipos de investigaciones que realizaremos serán exploratoria, descriptiva y por observación. Por la razón que, el tema que estamos trabajando ha sido tratado muy poco desde el punto de vista científico en nuestro país. Por tanto, comprendemos que llevar a cabo una investigación exploratoria nos garantizara poder indagar lo suficiente y alcanzar el éxito esperado en nuestro documento.

Además, se empleará la observación porque esta nos permitirá recolectar directamente los datos, aplicando las técnicas adecuadas y sin la manipulación de las posibles variables. Mientras que por su parte la investigación descriptiva nos ofrecerá una visión más amplia y detallada de nuestro objeto de estudio.

7.2.1 Técnica para recolección de la información

Las técnicas de recolección que aplicaremos serán encuestas y entrevista, las cuales se podrán realizar de manera presencial, comprendiendo que los personas que frecuentan el centro de salud son de escasos recursos y la tecnología no sería un medio viable. De igual manera, las entrevistas estarán enfocadas a un personal más capacitado en el tema.

8. FUENTES DE DOCUMENTACIÓN.

- Arcangel, J. L. (1 de febrero de 2019). La Fe es uno de los pocos hospitales españoles con un sistema de Big Data. *BigDataMagazine*. Obtenido de <https://bigdatamagazine.es/big-data-al-servicio-de-la-salud-de-los-pacientes-de-la-fe>
- Barbosa, J. G. (2016). *Big Data el poder de convertir datos en decisiones*. Madrid : Aunclidelastic .
- Desarrollo, B. I. (2017). *El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe*.
- Fe, L. (2018). *La Fe Departament de Salut* . Obtenido de http://www.lafe.san.gva.es/home/-/asset_publisher/Gv2P/content/la-fe-pone-en-marcha-una-plataforma-de-big-data-en-salud-para-hacer-mas-eficiente-la-atencion-al-paciente/maximized
- FUNDACIÓN VODAFONE ESPAÑA. (2016). *Big Data en salud digital*. ESPAÑA: Red.es.
- Instituto de Información Sanitaria. (s.f.). *El sistema de historia clinica digital* .
- Instituto de Ingenieria del conocimiento . (22 de octubre de 2019). *Instituto de Ingenieria del conocimiento* . Obtenido de Metodología para el análisis Big Data en salud: <http://www.iic.uam.es/lasalud/metodologia-analisis-big-data-en-salud/>
- Instituto de Ingenieria del Conocimiento. (2019). Big Data en Salud. Madrid. Obtenido de <http://www.iic.uam.es/soluciones/salud/>
- Mariscal, J. (2018). *Estudio sobre TIC y salud pública en America Latina* . ITU.
- Mendoza, J. B. (2018). *Oportunidades y retos de los macrodatos*. Madrid: © Fundación Gaspar Casal. Obtenido de https://www.fundacionmylan.com/-/media/mylanfoundaciones/files/libro_final_oportunidades_y_retos_macrodatos.pdf
- Mendoza, J. B. (s.f.). *OPORTUNIDADES Y* . Obtenido de https://www.fundacionmylan.com/-/media/mylanfoundaciones/files/libro_final_oportunidades_y_retos_macrodatos.pdf
- Organización Mundial de la Salud. (2016). *La eSalud en la Región de las Americas: derribando las barreras a la implementacion* . Washington D.C : Organización Panamericana de la Salud.

- Riskin, D. (1 de Octubre de 2012). The Next Revolution in Healthcare en Forbes. Obtenido de www.forbes.com/sites/singularity/2012/10/01/the-next-revolution-in-healthcare/
- Servicio Nacional de Salud. (2018). *Memoria Institucional*. Santo Domingo . Obtenido de <http://memorias.minpre.gob.do/api/documents/1245/download>
- Weiskopf N. y Weng C. (2013). *Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research*. *Oxford Journal*, (20), 144-151.doi. Obtenido de <http://dx.doi.org/10.1136/amiajnl-2011-000681>
- Morel, M. T. (1 de Mayo de 2019). A catorce años de inaugurado, el Hospital Taiwán sigue sin UCI. Obtenido de <https://m.elcaribe.com.do/2019/05/01/catorce-anos-de-inaugurado-el-hospital-taiwan-sigue-sin-uci/>
- Batista, L. (11 de 09 de 2009). En Azua tienen problemas con hospital Taiwán. Obtenido de <https://www.diariolibre.com/actualidad/en-azua-tienen-problemas-con-hospital-taiwn-CLDL215236>

9. ESQUEMA PRELIMINAR DE CONTENIDO DEL TRABAJO DE GRADO.

Dedicatorias

Agradecimientos

Resumen ejecutivo

Glosario de términos

Introducción

Capítulo I: Análisis teórica de los sistemas de información aplicados a la toma de decisiones médicas mediante la implementación de Big Data.

1.1 Conceptos de la Big Data

1.2 Antecedentes Históricos

1.3 Evolución de la Big Data.

1.4 Características de Big data

1.5 Modelos de Big Data

1.6 Beneficios de Big Data en la Salud

1.7 Estándares Internacionales aplicados a Big Data

1.8 Sistemas de Gestión de Información (SGI)

1.8.1 Conceptualización

1.8.2 Historia de un SGI

1.8.3 Elementos de un SGI

1.8.4 Características de un SGI

1.8.5 Ciclo de vida de un SGI

1.9 E-Salud

Capítulo II: Historiales médicos electrónicos

2.1 Historiales Médicos

2.2 Funciones y Beneficios

2.3 Servicios de Admisiones y Documentación Clínica

- 2.4 Historiales Médicos Electrónicos (HME)
- 2.5 Evolución de los Historiales Médicos Electrónicos
- 2.6 Estructura de los HME
- 2.7 Características de HME
- 2.8 Versiones de HME
- 2.9 Beneficios de HME
- 2.10 Seguridad de la información

Capítulo III: Diagnostico del sistema de gestión del historial médico de los pacientes del Hospital Regional Taiwán 19 de marzo de Azua.

- 3.1 Sistema actual de gestión de historiales médicos
- 3.2 Herramientas tecnologías usadas
 - 3.2.1 Motor de datos
 - 3.2.2 Sistema de Gestión Hospitalaria (Portal)
 - 3.2.3 Entorno y plataforma de desarrollo
- 3.3 Descripción del sistema (Versiones)
- 3.4 Propósito del sistema
- 3.5 Funcionamiento del sistema
- 3.6 Actores del sistema
- 3.7 Ámbitos y escenarios del sistema
- 3.8 Análisis FODA
- 3.9 Operación y manejo para pacientes Emergencia
- 3.10 Operación y manejo de pacientes Referidos
- 3.11 Técnicas de evaluación

Capítulo IV: Análisis y reingeniería del sistema de gestión de información actual.

- 4.1 Plan estratégico
- 4.2 Análisis de los procesos y propuestas
- 4.3 Implementación
- 4.4 Identificación del objetivo primordial del rediseño o reingeniería

- 4.5 Selección de los procesos fundamentales
- 4.6 Selección del líder y de los miembros del equipo
- 4.7 Formación y entrenamiento del equipo de rediseño de procesos
- 4.8 Elaboración del diagrama de procesos (mapa)
- 4.9 Identificación de los problemas
- 4.10 Análisis de los problemas
- 4.11 Propuesta de rediseño o reingeniería
- 4.12 Elaboración del nuevo diagrama del proceso (cómo debería ser)
- 4.13 Definición de formas de medición

Capítulo V: Propuesta de un diseño de un sistema de información para la toma de decisiones médicas y administrativas en el Hospital Regional Taiwán 19 de marzo de Azua, R.D. aplicando analítica de Big Data sobre el historial médico de los pacientes.

5.1 Planteamiento de la propuesta

5.2 Fundamentación del proyecto

5.3 Documento Visión

5.3.1 Propósito

5.3.2 Alcance

5.3.3 Posicionamiento

5.3.4 Descripción de Stakeholders y usuarios

5.3.5 Descripción Global del Producto

5.4 Especificaciones de Diseño

5.4.1 Requisitos

5.4.2 Requisitos funcionales

5.4.3 Requisitos no funcionales

5.4 Diagramas y especificaciones

5.5.1 Diagramas de Casos de Uso

5.5.2 Diagrama de Secuencia

5.5.3 Diagrama de Estado

5.5.4 Diagrama de Clases

5.5.5 Diagrama de Colaboración

5.6 Base de datos

5.6.1 Arquitectura de base de datos

5.6.2 Diagrama Entidad-Relación

5.6.3 Lenguaje de programación

5.7 Detalles técnicos de la propuesta

5.7.1 Herramienta de Big Data

5.7.2 Arquitectura de desarrollo Big Data

5.7.3 APIs implementados

5.7.4 Seguridad

5.8 Prototipo de Interfaz de usuario

5.9 Tecnologías implementadas

5.10 Estudio de Factibilidad de la Propuesta

5.10.1 Factibilidad Técnica

5.10.2 Factibilidad Operacional

5.10.3 Factibilidad Económica

5.11 Impacto estructural

5.12 Determinación de la factibilidad de diseño del sistema de información propuesto a partir de consultas a especialista

Conclusiones

Recomendaciones

Bibliografía

Anexos