

UNIVERSIDAD PRIVADA ANTENOR ORREGO
FACULTAD DE INGENIERÍA

ESCU3ELA PROFESIONAL DE INGENIERÍA DE COMPUTACIÓN Y
SISTEMAS



**MODELO PREDICTIVO DE MINERIA DE DATOS
DE APOYO A LA GESTION HOSPITALARIA
SOBRE LA MORBILIDAD DE PACIENTES
HOSPITALIZADOS**

Hospital de Huaraz Víctor Ramos Guardia

TESIS

PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE
COMPUTACIÓN Y SISTEMAS

LÍNEA DE INVESTIGACIÓN: MODELO DE MINERÍA DATOS

AUTOR: Bach. EDWIN JOHN SALDAÑA VALQUI.

ASESOR: Dr. ELMER HUGO GONZALEZ HERRERA

TRUJILLO – PERU

2015

**Modelo Predictivo de Minería de Datos de Apoyo a la Gestión
Hospitalaria sobre Morbilidad de Pacientes Hospitalizados**

Elaborado por:

Bach. Saldaña Valqui Edwin John

Aprobado por:

Ing. Caballero Alvarado Armando Javier
PRESIDENTE
CIP: 149181

Ing. Piminchumo Flores Jorge Luis
SECRETARIO
CIP: 137153

Ing. Abanto Cabrera Heber Gerson
VOCAL
CIP: 106421

Dr. González Herrera Elmer Hugo
ASESOR
CIP: 24721

Presentación

Sr. Miembros del Jurado:

De conformidad con los requisitos estipulados en el Reglamento de Grados y Títulos de la Universidad Privada Antenor Orrego, para optar el Título Profesional de Ingeniero de Computación y Sistemas, someto a vuestra consideración la Tesis titulada:

“Modelo Predictivo de Minería de Datos de Apoyo a la Gestión Hospitalaria sobre la morbilidad de Pacientes Hospitalizados”

Este trabajo de investigación es el resultado de mi esfuerzo, donde he plasmado todos los conocimientos y experiencias adquiridas a lo largo de mi formación profesional, completando además con la orientación y el apoyo de mi Asesor y todas aquellas personas que colaboraron durante el desarrollo del Proyecto.

Atentamente

Bach. Edwin John Saldaña Valqui.

Dedicatoria

Dedico esta Tesis a mi Dios quien supo guiarme por el buen camino, darme fuerzas para seguir adelante y no desmayar en los problemas que se me presentaban, sin perder nunca la dignidad ni desfallecer en el intento

A mis Padres Miriam Valqui A. y Jose Saldaña R., que me han dado todo lo que soy como persona, mis valores, mis principios, mi carácter, mi empeño, mi perseverancia, mi coraje para conseguir mis objetivos.

Gracias también a mi familia por el apoyo incondicional a mis hijos Danna, Thiago, Favio, Frank, Jessenia, Ana, esposa Jackeline, cuñada Merly, Jessica, Estela primos Gerson, Rebeca, Liliana, Arnaldo, Carlos, Cristian, hermano Stalin y demás familiares por su apoyo, confianza y compartir y estar conmigo en los buenos y malos momentos.

Agradecimientos

Mi agradecimiento al Dr. Elmer Hugo González Herrera, por su esfuerzo y dedicación. Que como asesor me ha orientado, apoyado, con mucho interés y entrega, que ha sobrepasado todas las expectativas, que como alumno, deposite en su persona.

Mi agradecimiento al jurado calificador Ing. Armando Caballero Alvarado, Ing. Jorge Piminchumo Flores, Ing. Heber Abanto Cabrera, por sus comentarios y sugerencias al presente trabajo.

Así mismo mis sinceros agradecimientos al Dr. Edwin Johny Asnate Salazar, especialista en Modelos Estadísticos, por atender mis consultas, material facilitado y las sugerencias recibidas.

Resumen

MODELO PREDICTIVO DE MINERIA DE DATOS DE APOYO A LA GESTION HOSPITALARIA SOBRE LA MORBILIDAD DE PACIENTES HOSPITALIZADOS.

Por:

Bach. Saldaña Valqui, Edwin John

La minería de datos en el sector salud permite identificar los perfiles de salud en los pacientes, ayuda a comprender el patrón de comportamiento, a través del historial de información almacenada que hace posible su gestión cotidiana, siendo así esta información diversa y compleja.

El presente trabajo de investigación, propone aplicar un marco estándar de actividades de minería de datos, creando un modelo predictivo, que sirva de apoyo a la Gestión Hospitalaria sobre la morbilidad con pacientes hospitalizados, basado en el algoritmo de análisis de serie de tiempo, Modelo **ARIMA** (AutoRegresive Integrated Moving Average) de **Box y Jenkins** (Box G.E.P. & Jenkins, 1973), con información histórica de los últimos 7 años de los pacientes del Hospital Víctor Ramos Guardia.

En la investigación, se tomó como referencia la metodología CRISP-DM (Cross Industry Standard Process For Data Mining), que consiste en la comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Por lo tanto se realizó la extracción de los datos, transformación de los datos, carga de datos, limpieza de datos, diseño del datamart “HEALTHMINING”, la selección y creación de variables que sirvieron como datos de entrada para mi modelo, para posteriormente crear un modelo de pronósticos, que me permitió conocer los casos de morbilidad en pacientes hospitalizados del hospital VRG para los próximos tres años.

Palabras clave: 1) MDHM: Modelo Predictivo Minería Datos – Health Mining, 2) AST-A: Algoritmo Serie Tiempo – Modelo Arima, 3) MD: Minería Datos, 4) DM: Datamart

Abstract

DATA MINING AND PREDICTIVE MODEL IN SUPPORT OF HEALTH CARE MANAGEMENT AND IT'S IN PATIENT MORBIDITY

by:

Bach.: Edwin John Saldaña Valqui

Data mining in the health care sector identifies the health profiles of patients, helps to understand the pattern of behavior of patients through history stored information from your transactional system.

This research proposes to apply a standard framework for data mining, creating a predictive model, which supports the Hospital Management on morbidity in hospitalized patients, based algorithm for time series analysis, **ARIMA** Model (AutoRegresive Integrated Moving Average) of **Box and Jenkins** (Box GEP & Jenkins, 1973), based on historical data to predict future or unknown values.

In research, reference was made to the CRISP-DM methodology (Cross Industry Standard Process for Data Mining), which consists of business understanding, data understanding, data preparation, modeling, evaluation and deployment. Therefore the data extraction , data transformation , data loading , data cleansing , data mart design " HEALTHMINING " the selection and creation of variables that were used as input to my model, was performed to create later forecasting model , which allowed me to hear cases of morbidity in hospitalized patients in the hospital VRG for the next three years.

Keywords: 1) MDHM: Model Data Mining-Mining Health, 2) AST-A: Time Series Algorithm-Model Arima, 3) MD: Data Mining, 4) DM: Datamart

Contenido

Índice General

UNIVERSIDAD PRIVADA ANTENOR ORREGO	i
Modelo Predictivo de Minería de Datos de Apoyo a la Gestión Hospitalaria sobre Morbilidad de Pacientes Hospitalizados	i
Presentación	ii
Dedicatoria.....	iii
Agradecimientos.....	iv
Resumen	v
Abstract.....	vi
Contenido.....	vii
Lista de figuras	x
INTRODUCCION.....	1
1.1 Antecedentes	1
1.2 Análisis y Formulación del Problema.....	2
1.3 Hipótesis	4
1.4 Objetivo General	4
1.5 Objetivos Específicos	4
CAPITULO I: FUNDAMENTO CONCEPTUAL	6
1.1 MODELO DE MINERIA DATOS.....	6
1.1.1 MODELO PREDICTIVO.....	8
1.1.2 MODELO DESCRIPTIVO.	9
1.2 METODOLOGIAS DE MINERIA DE DATOS	9
1.2.1 METODOLOGÍA CRISP-DM.....	9
1.2.1.1 COMPRENSION DEL NEGOCIO	10
1.2.1.2 COMPRENSION DE LOS DATOS	10
1.2.1.3 PREPARACION DE LOS DATOS	10
1.2.1.4 MODELADO	10
1.2.1.5 EVALUACION.....	11
1.2.1.6 DESPLIEGUE	11
1.2.2 METODOLOGÍA SEMMA	12
1.2.3 DATAMART	12
1.3 TECNICAS DE MINERIA DATOS	14
1.3.1 SERIES TEMPORALES.....	14
1.3.2 SERIES DE TIEMPO.	15

1.3.3	PRONOSTICOS DE SERIES DE TIEMPO.	15
1.3.4	MODELO ARIMA.	16
1.3.4.1	METODOLOGIA DEL MODELO ARIMA.	17
1.3.4.2	PASOS A SEGUIR PARA EL ANÁLISIS DE DATOS.	18
1.4	GESTION HOSPITALARIA.	20
1.4.1	INDICADORES DE GESTION HOSPITALARIA.	20
1.4.1.1	INDICADORES DEL SECTOR SALUD.	20
1.4.2	APOYO A LA TOMA DECISIONES.	21
1.4.3	SISTEMAS DE APOYO A LA TOMA DECISIONES.	22
1.4.4	SISTEMA INFORMACION GERENCIAL VS SISTEMA DE APOYO DE DECISIONES.	22
1.5	MORBILIDAD.	22
1.5.1	TASA DE MORBILIDAD.	23
1.5.2	DIAGNOSTICO EN PACIENTES.	24
1.5.3	CIE-10.	24
CAPITULO II: MATERIALES Y METODOS.		26
2.1	MATERIALES.	26
2.2	METODOLOGIA.	27
CAPITULO III: RESULTADOS.		29
3.1	COMPRESION DEL NEGOCIO.	29
3.1.1	ESTABLECIMIENTO DE LOS OBJETIVOS.	29
3.1.2	EVALUACION DE LA SITUACION.	30
3.2	COMPRESION DE LOS DATOS.	31
3.2.1	RECOPIACION INICIAL DE LOS DATOS.	31
3.2.2	DESCRIPCION DE LOS DATOS.	31
3.2.2.1	DIAGNOSTICO INGRESOS.	32
3.2.2.2	DIAGNOSTICO EGRESOS.	32
3.2.2.3	PACIENTES.	33
3.2.2.4	UBICACION GEOGRAFICA.	34
3.2.2.5	EPICRISIS.	35
3.3	PREPARACION DE LOS DATOS.	37
3.3.1	EXTRACCION DE LOS DATOS.	37
3.3.2	TRANSFORMACION DE LOS DATOS.	38
3.3.3	CARGA DE LOS DATOS.	38
3.3.4	CREACION DEL DATAMART.	41
3.3.5	LIMPIAR LOS DATOS.	42
3.3.6	SELECCIÓN DE LOS DATOS PARA LA CONSTRUCCION DEL MODELO.	46
3.4	MODELADO.	47
3.4.1	SELECCIÓN DE LA TECNICA DE MODELADO.	47
3.4.2	CONSTRUCCION Y EVALUACION DEL MODELO.	48
3.4.3	IDENTIFICACION DEL POSIBLE MODELO.	48
3.4.4	ESTIMACION DEL MODELO.	51
3.4.5	DIAGNOSTICO DEL MODELO.	57
3.4.6	PRONOSTICO DEL MODELO.	59
3.5	EVALUACION.	60
3.5.1	EVALUACION DE LOS RESULTADOS OBTENIDOS.	60

3.5.2	ESTIMACION DE COSTOS Y PRESUPUESTOS.....	61
3.5.3	SALIDA DE LOS RESULTADOS.	65
CAPITULO IV: DISCUSION		71
4.1	PLANTEAMIENTO DE LA HIPÓTESIS.....	71
4.2	CONTRASTACIÓN DE LA HIPÓTESIS	71
4.2.1	INDICADOR: COSTO	75
4.2.2	INDICADOR: SATISFACCIÓN DEL USUARIO	79
4.3	ANÁLISIS DE RESULTADOS	83
CAPITULO V: CONCLUSIONES Y RECOMENDACIONES		84
5.1	CONCLUSIONES	84
5.2	RECOMENDACIONES	86
ANEXOS		87
A.	RESULTADOS DE ENCUESTAS	87
B.	MATERIAL MEDICO Y MEDICAMENTOS VENCIDOS	90
C.	TENDENCIA DE LOS PRINCIPALES DIAGNOSTICOS DE PACIENTES HOSPITALIZADOS:.....	91
D.	CODIGO CONSULTA SQL DE PACIENTES CON DIAGNOSTICOS POR AÑO MES:.....	94
E.	BASE DATOS TRANSACCIONAL	95
F.	POBLACIÓN DEPARTAMENTO ANCASH, PROVINCIA HUARAZ, DISTRITO HUARAZ 2008-2015. (FUENTE INEI)	96
G.	NUMERO DE CASOS DE DIAGNOSTICOS POR PERIODO:.....	96
H.	TABLAS DE RESULTADOS	100

Lista de figuras

Pág.

Figura 1 Minería Datos. (Moxon's, 1996)	6
Figura 2. Metodología CRISP-DM. (http://www.crisp-dm.org/)	11
Figura 3. Algoritmo de Serie Temporales. Modelo Predictivo Sarima. (Mansilla, 2014)	14
Figura 4. Metodologías de Box-Jenkins. (Box G.E.P. & Jenkins, 1973)	19
Figura 5. Atributos - Diagnostico Ingresos	32
Figura 6. Atributos - Diagnostico Egresos	32
Figura 7. Atributos - Pacientes	33
Figura 8. Atributos - Ubicación geográfica	34
Figura 9. Extracción de Datos	37
Figura 10. Transformación de Ingresos de Pacientes con diagnósticos	39
Figura 11. Transformación, Egresos de Pacientes	39
Figura 12. Carga de Pacientes	40
Figura 13. Ingresos Epicrisis, Egresos Epicrisis	40
Figura 14. Carga Diagnósticos	40
Figura 15. Carga Ubicación Geográfica	40
Figura 16. Datamart - "HEALTHMINING"	41
Figura 17. Datos obtenidos del Datamart HEALTH MINING	48
Figura 18. Grafica del Diagnostico K35 Apendicitis con Tendencia creciente y con muchos picos	49
Figura 19. Procesando el test de Dickey Fuller	49
Figura 20. Resultado de la prueba de raíz unitaria de Dickey-Fuller	50
Figura 21. Correlograma. Representación gráfica de la muestra	51
Figura 22. Estimación del Modelo	52
Figura 23. Probabilidades del primer Modelo estimado	53
Figura 24. Estimación del modelo anterior excluyendo las probabilidades mayores a 5%	54
Figura 25. Parámetros del Modelo Final	55
Figura 26. Valores y Coeficientes del Modelo Final	55
Figura 27. Diagnóstico del Modelo	57
Figura 28. Distribución de errores	57
Figura 29. Correlograma del modelo final	58
Figura 30. Grafico Línea según Modelo ARIMA proyectado	59
Figura 31. Grafico Barras según Modelo ARIMA proyectado	59
Figura 32. No. Diagnósticos vs Costos proyectado 2015	64
Figura 33. No. Diagnósticos vs Costos proyectado 2016	64
Figura 34. No. Diagnósticos vs Costos proyectado 2017	64

Figura 35. Grafico Barras Pacientes Hospitalizados 2008-2014	66
Figura 36. Grafico Número de casos de pacientes hospitalizados 2008-2014	66
Figura 37. Casos de morbilidad de pacientes, excepto diagnostico O00-O99.....	67
Figura 38. Monitoreo de Diagnostico K35 últimos 7 años	68
Figura 39. Monitoreo de Diagnostico K35 a nivel distrital.....	68
Figura 40. Tasa de incidencia actual 2008 – 2014	69
Figura 41. Tasa de incidencia pronosticada 2008 - 2017	69
Figura 42. Grafico barras Tasa incidencia actual vs pronosticada.....	70
Figura 43. Gráfico de Líneas incidencias actual vs pronosticada	70
Figura 44: Región de aceptación o rechazo del indicador costo.	78
Figura 45: Región de aceptación o rechazo del indicador satisfacción del usuario.	82
Figura 46. Dx - Apendicitis	91
Figura 47. Dx – Colecistitis Crónica	92
Figura 48. Dx – Bronconeumonía	92
Figura 49. Dx – Diarreas y gastroenteritis	93
Figura 50. Dx – Traumatismo.....	93
Figura 51. Atenciones de Pacientes por Dx	100
Figura 52. Atenciones de Pacientes provincia Huaraz	101
Figura 53. Dx de pacientes en el distrito de Huaraz	103

Lista de tablas

Pág.

Tabla 1 Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud	25
Tabla 2: Número de diagnósticos más frecuentes de pacientes hospitalizados durante los últimos 7 años (2008-2014), excepto Parto, Embarazo y Puerperio.	47
Tabla 3: Número de diagnósticos sin modelo vs con modelo (K35 Apendicitis Aguda). 60	
Tabla 4: Modelo Propuesto vs método actual 1 y 2.....	61
Tabla 5: Costo promedio mensual por hora y minuto del recurso humano.....	61
Tabla 6: Farmacia – Receta de Material médico y medicamentos.	62
Tabla 7: Costo de insumos y material médico por consulta de paciente	62
Tabla 8: Resumen de Costos por paciente con diagnostico apendicitis	63
Tabla 9: Cuadro de Costos proyectados con el modelo para los próximos tres años.	63
Tabla 10: Número de casos de morbilidad en pacientes hospitalizados (2008-2014)	65
Tabla 11: Variación de datos históricos del indicador Costos.....	76
Tabla 12: Variación de datos históricos del indicador satisfacción del usuario final. 80	
Tabla 13: Medicamentos y material médico vencidos de periodo 2014.....	90
Tabla 14: Cuadro de cantidad de diagnósticos más frecuentes en los últimos 8 años 96	
Tabla 15: Detalle del número de diagnósticos en pacientes hospitalizados del hospital VRG.	100

INTRODUCCION

1.1 Antecedentes

La minería de datos entró en existencia a inicios de los años 90 y apareció como una poderosa herramienta para buscar sobre grandes volúmenes de datos patrones de comportamiento sobre información desconocida.

En EE.UU. la minería de datos apoya a las compañías prestadoras de salud en la detección de fraudes, solución médica a los pacientes a bajo costo, detección de las causas de las enfermedades, la identificación de los métodos de tratamiento médico, desarrollo de perfiles de salud de las personas, etc. (Tan, 2005).

Iván Gildo Tapia Rivas. Universidad Nacional Mayor de San Marcos Lima (2006), en su Tesis de Minería de Datos, sectorizo a los pacientes en el consumo de medicamentos, para descubrir y enumerar patrones presentes en los datos, utilizando algoritmos de segmentación o clasificación, para evaluar la forma con la que se consumen los medicamentos en un hospital en el Perú y poder identificar algunas realidades o características no observables que producirían desabastecimiento o insatisfacción del paciente, y para que sirva como una herramienta en la toma de decisión sobre el abastecimiento de medicamentos en el hospital.

La modelación ARIMA (Autorregresive Integrated Moving Average), desarrollada por *George Box* y *Gwilym Jenkins* en la década de los setenta del pasado siglo con el propósito de obtener mejores pronósticos en el control de la contaminación de la bahía de San Francisco, constituyó una verdadera revolución en el análisis de series de tiempo

La primera utilización de los modelos ARIMA en el ámbito sanitario se debe a Keewan Choi del Buró de Epidemiología del Centro para el Control de

Enfermedades (CDC) y Stephen Tacker de la Universidad de Emory, ambos en Atlanta, EE.UU., quienes propusieron este método para pronosticar el número de muertes esperadas por influenza y neumonía.

M. Pilar Muñoz G. Construyo un modelo utilizando algoritmo de series temporales, para la monitorización de las señales biomédicas obtenidas en un paciente con los valores de creatinina tras un trasplante renal. (Gracia, 2004)

Claudio Cárdenas Mansilla. Elaboro un modelo de pronóstico de serie temporal de (Box G.E.P. & Jenkins, 1973) para determinar la demanda total de consultas por emergencia para el consultorio de neumología de 5 centros hospitalarios de Santi. (Mansilla, 2014).

1.2 Análisis y Formulación del Problema

El modelo es parte fundamental de la minería de datos, que ayuda a comprender el patrón de comportamiento de los pacientes y de esta manera permite llegar a una toma de decisión rápida y óptima.

En la mayoría de centros hospitalarios existe una gestión hospitalaria deficiente, no utilizándose en forma adecuada tanto los recursos humanos como materiales. Esto trae consigo que se acentúe aún más el desbalance entre la capacidad de oferta y demanda. (Ramirez, 2012)

En el Hospital Víctor Ramos Guardia, como en la mayoría de centros hospitalarios, existe una programación deficiente en el abastecimiento de insumos, medicamentos,

provisiones, servicios y recurso humano de las diferentes especialidades de los diferentes servicios.

- La adquisición de medicamentos, insumos, provisiones, se adquieren basándose al consumos del mes anterior, consumo del año anterior y con riesgos a vencerse por sobre stock, o en quedar desabastecidos en el mes.
- El profesional especializado de los diferentes servicios, no se abastece en la atención al paciente.
- El médico de turno, tiene que asumir funciones y resolver problemas de otros especialistas, exponiéndose a errores de negligencia médica.

Si bien es cierto el Hospital cuenta con grandes volúmenes de informaron almacenada en su base datos transaccional es necesario tratarla, analizarla y extraer conocimiento que nos permita conocer los patrones de comportamiento para poder realizar trabajos de prevención.

En este contexto entre el uso de la herramienta tradicional “Sistema Transaccional” y la minería de datos nos planteamos el siguiente Problema:

¿Cómo apoyar la gestión hospitalaria sobre la morbilidad en pacientes hospitalizados del Hospital Víctor Ramos Guardia?

1.3 Hipótesis

La creación del modelo predictivo de minería de datos en el Hospital VRG de Huaraz, apoyara la gestión hospitalaria sobre la morbilidad en pacientes hospitalizados.

Para esto nos planteamos los siguientes objetivos:

1.4 Objetivo General

- Crear un modelo predictivo de minería de datos de apoyo a la gestión hospitalaria sobre la morbilidad de pacientes hospitalizados.

1.5 Objetivos Específicos

- Analizar el estado del arte en modelos predictivos minería de datos, de apoyo a la gestión hospitalaria, mediante un estudio de fuentes bibliográficas y modelos existentes para el sector salud.
- Diseñar el proceso de preparación los datos
- Aplicar la metodología de minería de datos de Box-Jenkins Modelo ARIMA.
- Diseñar un Modelo de Minería de Datos.

Por lo tanto esta investigación se desarrolla en los siguientes capítulos.

Capítulo 1: Comprende el fundamento conceptual sobre, toma de decisiones, pronósticos, metodologías y modelos de referencia que han servido como estudio para comprender y aplicar la metodología a desarrollar.

Capítulo 2: Se hace referencia a los materiales y metodología utilizada para esta investigación.

Capítulo 3: Capitulo de resultados, comprende desde la construcción del Datamart como fuente de información para poder crear un modelo predictivo de minería datos, que a la vez nos permita crear variables para la construcción del modelo de minería datos basado en el algoritmo de Análisis de Serie de Tiempo ARIMA, hasta la creación del modelo predictivo de minería datos.

Capítulo 4: Capítulo de Discusión consiste en contractar los resultados del modelo.

Capítulo 5: Se presentan las Conclusiones de la investigación y los trabajos futuros.

CAPITULO I: FUNDAMENTO CONCEPTUAL

1.1 MODELO DE MINERIA DATOS

La minería de datos trabaja con los datos con el objetivo de encontrar patrones de comportamiento ocultos en los mismos, con el fin de crear nuevas asociaciones, predecir resultados futuros y a partir de ello generar conocimiento útil. Es el proceso de que tiene como propósito descubrir, extraer y almacenar información. (Wong, Carpal tunnel syndrome: diagnostic usefulness of sonography, 2004)

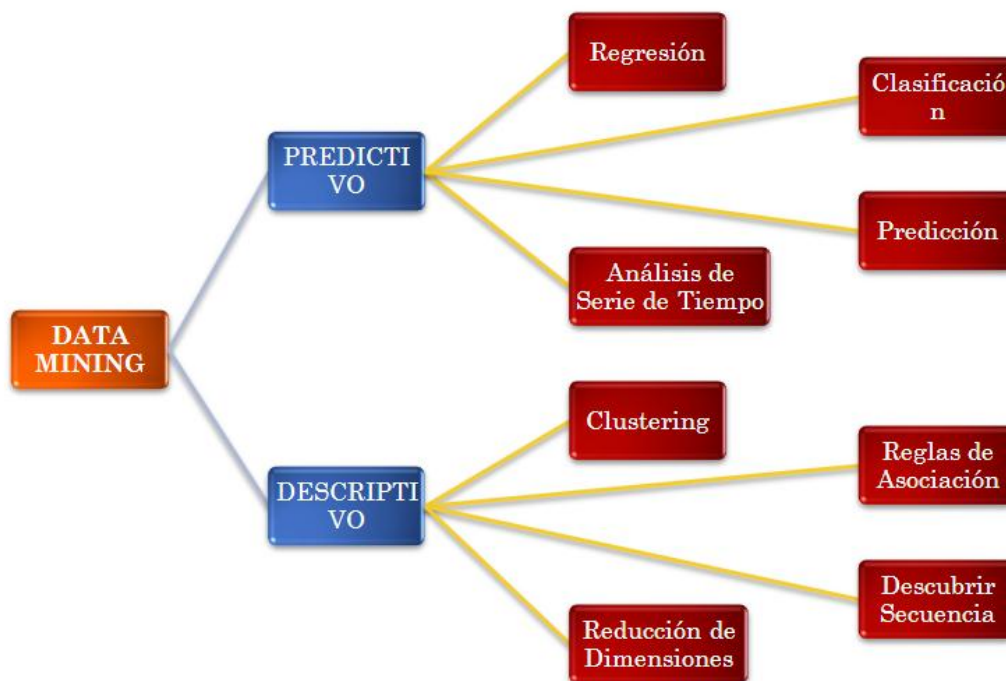


Figura 1 Minería Datos. (Moxon's, 1996)

Minería de Datos es un término genérico que engloba resultados de investigación, técnicas y herramientas usadas para extraer información útil de grandes bases de datos. Si bien Minería de Datos es una parte del proceso completo de KDD, en buena parte de la literatura los términos Minería de Datos y KDD se identifican como si fueran lo mismo. Concretamente, el término Minería de Datos es usado comúnmente por los estadísticos, analistas de datos, y por la comunidad de administradores de sistemas informáticos como todo el proceso del descubrimiento, mientras que el término KDD es utilizado más por los especialistas en Inteligencia Artificial.

Según William Frawley, Gregory Piatetsky y Christopher Matheus (1992): La minería de datos es la extracción no trivial de información implícita, desconocida previamente, y potencialmente útil desde los datos (Frawley, Piatetsky, & Matheus, 1991).

Según Evangelos Simoundis, Brian Livezey y Randy Kerber (1996): La minería es el proceso de extracción y refinamiento de conocimiento útil desde grandes bases de datos (Simoundis, Livezey, & Kerber, 1996).

Según P. Cabena, P. Hadjinian, R. Stadler, J. Verhees, A. Zanasi (1997): Los modelos de minería de datos es el proceso de extracción de información previamente desconocida, válida y procesable desde grandes bases de datos para luego ser utilizada en la toma de decisiones (Cabena, Hadjinian, Stadler, Verhees, & Zanasi, 1997).

Es la exploración y análisis, a través de medios automáticos y semiautomáticos, de grandes cantidades de datos con el fin de descubrir patrones y reglas significativos (G., 1997).

Es el proceso de planteamiento de distintas consultas y extracción de información útil, patrones y tendencias previamente desconocidas desde grandes cantidades de datos posiblemente almacenados en bases de datos (Thuraisingham, 1999).

Para Ian H. Witten y Eibe Frank (2000): La minería de datos es el proceso de descubrir modelos en los datos (Witten & Frank, 2000).

1.1.1 MODELO PREDICTIVO

Según J. Hernández, M.J. Ramírez y C. Ferri (2004): El modelo predictivo se emplea para estimar valores futuros de variables de interés. El proceso de basa en la información histórica de los datos, mediante las cuales se predice el comportamiento de los datos, ya sea mediante clasificaciones, categorizaciones o regresiones. El atributo a predecir se le conoce como variable dependiente u objetivo, mientras que los atributos utilizados para realizar la predicción se llaman variables independientes o de exploración. (Hernandez, Ramirez , & Ferri, 2004)

Según S. N. Sivanandam y S. Sumathi (2006): Los modelos predictivos siguen un aprendizaje supervisado, que consiste en aprender mediante el control de un supervisor o maestro que determina la respuesta que se desea generar del sistema (Sivanandam & Sumathi, 2006).

1.1.2 MODELO DESCRIPTIVO.

Según S. N. Sivanandam y S. Sumathi (2006): En el modelo descriptivo se identifican patrones que describen los datos mediante tareas. Destacan que mediante este modelo se identifican patrones que explican o resumen el conjunto de datos, siendo estos útiles para explorar las propiedades de los datos examinados. Los modelos descriptivos siguen un tipo de aprendizaje no supervisado, que consiste en adquirir conocimiento desde los datos disponibles, sin requerir influencia externa que indique un comportamiento deseado al sistema. (Sivanandam & Sumathi, 2006)

1.2 METODOLOGIAS DE MINERIA DE DATOS

Existen muchas formas de desarrollar proyectos de minería de datos, pero las más conocidas en el medio son: la metodología CRISP-DM y la metodología SEMMA.

El principal objetivo surge básicamente de la necesidad que se tiene en el medio de tener un proceso o pasos estándares para la resolución de problemas.

En este campo, cuando se desea solucionar uno o varios problemas, no es suficiente con tener las herramientas que me ayuden a resolver preguntas y predecir comportamientos, con un conjunto de métodos y algoritmos formulados para dar respuesta a estos, sino que además se requiere de un conjunto de pasos sistematizados que guíen el proceso que se debe seguir desde que se estudian los problemas que se desean tratar hasta que se tienen las respuestas a estos problemas formulados.

1.2.1 METODOLOGÍA CRISP-DM

El CRISP-DM (Cross Industry Standard Process for Data Mining) ofrece un marco estándar para la realización de actividades de minería de datos.

En 1997 con el apoyo financiero de la comisión Europea se formó el Consorcio, con el objetivo de lograr una herramienta para la industria lo más generalizada posible a fin de adaptarse a la mayor diversidad de industrias.

CRISP-DM representa el **ciclo de vida** de un proyecto DM en seis fases.

1.2.1.1 COMPRESION DEL NEGOCIO

- Establecimiento de los objetivos del negocio (Contexto inicial, objetivos, criterios de éxito)
- Evaluación de la situación (Inventario de recursos, requerimientos, supuestos, terminologías propias del negocio)
- Establecimiento de los objetivos de la minería de datos (objetivos y criterios de éxito)
- Generación del plan del proyecto (plan, herramientas, equipo y técnicas)

1.2.1.2 COMPRESION DE LOS DATOS

- Recopilación inicial de datos
- Descripción de los datos
- Exploración de los datos
- Verificación de calidad de datos

1.2.1.3 PREPARACION DE LOS DATOS

- Selección de los datos
- Limpieza de datos
- Construcción de datos
- Integración de datos
- Formateo de datos

1.2.1.4 MODELADO

- Selección de la técnica de modelado
- Diseño de la evaluación

- Construcción del modelo
- Evaluación del modelo

1.2.1.5 EVALUACION

- Evaluación de resultados
- Revisar el proceso
- Establecimiento de los siguientes pasos o acciones

1.2.1.6 DESPLIEGUE

- Planificación de despliegue
- Planificación de la monitorización y del mantenimiento
- Generación de informe final
- Revisión del proyecto

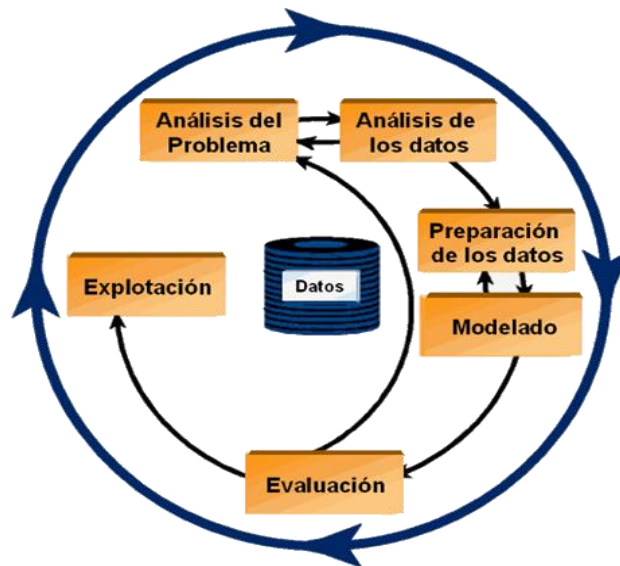


Figura 2. Metodología CRISP-DM. (<http://www.crisp-dm.org/>)

1.2.2 METODOLOGÍA SEMMA

SEMMA (Simple, Explore, Modify, Model, Assess), como metodología para procesos de Minería de Datos, fue desarrollada por SAS institute y la define como una herramienta que ayuda a los usuarios en los procesos de selección, exploración y modelación de cantidades significativas de datos almacenados, para así poder responder a preguntas o predecir eventos que pueden pasar.

Según SAS institute, más que una metodología de procesos de minería de datos, SEMMA se puede identificar como un conjunto de herramientas funcionales, enfocándose más en los aspectos del desarrollo del modelo de minería de datos.

1.2.3 DATAMART

Los usuarios a menudo realizan análisis de datos sobre un subconjunto de datos relativamente pequeño de todo el datawarehouse. Asimismo, era muy probable que los usuarios repitieran las mismas operaciones sobre el mismo subconjunto de datos cada vez que sea actualizado.

La ejecución repetida de tales operaciones sobre el mismo subconjunto de todo el almacén no es muy eficiente; por lo tanto, pareció buena idea construir algún tipo de "almacén" limitado de propósito general que estuviera hecho a la medida de ese propósito. Además, en algunos casos sería posible extraer y preparar los datos requeridos directamente a partir de las fuentes locales, lo que proporcionaba un acceso más rápido a los datos que si tuvieran que ser sincronizados con los demás datos cargados en todo el datawarehouse.

Se puede definir como "un almacén de datos especializado, orientado a un tema, integrado, volátil y variante en el tiempo para apoyar un subconjunto específico de decisiones de administración". La principal diferencia entre un datamart y un datawarehouse es que el datamart es especializado y volátil. Especializado quiere

decir que contiene datos para dar apoyo (solamente) a un área específica de análisis de negocios; por volátil se entiende que los usuarios pueden actualizar los datos e incluso, posiblemente, crear nuevos datos (es decir, nuevas tablas) para algún propósito.

C. MacGregor, C. Christina y J. Andrew (2012):

Señalaron que un datamart puede ser creado en forma independiente (es decir, no por medio de la extracción a partir del datawarehouse). Dicho enfoque puede ser adecuado si el datawarehouse es inaccesible por alguna causa: razones financieras, operacionales o incluso políticas (o puede ser que ni siquiera exista todavía el datawarehouse). Para mejorar la atención y el cuidado del paciente (C. MacGregor, 2012).

1.3 TECNICAS DE MINERIA DATOS

1.3.1 SERIES TEMPORALES.

Algoritmo de series de tiempo son datos estadísticos que se recopilan, observan o registran en intervalos de tiempo regulares (diario, semanal, semestral, anual, entre otros).

M. Pilar Muñoz G. Construyo un modelo utilizando algoritmo de series temporales, para la monitorización de las señales biomédicas obtenidas en un paciente con los valores de creatinina tras un trasplante renal. (Gracia, 2004)

Claudio Cárdenas Mansilla. Elaboro un modelo de pronóstico de serie temporal de (Box G.E.P. & Jenkins, 1973) para determinar la demanda total de consultas por emergencia para el consultorio de neumología de 5 centros hospitalarios. (Mansilla, 2014)

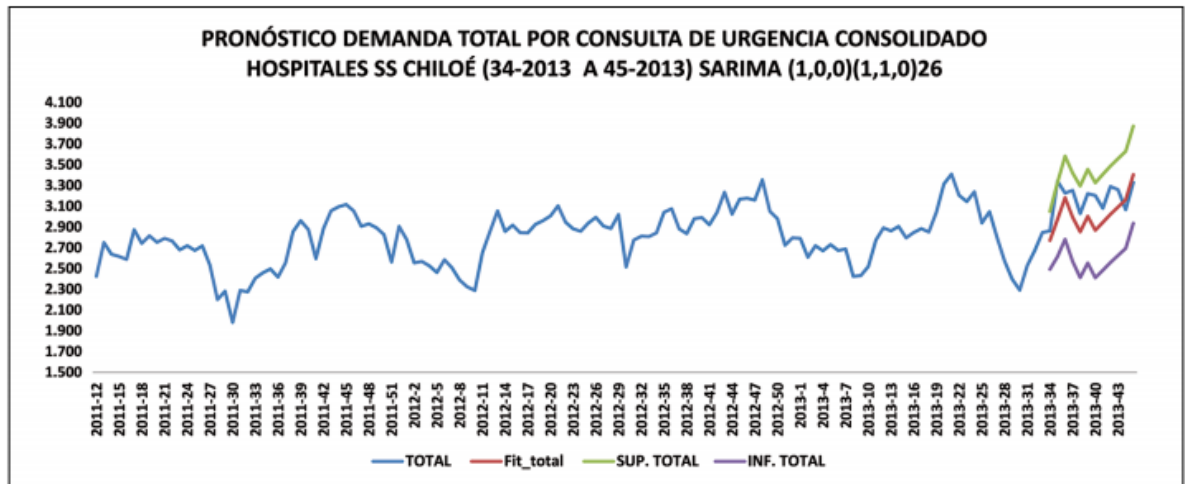


Figura 3. Algoritmo de Serie Temporales. Modelo Predictivo Sarima. (Mansilla, 2014)

1.3.2 SERIES DE TIEMPO.

Es una recolección de datos anteriores durante varios periodos. El conjunto de datos resultantes se les denomina Serie de Tiempo o Serie Temporal, porque contiene observaciones para alguna variable durante el tiempo, los periodos de tiempo varían en duración. Pueden ser anuales trimestrales, mensuales e incluso diario. El propósito del análisis del tiempo es predecir o proyectar los valores futuros de la variable, a partir de observaciones anteriores. (Douglas A. Lind, 2008).

Una serie de tiempo es una secuencia de datos numéricos, cada uno de los cuales se asocia con un instante específico del tiempo, es decir observaciones de la misma variable en tiempos $t_1, t_2, t_3, t_4, t_5 \dots t_i$. (G.S.Maddala, 1983)

Según Joannès Vermorel 2012: Una serie de tiempo es una lista de fechas, cada una de las cuales se asocia a un valor (un número). Las series de tiempo son un modo estructurado de representar datos. Visualmente, es una curva que evoluciona a lo largo del tiempo. Por ejemplo, las ventas diarias de un producto pueden representarse como una serie de tiempo. (Vermorel, 2012)

1.3.3 PRONOSTICOS DE SERIES DE TIEMPO.

Según Joannès Vermorel 2012: El pronóstico de las series de tiempo significa que extendemos los valores históricos al futuro, donde aún no hay mediciones disponibles. Existen dos variables estructurales principales que definen un pronóstico de serie de tiempo:

- El período, que representa el nivel de agregación. Los períodos más comunes son meses, semanas y días (para la optimización del inventario). En los centros de atención telefónica utilizan períodos en horas, medias o cuartos de hora.
- El horizonte, que representa la cantidad de períodos por adelantado que deben ser pronosticados.

1.3.4 MODELO ARIMA.

(Box G.E.P. & Jenkins, 1973). **Box y Jenkins** han desarrollado modelos estadísticos para series temporales que tienen en cuenta la dependencia existente entre los datos, esto es, cada observación en un momento dado es modelada en función de los valores anteriores. Los análisis se basan en un modelo explícito. Los modelos se conocen con el nombre genérico de **ARIMA** (AutoRegresive Integrated Moving Average), que deriva de sus tres componentes **AR** (Autoregresivo), **I** (Integrado) y **MA** (Medias Móviles).

El modelo **ARIMA** permite describir un valor como una función lineal de datos anteriores y errores debidos al azar, además, puede incluir un componente cíclico o estacional. Es decir, debe contener todos los elementos necesarios para describir el fenómeno. **Box y Jenkins** recomiendan como mínimo 50 observaciones en la serie temporal. (Box G.E.P. & Jenkins, 1973)

Un modelo **ARIMA** (0, d, 0), es una serie temporal que se convierte en ruido blanco (proceso puramente aleatorio) después de ser diferenciada d veces.

El modelo (0, d, 0) se expresa mediante: $(1-B)^d x_t = a_t$

El modelo general **ARIMA** (p, d, q) denominado proceso autorregresivo integrado de medias móviles de orden p, d, q, toma la expresión:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d X_t = (1 - \nu_1 B - \nu_2 B^2 - \dots - \nu_q B^q) a_t$$

Un modelo **ARIMA** (p,d,q) permite describir una serie de observaciones después de que hayan sido diferenciadas d veces, a fin de extraer las posibles fuentes de no estacionariedad. Esta fórmula se puede aplicar a cualquier modelo. Si hay alguna componente p, d, q, igual a cero, se elimina el término correspondiente de la fórmula general. Los modelos cíclicos o estacionales son aquellos que se caracterizan por

oscilaciones cíclicas, también denominadas variaciones estacionales. Las variaciones cíclicas a veces se superponen a una tendencia secular.

Las series con tendencia secular y variaciones cíclicas pueden representarse mediante los modelos **ARIMA** (p, d, q) (P, D, Q). El primer paréntesis (p, d, q) se refiere a la tendencia secular o parte regular de la serie y el segundo paréntesis (P, D, Q) se refiere a las variaciones estacionales, o parte cíclica de la serie temporal.

1.3.4.1 METODOLOGIA DEL MODELO ARIMA.

La metodología de Box y Jenkins se resume en cuatro fases:

- **Primera Fase:** Consiste en identificar el posible modelo **ARIMA** que sigue la serie, lo que requiere: Decidir qué transformaciones aplicar para convertir la serie observada en una serie estacionaria. Determinar un modelo ARMA para la serie estacionaria, es decir, los órdenes p y q de su estructura auto regresiva y de media móvil.
- **Segunda Fase:** Seleccionado provisionalmente un modelo para la serie estacionaria, se pasa a la segunda etapa de estimación, donde los parámetros AR y MA del modelo se estiman por máxima verosimilitud y se obtienen sus errores estándar y los residuos del modelo.
- **Tercera Fase:** es el diagnostico, donde se comprueba que los residuos no tienen estructura de dependencia y siguen un proceso de ruido blanco. Si los residuos muestran estructura se modifica el modelo para incorporarla y se repiten las etapas anteriores hasta obtener un modelo adecuado.
- **Cuarta Fase:** es la predicción, una vez que se ha obtenido un modelo adecuado se realizan predicciones con el mismo.

1.3.4.2 PASOS A SEGUIR PARA EL ANÁLISIS DE DATOS

- **Recogida de datos:** Es conveniente disponer de 50 o más datos, y en el caso de series mensuales, trabajar entre seis y diez años completos.
- **Representación gráfica:** Es de gran utilidad disponer de un gráfico de la serie para decidir sobre la estacionariedad. En ocasiones, se utilizan medias y desviaciones típicas por sub periodo para juzgar sobre la estacionariedad de la serie.
- **Transformación previa de la serie:** Cuando la serie no es estacionaria en varianza se requiere una transformación logarítmica. No obstante, la transformación logarítmica es muy frecuente incluso en series con dispersión relativamente constante en el tiempo. Una práctica habitual es ensayar con la serie original y en logaritmos y comprobar resultados.
- **Eliminación de la tendencia:** La observación del gráfico de la serie indica la existencia o no de tendencia. Una tendencia lineal será corregida tomando primeras diferencias, que será el caso más frecuente. Una tendencia no lineal suele llevar en la práctica al uso de dos diferencias como mucho.
- **Identificación del modelo:** Consiste en determinar el tipo de modelo más adecuado, esto es, el orden de los procesos auto regresivos y de medias móviles de las componentes regular y estacional. Técnicamente esta decisión se toma en base a las funciones de auto correlación (FAC) y auto correlación parcial (FAC parcial), tanto en la parte regular como estacional. Es habitual terminar eligiendo entre los procesos más simples AR(1), AR(2), MA(1), MA(2) y ARMA(1,1), tanto en la parte regular como estacional. En caso de duda pueden seleccionarse varios modelos alternativos que serán estimados y contrastados posteriormente, para definir finalmente el modelo adoptado.
- **Estimación de los coeficientes del modelo:** Decidido el modelo, se procede a la estimación de sus parámetros, dado que se trata de un procedimiento iterativo de cálculo, pueden sugerirse valores iniciales.
- **Contraste de validez del modelo:** Se utilizan distintos procedimientos para valorar el modelo o modelos inicialmente seleccionados: contraste de significación de

parámetros, covarianzas entre estimadores, coeficiente de correlación, suma de cuadrados de errores, etc.

- Análisis detallado de los errores: Se tendrán en cuenta las diferencias históricas entre valores reales estimados por el modelo para su valoración final. Hay que verificar un comportamiento no sistemático de los mismos, así como analizar la posible existencia de errores especialmente significativos.
- Selección del modelo: En base a los resultados de pasos anteriores, se decide sobre el modelo adoptado.
- Predicción: El modelo seleccionado se utilizará como fórmula inicial de predicción.



La metodología Box-Jenkins se aplica a series estacionarias (sin tendencia, sin estacionalidad)

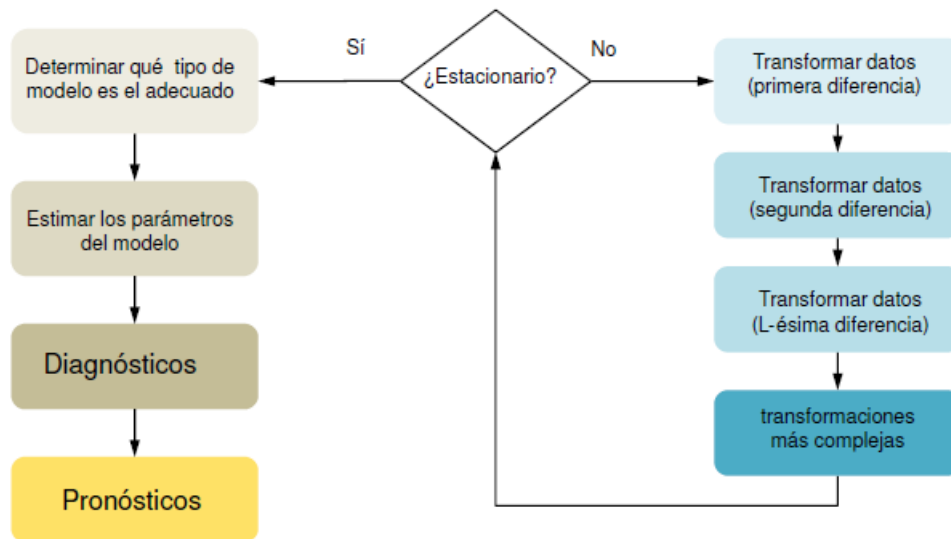


Figura 4. Metodologías de Box-Jenkins. (Box G.E.P. & Jenkins, 1973)

1.4 GESTION HOSPITALARIA

1.4.1 INDICADORES DE GESTION HOSPITALARIA.

Según: Sánchez, Guzmán Mariano I. (2005):

Hoy en día la medición es una herramienta que el ser humano ha utilizado para casi cualquier actividad por muy trivial que ésta sea. En el caso de la alta gerencia, todas las teorías administrativas que se puedan o quieran implementar en cualquier empresa, sin importar su giro, terminan con la medición de su estructura, de sus procesos o de sus resultados. En el caso de la vida hospitalaria, estos indicadores de gestión administrativa se han convertido en una herramienta esencial, porque todo lo que no se mide no puede ser mejorado y por tanto, deja de ser útil para una mejor toma de decisiones. Los indicadores de gestión hospitalaria, así son un instrumento de cambio en la estructura de la institución y miden el cumplimiento de los objetivos institucionales. (Sanchez, 2005)

1.4.1.1 INDICADORES DEL SECTOR SALUD

Los indicadores son los más usados en el sector salud, se pueden distinguir operacionalmente al menos en cuatro tipos:

- 1) Indicadores de Mortalidad: son ampliamente utilizados ya que la muerte es un fenómeno universal, ocurre una sola vez y se registra habitualmente en forma sistemática.
- 2) Natalidad: en este rubro son importantes los indicadores que miden la capacidad de reproducción de una población. Existe una asociación positiva entre altas tasas de natalidad y nivel sanitario, socioeconómico y cultural. Son importantes también, los indicadores que reflejan el estado de la salud materno-infantil, como por ejemplo, la tasa de bajo peso al nacer, la que requiere de un registro confiable de nacimientos (vivos o muertos o defunciones fetales).

- 3) Morbilidad: son indicadores que intentan estimar el riesgo de enfermedad (carga de morbilidad), cuantificar su magnitud e impacto. Los eventos de enfermedad pueden no ser fáciles de definir y pueden prolongarse y repetirse en el tiempo, lo que plantea dificultades en la elaboración de indicadores de morbilidad.
- 4) Calidad de vida: son indicadores generalmente compuestos que intentan objetivar un concepto complejo que considera aspectos como: capacidad funcional de las personas, expectativa de vida, y nivel de adaptación del sujeto en relación con su medio. Son ejemplos sencillos de este tipo de mediciones la "calidad material de la vida" que se construye a partir de la mortalidad infantil, la expectativa de vida al nacer y la capacidad de leer y escribir.

1.4.2 APOYO A LA TOMA DE DECISIONES

Según Lee, McCosh y Migliarese (1988):

Los sistemas de apoyo a la toma de decisiones son herramientas que usan los directivos para tomar decisiones eficaces, y se basan en la teoría de la decisión. Se puede considerar a las herramientas de Minería de Datos como tipos especiales de herramientas de apoyo a la toma de decisiones. Las herramientas de apoyo a la toma de decisiones pertenecen a una amplia categoría (Lee, McCosh, & Migliarese, Decision Support Journal, 1988)

En general, las herramientas de apoyo a la toma de decisiones podrían utilizarse también como herramientas para eliminar los resultados innecesarios e irrelevantes obtenidos de la Minería de Datos. También pueden ser consideradas de este tipo, herramientas tales como las hojas de cálculo, sistemas expertos, sistemas de hipertexto, sistemas de gestión de información de web, y cualquier otro sistema que ayude a analistas y gestores a manejar eficazmente grandes cantidades de datos e información.

Se puede pensar que el apoyo a la toma de decisiones es una tecnología que se solapa con la Minería de Datos, almacenamiento de datos, gestión del conocimiento,

aprendizaje automático, estadística, y otras tecnologías que ayudan gestionar el conocimiento de una organización y los datos.

1.4.3 SISTEMAS DE APOYO A LA TOMA DECISIONES

Los Sistemas de Apoyo a la Toma de Decisiones, son Sistemas de Información que sirven al nivel de dirección para tomar decisiones semi estructuradas, únicas o rápidamente cambiantes y no fácilmente especificadas por adelantado, combinando datos y modelos de análisis, “Apoyar” la toma de decisiones implica el uso de los medios computacionales para ayudar a los directivos en un proceso de decisión, lejos de reemplazar su fallo se trata de perfeccionar la eficacia de la toma de decisión en vez de su eficiencia (Ralph, 1996).

1.4.4 SISTEMA INFORMACION GERENCIAL VS SISTEMA DE APOYO DE DECISIONES.

- **SIG:** Los sistemas de información gerencial, es toda metodología que se usa dentro de la organización, para que la gerencia acceda a cualquier información que esta necesite.
- **SAD:** Los sistemas de apoyo de decisiones, es un sistema de información de la gerencia, cuyo propósito es proporcionar a los gerentes la información necesaria para tomar decisiones inteligentes, en el presente y para el futuro.

1.5 MORBILIDAD

Según la Real Academia Española (2012):

Se le entiende por morbilidad a la cantidad de individuos considerados enfermos o que son víctimas de enfermedad en un espacio y tiempo determinado. La morbilidad es un dato estadístico importante para comprender la evolución o retroceso de alguna

enfermedad, las razones de su surgimiento y las posibles soluciones. En el sentido de la epidemiología se puede ampliar al estudio y cuantificación de la presencia y efectos de alguna enfermedad en una población. (Española, 2012)

1.5.1 TASA DE MORBILIDAD

Las tasas de morbilidad es un concepto epidemiológico que se refiere al número de personas que tienen una enfermedad respecto a una población y periodo determinado.

- Prevalencia: Frecuencia de todos los casos (antiguos y nuevos) de una enfermedad patológica en un momento dado del tiempo o durante un período definido.
- Incidencia: Es la rapidez con la que ocurre una enfermedad. También, la frecuencia con que se agregan nuevos casos de una enfermedad/afección durante un período específico y en un área determinada. (Española, 2012)

Cálculo

La tasa de incidencia se calcula como el cociente entre el número de casos nuevos (incidencia) y el número de habitantes de la población en un período dado:

$$TI = \frac{I}{PT}$$

Dónde:

- TI es la tasa de incidencia
- I es la incidencia (número de casos nuevos)
- PT es la población total expuesta en el período de tiempo T

1.5.2 DIAGNOSTICO EN PACIENTES

Según: Dr. Juan Gervas.

El diagnóstico es un juicio clínico sobre el estado psicofísico de una persona; representa una manifestación en respuesta a una demanda para determinar tal estado. Diagnosticar es dar nombre al sufrimiento del paciente. (Gervas, 2010)

Según: David Ben y K. Sarosi (2010).

El diagnóstico de la apendicitis se basa en la exploración física y en la historia clínica, complementado con análisis de sangre, orina y otras pruebas como las radiografías simples de abdomen, de pie y en decúbito, ecografía abdominal y TAC abdominal simple entre otros. La exploración física se basa en la palpación abdominal. La presentación de la apendicitis suele comenzar como un dolor repentino localizado en la mitad inferior de la región epigástrica o región umbilical, caracterizado por ser un dolor de moderada intensidad y constante al que ocasionalmente se le superponen calambres intermitentes. (Ben & Sarosi, 2010)

1.5.3 CIE-10

Es el acrónimo de *CLASIFICACION INTERNACIONAL DE ENFERMEDADES*, determina la clasificación y codificación de las enfermedades y de una amplia variedad de signos, síntomas, hallazgos anormales, causas externas de daños y/o enfermedad.

Capítulo I

Tabla 1 Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud

Cap	Código	Título
I	A00-B99	Ciertas enfermedades infecciosas y parasitarias
II	C00-D48	Neoplasias
III	D50-D89	Enfermedades de la sangre y de los órganos hematopoyéticos y otros trastornos que afectan el mecanismo de la inmunidad
IV	E00-E90	Enfermedades endocrinas, nutricionales y metabólicas
V	F00-F99	Trastornos mentales y del comportamiento
VI	G00-G99	Enfermedades del sistema nervioso
VII	H00-H59	Enfermedades del ojo y sus anexos
VIII	H60-H95	Enfermedades del oído y de la apófisis mastoides
IX	I00-I99	Enfermedades del sistema circulatorio
X	J00-J99	Enfermedades del sistema respiratorio
XI	K00-K93	Enfermedades del aparato digestivo
XII	L00-L99	Enfermedades de la piel y el tejido subcutáneo
XIII	M00-M99	Enfermedades del sistema osteomuscular y del tejido conectivo
XIV	N00-N99	Enfermedades del aparato genitourinario
XV	O00-O99	Embarazo, parto y puerperio
XVI	P00-P96	Ciertas afecciones originadas en el periodo perinatal
XVII	Q00-Q99	Malformaciones congénitas, deformidades y anomalías cromosómicas
XVIII	R00-R99	Síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte
XIX	S00-T98	Traumatismos, envenenamientos y algunas otras consecuencias de causa externa
XX	V01-Y98	Causas externas de morbilidad y de mortalidad
XXI	Z00-Z99	Factores que influyen en el estado de salud y contacto con los servicios de salud
XXII	U00-U99	Códigos para situaciones especiales

CAPITULO II: MATERIALES Y METODOS

2.1 MATERIALES.

INSTRUMENTOS DE RECOLECCIÓN DE DATOS.

- **Entrevistas:** se realizaron entrevistas de tipo no estructuradas, en forma espontaneas al personal de los diferentes servicios del hospital.
- **Fuentes Documentales:** recopilación de información, a través de fichas bibliográficas, estado del arte sobre el tema de tesis.
- **Fuentes Datos:** la recopilación de la información de los registros transaccionales de los últimos 8 años registrados en la base datos del hospital.

SOFTWARE DE APLICACIONES.

- **EIEWS:** Es una herramienta estadística poderosa, de pronóstico y modelación, especialmente útil para realizar análisis econométricos, como modelos de estimación y predicción con modelos de series de tiempo. Utilizado por grandes investigadores, empresas y entes gubernamentales. Interfaz gráfica intuitiva y de fácil uso.
- **QLIK VIEW:** QlikView es una herramienta de Business Intelligence (BI o Inteligencia de negocio) que le permite generar cuadros de mandos, informes y gráficas con las que puede interactuar a tiempo real. Gracias a ello, podrá conocer cómo va su negocio al instante y generar informes y todo tipo de gráficas perfectamente detalladas, según sus necesidades.
- **SQL SERVER - Analysis Services:** Analysis Services proporciona una plataforma integrada para las soluciones que incorporan la minería de datos, permitiéndonos realizar ETL entre distintas fuentes de datos, además utilizar datos relacionales o de cubo para crear soluciones de Business Intelligence para su análisis.

2.2 METODOLOGIA.

La presente investigación se desarrolló basándose en la metodología ARIMA para el proceso de Modelamiento y siguiendo la metodología CRISP-DM.

METODOLOGIA CRISP-DM:

Tiene como propósito **construir variables** que sirva como fuente de información para crear el modelo, siguiendo un conjunto de pasos que guíen el proceso que se debe seguir, como la metodología CRISP-DM que comprende las siguientes fases:

1. COMPRESION DEL NEGOCIO

- Objetivos del negocio
- Evaluación de la situación

2. COMPRESION DE LOS DATOS

- Entender el problema existente en la información transaccional, analizándola y seleccionando los campos pertinentes de las tablas seleccionadas

3. PREPARACION DE LOS DATOS

- ETL extracción, transformación y carga de los datos de la muestra seleccionada.
- Limpiar los datos de la muestra seleccionada.
- Diseñar el Esquema Dimensional del Datamart.
- Llevar la muestra hacia un modelo dimensional.
- Selección de Atributos para el análisis del algoritmo.

4. MODELADO

- Identificación del posible modelo
- Estimación del modelo
- Diagnóstico del modelo
- Pronostico del modelo

METODOLOGIA ARIMA:

Tiene como propósito **crear un modelo predictivo de minería de datos** basado en pronósticos de series de tiempo MODELO ARIMA de **Box y Jenkins** (Box G.E.P. & Jenkins, 1973) a partir de los datos históricos del hospital Víctor Ramos Guardia.

Esta metodología consta de cuatro fases:

1. PRIMERA FASE:

Consiste en identificar el posible Modelo Arima.

2. SEGUNDA FASE:

Seleccionado provisionalmente un modelo para la serie estacionaria, se pasa a la segunda etapa de estimación.

3. TERCERA FASE:

Es la fase del diagnóstico, donde se comprueba que los residuos no tienen estructura de dependencia y siguen un proceso de ruido blanco.

4. CUARTA FASE:

Es la fase de predicción, una vez obtenido el modelo.

CAPITULO III: RESULTADOS

A continuación se muestra el desarrollo de la metodología CRISP-DM y ARIMA.

3.1 COMPRESION DEL NEGOCIO

El Hospital Víctor Ramos Guardia de Huaraz fue inaugurado el 7 de diciembre de 1963, como “Hospital Centro de Salud de Huaraz”. Proyecto que nació en el periodo presidencial del doctor Manuel Prada Ugarteche, que con el aporte del gobierno Alemán. Para aquel entonces la población de Huaraz era de aproximadamente de 30 mil habitantes. Posteriormente fue rehabilitado y ampliado, con un pabellón de residencia médica, que actualmente es ocupada por la Dirección Regional de Salud. El hospital contaba con 120 camas y servicios básicos de Cirugía, Medicina General, Ginecología, Obstetricia y Pediatría.

Sin embargo a pesar que la población de la provincia de Huaraz supera actualmente los 160 mil habitantes, la infraestructura del hospital en el año 2000 fue ampliado con el pabellón de Emergencia y la Unidad de Cuidados Intensivos, aun así resulta insuficiente.

3.1.1 ESTABLECIMIENTO DE LOS OBJETIVOS

- Mejorar la eficiencia de los servicios y la calidad de las prestaciones
- Establecer políticas que faciliten la satisfacción de los pacientes.
- Realizar trabajos de prevención para reducir la tasa de morbilidad en la ciudadanía.
- Pronosticar la tasa de morbilidad durante los próximos 12 meses.

3.1.2 EVALUACION DE LA SITUACION

El Hospital Víctor Ramos Guardia, cuenta con información histórica desde año 2002, información registrada a través de sus sistemas transaccionales que hacen posible su gestión cotidiana, a la vez es sumamente valiosa a la hora de tomar decisiones que necesiten conocimientos sobre datos históricos y tendencias.

- En la actualidad para obtener información sobre tasas de morbilidad, mortalidad, natalidad y otros, se los solicita a todos los servicios de hospitalización como en Maternidad, Pediatría, Cirugía, UCI, Neonatología, etc. la presentación consolidada mensual sobre las atenciones, consumos y diagnósticos.
- Esta información es consolidada por el departamento de informática y presentada en forma de resumen mensual.
- Para programar de la necesidad del personal médico especialistas necesario en los próximos 12 meses se basan en la información del año anterior.
- La programación del personal médico de turno está basado con respecto al mes anterior.
- El hospital no cuenta con información sobre pronósticos de los casos más frecuentes para los próximos 12 meses

3.2 COMPRESION DE LOS DATOS

3.2.1 RECOPIACION INICIAL DE LOS DATOS.

Como parte de la recopilación de los datos se cuenta con las siguientes tablas que se encuentran en una Base Datos SQL SERVER 2008, que se emplearan para el análisis de la información:

3.2.2 DESCRIPCION DE LOS DATOS.

- **EPICRISIS:** Contiene resumen de la historia clínica del paciente al ingresar y salir del hospital.
- **DIAGNOSTICO:** Contiene información de los diagnósticos según CIE10.
- **DIAGNOSTICOS INGRESO:** Contiene información de los diagnósticos del paciente al ingresar al hospital
- **DIAGNOSTICOS EGRESO:** Contiene información de los diagnósticos del paciente al dar de alta al hospital
- **ESTABLECIMIENTO:** Contiene lista de los establecimientos de salud a nivel nacional.
- **PACIENTES:** Contiene datos personales de los pacientes, historia clínica, datos de su familiares, ubicación y otros.
- **DEPARTAMENTO:** Contiene lista de departamento
- **PROVINCIA:** Contiene lista de provincias
- **DISTRITO:** Contiene lista de distritos
- **CENTRO POBLADO:** Contiene lista de centros poblados

3.2.2.1 DIAGNOSTICO INGRESOS

En ella se registran los diagnósticos de los pacientes al ingresar al hospital.

- IdEpicrisis: Llave primaria.
- CodigoCIE10: Código de Clasificación internacional de enfermedades.
- DetalleDxIngresos: Descripción del código de enfermedades.

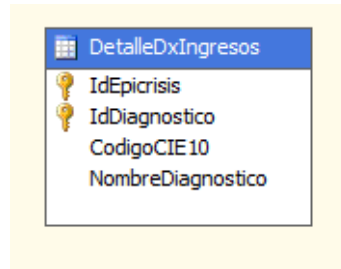


Figura 5. Atributos - Diagnostico Ingresos

3.2.2.2 DIAGNOSTICO EGRESOS

En ella se registran los diagnósticos de los pacientes al salir de alta del hospital.

- IdEpicrisis: Llave primaria.
- CodigoCIE10: Código de Clasificación internacional de enfermedades.
- DetalleDxEgresos: Descripción del código de enfermedades.

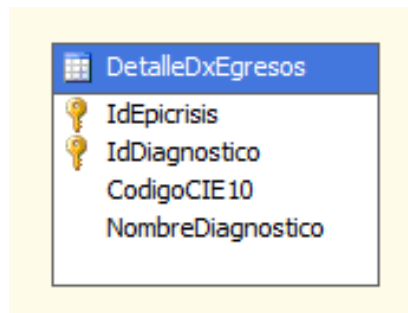


Figura 6. Atributos - Diagnostico Egresos

3.2.2.3 PACIENTES

- HistoriaClinica: Llave primaria.
- NombreCompleto: Nombre completo del paciente
- Departamento: código de ubicación geográfica del departamento
- Provincia: código de ubicación geográfica de la provincia
- Distrito: código de ubicación geográfica del distrito
- CentroPoblado: código de ubicación geográfica del territorio donde residen.
- EstadoCivil: Estado civil del paciente
 - 1 Casado
 - 2 Soltero
 - 3 Conviviente
 - 4 Divorciado
 - 5 Viudo
 - 0 No responde / Ignorado
- GradoInstrucción: Grado de instrucción del paciente
 - 0 No Responde
 - 1 Sin Instrucción
 - 2 Inicial o Preescolar
 - 3 Primaria Incompleta
 - 4 Primaria Completa
 - 5 Secundaria Incompleta
 - 6 Secundaria completa
 - 7 Superior Universitaria Incompleta
 - 8 Superior Universitaria Completa
 - 9 Superior No Universitaria Incompleta
 - 10 Superior Universitaria Completa
- Sexo: Formato texto
 - 1 Masculino
 - 2 Femenino
- FechaNacimiento: Formato Date.

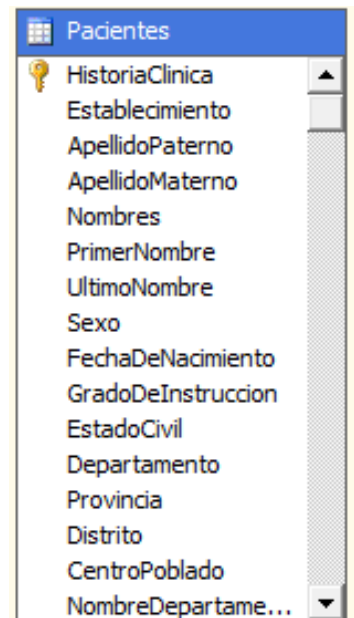


Figura 7. Atributos - Pacientes

3.2.2.4 UBICACION GEOGRAFICA

En ella se encuentra la ubicación geográfica de los pacientes.

- Departamento: Código de ubicación geográfica del departamento
- DepartamentoNombre: Nombre del departamento
- DepartamentoPoblacion: Cantidad de pobladores del departamento.
- Provincia: código de ubicación geográfica de la provincia
- ProvinciaNombre: nombre de la provincia
- ProvinciaPoblacion: Representa la cantidad de pobladores de la provincia.
- Distrito: código de ubicación geográfica del distrito
- DistritoNombre: nombre del distrito
- DistritoPoblacion: Cantidad de pobladores del distrito..
- CentroPoblado: Código de ubicación geográfica del territorio donde residen.
- CentroPobladoNombre: nombre centro poblado

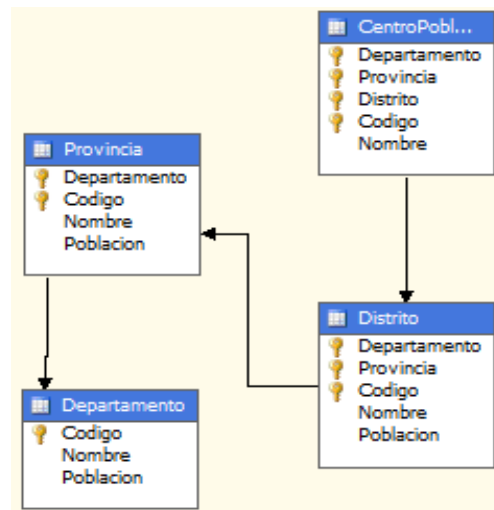


Figura 8. Atributos - Ubicación geográfica

3.2.2.5 EPICRISIS

En ella se encuentra la información del paciente al momento de ingresar al hospital y al momento de salir de alta del paciente.

- IdEpicrisis: Llave primaria
- Establecimiento: Código del establecimiento Minsa donde se atendió
- HistoriaClinica: Código autogenerado del paciente
- FechadeIngreso: Fecha de ingreso al establecimiento
- HoradeIngreso: Hora de ingreso al establecimiento
- NrodeCamadeIngreso: número de la cama del paciente
- Edad en Años Cumplidos: Edad del paciente en años
- Periodo: Periodo de ingreso del paciente en el formato año mes “YYYYMM”
- Vía de Admisión: Es la vía de ingreso del paciente
 - 1 Consulta Externa
 - 2 Referencia
 - 3 Emergencia
- Categoría Socio Económica: Es el nivel o estatus socio económico del paciente, que se clasifica por lo general en 3 categorías.
 - 1 Alto
 - 2 Medio
 - 3 Bajo
- Condición Alta: Es el resumen de la alta médica del paciente.
 - 1 Paciente Fallecido
 - 2 Paciente Curado
 - 3 Paciente mejorado
 - 4 Paciente Inalterado
- Tipos Alta: Es la modalidad de la alta médica del paciente.
 - 1 Alta por indicación medica
 - 2 Alta voluntaria
 - 3 Fuga

Capítulo III

- 4 Alta por referencia
 - 5 Refuerzo de vacuna
- Tipos de Financiamiento: Tipos de financiamiento del paciente.
 - 1 Paciente Normal y/o paciente común
 - 2 Paciente de Aseguradora Particular(Rímac, Pacifico etc)
 - 3 Paciente de Seguro Nacional (Sis – Soat).
 - 4 Paciente Particular y/ o Clínica.
 - 5 Paciente de Convenios (Essalud, Fospolis)
- Tipos de Seguro: Tipos de seguro de afiliación del paciente.
 - 1 Sis – Seguro Integral de Salud
 - 2 Soat – Seguro Obligatorio de accidentes de transito
- FechadeEgreso: Fecha de egreso del establecimiento del paciente
- HoradeEgreso: Hora de egreso del establecimiento del paciente
- NrodeCamadeEgreso: número de la cama al egresar el paciente
- EstanciaTotal: Estancia del paciente en horas
- EstanciaMasDe24: Indicador del paciente con más de 24 horas de estancia.
- EstanciaMasDe48: Indicador del paciente con más de 48 horas de estancia.
- TotaldeDiasdelPaciente: Estancia en Número de días del paciente
- DiagnosticosPaciente: Código CIE10 del diagnóstico del paciente
- NumeroDiagnosticos: Cantidad de Diagnósticos

3.3 PREPARACION DE LOS DATOS.

3.3.1 EXTRACCION DE LOS DATOS.

En este primer paso se consistió en extraer las tablas seleccionadas del sistema transaccional a la nueva base datos para la creación del datamart.



Figura 9. Extracción de Datos

3.3.2 TRANSFORMACION DE LOS DATOS.

En esta etapa, los datos existentes en la DSA son procesados con la finalidad de ser llevados a los formatos utilizados en el repositorio, como también para poder extraer información, en esta fase los datos fuentes requieren de una pequeña manipulación en los datos.

Se han seleccionado solo columnas o campos necesarios para su carga.

En los datos origen se almacenan en el campo sexo “H” para hombre, “M” para mujer, pero el destino tiene que guardar 1 para hombre y 2 para mujer.

En el campo vía de admisión se almacén “C” para consulta “H” para hospitalización y “E” para emergencia tiene que guardar 1 consulta externa, 2 hospitalización 3 emergencia.

En el campo categoría socioeconómica de los datos fuentes se almacenan A, B, C en el destino tiene que guardar 1, 2, y 3 respectivamente. Figura 10 y 11.

3.3.3 CARGA DE LOS DATOS.

Se trata de introducir los datos, ya adaptados al formato deseado. En algunos casos se sobre escribe la información antigua con la nueva, mientras que en otros se guarda un historial de cambios que permite consultas retrospectivas en el tiempo, así como revertir modificaciones. Para la carga masiva de datos suele ser necesario desactivar temporalmente la integridad referencial de la base de datos destino. Figura 12,13, 14, y 15.

TRANSFORMACION DE LOS DATOS

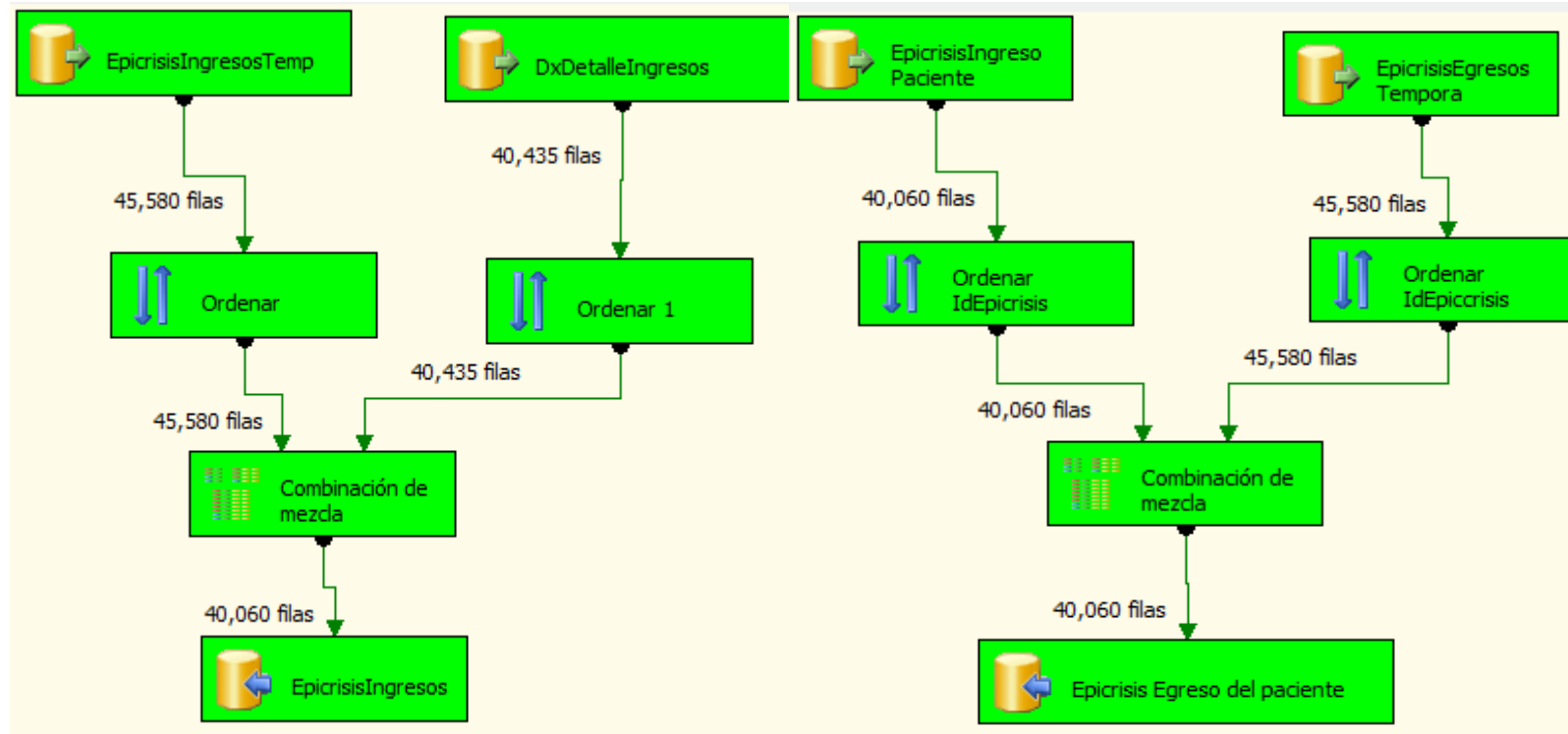


Figura 10. Transformación de Ingresos de Pacientes con diagnósticos

Figura 11. Transformación, Egresos de Pacientes

CARGA DE DATOS

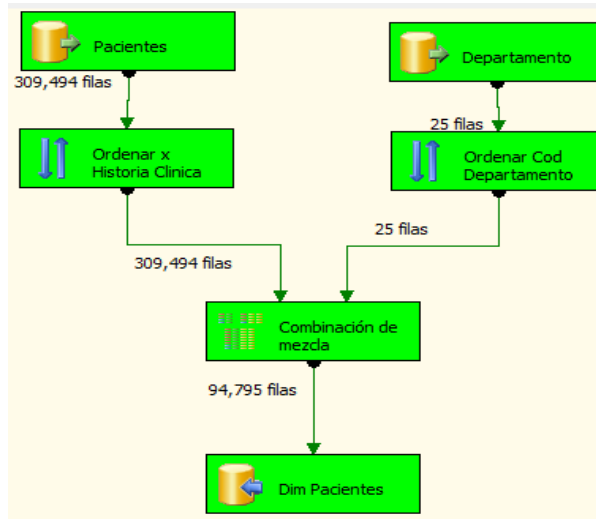


Figura 12. Carga de Pacientes

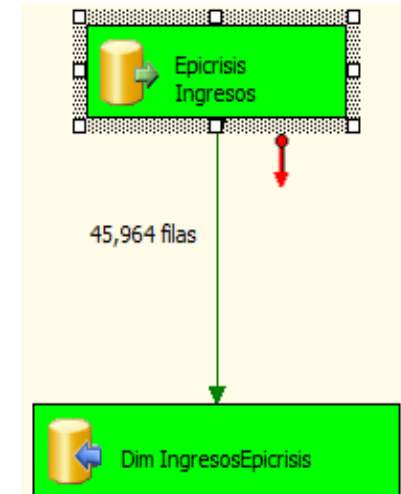
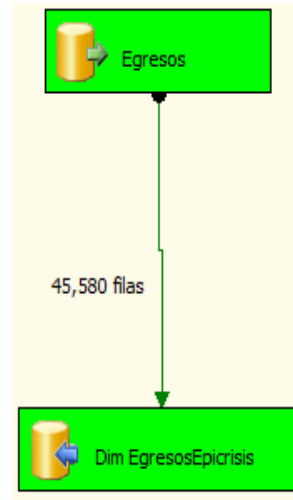


Figura 13. Ingresos Epicrisis, Egresos Epicrisis

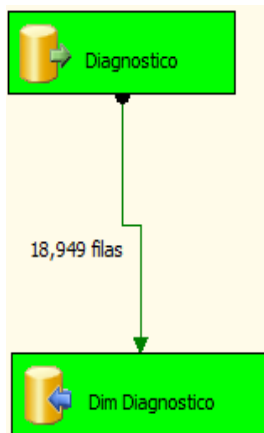


Figura 14. Carga Diagnósticos

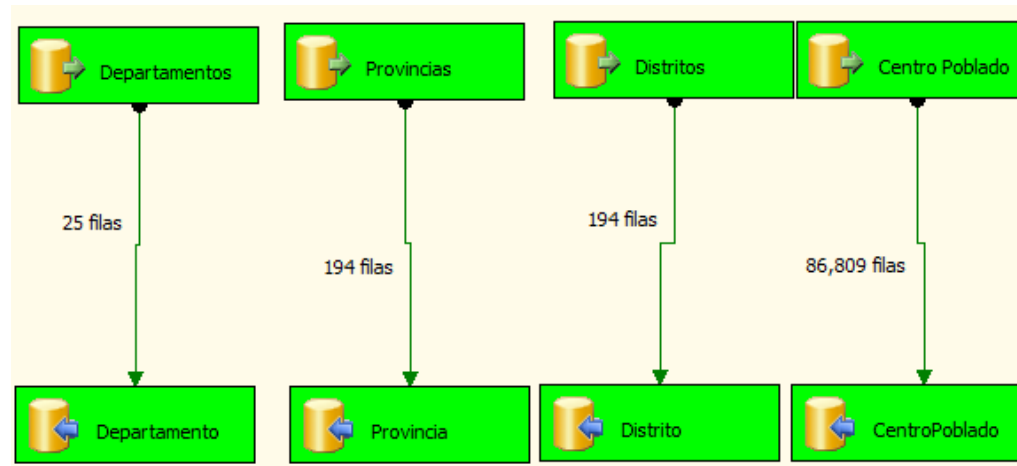


Figura 15. Carga Ubicación Geográfica

3.3.4 CREACION DEL DATAMART

Creación del DATAMART “HEALTHMINING” es con el propósito de almacenar la información histórica, para poder ser agrupados, explorados y analizados

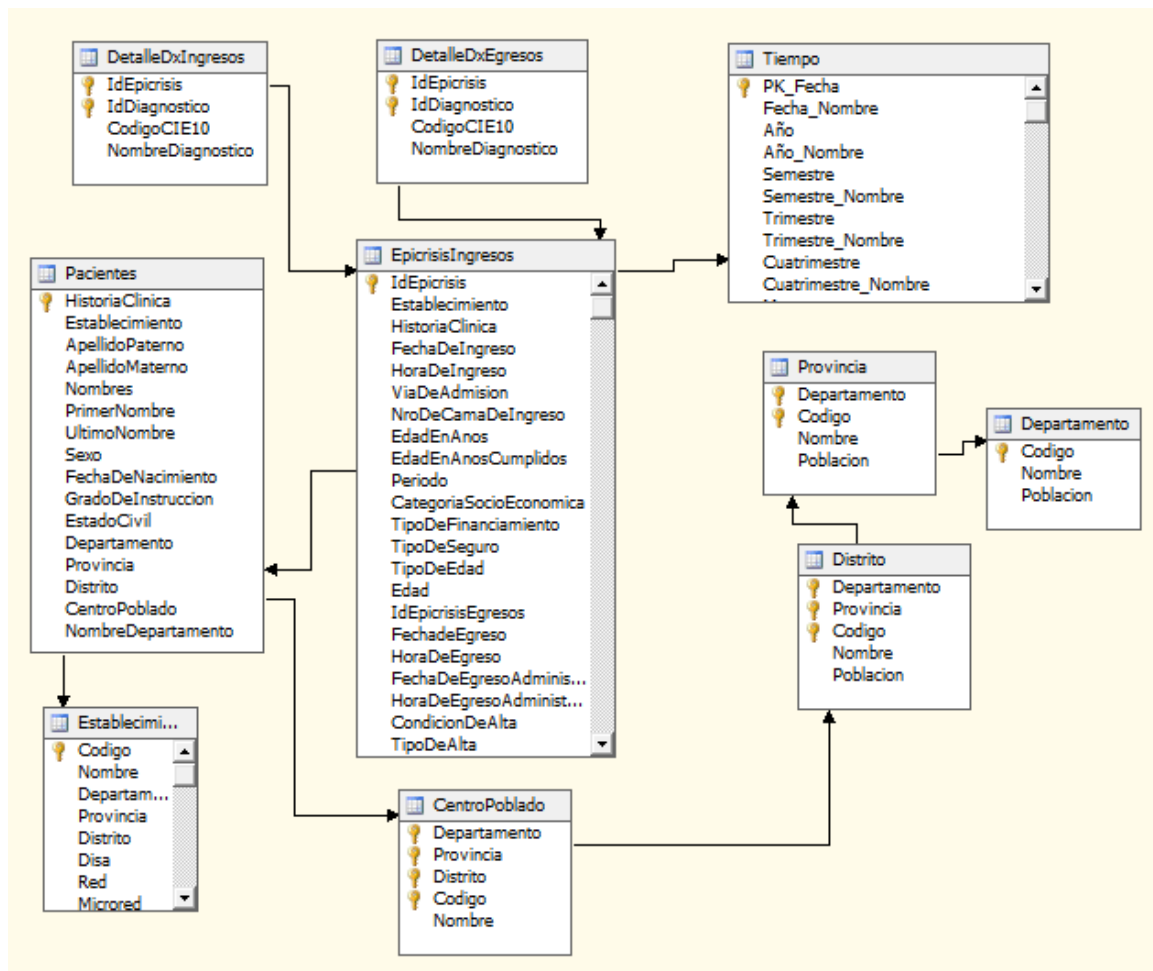


Figura 16. Datamart - “HEALTHMINING”

3.3.5 LIMPIAR LOS DATOS.

Este paso es el que tomo más tiempo, porque hay que verificar que todos los campos estén correctamente ingresados, que no existan datos nulos, que no contengan fechas erróneas o mal ingresadas etc.

Es por eso que hay que verificar minuciosamente creando consultas con Transact-Sql., para detectar campos de fechas fuera del rango, fechas nulas, fechas incorrectas, fechas mal calculadas, campos de tipos numéricos que no sean negativos, que no contengan datos nulos, etc.

SCRIPT PARA LA LIMPIEZA y FORMATEO DE LOS DATOS

```
-- creacion de campo clave iddiagnostico
update [DetalleDxEgreso]
set
  IdDiagnostico = ( select IdDiagnostico
                   from [Diagnostico] a where a.codigodiagnostico =
                   [DetalleDxEgreso].diagnostico )

go
-- Limpieza de campos Nulos
DELETE
FROM [DetalleDxEgreso]
where [IdDiagnostico] is null

go
-- Limpieza de ingresos duplicados
select distinct IdEpicrisis,iddiagnostico
into [DetalleDxEgresos]
FROM [DetalleDxEgreso]

-- Creacion de clave primaria tabla [DetalleDxEgresos]
Alter table [DetalleDxEgresos]
ADD Primary Key (IdEpicrisis,IdDiagnostico)

-- /// -----
-- creacion de campo clave iddiagnostico
update [DetalleDxIngresos]
set
  IdDiagnostico = ( select IdDiagnostico
                   from [Diagnostico] a where a.codigodiagnostico =
                   [DetalleDxIngresos].diagnostico )

go
-- verifica si hay nulos
select COUNT(*)
FROM [DetalleDxIngresos]
where [IdDiagnostico] is null
```

```
-- Limpieza de campos Nulos
DELETE
FROM [DetalleDxIngresos]
where [IdDiagnostico] is null

go
-- Limpieza de ingresos duplicados
select distinct IdEpicrisis,iddiagnostico
into [DetalleDxIngresos]
FROM [DetalleDxIngresos]

-- Creacion de clave primaria tabla [DetalleDxIngresos]
Alter table [DetalleDxIngresos]
ADD Primary Key (IdEpicrisis,IdDiagnostico)

go
--- Aseguramos que exista un registro de Historial de ingresos del
paciente
--- por cada historia de salida del paciente
--- puede existir pacientes que todavia se encuentren
hospitalizados por lo tanto no se conoce su
--- historial de como esta saliendo el paciente

select *
into EpicrisisIngresos
from IngresosEpicrisis a
inner join EpicrisisEgresos b on
a.IdEpicrisisIngresos=b.IdEpicrisisEgresos

-- Creacion de clave primaria tabla [EpicrisisIngresos]
Alter table [EpicrisisIngresos]
ADD Primary Key (IdEpicrisis)

go

drop table IngresosEpicrisis
---

select *
into EpicrisisIngresos
from EpicrisisIngresosXX a
--inner join EpicrisisIngresos b on a.IdEpicrisis=b.IdEpicrisis
--order by 1

---
select a.*
into EpicrisisEgresos
from EpicrisisEgresosXX a
inner join EpicrisisIngresos b on a.IdEpicrisis=b.IdEpicrisis
order by 1
go
-- Creacion de clave primaria tabla [EpicrisisEgresos]
Alter table [EpicrisisEgresos]
ADD Primary Key (IdEpicrisis)
go
Alter table [EpicrisisIngresos]
ADD Primary Key (IdEpicrisis)
```

```
---- transformacion via de admision.
update dbo.EpicrisisIngresos
set ViaDeAdmision = case
    when ViaDeAdmision = 'E' then 3
    when ViaDeAdmision = 'RE' then 2
    when ViaDeAdmision = 'CE' then 1
end

---- transformacion Categoria Socio Economico.
update dbo.EpicrisisIngresos
set CategoriaSocioEconomica= case
    when CategoriaSocioEconomica= 'A' then 1
    when CategoriaSocioEconomica= 'B' then 2
    when CategoriaSocioEconomica= 'C' then 3
end

---- transformacion Condicion de alta medica.
update dbo.EpicrisisIngresos
set CondicionDeAlta = case
    when CondicionDeAlta= 'PF' then 1
    when CondicionDeAlta= 'PAR' then 2
    when CondicionDeAlta= 'PAM' then 3
    when CondicionDeAlta= 'I' then 4
end

---- transformacion Tipos de alta medica.
update dbo.EpicrisisIngresos
set TipoDeAlta = case
    when TipoDeAlta= 'AM' then 1
    when TipoDeAlta= 'AV' then 2
    when TipoDeAlta= 'F' then 3
    when TipoDeAlta= 'RE' then 4
    when TipoDeAlta= 'RV' then 5
end

---- transformacion Tipos de seguro de pacientes.

update dbo.EpicrisisIngresos
set TipoDeSeguro = case
    when TipoDeSeguro= '01' then 1
    when TipoDeSeguro= '02' then 2
    when TipoDeSeguro= '' or TipoDeSeguro= '|' then null
end

---- transformacion Tipos de financiamiento de pacientes.

update dbo.EpicrisisIngresos
set TipoDeFinanciamiento = case
    when TipoDeFinanciamiento= 'N' then 1
    when TipoDeFinanciamiento= 'A' then 2
    when TipoDeFinanciamiento= 'S' then 3
    when TipoDeFinanciamiento= 'P' then 4
    when TipoDeFinanciamiento= 'C' then 5
end

---- transformacion Estado Civil de pacientes.

select * from [Pacientes]
select estadocivil from [Pacientes]
group by estadocivil
```

```
update [Pacientes]
set
estadocivil= 0
where estadocivil is null
```

```
update [Pacientes]
set
estadocivil=
case estadocivil
when '01' then 1
when '02' then 2
when '03' then 3
when '04' then 4
when '05' then 5
when '99' then 0
end
```

3.3.6 SELECCIÓN DE LOS DATOS PARA LA CONSTRUCCION DEL MODELO

A continuación definiremos los atributos a ser evaluados para la construcción del modelo.

- Recogida de datos: Es conveniente disponer de 50 o más datos, y en el caso de series mensuales, trabajar entre seis y diez años completos. Para nuestro caso de estudio contamos con información histórica de 8 años.
 - Fecha: Formato Date
 - Año: Representa al año que ingreso el paciente al hospital por hospitalización.
 - Mes: Representa al mes que ingreso el paciente al hospital por hospitalización.
 - DiagnosticosPaciente: Código CIE10 del diagnóstico del paciente
 - NumeroDiagnosticos: Cantidad de Diagnósticos

3.4 MODELADO.

3.4.1 SELECCIÓN DE LA TECNICA DE MODELADO

Para la creación del modelo de predicción vamos a tomar en cuenta el algoritmo de análisis de serie de tiempo **ARIMA** de *George Box* y *Gwilym Jenkin*. Tomando como datos de entrada el Datamart HEALTH MINING.

Para la utilización de la metodología **ARIMA** es necesario que la serie de la muestra sea **estacionaria en varianza y en media**, esto quiere decir que la serie no tenga tendencia y que presente un grado de dispersión similar. La no estacionalidad en media recibe el nombre de tendencia, cambian a lo largo del tiempo.

Con la ayuda de la herramienta **EIEWS** obtenemos los datos de los 5 diagnósticos más frecuentes de pacientes hospitalizados en el Hospital Víctor Ramos Guardia – Huaraz.

El presente modelo predictivo a modelar será el grupo de diagnóstico (**K00-K95 Enfermedades del Aparato Digestivo como Apendicitis, Peritonitis, Colestitis crónica**).

Por ser el más frecuente durante los últimos 8 años en el hospital Víctor Ramos Guardia.

Tabla 2: Número de diagnósticos más frecuentes de pacientes hospitalizados durante los últimos 7 años (2008-2014), excepto Parto, Embarazo y Puerperio.

Dx.	Descripción	Nro.
K00-K93	Enfermedades del Aparato Digestivo como Apendicitis, Peritonitis, Colestitis crónica etc.	3088
S00-T98	Traumatismo, Envenenamiento y otras consecuencias de causa externa	2839
J00-J99	Ciertas enfermedades del sistema respiratorio	2006
A00-B99	Ciertas enfermedades infecciosas y parasitarias	1514
P00-P96	Ciertas afecciones originadas en el periodo perinatal.	1165

3.4.2 CONSTRUCCION Y EVALUACION DEL MODELO

3.4.3 IDENTIFICACION DEL POSIBLE MODELO

Esta fase nos permite identificar el posible modelo, para esto se cuenta con una muestra de datos procesados en el datamart HEALTH MINING en el capítulo II.

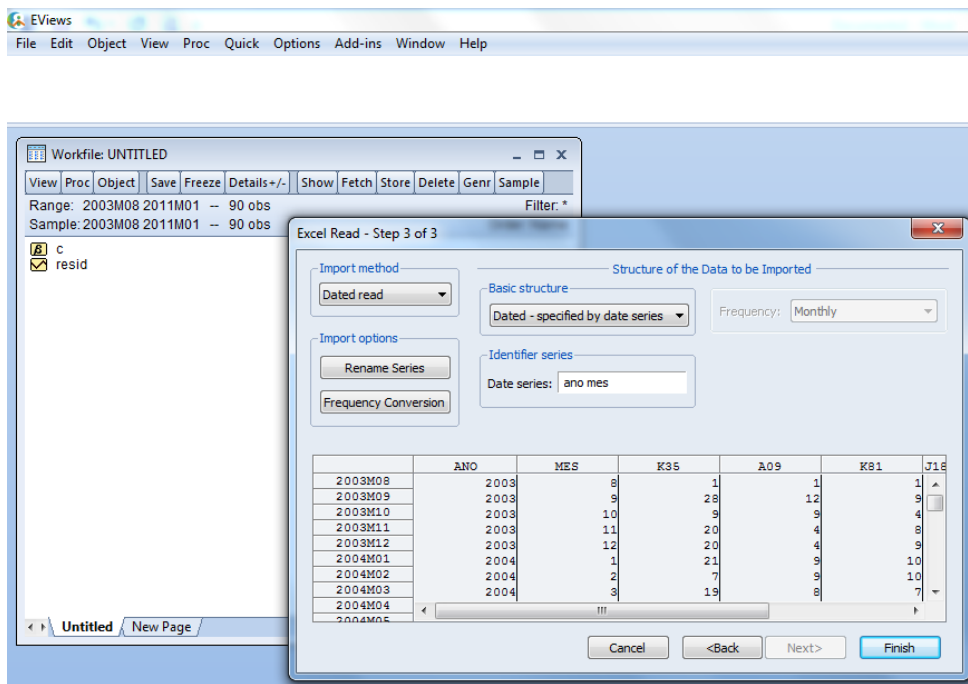


Figura 17. Datos obtenidos del Datamart HEALTH MINING

La fase de identificación comienza realizando una identificación gráfica de la variable con el objetivo de observar la estacionalidad.

En la gráfica del diagnóstico K35 – Apendicitis se observa que la tendencia es creciente, por lo tanto para aplicar la metodología de Box y Jenkins es necesario que la serie sea estacionaria en media y varianza

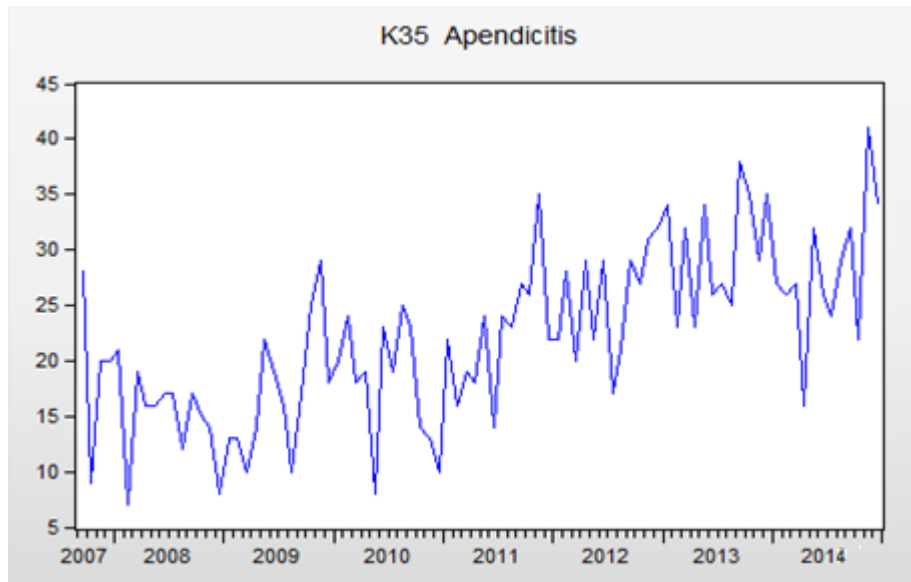
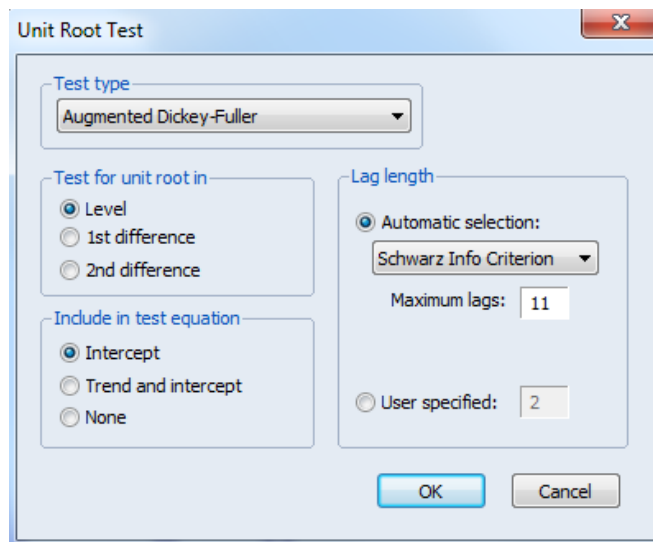


Figura 18. Grafica del Diagnostico K35 Apendicitis con Tendencia creciente y con muchos picos



Para observar si la serie es estacionaria se tendrá que realizar el test de Dickey Fuller.

Figura 19. Procesando el test de Dickey Fuller.

Test de Dickey Fuller:

La prueba de la raíz unitaria de Dickey Fuller nos permite verificar si la serie es o no estacionaria., según Figura 33, podemos concluir lo siguiente.

Si $|t\text{-statistic}| > |VC5\%|$ y la probabilidad de los coeficientes de intercepto y tendencia resulta ser significativo (probabilidad $< 5\%$), rechazamos la hipótesis nula de no estacionaria.

Como $|-3.1698| > |-2.8947|$ y probabilidad de coeficientes e intercepto 0.0252 ($2.52\% < 5\%$), **rechazamos la hipótesis nula de no estacionaria** y se concluye que el **diagnóstico** es estacionario con 11 rezagos a un nivel de confianza del 95%.

Null Hypothesis: K35 has a unit root
 Exogenous: Constant
 Lag Length: 1 (Automatic - based on SIC, maxlag=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.169835	0.0252
Test critical values:		
1% level	-3.506484	
5% level	-2.894716	
10% level	-2.584529	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(K35)
 Method: Least Squares
 Date: 11/08/14 Time: 14:22
 Sample (adjusted): 2003M10 2011M01
 Included observations: 88 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
K35(-1)	-0.324341	0.102321	-3.169835	0.0021
D(K35(-1))	-0.390542	0.096065	-4.065384	0.0001
C	7.170466	2.346625	3.055650	0.0030

R-squared	0.384344	Mean dependent var	-0.170455
Adjusted R-squared	0.369858	S.D. dependent var	7.806896
S.E. of regression	6.197230	Akaike info criterion	6.519578
Sum squared resid	3264.481	Schwarz criterion	6.604033
Log likelihood	-283.8614	Hannan-Quinn criter.	6.553603
F-statistic	26.53207	Durbin-Watson stat	1.957338
Prob(F-statistic)	0.000000		

Figura 20. Resultado de la prueba de raíz unitaria de Dickey-Fuller

3.4.4 ESTIMACION DEL MODELO

Al observar la gráfica del correlograma de la serie, vemos que las barras laterales “rezagos” exceden los límites establecidos o intervalos de confianza. Por lo tanto existe autocorrelación simple que determinan los parámetros MA (media móvil) y autocorrelación parcial que determinan los parámetros AR (auto regresivo).

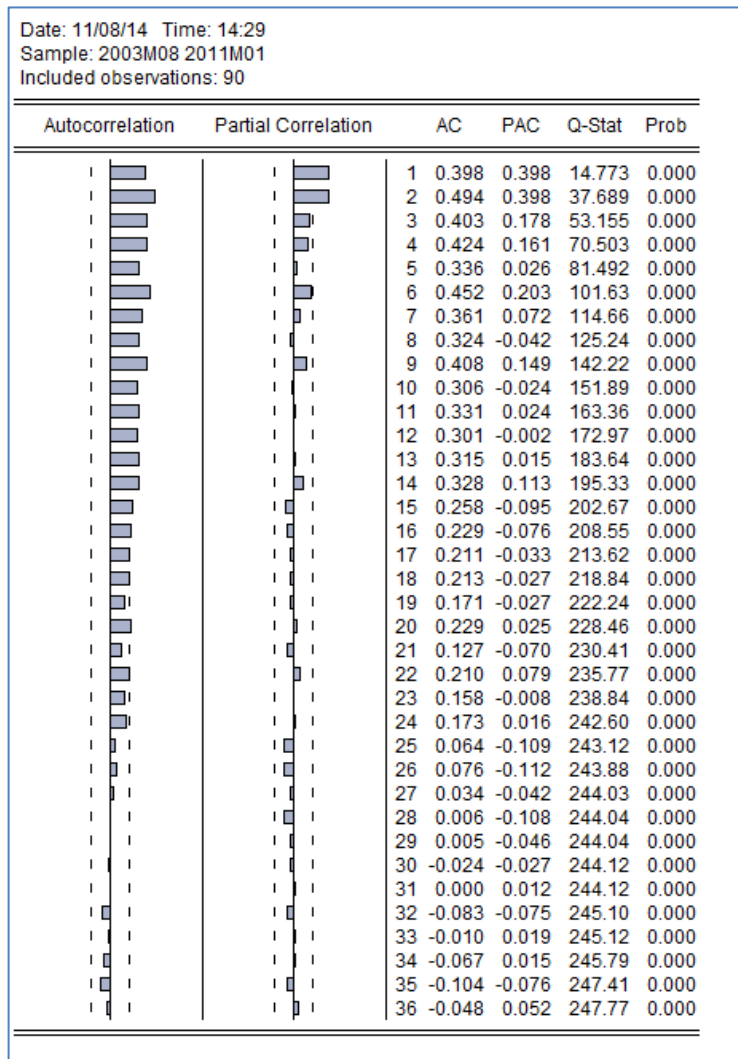


Figura 21. Correlograma. Representación gráfica de la muestra

Estimamos el modelo incluyendo los rezagos mayores al 5% Auto regresivos y media móvil.

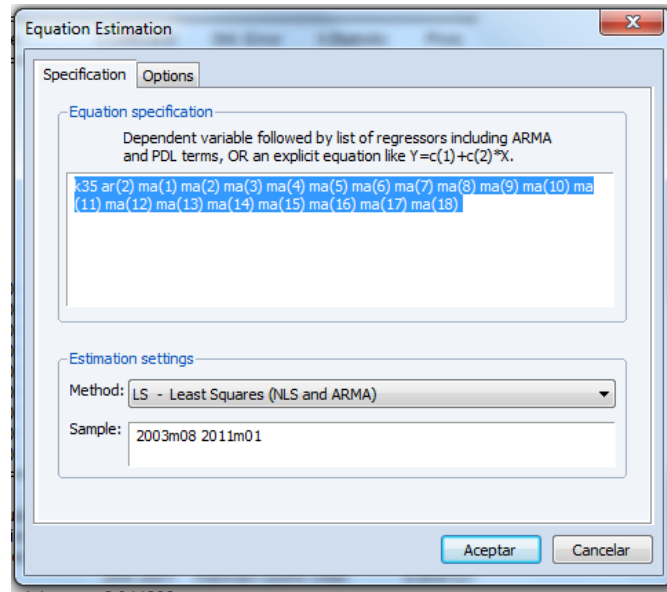


Figura 22. Estimación del Modelo

Capítulo III

La variable esta estacionalizado y depende de un AR(2) MA(1) MA(2) MA(3) MA(4) MA(5) MA(6) MA(7) MA(8) MA(9) MA(10) MA(11) MA(12) MA(13) MA(14) MA(15) MA(16) MA(17) MA(18), pero todos ellos en realidad no son significativos, porque ahora tiene que ver la probabilidad, si la probabilidad es mayor al 5% debemos excluir esos parámetros, empezando por el que tenga mayor probabilidad, eliminando elemento por elemento.

Dependent Variable: K35				
Method: Least Squares				
Date: 11/08/14 Time: 14:36				
Sample (adjusted): 2003M10 2011M01				
Included observations: 88 after adjustments				
Convergence achieved after 30 iterations				
MA Backcast: 2002M04 2003M09				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(2)	0.994460	0.022125	44.94655	0.0000
MA(1)	0.310594	0.124528	2.494175	0.0150
MA(2)	-0.712322	0.134411	-5.299581	0.0000
MA(3)	-0.100379	0.137623	-0.729381	0.4682
MA(4)	-0.095001	0.144171	-0.658946	0.5121
MA(5)	-0.131230	0.104143	-1.260094	0.2119
MA(6)	-0.099806	0.101065	-0.987548	0.3268
MA(7)	0.002042	0.100102	0.020403	0.9838
MA(8)	0.043117	0.101586	0.424436	0.6726
MA(9)	0.181884	0.105195	1.729009	0.0883
MA(10)	-0.105626	0.104077	-1.014887	0.3137
MA(11)	1.67E-05	0.096131	0.000174	0.9999
MA(12)	0.095968	0.106594	0.900306	0.3711
MA(13)	-0.152669	0.103426	-1.476120	0.1445
MA(14)	0.653250	0.103882	6.288385	0.0000
MA(15)	0.415391	0.131088	3.168787	0.0023
MA(16)	-0.493717	0.138185	-3.572865	0.0006
MA(17)	-0.177645	0.136625	-1.300237	0.1978
MA(18)	0.179812	0.119389	1.506100	0.1366
R-squared	0.578649	Mean dependent var	22.01136	
Adjusted R-squared	0.468732	S.D. dependent var	7.555913	
S.E. of regression	5.507366	Akaike info criterion	6.438638	
Sum squared resid	2092.844	Schwarz criterion	6.973517	
Log likelihood	-264.3001	Hannan-Quinn criter.	6.654127	
Durbin-Watson stat	2.044993			
Inverted AR Roots	1.00	-1.00		
Inverted MA Roots	.98-.18i	.98+.18i	.73+.59i	.73-.59i
	.56+.21i	.56-.21i	.43-.89i	.43+.89i
	.02-.98i	.02+.98i	-.39-.92i	-.39+.92i
	-.75-.30i	-.75+.30i	-.76+.62i	-.76-.62i
	-.98-.09i	-.98+.09i		

Figura 23. Probabilidades del primer Modelo estimado

Capítulo III

Volvemos a estimar el modelo, excluyendo las probabilidades mayores al 5%, de forma **iterativa**, hasta determinar que todos los coeficientes o parámetros **auto regresivos y medias móviles** sean significativos.

Dependent Variable: K35 Method: Least Squares Date: 11/08/14 Time: 14:38 Sample (adjusted): 2003M10 2011M01 Included observations: 88 after adjustments Convergence achieved after 36 iterations MA Backcast: OFF (Roots of MA process too large)				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(2)	0.986053	0.015819	62.33189	0.0000
MA(1)	-0.176864	0.076413	-2.314573	0.0232
MA(2)	-0.748879	0.075965	-9.858199	0.0000
MA(14)	0.335234	0.116445	2.878910	0.0051
MA(15)	0.124383	0.120364	1.033397	0.3045
MA(16)	-0.295815	0.144396	-2.048634	0.0438
MA(17)	0.018557	0.122140	0.151930	0.8796
MA(18)	0.286080	0.132255	2.163100	0.0335
R-squared	0.388769	Mean dependent var	22.01136	
Adjusted R-squared	0.335286	S.D. dependent var	7.555913	
S.E. of regression	6.160336	Akaike info criterion	6.560648	
Sum squared resid	3035.979	Schwarz criterion	6.785860	
Log likelihood	-280.6685	Hannan-Quinn criter.	6.651380	
Durbin-Watson stat	1.828029			
Inverted AR Roots	.99	-.99		
Inverted MA Roots	1.02-.15i	1.02+.15i	.79+.43i	.79-.43i
	.64+.60i	.64-.60i	.39+.86i	.39-.86i
	.01-.96i	.01+.96i	-.37+.88i	-.37-.88i
	-.67-.64i	-.67+.64i	-.80+.37i	-.80-.37i
	-.93-.14i	-.93+.14i		
	Estimated MA process is noninvertible			

Dependent Variable: K35 Method: Least Squares Date: 11/08/14 Time: 14:40 Sample (adjusted): 2003M10 2011M01 Included observations: 88 after adjustments Convergence achieved after 22 iterations MA Backcast: 2002M04 2003M09				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(2)	0.996539	0.013932	71.52926	0.0000
MA(1)	-0.011145	0.032223	-0.345885	0.7303
MA(2)	-0.736399	0.050381	-14.61654	0.0000
MA(14)	0.415687	0.071112	5.845548	0.0000
MA(16)	-0.683196	0.050619	-13.49683	0.0000
MA(18)	0.585429	0.061301	9.550048	0.0000
R-squared	0.489082	Mean dependent var	22.01136	
Adjusted R-squared	0.457928	S.D. dependent var	7.555913	
S.E. of regression	5.563080	Akaike info criterion	6.335927	
Sum squared resid	2537.724	Schwarz criterion	6.504836	
Log likelihood	-272.7808	Hannan-Quinn criter.	6.403976	
Durbin-Watson stat	1.641551			
Inverted AR Roots	1.00	-1.00		
Inverted MA Roots	.98-.14i	.98+.14i	.85+.39i	.85-.39i
	.70+.67i	.70-.67i	.39-.91i	.39+.91i
	.00-1.00i	.00+1.00i	-.39-.91i	-.39+.91i
	-.69+.67i	-.69-.67i	-.84+.39i	-.84-.39i
	-.98+.14i	-.98-.14i		

Figura 24. Estimación del modelo anterior excluyendo las probabilidades mayores a 5%

PARAMETROS DEL MODELO FINAL: ar(2) ma(2) ma(16) ma(18)

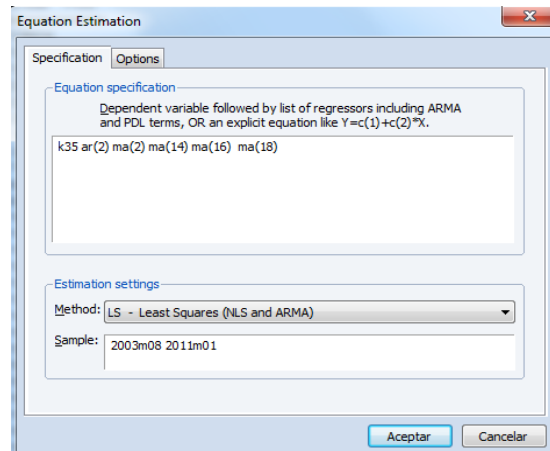


Figura 25. Parámetros del Modelo Final

Dependent Variable: K35				
Method: Least Squares				
Date: 11/08/14 Time: 14:41				
Sample (adjusted): 2003M10 2011M01				
Included observations: 88 after adjustments				
Convergence achieved after 19 iterations				
MA Backcast: 2002M04 2003M09				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(2)	1.013975	0.005458	185.7639	0.0000
MA(2)	-0.857722	0.018407	-46.59743	0.0000
MA(14)	0.103569	0.041167	2.515793	0.0138
MA(16)	-0.847417	0.035754	-23.70099	0.0000
MA(18)	0.832270	0.027422	30.35021	0.0000
R-squared	0.539576	Mean dependent var	22.01136	
Adjusted R-squared	0.517386	S.D. dependent var	7.555913	
S.E. of regression	5.249122	Akaike info criterion	6.209139	
Sum squared resid	2286.923	Schwarz criterion	6.349897	
Log likelihood	-268.2021	Hannan-Quinn criter.	6.265847	
Durbin-Watson stat	1.588689			
Inverted AR Roots	1.01	-1.01		
Estimated AR process is nonstationary				
Inverted MA Roots	.98-.08i	.98+.08i	.90-.39i	.90+.39i
	.70+.70i	.70-.70i	.38-.92i	.38+.92i
	.00+1.00i	.00-1.00i	-.38-.92i	-.38+.92i
	-.70+.70i	-.70-.70i	-.90+.39i	-.90-.39i
	-.98-.08i	-.98+.08i		

Figura 26. Valores y Coeficientes del Modelo Final

Capítulo III

Siendo el modelo estimado con el que nos quedamos:

$$y_t = 1.013975y_{t-2} - 0.857722e_{t-2} + 0.103569e_{t-14} - 0.847417e_{t-16} + 0.832270e_{t-18}$$

(0,005458) (0,018407) (0,041167) (0,035754) (0,027422)

Donde y_t es la serie estacionaria

Entre paréntesis se indica el valor crítico del Std. error

3.4.5 DIAGNOSTICO DEL MODELO

Como se observa los valores del modelo está dentro del intervalo y su error cuadrático medio es 8.937636 (bajo)

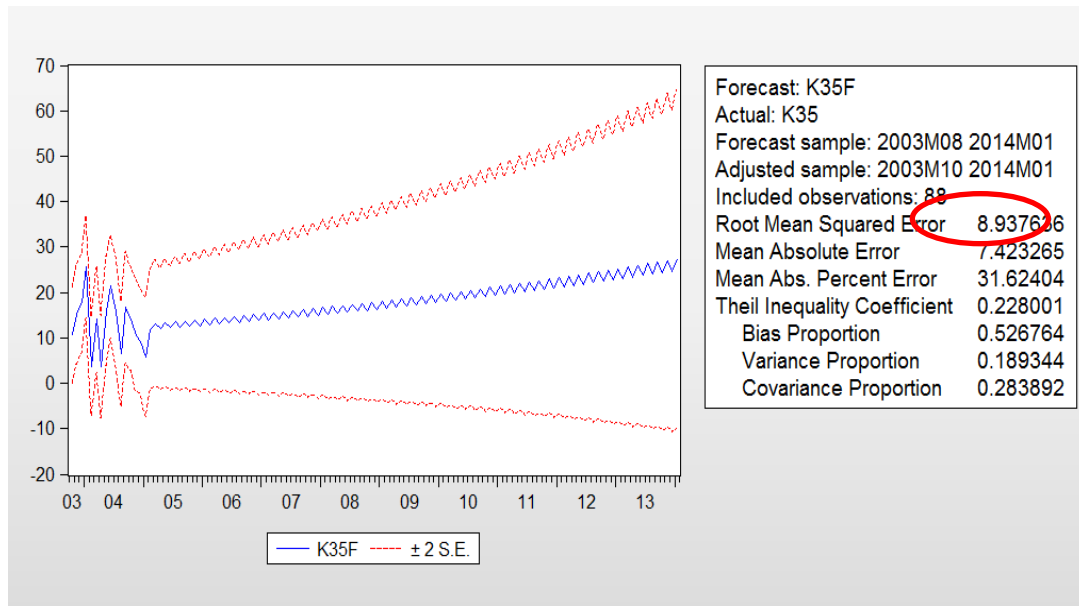


Figura 27. Diagnóstico del Modelo

Podemos observar que la distribución de los errores es normal porque la probabilidad de Jarque – Bera es mayor al 5% y no se rechaza la hipótesis nula de normalidad de errores a una confianza de 95%.

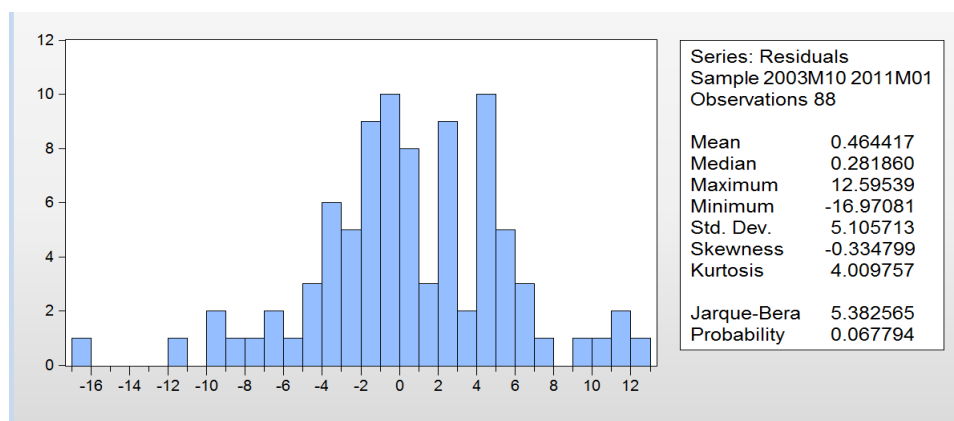


Figura 28. Distribución de errores

Observando el correlograma de los errores afirmamos que todos los errores están dentro de la distribución normal y las correlaciones entre si también están dentro del intervalo de confianza de los errores.

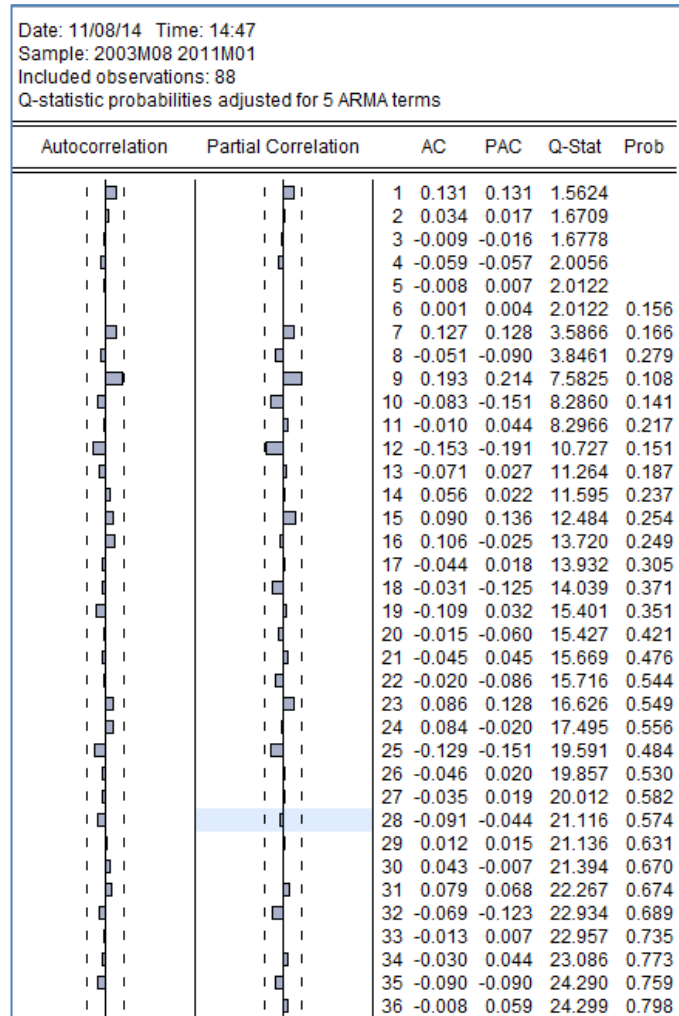


Figura 29. Correlograma del modelo final

3.4.6 PRONOSTICO DEL MODELO

COMPORTAMIENTO PREDICTIVO DE LA MORBILIDAD APENDICITIS

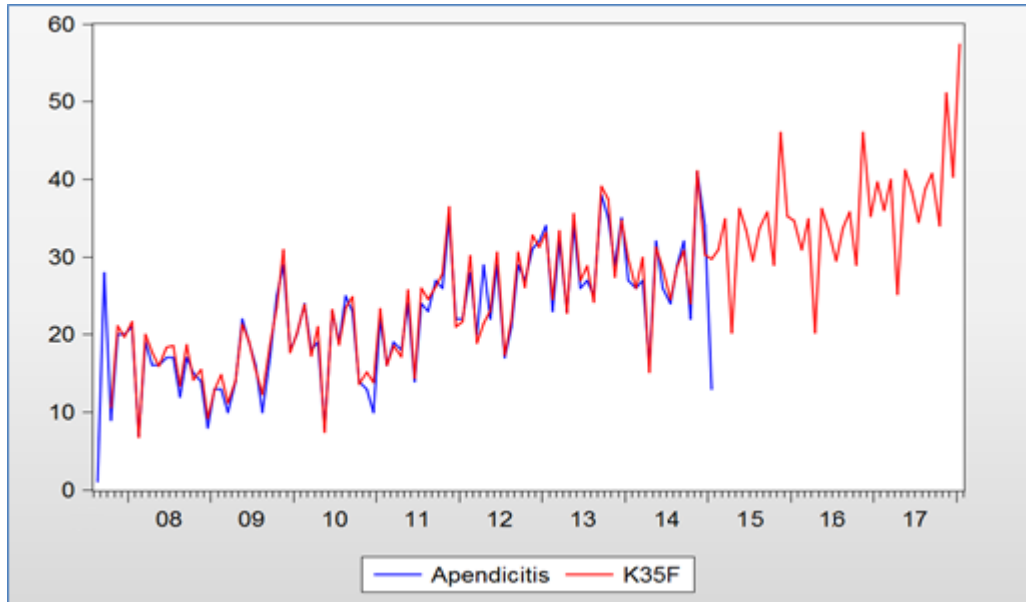


Figura 30. Grafico Línea según Modelo ARIMA proyectado

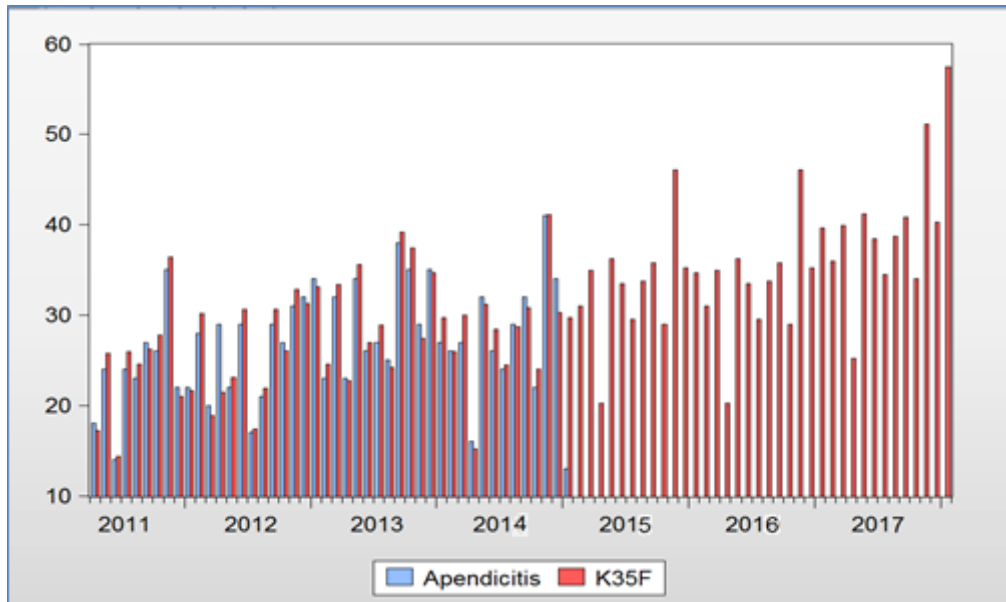


Figura 31. Grafico Barras según Modelo ARIMA proyectado

3.5 EVALUACION.

3.5.1 EVALUACION DE LOS RESULTADOS OBTENIDOS.

Con series de tiempo, me permitió crear un modelo de minería de datos, desde año 2008 al 2017.

Tabla 3: Número de diagnósticos sin modelo vs con modelo (K35 Apendicitis Aguda).

	2008		2009		2010		2011		2012	
Mes	K35	K35F	K35	K35F	K35	K35F	K35	K35F	K00	K35F
Enero	7	3.89	13	14.81	24	19.00	16	19.27	28	25.55
Febrero	19	14.01	10	10.96	18	18.06	19	11.39	20	29.36
Marzo	16	4.28	14	11.55	19	21.30	18	18.09	29	30.85
Abril	16	16.69	22	20.15	8	15.82	24	23.19	22	25.51
Mayo	17	23.88	19	12.77	23	20.56	14	17.24	29	26.67
Junio	17	16.20	16	11.92	19	20.44	24	22.90	17	18.76
Julio	12	7.85	10	5.87	25	18.28	23	23.34	21	24.42
Agosto	17	17.49	17	17.54	23	18.56	27	29.82	29	24.51
Septiembre	15	16.29	25	22.61	14	16.78	26	20.32	27	26.87
Octubre	14	11.21	29	16.40	13	17.54	35	24.97	31	25.01
Noviembre	8	10.97	18	13.92	10	18.51	22	17.55	32	25.44
Diciembre	13	7.46	20	19.97	22	22.23	22	21.45	34	30.64

	2013		2014		2015		2016		2017	
Mes	K35	K35F	K35	K35F	K35	K35F	K35	K35F	K35	K35F
Enero	23	22.19	26	30.85		22.43		27.55		36.64
Febrero	32	20.80	27	27.32		35.45		32.89		28.19
Marzo	23	23.84	16	25.16		30.64		33.25		26.80
Abril	34	29.75	32	30.95		22.19		29.32		25.84
Mayo	26	25.48	26	28.77		20.80		27.16		33.75
Junio	27	38.85	24	25.29		23.84		29.95		29.48
Julio	25	29.76	29	30.17		29.75		31.77		40.85
Agosto	38	32.55	32	29.94		25.48		28.29		32.76
Septiembre	35	25.57	22	28.78		38.85		32.17		30.55
Octubre	29	32.89	41	44.83		29.76		31.94		29.57
Noviembre	35	32.25	34	33.57		32.55		33.78		35.89
Diciembre	27	28.09	13	29.97		25.57		35.94		38.25

Tabla 4: Modelo Propuesto vs método actual 1 y 2

PERIODO	VALOR REAL	PRONOSTICO	M1	M2	% A.P	% A.M1	% A.M2
201404	32.00	30.95	16.00	23.00	96.73%	50.00%	71.87%
201405	26.00	28.77	32.00	25.25	89.36%	76.92%	97.12%
201406	24.00	25.29	26.00	25.40	94.61%	91.67%	94.17%
201407	29.00	30.17	24.00	25.17	95.97%	82.76%	86.78%
201408	32.00	29.94	29.00	25.71	93.57%	90.62%	80.36%
201409	22.00	28.78	32.00	26.50	69.20%	54.55%	79.55%
201410	41.00	44.83	22.00	26.00	90.67%	53.66%	63.41%
201411	34.00	33.57	41.00	27.50	98.73%	79.41%	80.88%
201412	33.00	29.97	34.00	28.09	90.82%	96.97%	85.12%
Nivel de Confianza					91.07%	75.17%	82.14%

VALOR REAL: Numero de diagnósticos atendidos en el periodo

PRONÓSTICO: Numero de diagnósticos pronosticados con el modelo propuesto

M1: Numero de diagnósticos con la metodología actual (método ultimo valor)

M2: Numero de diagnósticos con la metodología actual (método promedio simple)

%A.P: Porcentaje de acierto con el modelo propuesto

%A.M1 y %A.M2: Porcentaje de acierto actual sin el modelo propuesto.

3.5.2 ESTIMACION DE COSTOS Y PRESUPUESTOS.

Tabla 5: Costo promedio mensual por hora y minuto del recurso humano

Grupo Ocupacional	Ingreso promedio Bruto mensual S/. (a)	Jornada laboral Mensual 150 horas = 9000 minuto (b)		Costo promedio por minuto del RH S/. (c)=(a)/(b)	
		No. Horas	No. Minutos	C/hora	C/minuto
Médico Cirujano	4500	150	9000	30.00	0.50
Enfermera	2000	150	9000	13.33	0.22
Auxiliar Técnico	1900	150	9000	12.67	0.21
Técnico Laboratorio	1800	150	9000	12.00	0.20
Técnico Radiología	1800	150	9000	12.00	0.20
Anestesiista	2000	150	9000	13.33	0.22

Fuente: Oficina General de Gestión de Recursos Humanos – Minsa.

Tabla 6: Farmacia – Receta de Material médico y medicamentos.

Material médico / Medicamento	Unidad Medida	Cantidad	Costo	Total
Apósito adhesivo transparente	Und	1	1.05	1.05
Cateter IV No 18 G x 1 1/4"	Und	1	4.00	4.00
Ceftriaxona 1g INY	Amp	8	1.70	13.60
Dimenhidrinato 50 mg INY	Amp	2	1.10	2.20
Equipo de Venoclisis	Und	1	0.68	0.68
Gasa 15 x 42	Und	1	3.00	3.00
Gasa 25 x 42 SOP	Und	4	3.00	12.00
Gasa dressing o gasa grande	Und	2	4.50	9.00
Glucosa al 5% AD x 1 Litro	Fco	3	3.30	9.90
Guantes quirúrgicos no. 7	Par	2	0.75	1.50
Guantes quirúrgicos no. 7 ½	Par	6	0.75	4.50
Jeringa descartable 10cc C/A 21 x 1 1/2	Und	12	0.21	2.52
Llave 3 vías sin extensión	Und	1	0.65	0.65
Metamizol IV IM 1 gr	Und	7	0.28	1.96
Metronidasol 500mg	Amp	5	1.50	7.50
Nylon azul 0 mr 35	Und	2	3.85	7.70
Nylon azul 3/0 tc 30	Und	2	3.85	7.70
Paquete anestesia	Und	1	39.54	39.54
Potasio Cloruro 20% x 10ml	Amp	1	0.28	0.28
Sodio Cloruro al 20% x 20 ml	Amp	3	0.40	1.20
Yodapovidona Espuma x 120 ml	Fco	1	3.70	3.70
Yodapovidona Solucion x 120 ml	Fco	1	4.75	4.75
Total				138.93

El costo promedio por paciente con diagnostico apendicitis, en medicamentos y material médico, recetados en sala de operaciones y después de la operación es: S/. 138.93

Tabla 7: Costo de insumos y material médico por consulta de paciente

Insumos y Material Medico	Unidad Medida	Cantidad que se consume (a)	Forma de Presentación		Precio compra (c)	Costo unitario (d)=(a)x(c)/(b)
			Unidad medida de compra	Equivalencia (b)		
Jabón antiséptico	cc	5	Galón	4000 cc	49.6	0.06
Algodón hidrófilo 500g	gr	2	Rollo	500grs	8.6	0.03
Mascarilla aséptica descartable	unidad	1	unidad	1 und	0.2	0.20
Guante descartable de polietileno	par	1	par	1 und	0.2	0.20
baja lengua de madera	unidad	1	unidad	1 und	0.1	0.10
Costo de insumos y material medico						0.60

Tabla 8: Resumen de Costos por paciente con diagnostico apendicitis

centros de costos			Unidad Medida	Tratamiento		Total
				Costo S/.	Cant.	
Costos Variables (cv)	Ambulatoria y/o Emergencia	Consulta	diario	10.00	1	10.00
		Insumos	diario	0.60	1	0.60
Costos Fijos Directos (cf)	CIRUGIA	Días cama / Estancia Hosp.	diario	30.00	5	150.00
	FARMACIA	Medicamentos	Paciente	138.93	1	138.93
	LABORATORIO	Exámenes	Paciente	30.00	1	30.00
	RADIOLOGIA	Placa Radiográfica 14 x 14	Paciente	31.00	1	31.00
	CENTRO QUIRURGICO	Instrumental medico	Paciente	70.00	1	70.00
	SALA OPERACIONES	Cirugía mayor	Paciente	240.00	1	240.00
Costos Indirectos (ci)	Recurso Humano	Médico Cirujano	Horas	30.00	7	210.00
		Enfermera	Horas	13.33	5	66.67
		Auxiliar Técnico	Horas	12.67	3	38.00
		Técnico Laboratorio	Horas	12.00	1	12.00
		Técnico Radiología	Horas	12.00	1	12.00
		Anestesiista	Horas	13.33	2	26.67
Total						975.86

Calculo del costo por paciente, durante la estancia promedio de 3 días.

$$\text{Costo Total} = cv + cf + ci = S/. 975.86$$

Tabla 9: Cuadro de Costos proyectados con el modelo para los próximos tres años.

Mes	No. Dx. con el Modelo			Costo Total proyectado		
	2015	2016	2017	2015	2016	2017
Enero	22	28	37	21,468.92	27,324.08	36,106.82
Febrero	35	33	28	34,155.10	32,203.38	27,324.08
Marzo	31	33	27	30,251.66	32,203.38	26,348.22
Abril	22	29	26	21,468.92	28,299.94	25,372.36
Mayo	21	27	34	20,493.06	26,348.22	33,179.24
Junio	24	30	29	23,420.64	29,275.80	28,299.94
Julio	30	32	41	29,275.80	31,227.52	40,010.26
Agosto	25	28	33	24,396.50	27,324.08	32,203.38
Septiembre	39	32	31	38,058.54	31,227.52	30,251.66
Octubre	30	32	30	29,275.80	31,227.52	29,275.80
Noviembre	33	34	36	32,203.38	33,179.24	35,130.96
Diciembre	26	36	38	25,372.36	35,130.96	37,082.68
Total	338	374	390	329,840.68	366,987.64	382,602.40

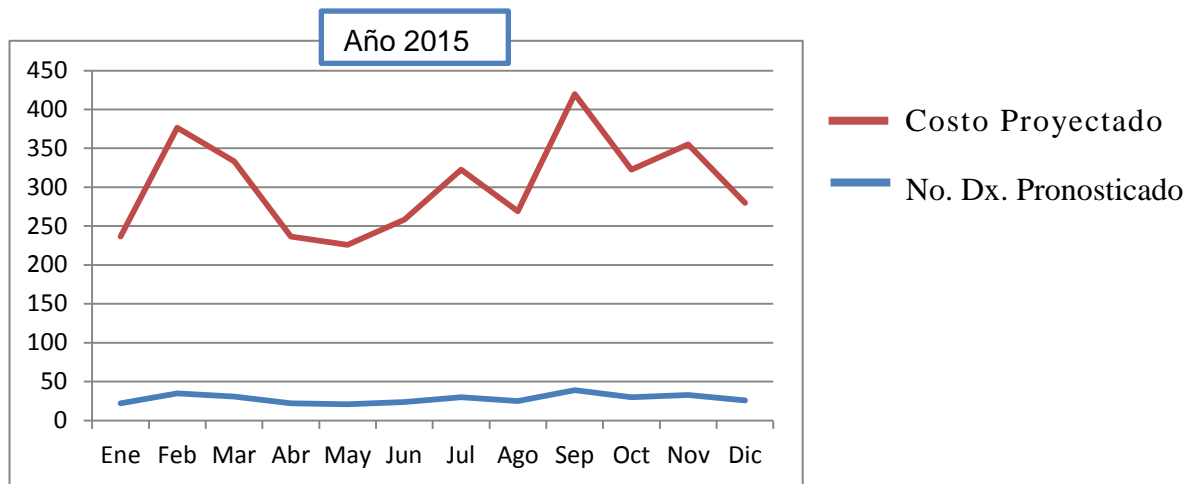


Figura 32. No. Diagnósticos vs Costos proyectado 2015

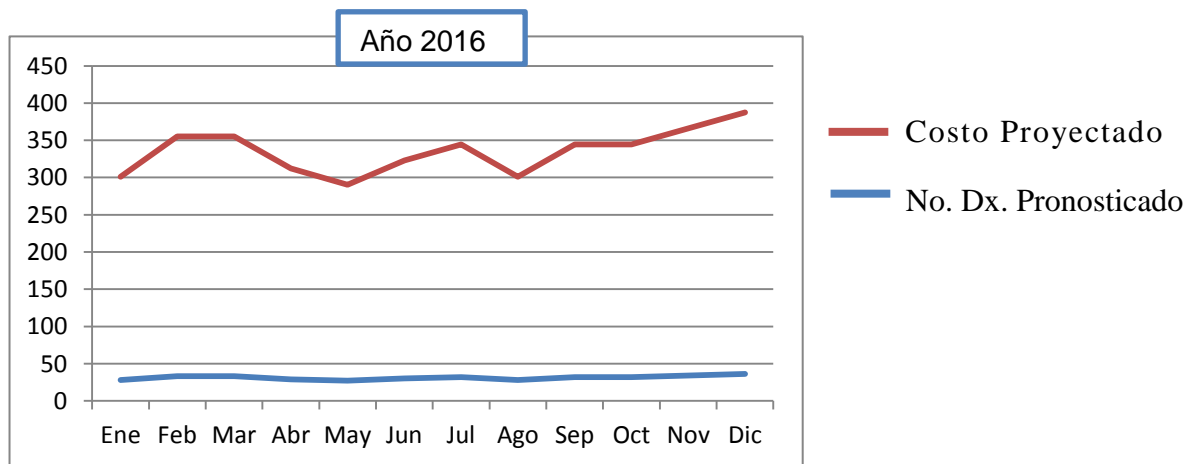


Figura 33. No. Diagnósticos vs Costos proyectado 2016

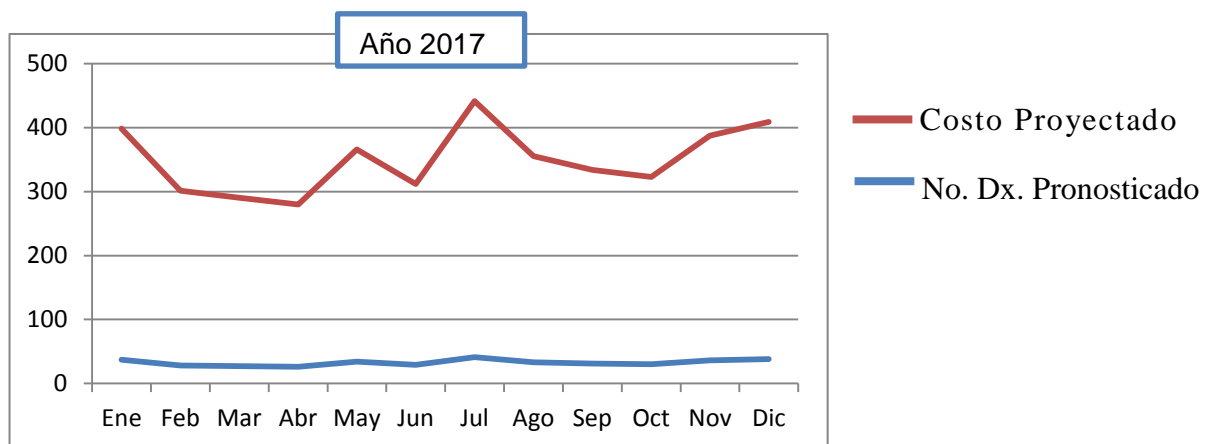


Figura 34. No. Diagnósticos vs Costos proyectado 2017

3.5.3 SALIDA DE LOS RESULTADOS.

Para **crear el modelo predictivo de minería de datos** basado en pronósticos de series de tiempo se recopiló los datos históricos registrados a partir del año 2000, registrados a través del sistema transaccional con Base Datos SQL Server 2008.

- La creación del DATAMART “**HEALTHMINING**” tiene como objetivo servir como fuente de información para el presente y futuros modelos, además que sirva de apoyo a la gestión, monitoreando a través del panel de control.
- Durante el proceso de preparación de los datos, se realizó la extracción, transformación, carga, limpieza, y para luego construir las **variables** necesarias para la creación del modelo.
- En los últimos 7 años (enero 2008- diciembre 2014) se han registrado **31,289** casos de morbilidad en pacientes hospitalizados.

Tabla 10: Número de casos de morbilidad en pacientes hospitalizados (2008-2014)

Código	Diagnostico	NoDx
O00-O99	Embarazo, Parto y Puerperio	15,615.00
K00-K93	Enfermedades del aparato digestivo	3,088.00
S00-T98	Traumatismos, envenenamientos y otras consecuencias de causa externa	2,839.00
J00-J99	Enfermedades del sistema respiratorio	2,006.00
A00-B99	Ciertas enfermedades infecciosas y parasitarias	1,514.00
P00-P96	Ciertas afecciones originadas en el periodo perinatal	1,165.00
N00-N99	Enfermedades del aparato genitourinario	1,138.00
Z00-Z99	Factores que influyen en el estado de salud y contacto con los servicios de salud	858.00
I00-I99	Enfermedades del sistema circulatorio	681.00
C00-D48	Neoplasias	481.00
L00-L99	Enfermedades de la piel y el tejido subcutáneo	427.00
R00-R99	Síntomas, signos y hallazgos anormales clínicos y de laboratorio	413.00
E00-E90	Enfermedades endocrinas, nutricionales y metabólicas	302.00
M00-M99	Enfermedades del sistema osteomuscular y del tejido conectivo	173.00
D50-D89	Enfermedades de la sangre y de los órganos hematopoyéticos	127.00
G00-G99	Enfermedades del sistema nervioso	123.00
F00-F99	Trastornos mentales y del comportamiento	122.00
Q00-Q99	Malformaciones congénitas, deformidades y anomalías cromosómicas	114.00
H60-H95	Enfermedades del oído y de la apófisis mastoides	50.00
V01-Y98	Causas externas de morbilidad y de mortalidad	27.00
H00-H59	Enfermedades del ojo y sus anexos	26.00

En el presente investigación vemos que el caso más común por el que un paciente ingresa a hospitalización es por el diagnóstico **O00-O99 Embarazo, Parto y Puerperio**, siendo el 49.9% de casos registrados, por ser este tipo de diagnóstico no tipificado como un tipo de enfermedad, no se tomó a esta población como caso de estudio para el modelamiento.

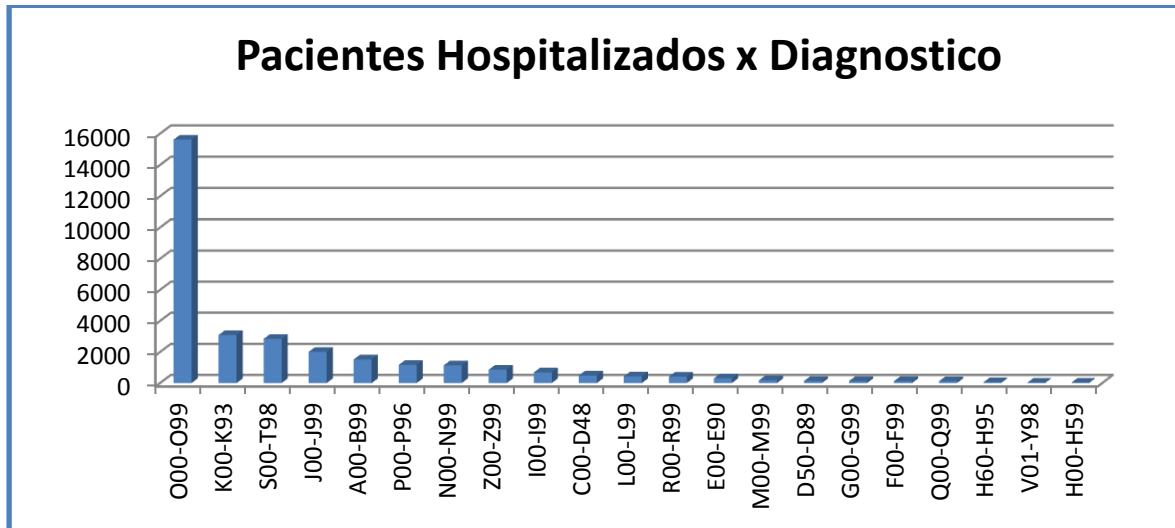


Figura 35. Grafico Barras Pacientes Hospitalizados 2008-2014

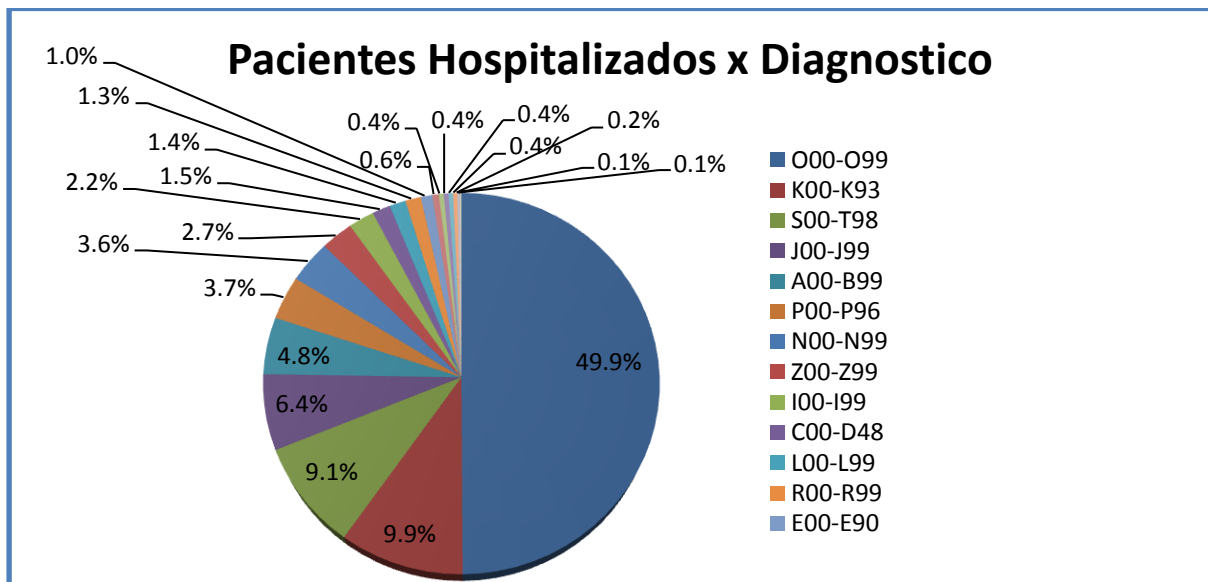


Figura 36. Grafico Número de casos de pacientes hospitalizados 2008-2014

- En la presente grafico **Figura 37** presento los casos de morbilidad en la hospitalización excepto el código del diagnóstico O00-O99 Embarazo Parto y Puerperio, donde se observa que el diagnostico cuyo código es **K00-K93 Enfermedades del Aparato Digestivo como Apendicitis, Peritonitis, Colestitis crónica**, es el de mayor frecuencia, con un 20% de casos en pacientes hospitalizados, por lo tanto se tomó este diagnóstico como caso de estudio para la creación del modelo.

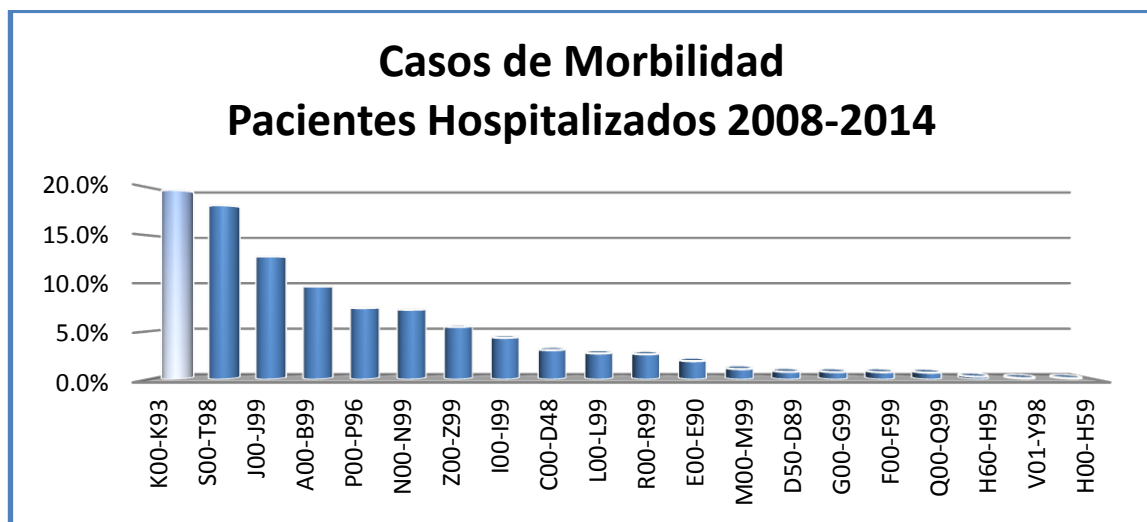


Figura 37. Casos de morbilidad de pacientes, excepto diagnostico O00-O99

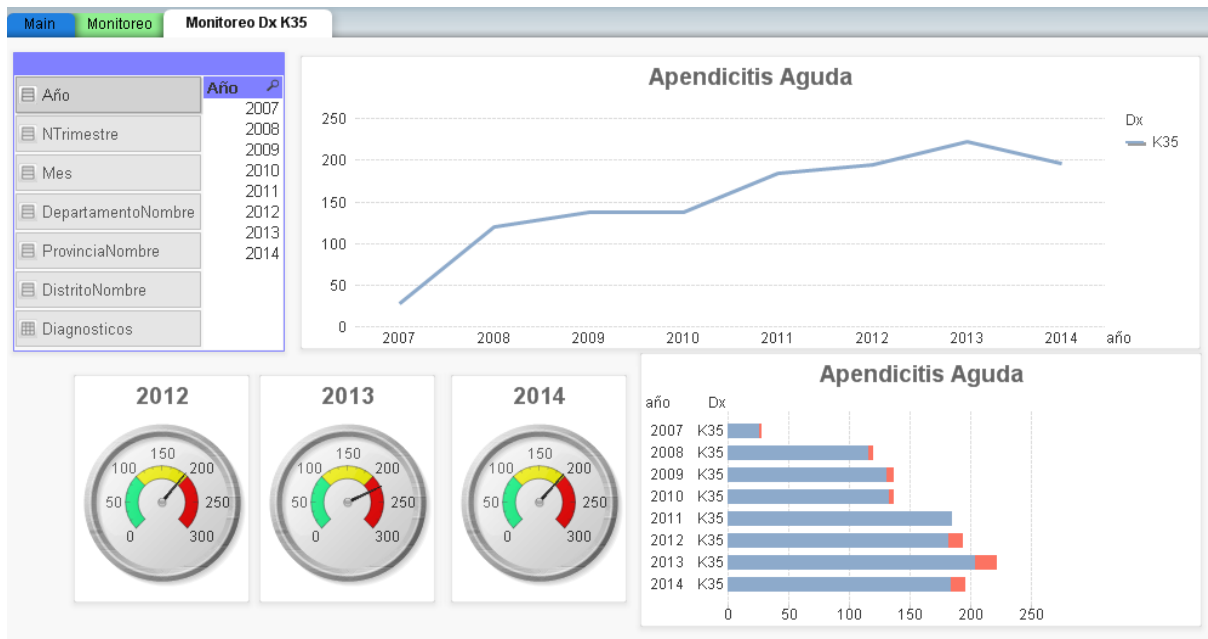


Figura 38. Monitoreo de Diagnostico K35 últimos 7 años

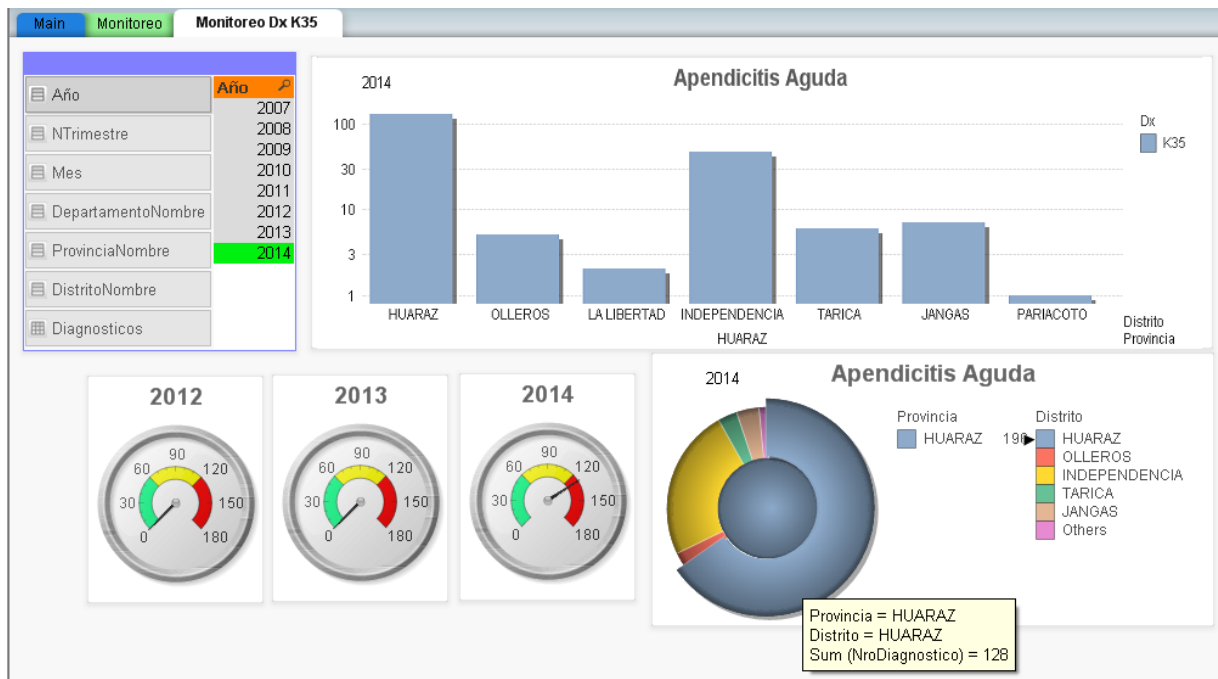


Figura 39. Monitoreo de Diagnostico K35 a nivel distrital

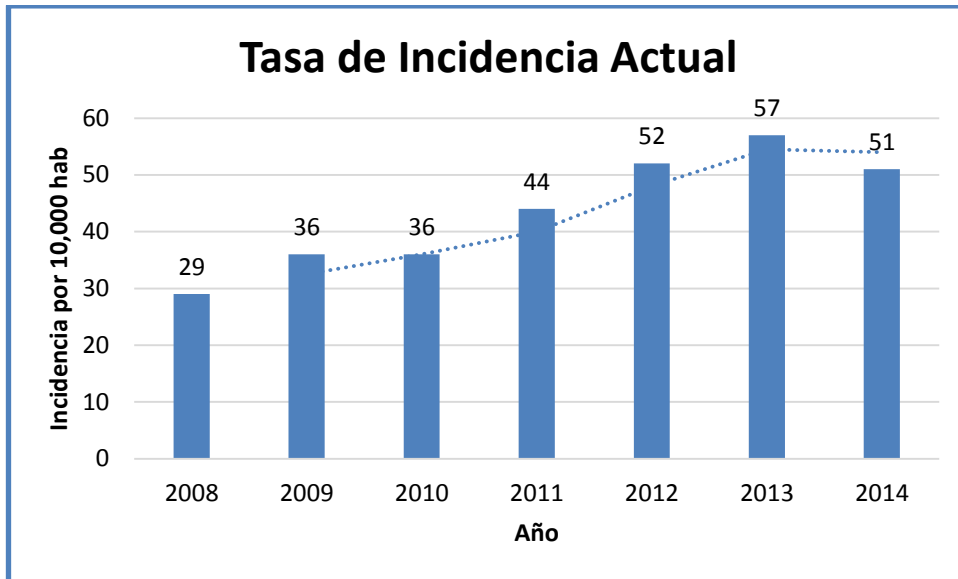


Figura 40. Tasa de incidencia actual 2008 – 2014

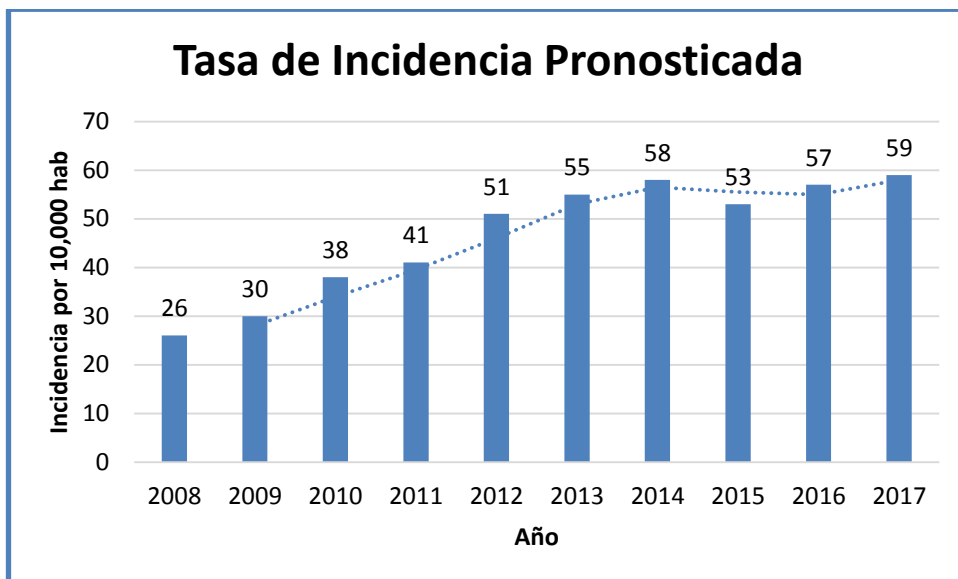


Figura 41. Tasa de incidencia pronosticada 2008 - 2017

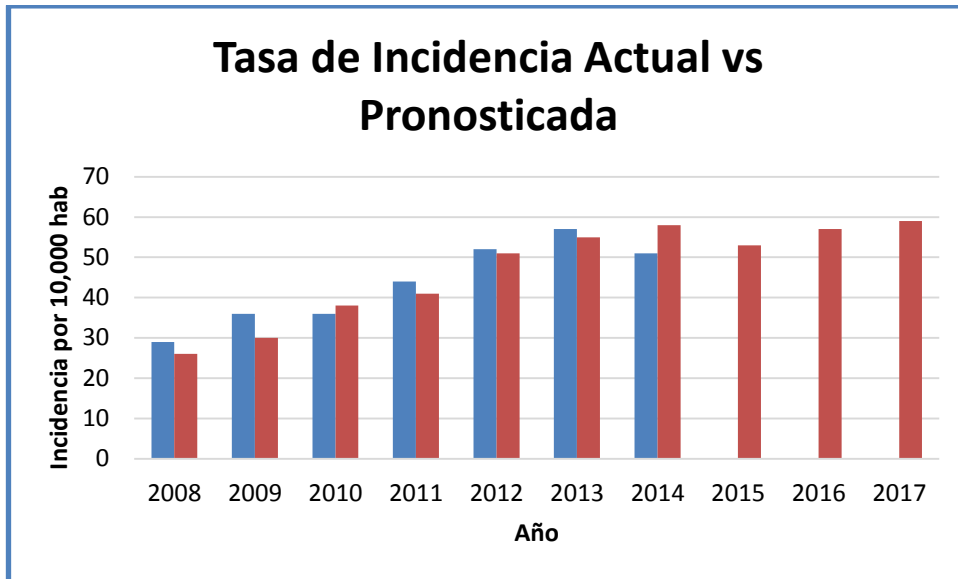


Figura 42. Grafico barras Tasa incidencia actual vs pronosticada

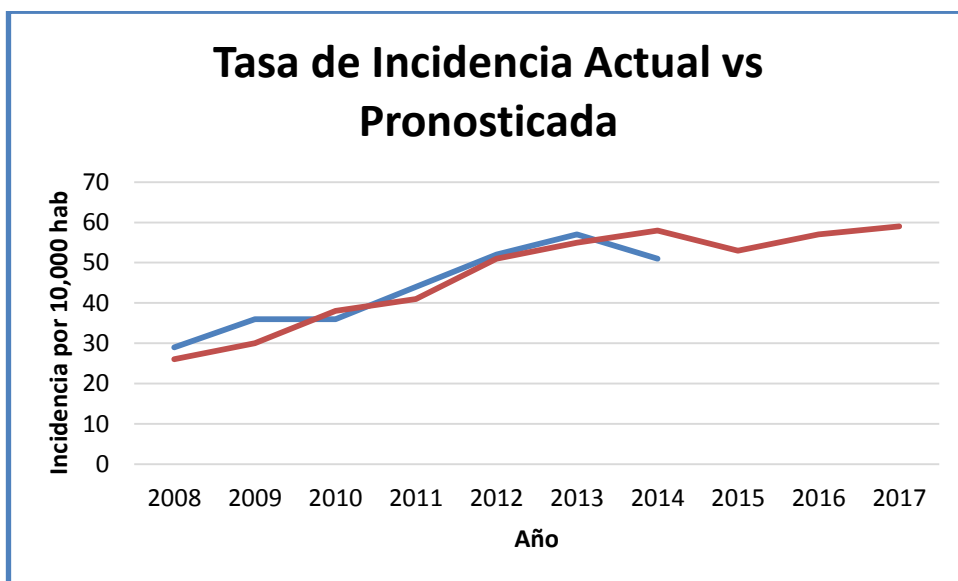


Figura 43. Gráfico de Líneas incidencias actual vs pronosticada

CAPITULO IV: DISCUSION

En este capítulo se evaluará si la hipótesis planteada es este proyecto cumple con dar solución al problema resolver, esto se realiza a través de contrastación de hipótesis que para este proyecto es apoyado por indicadores como: Costos y Satisfacción del usuario final, usando el Método Pre-Test y Post-Test.

4.1 PLANTEAMIENTO DE LA HIPÓTESIS

Para nuestro proyecto se planteó la siguiente hipótesis:

“La creación del modelo predictivo de minería de datos, en el Hospital VRG de Huaraz, apoyara la gestión hospitalaria, sobre morbilidad de pacientes hospitalizados.”

Variable Independiente (VI): Modelo Predictivo de Minería de Datos.

Variable Dependiente (VD): Gestión Hospitalaria sobre la morbilidad de pacientes hospitalizados.

Estas variables van a ser medidas a través de los siguientes indicadores:

- Costos
- Satisfacción del usuario

La inferencia de la validez de la hipótesis será comprobada si del total de indicadores medidos (costos y satisfacción del usuario), por lo menos uno de estos son aceptados.

4.2 CONTRASTACIÓN DE LA HIPÓTESIS

Para todas los **Indicadores Cuantitativos** se procederá de la siguiente manera:

- Paso 1: Definición de la variable a evaluar.
- Paso 2: Planteamiento de la hipótesis estadística.
- Paso 3: Definición del nivel de significancia, para todos será del 5%. Por lo tanto el Nivel de Confianza ($1-\alpha = 0.95$) será del 95%.
- Paso 4: Definición del tipo de prueba a aplicar, para todos será la distribución Normal Z.

- Paso 5: Tabulación de valores obtenidos antes y después de la implementación del sistema web.
- Paso 6: Cálculo del Promedio Muestral usando las expresiones:

$$\bar{X}_A = \frac{\sum_{i=1}^n X_{Ai}}{n}$$

$$\bar{X}_D = \frac{\sum_{i=1}^n X_{Di}}{n}$$

Se entiende que A significa Antes y D significa Después de la implementación del sistema web, aquí como en las siguientes expresiones.

- Paso 7: Cálculo de la Varianza Muestral, que es el promedio de todos los valores obtenidos antes y después del estímulo, usando las expresiones:

$$\sigma_A^2 = \frac{\sum_{i=1}^n (X_{Ai} - \bar{X}_A)^2}{n - 1}$$

$$\sigma_D^2 = \frac{\sum_{i=1}^n (X_{Di} - \bar{X}_D)^2}{n - 1}$$

- Paso 8: Cálculo Estadístico de la Prueba, que es la diferencia al cuadrado de las diferencias obtenidas, entre el valor observado y la media, antes y después del estímulo, usando la expresión:

$$Z_c = \frac{\bar{X}_A - \bar{X}_D}{\sqrt{\left(\frac{\sigma_A^2}{n_A} + \frac{\sigma_D^2}{n_D}\right)}}$$

- Paso 9: Búsqueda del Valor Crítico de Z en las tablas estadísticas de la Distribución Z.
- Paso 10: Redacción de la conclusión de la prueba estadística.

El método de Análisis para los **Indicadores Cualitativos** será el siguiente:

- Paso 1: Definición de la Variable a evaluar.
- Paso 2: Planteamiento de la Hipótesis Estadística.
- Paso 3: Formulación de la encuesta a aplicar y procesamiento de los resultados utilizando las tablas correspondientes ver Anexo I.
- Paso 4: Cálculo del Promedio Muestral, tomando los resultados del respectivo subtítulo del Anexo 2 se hacen uso de las expresiones:

$$PN_A = \frac{\sum_{i=1}^n NA_i}{n}$$

$$PN_D = \frac{\sum_{i=1}^n ND_i}{n}$$

Dónde:

PN_A = Promedio del antes de la implementación del modelo.

NA_i = Puntaje total de la pregunta i – ésima.

n = Número de preguntas

PN_D = Promedio después de la implementación del modelo.

ND_i = Puntaje total de la pregunta i – ésima.

- Paso 5: Cálculo de la Desviación Estándar Muestral, que es el promedio de todos los valores obtenidos antes y después del estímulo, usando las expresiones:

$$S_D^2 = \int d = \frac{n \sum_{i=1}^n D^2 - (\sum_{i=1}^n D_i)^2}{n(n-1)}$$

Dónde:

S_D^2 = Desviación Estándar

D = Diferencia

D_i = Diferencia de promedio de la pregunta i – ésima.

n = Numero de preguntas

- Paso 6: Definición del Tipo de Prueba a aplicar, para todos será la distribución t de Student para SEIS (6) preguntas de la encuesta.
- Paso 7: Cálculo Estadístico de la prueba usando la expresión:

$$t_c = \frac{\bar{D}\sqrt{n}}{\sqrt{S_D^2}}$$

- Paso 8: Cálculo del Valor Crítico de t, se busca en las tablas de Distribución de t de Student con Nivel de Significancia, para todos será del $\alpha= 5\%$, por lo tanto el Nivel de Confianza ($1-\alpha = 0.95$) será del 95%, y con $n-1=6-1=5$ grados de libertad y se obtiene:

$$t(1- \alpha) (n-1) = t(1-0.05)(6-1) = 2.0150$$

- Paso 10: Redacción de la Conclusión de la prueba estadística.

4.2.1 INDICADOR: COSTO

Se mide los costos que resultan de los recursos utilizados en la gestión de procesos importantes.

Pre-Test (C_1): Medición previa de la variable dependiente a ser utilizada.

Post-Test (C_2): Corresponde a la nueva medición de la variable dependiente a ser utilizada.

Dónde:

$$C_1 \text{_____} O \text{_____} C_2$$

C_1 : Costos que resultan de los recursos utilizados en la gestión de procesos importantes con el proceso anterior.

O: (Aplicación de la variable independiente) Modelo Predictivo de minería de datos de apoyo a la gestión hospitalaria sobre la morbilidad en pacientes hospitalizados.

C_2 : Costos que resultan de los recursos utilizados en la gestión de procesos importantes con el Modelo Predictivo de minería de datos de apoyo a la gestión hospitalaria sobre la morbilidad en pacientes hospitalizados.

A continuación, se muestran las mediciones hechas a la variable correspondiente:

Dónde:

X_{Ai} : Costos en soles que resultan de los recursos utilizados en la gestión de procesos importantes con la forma habitual del proceso de estimación de costos.

X_{Di} : Costos en soles que resultan de los recursos utilizados en la gestión de procesos importantes después de la implementación del modelo.

Capítulo VI

Procesos importantes	Perfil del responsable	Ratio (S.)/H H	Tiempo promedio sin el Modelo	Tiempo promedio con el Modelo	Ratio (S.)/M M	Tiempo actual (Min)	Tiempo propuesto (Min)	XAi	XDi	(XAi) ²	(XDi) ²
Proceso de planificación del abastecimiento	Jefe de Servicio	57.19	03:25:57	01:43:09	0.953	205.95	103.15	196.30	98.32	38,535.53	9,666.65
Proceso de planificación del recurso	Jefe de Servicio	57.19	03:57:49	01:58:12	0.953	237.82	118.20	226.68	112.66	51,383.33	12,693.24
Proceso de consolidación de la información	Jefe de Informática	57.19	01:53:56	00:46:15	0.953	113.93	46.25	108.60	44.08	11,793.41	1,943.40
Proceso de cálculo y monitoreo de indicadores	Director Administrativo	115.56	00:55:52	00:18:07	1.926	55.87	18.12	107.60	34.89	11,577.59	1,217.50
Totales								639.18	289.96	113,289.85	25,520.794

Tabla 11: Variación de datos históricos del indicador Costos.

Fuente: [Elaboración Propia].

Cálculo del Promedio Muestral:

$$\bar{X}_A = \frac{\sum_{i=1}^n X_{Ai}}{n} = \frac{639.18}{4} = 159.7951$$

$$\bar{X}_D = \frac{\sum_{i=1}^n X_{Di}}{n} = \frac{289.96}{4} = 72.49$$

Cálculo de la Varianza Muestral:

$$\sigma_A^2 = \frac{\sum_{i=1}^n (X_{Ai} - \bar{X}_A)^2}{n-1} = \frac{\sum_{i=1}^n (X_{Ai})^2 - n(\bar{X}_A)^2}{n-1} =$$

$$\frac{113289.85 - 4(159.7951)^2}{3} = \sigma_A^2 = 3717.335$$

$$\sigma_D^2 = \frac{\sum_{i=1}^n (X_{Di} - \bar{X}_D)^2}{n-1} = \frac{\sum_{i=1}^n (X_{Di})^2 - n(\bar{X}_D)^2}{n-1} =$$

$$\frac{25520.794 - 4(72.49)^2}{3} = \sigma_D^2 = 1500.5264$$

Cálculo Estadístico de la Prueba:

$$\begin{aligned} Z_c &= \frac{\bar{X}_A - \bar{X}_D}{\sqrt{\left(\frac{\sigma_A^2}{n_A} + \frac{\sigma_D^2}{n_D}\right)}} = \frac{159.7951 - 72.49}{\sqrt{\left(\frac{3717.335}{4} + \frac{1500.5264}{4}\right)}} = \frac{87.3051}{\sqrt{(929.3338 + 375.1316)}} \\ &= \frac{87.3051}{36.1174} = Z_c = 2.4173 \end{aligned}$$

Valor Crítico de Z en las Tablas estadísticas de la Distribución Z:

$$Z_{(\alpha)} = Z_{(0.05)} = Z_{(0.95)} = 1.6449$$

Hipótesis estadística:

- Hipótesis H_0 : Los Costos en soles que resultan de los recursos utilizados en la gestión de procesos importantes antes de la implementación del modelo es menor o igual que los costos en soles que resultan de los recursos utilizados en la gestión de procesos importantes después de la implementación del modelo antes mencionado.
- Hipótesis H_i : Los costos en soles que resultan de los recursos utilizados en la gestión de procesos importantes antes de la implementación del modelo es mayor o igual que los costos en soles que resultan de los recursos utilizados en la gestión de procesos importantes después de la implementación del modelo antes mencionado.

$H_0: O_1 - O_2 = 0$ Hipótesis nula

No hay diferencia alguna entre dos muestras.

$H_i: O_1 - O_2 > 0$ Hipótesis alternativa

Si hay diferencia

Región de aceptación o rechazo:

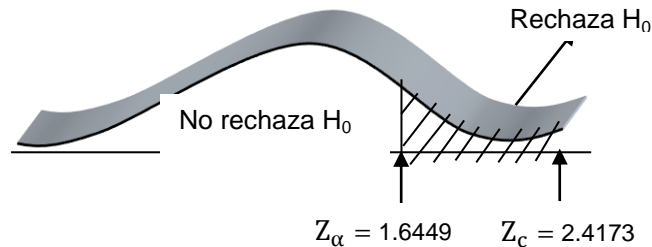


Figura 44: Región de aceptación o rechazo del indicador costo.

Fuente: [Elaboración Propia].

Conclusión

Puesto que $Z_c = 2.4173$ calculado es mayor que $Z_\alpha = 1.6449$ y estando este valor dentro de la región de rechazo, entonces se rechaza H_0 y por consiguiente se acepta H_i , se concluye entonces que los costos que resultan de los recursos utilizados en la gestión de procesos importantes con el uso del modelo predictivo *son menores* a los costos que resultan de los recursos utilizados en la gestión sin el modelo.

4.2.2 INDICADOR: SATISFACCIÓN DEL USUARIO

Se mide el nivel de satisfacción del usuario con el modelo predictivo de minería de datos. Esta medición se realiza, mediante la aplicación de encuestas a los principales involucrados (“50 usuarios”) del área de Estadística e Informática, dirección, administración y Jefaturas de los servicios.

Pre-Test (S1): Medición previa de la variable dependiente a ser utilizada.

Post-Test (S2): Corresponde a la nueva medición de la variable dependiente a ser utilizada.

Dónde:

S_1 _____ O _____ S_2

S_1 : Satisfacción del usuario con el proceso actual.

O : (Aplicación de la variable independiente) Modelo Predictivo de minería de datos de apoyo a la gestión hospitalaria sobre la morbilidad en pacientes hospitalizados.

S_2 : Satisfacción del usuario final con el Modelo Predictivo de minería de datos de apoyo a la gestión hospitalaria sobre la morbilidad en pacientes hospitalizados.

Capítulo VI

A continuación, se muestran las mediciones hechas a la variable correspondiente:

Dónde:

d_i : diferencia medida del promedio de la satisfacción del usuario.

d_i^2 : diferencia al cuadrado, medida del promedio de la satisfacción del usuario.

Descripción	Sin el modelo propuesto %	Con el modelo Propuesto %	d_i %	d_i^2 %
¿Considera usted que el tiempo de respuesta a los datos históricos de los últimos doce años, no será muy excesivo?	60	90	30	900
¿Considera usted que la recolección de datos, no tomara demasiado tiempo?	76	82	6	36
¿Considera usted que la información presentada a la gestión para la toma de decisiones es confiable?	56	88	32	1024
¿Considera usted que el tiempo de entrega de resultados, se dan dentro de los plazos previstos?	44	80	36	1296
¿Considera usted que no existen datos inconsistentes o mal registrados que tengan injerencia en las decisiones?	66	86	20	400
¿Considera usted que el nivel de confianza de acierto en los pronósticos es confiable?	54	94	40	1600
Total			164	5256

Tabla 12: Variación de datos históricos del indicador satisfacción del usuario final.

Fuente: [Elaboración Propia].

$$\sum d_i = 164$$

$$\sum d_i^2 = 5256$$

$$\bar{d}_i = \frac{\sum d_i}{n} = \frac{164}{6} = 27.3333$$

Encontrando la desviación estándar:

$$s_d = \sqrt{\frac{n \sum d_i^2 - (\sum d_i)^2}{n(n-1)}} = \sqrt{\frac{6(5256) - (164)^2}{6(6-1)}} = 12.4365$$

Hipótesis estadística:

$H_0: O_1 - O_2 = 0$ Hipótesis nula

No hay diferencia alguna entre dos muestras.

$H_1: O_2 - O_1 > 0$ Hipótesis alternativa

Si hay diferencia

Nivel de significancia $\alpha = 0.05$

Función de prueba:

$$t_c = \frac{\bar{d}_i}{s_d} \sqrt{n}$$

Valor Crítico de “t” de student:

$$t_t(1-\alpha)(n-1)$$

$$t_t(1-0.05)(6-1)$$

$$t_t(0.95)(5) = 2.0150$$

$$t_t = 2.0150$$

$$t_c = \frac{27.3333}{12.4365} \sqrt{6} = 5.3836$$

Región de aceptación o rechazo:

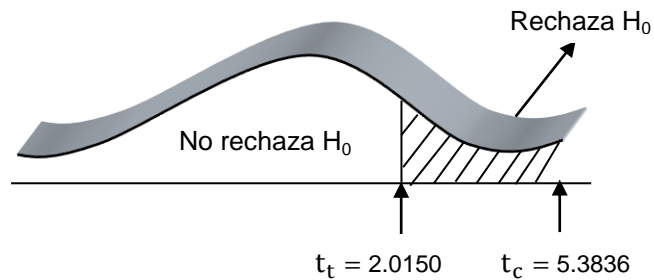


Figura 45: Región de aceptación o rechazo del indicador satisfacción del usuario.

Fuente: [Elaboración Propia].

Se concluye $t_c = 5.3836$ es mayor que $t_t = 2.0150$

Se rechaza H_0 y se acepta $H_1: O_2 - O_1 > 0$

Es decir, se concluye que los usuarios *están satisfechos* con el modelo propuesto y opinan que con el uso de éste se van a solucionar los principales problemas que ocurrían, además de hacerles más eficiente el desarrollo de sus actividades durante el proceso.

4.3 ANÁLISIS DE RESULTADOS

Los resultados obtenidos en las mediciones hechas a los indicadores son los siguientes:

Indicadores:

- Costos:

Se concluye $C_1 > C_2$, por tanto SE ACEPTA $H_i: O_1 - O_2 > 0$

- Satisfacción del Usuario:

Se concluye $S_2 > S_1$, por tanto SE ACEPTA $H_i: O_2 - O_1 > 0$

Por lo tanto, de acuerdo a estos resultados se cumple la regla de inferencia establecida, ya que los 2 indicadores son aceptados y se concluye que la HIPÓTESIS: “La creación del modelo predictivo de minería de datos, en el Hospital VRG de Huaraz, apoyara la gestión hospitalaria, sobre morbilidad de pacientes hospitalizados.”

ES ACEPTADA.

CAPITULO V: CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

- A lo largo de esta investigación se ha llevado a cabo una importante recopilación bibliográfica y revisión teórica sobre aspectos relacionados con mi tema, que me han permitido conocer técnicas predictivas de series de tiempo estacionarias y no estacionarias, así como los métodos de pronósticos y suavizamiento de Box & Jenkins que incluye los Modelos AR (Auto-Regresivos), Modelos MA (Media Móvil), Modelos ARIMA (Auto regresivo Integrado con Media Móvil) y Modelos SARIMA (Auto regresivos Integrados con media móvil estacional).
- Durante del proceso de preparación de los datos, se identificó los datos de origen, en dos bases datos transaccionales SQL Server (SYSFAR y GALENHOS), luego del proceso de Extracción Transformación y Carga (ETL), a través de consultas rápidas y técnicas de muestreo se detectó datos anómalos, eliminando o separando las tuplas, para posteriormente crear y/o cargar el Datamart con seis dimensiones y una tabla de hecho formando un modelo estrella, que sirve como repositorio para que finalmente durante un proceso de selección y transformación de variables, obtener los datos de entrada para mi modelo.
- Para aplicar la metodología de Box-Jenkins Modelo **ARIMA**, se crearon las variables, periodo y número de diagnósticos en el mes, basados en la recolección de los datos de los últimos siete años, identificando la serie, estimando los coeficientes de forma iterativa hasta obtener un mejor modelo, validando el modelo estimado, cuantificando a futuro los valores de la serie, para que finalmente obtener los mejores pronósticos para casos de morbilidad en enfermedades del aparato digestivo “Apendicitis” en los pacientes hospitalizados del Hospital VRG.

- El crear un modelo predictivo, utilizando técnica de análisis de series de tiempo y aplicando la metodología de Box-Jenkins, en su modelo ARIMA me permite obtener pronósticos de nuevos casos a corto, mediano y largo plazo.
- La creación de un modelo predictivo de minería de datos, permitirá entregar información valiosos que pueden ser utilizados por la gestión, de tal manera que puedan realizar una mejor distribución de los recursos, permitiéndole una mejora en la calidad del servicio e intentar reducir la tasa de morbilidad en la región, futuras planificaciones en medicamentos, material médico e insumos y entrega de resultados confiables a menor costo y tiempo posible.

5.2 RECOMENDACIONES

- Para futuros trabajos de investigación sería muy beneficioso para la gestión la creación de modelos relacionado al grupo de O00-O99 Embarazos, partos y puerperios. Que cubren aprox. el 50% de las atenciones en hospitalización y poder conocer los centros poblados con mayor tasa de Natalidad en la región, realizar campañas de prevención para prevenir el embarazo adolescente.
- Como línea futura de trabajo, para intentar medir de forma cuantitativa la gestión, evaluar logros y metas a un casi 100%, se propone analizar modelos predictivos sobre mortalidad, natalidad y calidad de vida de los pacientes.
- Tomar en consideración hasta 12 meses de pronóstico, a mucho más largo plazo la certeza de la predicción disminuye.
- Aplicar modelos para grupo de diagnóstico en pacientes hospitalizados. No existe un mejor modelo que predomine sobre los demás, lo que sí existe es un modelo más adecuado para cada caso de morbilidad y/o diagnóstico.

ANEXOS

A. RESULTADOS DE ENCUESTAS

Encuestas en la Situación Actual

1.- ¿Considera usted que el tiempo de respuesta a los datos históricos de los últimos doce años, no sera excesivo?

Rpta	Nro Personal	%
SI	30	60
NO	20	40
TOTAL	50	100

2.- ¿Considera usted que la recolección de datos, no tomara demasiado tiempo?

Rpta	Nro Personal	%
SI	38	76
NO	12	24
TOTAL	50	100

3.- ¿Considera usted que la información presentada a la gestión para la toma de decisiones es confiable?

Rpta	Nro Personal	%
SI	28	56
NO	22	44
TOTAL	50	100

4.- ¿Considera usted que el tiempo de entrega de resultados, se dan dentro de los plazos previstos?

Rpta	Nro Personal	%
SI	22	44
NO	28	56
TOTAL	50	100

5.- ¿Considera usted que no existen datos inconsistentes o mal registrados que tengan injerencia en las decisiones?

Rpta	Nro Personal	%
SI	33	66
NO	17	34
TOTAL	50	100

6.- ¿Considera usted que el nivel de confianza de acierto en los pronósticos no es confiable?

Rpta	Nro Personal	%
SI	27	54
NO	23	46
TOTAL	50	100

Encuestas con el modelo propuesto.

1.- ¿Considera usted que el tiempo de respuesta a los datos históricos de los últimos con el modelo propuesto no será excesivo?

Rpta	Nro Personal	%
SI	45	90
NO	5	10
TOTAL	50	100

2.- ¿Considera usted que la recolección de datos, no tomara demasiado tiempo?

Rpta	Nro Personal	%
SI	41	82
NO	9	18
TOTAL	50	100

Anexos

3.- ¿Considera usted que la información presentada a la gestión para la toma de decisiones es confiable?

Rpta	Nro Personal	%
SI	44	88
NO	6	12
TOTAL	50	100

4.- ¿Considera usted que el tiempo de entrega de resultados, se dan dentro de los plazos previstos?

Rpta	Nro Personal	%
SI	40	80
NO	10	20
TOTAL	50	100

5.- ¿Considera usted que no existen datos inconsistentes o mal registrados que tengan injerencia en las decisiones?

Rpta	Nro Personal	%
SI	43	86
NO	7	14
TOTAL	50	100

6.- ¿Considera usted que el nivel de confianza de acierto en los pronósticos no es confiable?

Rpta	Nro Personal	%
SI	47	94
NO	3	6
TOTAL	50	100

B. MATERIAL MEDICO Y MEDICAMENTOS VENCIDOS

Tabla 13: Medicamentos y material médico vencidos de periodo 2014

Lista de medicamentos y material médico asignados al servicio de cirugía, para el tratamiento de diagnóstico de K00-K79, enfermedades del aparato digestivo (Apendicitis)

MES	MEDICAMENTO E INSUMO	UND	PV	CANTIDAD	TOTAL
ENE	DICLOXACILINA (COMO SAL SODICA) 250 mg/5 mL SUS 60 mL	FCO	3.50	67.00	234.50
ENE	HIERRO SACARATO 100 mg INY	AMP	3.80	90.00	342.00
ENE	CINTA DE CATGUT CROMADA SUTURA ABSORBIBLE 4 mm C/A 1/2 CIRCULO REDONDA 65 mm X 60 cm	UND	68.75	50.00	3437.50
FEB	CATGUT CROMICO 6/0 C/A 1/2 CR 15 mm x 70 cm	UND	3.50	125.00	437.50
MAR	SUTURA CATGUT CROMICO 0 S/A X 150 cm	UND	3.74	420.00	1569.75
ABR	BACLOFENO 10 mg TAB	TAB	1.63	662.00	1075.75
ABR	DESLANOSIDO 200 µg/mL INY 2 mL	AMP	2.33	106.00	246.45
ABR	ETILEFRINA 10 mg/mL INY 1 mL	AMP	1.63	40.00	65.00
MAY	TIAMAZOL 5 mg TAB	TAB	1.54	500.00	768.75
MAY	SUTURA CATGUT SIMPLE 1 C/A 1/2 CIRCULO REDONDA 35 mm X 70 cm	UND	2.88	80.00	230.00
MAY	SUTURA NYLON AZUL MONOFILAMENTO 1 C/A 1/2 CIRCULO REDONDA 30 mm X 75 cm	UND	3.90	71.00	276.90
JUN	SUERO ANTIBOTROPICO POLIVALENTE 40 mg/10 mL INY	INY	74.31	20.00	1486.20
JUN	SUTURA CATGUT CROMICO 2 S/A X 150 cm	UND	2.88	24.00	69.00
JUN	SUTURA NYLON AZUL MONOFILAMENTO 1 C/A 1/2 CIRCULO REDONDA 30 mm X 75 cm	UND	3.90	15.00	58.50
JUN	SUTURA SEDA NEGRA TRENZADA MULTIEMPAQUE 3/0 S/A X 70 cm	UND	2.88	31.00	89.13
JUL	NEVIRAPINA 200 mg TAB	TAB	1.38	200.00	275.00
AGO	TIMOLOL (COMO MALEATO) 5 mg/mL (0.5 %) SOL OFT 5 mL	FCO	3.49	150.00	523.13
SET	CARVEDILOL 6.25 mg TAB	TAB	1.13	503.00	565.88
SET	BUDESONIDA 100 ug x 300 Dosis	INH	20.00	200.00	4000.00
SET	BUPIVACAINA 0.5 % x 4 ml pesado	AMP	13.00	105.00	1365.00
OCT	CARBAMAZEPINA 100 mg / 5 ml x 100 ml	FCO	14.50	90.00	1305.00
OCT	CEFOTAXIMA 1 g / vial	AMP	1.50	300.00	450.00
NOV	SUTURA CATGUT SIMPLE 1 C/A 1/2 CIRCULO REDONDA 15 mm X 70 cm	UND	2.88	120.00	345.00
NOV	SONDA DE ALIMENTACION Nº 14	UND	1.00	120.00	120.00
DIC	SUTURA CATGUT SIMPLE 0 C/A 1/2 CIRCULO REDONDA 15 mm X 70 cm	UND	2.88	80.00	230.00

**C. TENDENCIA DE LOS PRINCIPALES DIAGNOSTICOS DE
PACIENTES HOSPITALIZADOS:**

APENDICITIS: En la gráfica se observa que la tendencia es creciente.

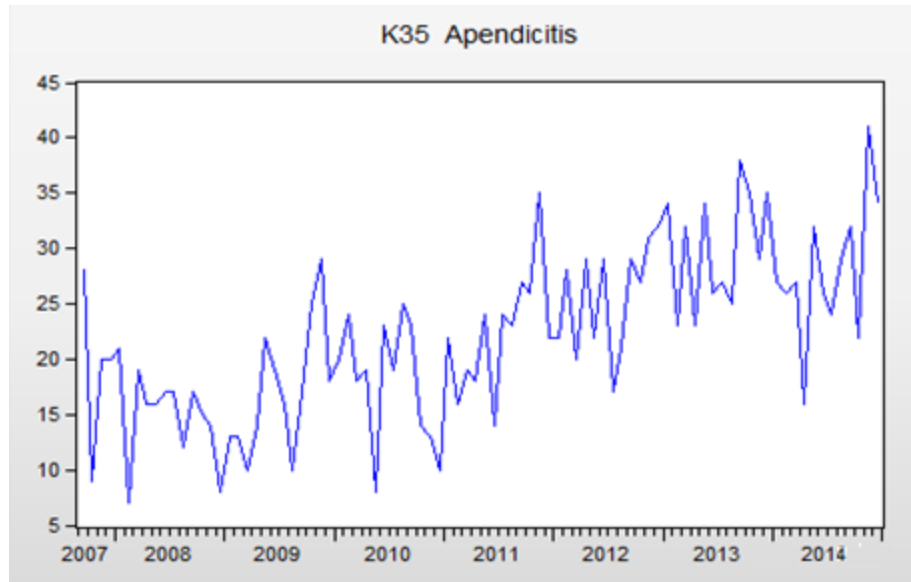


Figura 46. Dx - Apendicitis

COLECISTITIS CRONICA: En la gráfica se observa que la tendencia a partir del 2009 es aparentemente estacionaria.

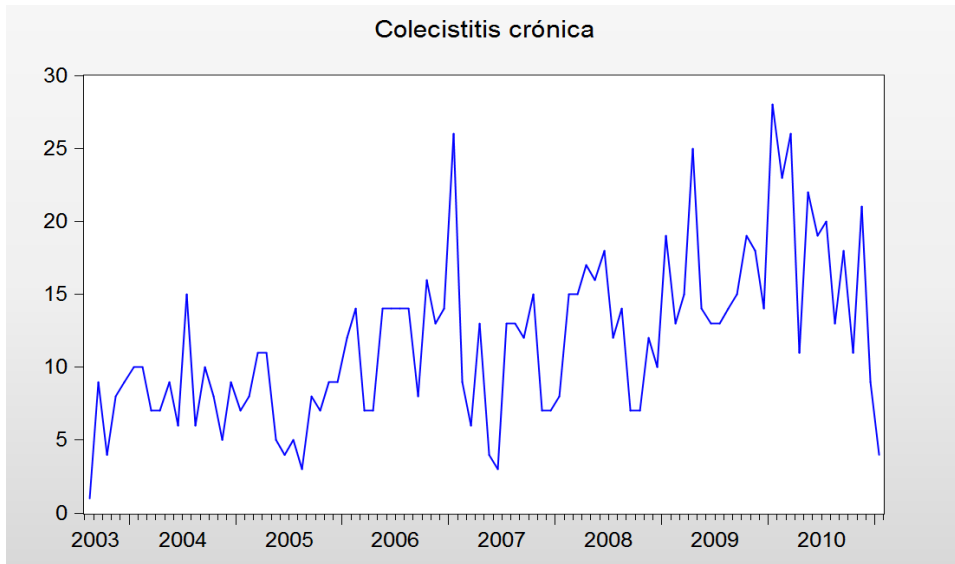


Figura 47. Dx – Colecistitis Crónica

BRONCONEUMONIA: En la gráfica se observa que la tendencia es aparentemente estacionaria.

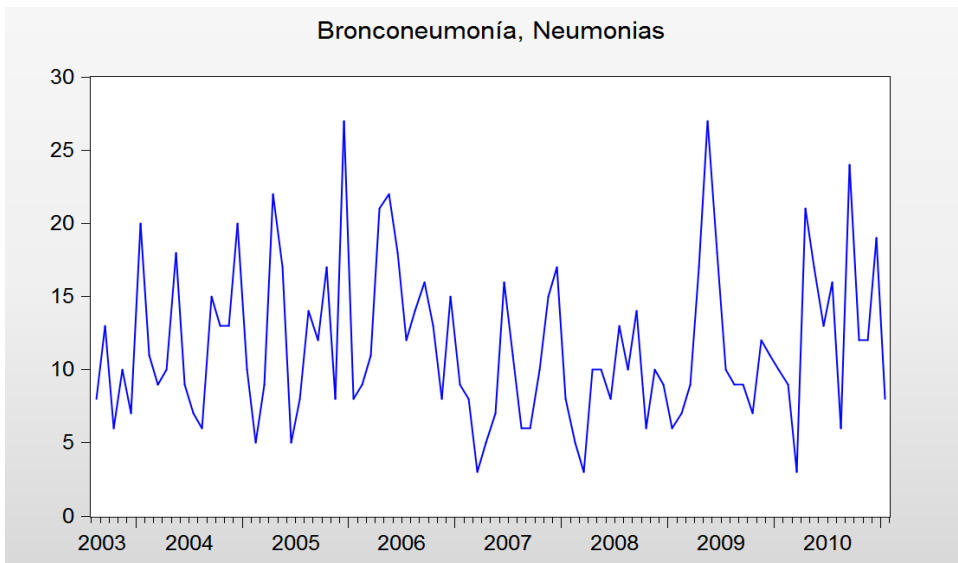


Figura 48. Dx – Bronconeumonía

DIARREA Y GASTROENTERITIS: En la gráfica se observa que la tendencia es aparentemente estacionaria con picos muy altos en algunos meses del año.

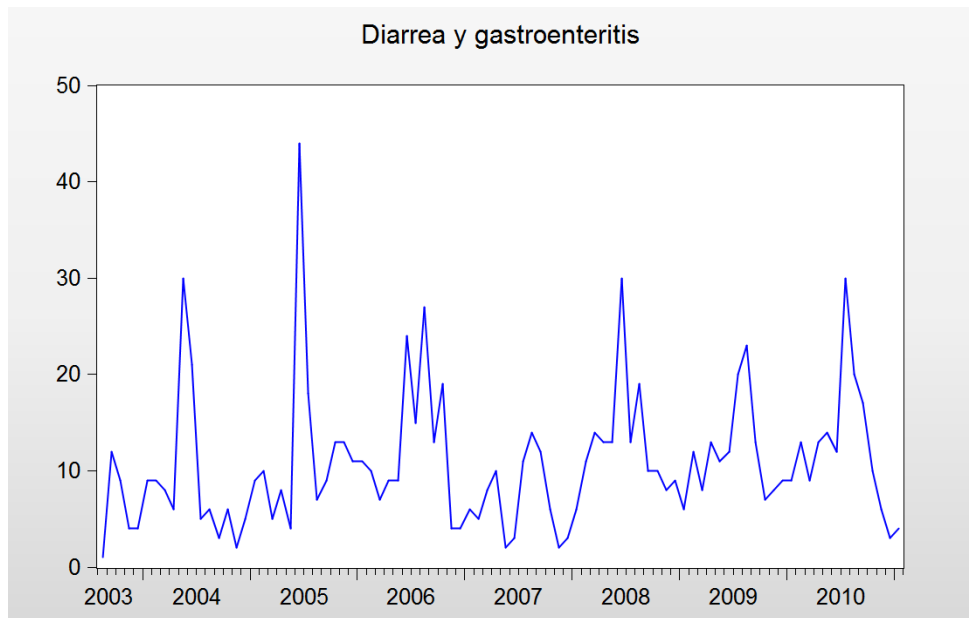


Figura 49. Dx – Diarreas y gastroenteritis

TRAUMATISMO: En la gráfica se observa una tendencia no estacionaria.

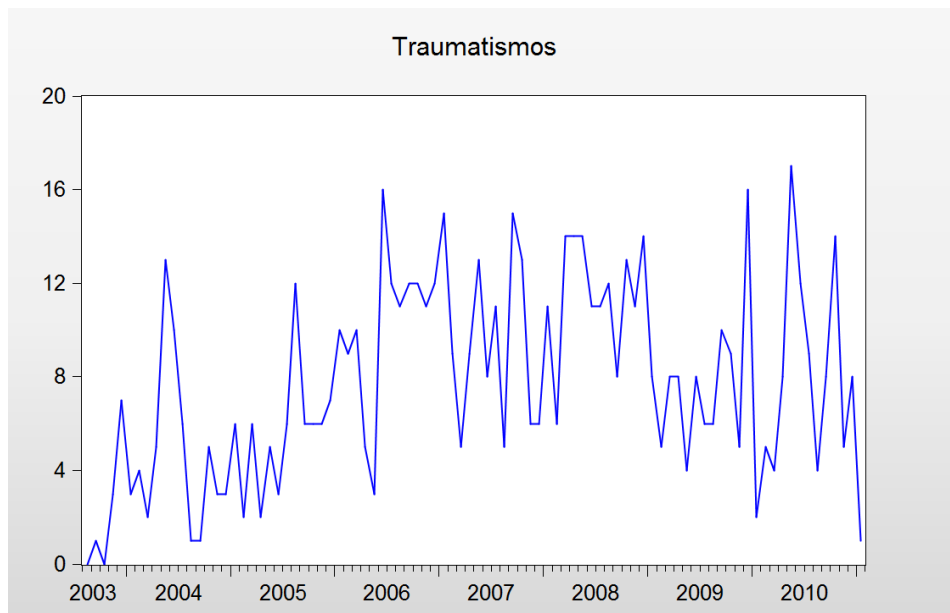


Figura 50. Dx – Traumatismo

D. CODIGO CONSULTA SQL DE PACIENTES CON DIAGNOSTICOS POR AÑO MES:

```
SELECT
    year(FechaDeIngreso) as año,
    month(FechaDeIngreso) as mes
    ,b.EstadoCivil
    ,b.Sexo
    ,EdadEnAnosCumplidos as EdaPaciente
    ,Departamento
    ,provincia as codprovincia
    ,(Select Nombre from Provincia where Codigo=provincia and
Departamento = '02') as Provincia
    ,distrito as coddistrito
    ,(Select Nombre from Distrito where Codigo=distrito and
Provincia='01' and Departamento = '02') as Distrito
    ,[DxEgresoCod]
    ,[DxEgresoDes], COUNT(*) as NroDiagnostico
    into Resultado1
FROM [HEALTHMINING].[dbo].[EpicrisisIngresos] a
    inner join dbo.Pacientes b on a.HistoriaClinica =
b.HistoriaClinica
where departamento = '02' and provincia = '01'
group by
    year(FechaDeIngreso),
    month(FechaDeIngreso)
    ,EstadoCivil
    ,Sexo
    ,EdadEnAnosCumplidos
,departamento,provincia,distrito
,[DxEgresoCod], [DxEgresoDes]
order by 1,2,13 desc
```


F. POBLACIÓN DEPARTAMENTO ANCASH, PROVINCIA HUARAZ, DISTRITO HUARAZ 2008-2015. (FUENTE INEI)

	Año	Población
HUARAZ	2008	58,593
	2009	59,370
	2010	60,150
	2011	60,940
	2012	61,736
	2013	62,534
	2014	63,327
	2015	64,109

G. NUMERO DE CASOS DE DIAGNOSTICOS POR PERIODO:

Tabla 14: Cuadro de cantidad de diagnósticos más frecuentes en los últimos 8 años

Año	mes	K35	A09	K81	J18	S06
		Apendicitis	Diarreas	Colesistiti	Bronconeumonia	Traumatismo
2014	12	34	3	9	19	8
2014	11	41	6	21	12	5
2014	10	22	10	11	12	14
2014	9	32	17	18	24	8
2014	8	29	20	13	6	4
2014	7	24	30	20	16	9
2014	6	26	12	19	13	12
2014	5	32	14	22	17	17
2014	4	16	13	11	21	8
2014	3	27	9	26	3	4
2014	2	26	13	23	9	5
2014	1	27	9	28	10	2

Anexos

2013	12	35	9	14	11	16
2013	11	29	8	18	12	5
2013	10	35	7	19	7	9
2013	9	38	13	15	9	10
2013	8	25	23	14	9	6
2013	7	27	20	13	10	6
2013	6	26	12	13	18	8
2013	5	34	11	14	27	4
2013	4	23	13	25	17	8
2013	3	32	8	15	9	8
2013	2	23	12	13	7	5
2013	1	34	6	19	6	8
2012	12	32	9	10	9	14
2012	11	31	8	12	10	11
2012	10	27	10	7	6	13
2012	9	29	10	7	14	8
2012	8	21	19	14	10	12
2012	7	17	13	12	13	11
2012	6	29	30	18	8	11
2012	5	22	13	16	10	14
2012	4	29	13	17	10	14
2012	3	20	14	15	3	14
2012	2	28	11	15	5	6
2012	1	22	6	8	8	11
2011	12	22	3	7	17	6
2011	11	35	2	7	15	6
2011	10	26	6	15	10	13
2011	9	27	12	12	6	15
2011	8	23	14	13	6	5
2011	7	24	11	13	11	11

Anexos

2011	6	14	3	3	16	8
2011	5	24	2	4	7	13
2011	4	18	10	13	5	9
2011	3	19	8	6	3	5
2011	2	16	5	9	8	9
2011	1	22	6	26	9	15
2010	12	10	4	14	15	12
2010	11	13	4	13	8	11
2010	10	14	19	16	13	12
2010	9	23	13	8	16	12
2010	8	25	27	14	14	11
2010	7	19	15	14	12	12
2010	6	23	24	14	18	16
2010	5	8	9	14	22	3
2010	4	19	9	7	21	5
2010	3	18	7	7	11	10
2010	2	24	10	14	9	9
2010	1	20	11	12	8	10
2009	12	18	11	9	27	7
2009	11	29	13	9	8	6
2009	10	25	13	7	17	6
2009	9	17	9	8	12	6
2009	8	10	7	3	14	12
2009	7	16	18	5	8	6
2009	6	19	44	4	5	3
2009	5	22	4	5	17	5
2009	4	14	8	11	22	2
2009	3	10	5	11	9	6
2009	2	13	10	8	5	2
2009	1	13	9	7	10	6

Anexos

2008	12	8	5	9	20	3
2008	11	14	2	5	13	3
2008	10	15	6	8	13	5
2008	9	17	3	10	15	1
2008	8	12	6	6	6	1
2008	7	17	5	15	7	6
2008	6	17	21	6	9	10
2008	5	16	30	9	18	13
2008	4	16	6	7	10	5
2008	3	19	8	7	9	2
2008	2	7	9	10	11	4
2008	1	21	9	10	20	3

H. TABLAS DE RESULTADOS

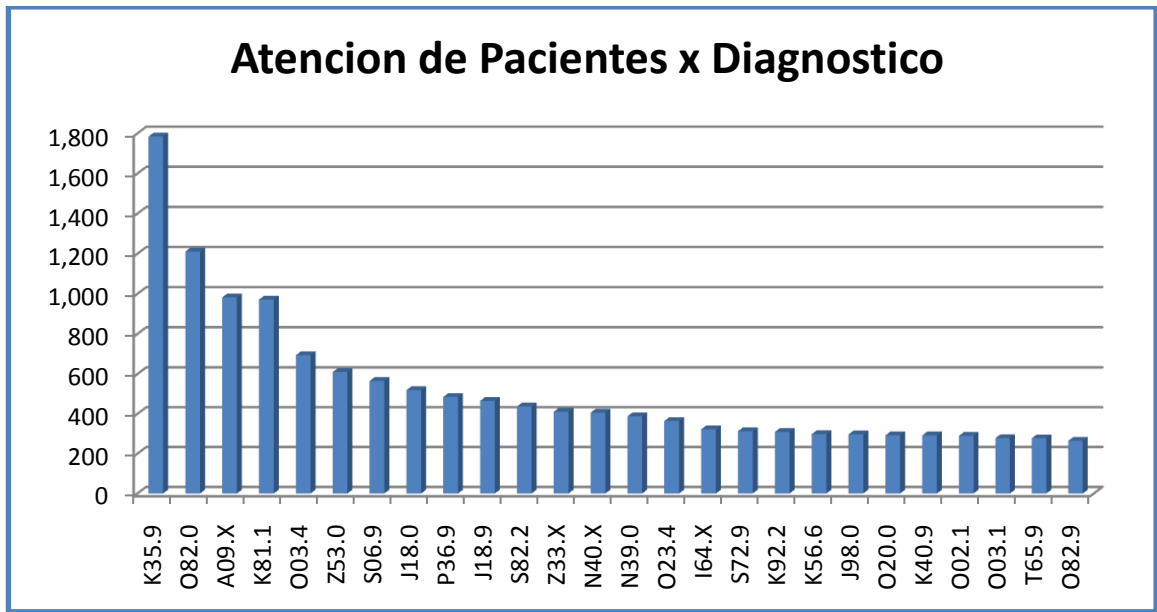


Figura 51. Atenciones de Pacientes por Dx

Tabla 15: Detalle del número de diagnósticos en pacientes hospitalizados del hospital VRG.

CIE10	Diagnostico	Pacientes
K35.9	Apendicitis aguda, no especificada	1,784
O82.0	Parto por cesárea electiva	1,211
A09.X	Diarrea y gastroenteritis de presunto origen infeccioso	982
K81.1	Colecistitis crónica	971
O03.4	Aborto espontáneo, incompleto, sin complicación	694
Z53.0	Procedimiento no realizado por contraindicación	610
S06.9	Traumatismo intracraneal, no especificado	565
J18.0	Bronconeumonía, no especificada	519
P36.9	Sepsis bacteriana del recién nacido, no especificada	485
J18.9	Neumonía, no especificada	465
S82.2	Fractura de la diáfisis de la tibia	437

Anexos

Z33.X	Estado de embarazo, incidental	411
N40.X	Hiperplasia de la próstata	406
N39.0	Infección de vías urinarias, sitio no especificado	388
O23.4	Infección no especificada de las vías urinarias en el embarazo	364
I64.X	Accidente vascular encefálico agudo, no especificado como hemorrágico o isquémico	322
S72.9	Fractura del fémur, parte no especificada	313
K92.2	Hemorragia gastrointestinal, no especificada	309
K56.6	Otras obstrucciones intestinales y las no especificadas	298
J98.0	Enfermedades de la tráquea y de los bronquios, no clasificadas en otra parte	297
O20.0	Amenaza de aborto	292
K40.9	Hernia inguinal unilateral o no especificada, sin obstrucción ni gangrena	292
O02.1	Aborto retenido	290
O03.1	Aborto espontáneo, incompleto, complicado por hemorragia excesiva o tardía	278
T65.9	Efecto tóxico de sustancia no especificada	277
O82.9	Parto por cesárea, sin otra especificación	264

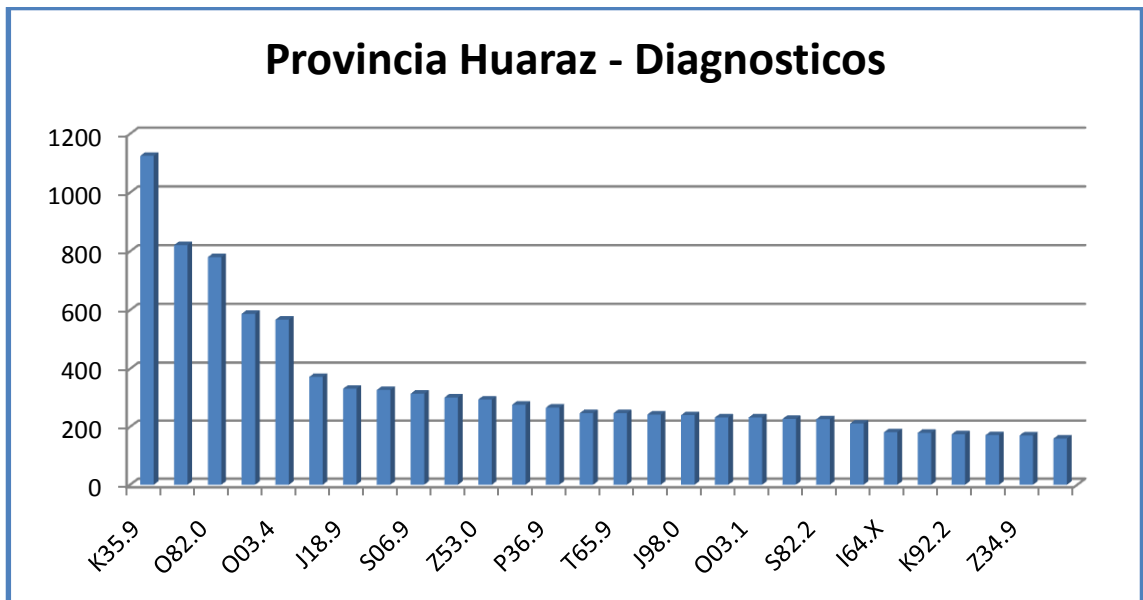


Figura 52. Atenciones de Pacientes provincia Huaraz

Anexos

Provincia	CIE10	Diagnostico	Pacientes
HUARAZ	K35.9	Apendicitis aguda, no especificada	1125
HUARAZ	A09.X	Diarrea y gastroenteritis de presunto origen infeccioso	821
HUARAZ	O82.0	Parto por cesárea electiva	780
HUARAZ	K81.1	Colecistitis crónica	587
HUARAZ	O03.4	Aborto espontáneo, incompleto, sin complicación	567
HUARAZ	J18.0	Bronconeumonía, no especificada	372
HUARAZ	J18.9	Neumonía, no especificada	331
HUARAZ	Z33.X	Estado de embarazo, incidental	327
HUARAZ	S06.9	Traumatismo intracraneal, no especificado	314
HUARAZ	N39.0	Infección de vías urinarias, sitio no especificado	301
HUARAZ	Z53.0	Procedimiento no realizado por contraindicación	294
HUARAZ	O23.4	Infección no especificada de las vías urinarias en el embarazo	276
HUARAZ	P36.9	Sepsis bacteriana del recién nacido, no especificada	266
HUARAZ	O20.0	Amenaza de aborto	247
HUARAZ	T65.9	Efecto tóxico de sustancia no especificada	247
HUARAZ	O02.1	Aborto retenido	242
HUARAZ	J98.0	Enfermedades de la tráquea y de los bronquios, no clasificadas	240
HUARAZ	N40.X	Hiperplasia de la próstata	232
HUARAZ	O03.1	Aborto espontáneo, incompleto, complicado por hemorragia	232
HUARAZ	J45.9	Asma, no especificada	227
HUARAZ	S82.2	Fractura de la diáfisis de la tibia	226
HUARAZ	O80.1	Parto único espontáneo, presentación de nalgas o podálica	211
HUARAZ	I64.X	Accidente vascular encefálico agudo, no especificado como hemorrágico o isquémico	181
HUARAZ	O06.0	Aborto no especificado, incompleto, complicado con infección genital y pelviana	179
HUARAZ	K92.2	Hemorragia gastrointestinal, no especificada	174
HUARAZ	O21.0	Hiperemesis gravídica leve	171
HUARAZ	Z34.9	Supervisión de embarazo normal no especificado	170
HUARAZ	O82.9	Parto por cesárea, sin otra especificación	159

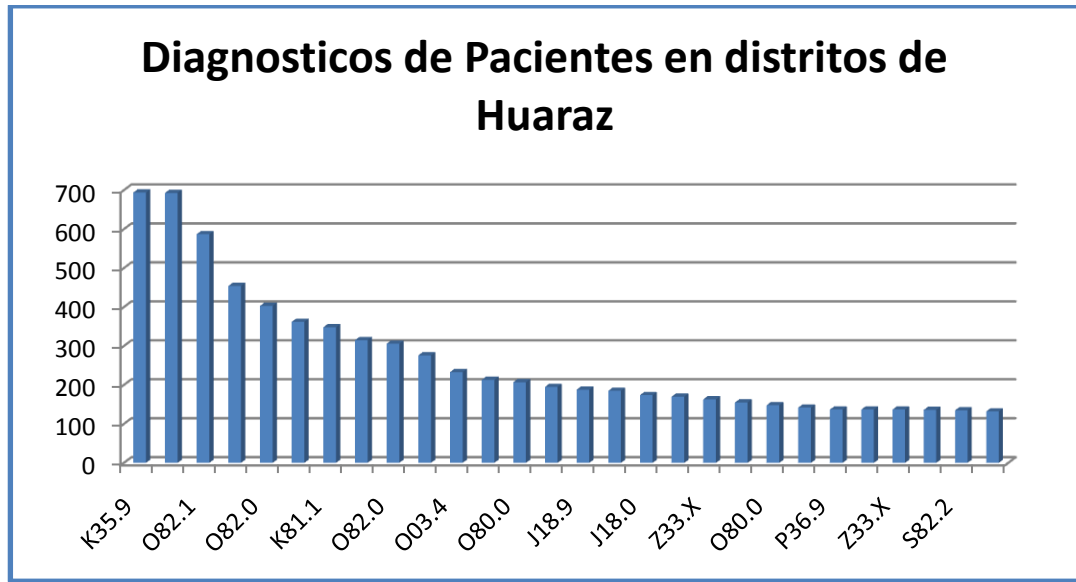


Figura 53. Dx de pacientes en el distrito de Huaraz

Provincia	Distrito	CIE10	Diagnostico	Pacientes
HUARAZ	HUARAZ	K35.9	Apendicitis aguda, no especificada	694
HUARAZ	HUARAZ	O82.1	Parto por cesárea de emergencia	693
HUARAZ	INDEPENDENCIA	O82.1	Parto por cesárea de emergencia	587
HUARAZ	HUARAZ	A09.X	Diarrea y gastroenteritis de presunto origen infeccioso	454
HUARAZ	HUARAZ	O82.0	Parto por cesárea electiva	403
HUARAZ	INDEPENDENCIA	K35.9	Apendicitis aguda, no especificada	362
HUARAZ	HUARAZ	K81.1	Colecistitis crónica	348
HUARAZ	INDEPENDENCIA	A09.X	Diarrea y gastroenteritis de presunto origen infeccioso	315
HUARAZ	INDEPENDENCIA	O82.0	Parto por cesárea electiva	306
HUARAZ	HUARAZ	O03.4	Aborto espontáneo, incompleto, sin complicación	276
HUARAZ	INDEPENDENCIA	O03.4	Aborto espontáneo, incompleto, sin complicación	233
HUARAZ	HUARAZ	S06.9	Traumatismo intracraneal, no especificado	213
HUARAZ	TARICA	O80.0	Parto único espontáneo, presentación cefálica de vértice	207
HUARAZ	INDEPENDENCIA	K81.1	Colecistitis crónica	195

Anexos

HUARAZ	HUARAZ	J18.9	Neumonía, no especificada	188
HUARAZ	HUARAZ	N39.0	Infección de vías urinarias, sitio no especificado	185
HUARAZ	HUARAZ	J18.0	Bronconeumonía, no especificada	174
HUARAZ	INDEPENDENCIA	J18.0	Bronconeumonía, no especificada	170
HUARAZ	INDEPENDENCIA	Z33.X	Estado de embarazo, incidental	163
HUARAZ	HUARAZ	T65.9	Efecto tóxico de sustancia no especificada	155
HUARAZ	JANGAS	O80.0	Parto único espontáneo, presentación cefálica de vértice	148
HUARAZ	HUARAZ	N40.X	Hiperplasia de la próstata	142
HUARAZ	HUARAZ	P36.9	Sepsis bacteriana del recién nacido, no especificada	137
HUARAZ	HUARAZ	O23.4	Infección no especificada de vías urinarias del embarazo	137
HUARAZ	HUARAZ	Z33.X	Estado de embarazo, incidental	137
HUARAZ	HUARAZ	Z53.0	Procedimiento no realizado por contraindicación	136
HUARAZ	HUARAZ	S82.2	Fractura de la diáfisis de la tibia	135
HUARAZ	HUARAZ	I64.X	Accidente vascular encefálico agudo, no especificado como hemorrágico o isquémico	132

LISTA DE REFERENCIAS

- Ben, D. K., & Sarosi, G. J. (2010). Appendicitis. *Gastrointestinal and Liver Disease*, Cap. 116.
- Ben-David K, S. G. (2010). Appendicitis. *Gastrointestinal and Liver Disease*, Cap. 116.
- BERRY, M. a. (USA: 1997). Data Mining Techniques. *John Wiley & Sons*, 5.
- Box G.E.P. & Jenkins, G. (1973). Some comments on a paper by Chatfield and Prothero and on a review by Kendall. *Journal of the Royal Statistical Society*, Series A (General), 136(3), 337-352.
- C. Gennings, R. E. (2011). "Linking empirical estimates of body burden of environmental chemicals and wellness using NHANES data".
- C. Gennings, R. E. (2011). Linking empirical estimates of body burden of environmental chemicals and wellness using NHANES data.
<http://dx.doi.org/10.1016/j.envint.2011.09.002,2011>.
- C. Gennings, R. E. (02 de 09 de 2011). *Linking empirical estimates of body burden of environmental chemicals and wellness using NHANES data*. Obtenido de dx.doi.org: <http://dx.doi.org/10.1016/j.envint.2011.09.002>.
- C. MacGregor, C. C. (2012). A Process Mining Driven Framework for Clinical guideline improvement in critical care. *Artificial Intelligence in Medicine (LEMEDS)*. (pág. Vol. 765). Learning from Medical Data Streams 13th Co.
- C.H.Jena, C. W. (2012). Application of classification techniques on development an early-warning systemfor chronic illnesses. *Expert Systems with Applications*, 8852-8858.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1997). Discovering Data Mining From concept to implementation. *Prentice Hall*.
- Chaudhuri & Dayal Chawatte, G. M. (1997).
- D. Hand, H. M. (2001). "Principles of data mining". *MIT*.
- Douglas A. Lind, W. G. (2008). *Estadística Aplicada a los Negocios y la Economía*. 213-220.

- E. Simoundis, B. L. (1996). Integrating Inductive and Deductive Reasoning for Data Mining. *Advances in Knowledge Discovery and Data Mining*, Pag. 353-373. Elsevier. (s.f.). Decision Support Journal. *North Holland Publications*.
- Española, R. A. (2012). Morbilidad. *Diccionario lengua Española*.
- Fayyad, U. P.-S. (1996). *Advances in Knowledge Discovery and Data Mining*. pag. 12-16.
- Fayyad., U. (1996). *Advanced in Knowledge Discovery and Data Mining*. MIT Press, MA.
- Frank, H. W. (2000). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco CA: Morgan Kaufmann.
- Frawley, G. P.-S. (1991). *Knowledge Discovery in Databases*. Cambridge, MA.
- Frawley, W., Piatetsky, G., & Matheus, C. (1991). *Knowledge Discovery in Databases*. Cambridge, MA.
- G., B. M. (1997). *Data Mining Techniques for Marketing, Sales and Customer Support*. John Wiley NY.
- G.S.Maddala. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge.
- Gervas, D. J. (2010). *Acta Sanitaria*.
- Gracia, M. P. (2004). *Previsión de la evolución de un paciente*.
- HAND, D. M. (USA: 2001). *Principles of Data Mining*. The MIT Press, 1.
- Hernandez J., R. M. (2004). *Introduccion a la Minería Datos*. Pearson Educacion, pp. 680.
- Hernandez, J., Ramirez , M., & Ferri, C. (2004). *Introduccion a la Minería Datos*. Pearson Educacion, pp. 680.
- <http://www.crisp-dm.org/>. (s.f.). Obtenido de <http://www.crisp-dm.org/>
- Inmon, W. H. (2005). *Building the Data Warehouse*. 4ta Edicion.
- J. Yanqing, H. Y. (2011). *Mining Infrequent Causal Associations in Electronic Health Databases*. *International Conference on Data Mining Workshops*. 11th IEEE.

- Kamber, J. H. (2006). "Data Mining: Concepts and techniques". *The Morgan Kaufmann Series*.
- KANTARDZIC, M. (USA: 2001). Data mining: concepts, models, methods, and algorithms. *Wiley - Interscience*, 2, 117, 165.
- L. Lenert, A. L. (2010). *Clustering in the Service of the Public's Health*. Obtenido de www.stat.stanford.edu:
<http://www.stat.stanford.edu/~olshen/manuscripts/helsinki.pdf>
- Lee, R., McCosh, A., & Migliarese, P. (1988). Decision Support Journal. *North Holland Publications*.
- Lee, R., McCosh, A., & Migliarese, P. (1988). Organizational Decision Support Systems. *North Holland Publications*.
- Lu, K. F. (2009). BBN-Based Decision Support for Health Risk Analysis. *Fifth International Joint Conference on INC, IMS and IDC*.
- M. Shouman, T. T. (2012). "Applying K-Nearest Neighbour in Diagnosing Heart Disease Patients". *International Conference on Knowledge Discovery*. ICKD-2012.
- Mansilla, C. C. (2014). *Modelo Predictivo del Servicio de Salud de Consultas de urgencia general y por causa respiratoria*. Chile.
- MENA, J. (USA: 1999). Data Mining Your. *Website Digital Press*, 5.
- Moxon, B. (01 de 08 de 1996). *Defining Data Mining*. (Vol. 9). San Francisco, USA.: Miller Freeman, Inc.
- Moxon's, B. (1996). Defining Data Mining. *DBMS On Line*.
- O. Er, N. Y. (2010). "Chest diseases diagnosis using artificial neural networks". *vol 37 Expert Systems with Applications*, 7648-7655.
- O. Er, N. Y. (2010). Chest diseases diagnosis using artificial neural networks. *Expert Systems with Applications*, 7648-7655.
- O. Er, N. Y. (2010). Chest diseases diagnosis using artificial neural networks. *Expert Systems with Applications*, 37, 7648-7655.
- P. Cabena, P. H. (1997). Discovering Data Mining From concept to implementation. *Prentice Hall*.

- R.S.Michalski, I. B. (1998). Machine Learning and Data Mining, Methods and Applications.
- Ralph, M. S. (1996). Principles of Information Systems: a Management Approach. *Thompson Publishing (2da ed.)*.
- Ramirez, D. C. (2012). *Diagnostico Fisico Funcional - Plan Hospitalario Integral*. Recuperado el 01 de Mayo de 2014, de <http://www.minsa.gob.pe>: http://www.minsa.gob.pe/portal/Comisiones/Hospitales/archivos/Diagnostico_CPH.pdf
- Rodriguez, J. E. (25 de 01 de 2012). *Development of tools for data mining, "UDMiner"*. Mexio: Vinculos.
- Sanchez, G. M. (2005). Indicadores de Gestion hospitalaria. *Instituto Nacional de Enfermedades Respiratorias*, Vol. 18. Pag. 2.
- Shouman, M., Turner, T., & Stocker, R. (2012). "Applying K-Nearest Neighbour in Diagnosing Heart Disease Patients". *International Conference on Knowledge Discovery*. ICKD-2012.
- Simoundis, E., Livezey, B., & Kerber, R. (1996). Integrating Inductive and Deductive Reasoning for Data Mining. *Advances in Knowledge Discovery and Data Mining*, Pag. 353-373.
- Sivanandam, S. S. (2006). Introduction to Data Mining and its Applications. *Studies in Computational Intelligence*.
- Sivanandam, S., & Sumathi, S. (2006). Introduction to Data Mining and its Applications. *Studies in Computational Intelligence*.
- T. H. A. Soliman, A. A. (2010). "A Gene Selection Approach for Classifying Diseases Based on Microarray Datasets". *nd International Conference on Computer Technology and Development*. ICCTD 2010.
- Tan, H. C. (2005). Data Mining Application in Healthcare. *Journal of Healthcare Information Management*, Vol. 19, No 2.
- Taylor, N. C. (2000). "An Introduction to Support Vector Machines and other Kernel - Based learning methods". *Cambride University Press*.
- Taylor, N. C.-T. (2000). "An Introduction to Support Vector Machines". *Cambridge University Press*.

- Thuraisingham, B. (1999). *Data Mining Technologies Techniques Tools and Trends*. CRC Press.
- Toshniwal, B. M. (2010). "Association rule for classification of type-2 diabetic patients". *Second International Conference on Machine Learning and Computing*.
- U. Fayyad, G. P.-S. (1996). The KDD process of extracting useful knowledge form volumnes of data commun. *ACM, vol. 39, no. 11, 27-34*.
- U. Fayyad, G. P.-S. (1996). The KDD process of extracting useful knowledge form volumnes of data commun. *ACM, vol. 39, no. 11, 27-34*.
- U. Fayyod, G. P.-S. (1996). "From Data Mining Knowledge discovery in Databases". *Commun. ACM vol 39, no. 11, 24-26*.
- U.M.Fayyad, G. P.-S. (1996). *Advances in knowledge Discovery and Data Mining*. AAAI/MT Press.
- V., V. (1998). *Statistical Learning Theory*. Wiley.
- V., V. (1998). The support vector method of function estimation.
- Vapnik, V. (1998). "Statistical Learning Theory". Wiley.
- Vapnik, V. (1998). "The support vector method of function estimation".
- Vermorel, J. (01 de 01 de 2012). <http://www.lokad.com/>. Recuperado el 01 de 07 de 2014, de <http://www.lokad.com/>: <http://www.lokad.com/es/que-es-el-pronostico-de-series-de-tiempo>
- W.J.Frawley, G. P.-S. (1991). *Knoledge Discovery in Databases*. Cambridge.
- Witten, I., & Frank, E. (2000). *Data Mining: Practical Machine Larning Tools and Techniques with Java Implementations*. San Francisco CA: Morgsn Kaufmann.
- Wong SM, e. a. (2004). Carpal tunnel syndrome: diagnostic usefulness of sonography. *Radiology 2004; 231(1), 93-99*.
- Wong, S. (2004). Carpal tunnel syndrome: diagnostic usefulness of sonography. *Radiology, 93-99*.

Wong, S. (2004). Carpal tunnel syndrome: diagnostic usefulness of sonography.
Radiology 2004, 1(231), 93-99.