



EDUCACIÓN

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO NACIONAL DE MÉXICO

Instituto Tecnológico de Orizaba

“2020, Año de Leona Vicario, Benemérita Madre de la Patria”

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

OPCIÓN I.- TESIS

TRABAJO PROFESIONAL

“ANÁLISIS DE ENCUESTAS SOBRE LA DISMINUCIÓN
DE AUTOPSIAS EN HOSPITALES MEXICANOS USANDO
TÉCNICAS DE MINERÍA DE DATOS”

QUE PARA OBTENER EL GRADO DE:
MAESTRA EN SISTEMAS
COMPUTACIONALES

PRESENTA:

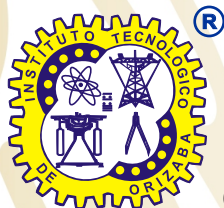
I.S.C. Ingrid Aylin Ríos Méndez

DIRECTOR DE TESIS:

Dra. Lisbeth Rodríguez Mazahua

CODIRECTOR DE TESIS:

M.C. Silvestre Gustavo Sergio Peláez Camarena



ORIZABA, VERACRUZ, MÉXICO.

NOVIEMBRE 2020



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

Orizaba, Veracruz, 13/11/2020
Dependencia: División de Estudios de
Posgrado e Investigación
Asunto: Autorización de Impresión
OPCION: I

C. INGRID AYLIN RÍOS MÉNDEZ
Candidato a Grado de Maestro en:
SISTEMAS COMPUTACIONALES
P R E S E N T E.-

De acuerdo con el Reglamento de Titulación vigente de los Centros de Enseñanza Técnica Superior, dependiente de la Dirección General de Institutos Tecnológicos de la Secretaría de Educación Pública y habiendo cumplido con todas las indicaciones que la Comisión Revisora le hizo respecto a su Trabajo Profesional titulado:

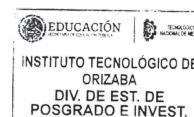
" ANÁLISIS DE ENCUESTAS SOBRE LA DISMINUCIÓN DE AUTOPSIAS EN HOSPITALES MEXICANOS USANDO TÉCNICAS DE MINERÍA DE DATOS"

Comunico a Usted que este Departamento concede su autorización para que proceda a la impresión del mismo.

ATENTAMENTE

Excelencia en Educación Tecnológica®
CIENCIA - TÉCNICA - CULTURA®

DR. MARIO L. ARRIOJA RODRÍGUEZ
JEFE DE LA DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN





"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

FECHA: 22/07/2020

ASUNTO: Revisión del Trabajo Escrito

C. MARIO LEONCIO ARRIJOJA RODRIGUEZ
JEFE DE LA DIVISION DE ESTUDIOS
DE POSGRADO E INVESTIGACION
P R E S E N T E.

Los que suscriben miembros del jurado, han realizado la revisión de la Tesis del (la) C.:

INGRID AYLIN RÍOS MÉNDEZ
No. DE CONTROL: M13011220

La cual lleva el título de:

**"ANÁLISIS DE ENCUESTAS SOBRE LA DISMINUCIÓN DE AUTOPSIAS EN HOSPITALES
MEXICANOS USANDO TÉCNICAS DE MINERÍA DE DATOS"**

Y concluyen que se acepta.

A T E N T A M E N T E

PRESIDENTE: DRA. LISBETH RODRÍGUEZ MAZAHUA

SECRETARIO: M.C. S. GUSTAVO PELÁEZ CAMARENA

VOCAL: M.C. MA. ANTONIETA ABUD FIGUEROA

VOCAL SUP.: DR. GINER ALOR HERNÁNDEZ

[Handwritten signatures and names]

ma. Antonieta Abud F.
Giner Alor Hernández

EGRSADO (A) DE LA MAESTRIA EN **SISTEMAS COMPUTACIONALES**

OPCION: **1 Tesis**



Agradecimientos

A Dios.

Por darme esta oportunidad y todo lo necesario para llegar a culminar mi maestría, por ser mi fortaleza en todo momento, por su infinito amor y misericordia.

A mi madre Xóchitl.

Por apoyarme, guiarme y motivarme cuando estaba a punto de tirar la toalla, por las noches de desvelo enseñándome o explicándome lo que no entendía, por ser mi amiga, confidente y gran ejemplo, por hacerme independiente dejándome tomar mis propias decisiones, por enseñarme que todo es posible cuando Dios está en nuestras vidas.

A mi padre César.

Por ser un hombre muy trabajador, comprensivo y de un corazón noble, por ser mi motivación cuando siento que no puedo, por haber mostrado fortaleza cuando no tenía, con tal de que mis hermanos y yo estemos bien, por ser un hombre ejemplar, por darme estudios y amor.

A mis hermanos.

Por motivarme para ser un buen ejemplo para ellos, por decirme palabras de aliento cuando no me sentía capaz de hacer algo y por todo el amor que me dan.

A mis abuelos.

Camilo, por haber sido un gran pilar para toda la familia, por acercarnos a Dios y hacer que nuestra fe creciera, por motivarme a estudiar y siempre confiar en mí, por cuidarme, por enseñarme a ser bondadosa, por la fuerza y amor que demostró cuando el mundo se le venía encima, por volver mis sueños suyos.

Olivia, por cuidarme, por estar pendiente de mí, darme amor y cariño.

Luis, por las largas caminatas y charlas descubriendo paisajes y caminos nuevos, por el amor y dedicación de tiempo.

Silvia, por cuidarme, por nunca negarme nada, por su amor y cariño.

A mi asesora de tesis.

Lisbeth por el apoyo que me dio durante el desarrollo de mi tesis y por estar siempre pendiente de mi trabajo.

También quiero agradecer al Dr. Palet por el apoyo y contribución en este proyecto.

A mi mejor amiga.

Sendy Cruz por apoyarme en momentos difíciles y personales, por el apoyo y buen equipo durante el tiempo de conocernos, por hacer todo un poco más ligero, por todas las risas y carcajadas, por siempre motivarme y confiar en mí.

Índice General

Capítulo 1. Antecedentes	11
Marco Teórico	11
1.1.1 Minería de datos	11
1.1.2 Tareas de la minería de datos	11
1.1.2.1 Tareas descriptivas	11
1.1.2.2 Tareas predictivas	13
1.1.2.3 Descubrimiento supervisado de reglas descriptivas	14
1.1.3 Autopsia	16
1.2 Situación tecnológica, económica y operativa de la empresa	16
1.3 Planteamiento del problema	17
1.4 Objetivo general y específicos	18
Objetivo general	18
Objetivos específicos	18
1.5 Justificación	19
Capítulo 2. Estado de la práctica	20
2.1 Trabajos relacionados	20
2.2 Análisis Comparativo	29
2.3 Propuesta de solución	41
Capítulo 3. Aplicación de la metodología	43
3.1 Análisis de requerimientos	44
3.2 Diseño de la aplicación	52
3.2.1 Arquitectura	53
3.2.2 Modelo conceptual	54
3.2.3 Modelo de navegación	58
3.2.4 Modelo de presentación	59
3.2.5 Modelo de procesos	64
3.3. Desarrollo de los modelos de minería de datos utilizando KDD	68
3.3.1 Selección de Datos	69
3.3.2 Pre-procesamiento de Datos	71
3.3.3 Minería de Datos	73
3.3.3.1 Análisis de minería de patrones emergentes	73
Capítulo 4. Resultados	95

4.1 Presentación del sistema y caso de estudio	96
4.2 Evaluación de los resultados	110
Capítulo 5. Conclusiones y recomendaciones	111
5.1. Conclusiones	111
5.2. Recomendaciones	111
Productos Académicos	113
Referencias	114
Anexos	117
Gráficas de las encuestas.	117

Índice de Tablas

Tabla 2.1 Análisis comparativo de trabajos relacionados.	29
Tabla 3.1 Actores de la aplicación Web	44
Tabla 3.2 Resumen de la encuesta aplicada al personal médico	70
Tabla 3.3 Características de los conjuntos de datos C y D.....	73
Tabla 3.4 Aplicación de los algoritmos en los conjuntos de datos.....	77
Tabla 3.5 Clase Med_aut	77
Tabla 3.6 Comparación de algoritmos de la clase Med_aut.....	78
Tabla 3.7 Comparación considerando la confianza.....	78
Tabla 3.8 EPs obtenidos con DGCP-Tree para Med_aut.....	79
Tabla 3.9 Clase Per_sol_aut	80
Tabla 3.10 Comparación de algoritmos de la clase Per_sol_aut	80
Tabla 3.11 Comparación considerando la confianza.....	81
Tabla 3.12 EPs obtenidos con DGCP-Tree para Per_sol_aut	81
Tabla 3.13 Clase No_hosp.....	82
Tabla 3.14 Comparación de algoritmos de la clase No_hosp.....	82
Tabla 3.15 Comparación considerando la confianza.....	83
Tabla 3.16 EPs obtenidos con iEPMiner para No_hosp.....	84
Tabla 3.17 Clase Rechazo_fam	85
Tabla 3.18 Comparación de algoritmos de la clase Rechazo_fam.....	85
Tabla 3.19 Comparación considerando confianza.....	86
Tabla 3.20 EPs obtenidos con DGCP-Tree para Rechazo_fam	86
Tabla 3.21 Clase Mcc_no_aut	87
Tabla 3.22 Comparación de algoritmos de la clase Mcc_no_aut.....	88
Tabla 3.23 Comparación considerando confianza.....	88
Tabla 3.24 EPs obtenidos con DGCP-Tree para Mcc_no_aut	89
Tabla 3.25 Clase Mcc_aut	90
Tabla 3.26 Comparación de algoritmos de la clase.....	90
Tabla 3.27 Comparación considerando confianza.....	91
Tabla 3.28 EPs obtenidos con DGCP-Tree para Mcc_aut	91
Tabla 3.29 Los mejores algoritmos para cada atributo de clase.....	92
Tabla 3.30 Comparación para la clase del hospital {1,2}	93
Tabla 3.31 Comparación considerando confianza.....	93
Tabla 3.32 EPs obtenidos con iEPMiner	93
Tabla 3.33 EPs obtenidos con DGCP-Tree	94

Índice de Figuras

Figura 3.1 Esquema de solución	43
Figura 3.2 Diagrama de casos de uso de la aplicación web	45
Figura 3.3 Diagrama de actividad del caso de uso Iniciar Sesión	46
Figura 3.4 Diagrama de actividad del caso de uso Cerrar Sesión.	46
Figura 3.5 Diagrama de actividad del caso de uso Responder Encuesta.	47
Figura 3.6 Diagrama de actividad del caso de uso Consultar Encuesta	47
Figura 3.7 Diagrama de actividad del caso de uso Eliminar Encuesta	48
Figura 3.8 Diagrama de actividad del caso de uso Generar Modelo	49
Figura 3.9 Diagrama de actividad del caso de uso Interpretar Resultados de Minería	49
Figura 3.10 Diagrama de actividad del caso de uso Buscar centro estudio y Agregar centro estudio	50
Figura 3.11 Diagrama de actividad del caso de uso Insertar Usuario.	50
Figura 3.12 Diagrama de actividad del caso de uso Consultar Usuario.	51
Figura 3.13 Diagrama de actividad del caso de uso Actualizar Usuario	51
Figura 3.14 Diagrama de actividad del caso de uso Eliminar Usuario.	52
Figura 3.15 Arquitectura propuesta	53
Figura 3.16 Diagrama conceptual de la aplicación	55
Figura 3.17 Diagrama lógico de la aplicación	55
Figura 3.18 Diagrama físico de la aplicación	56
Figura 3.19 Diagrama de navegación de la aplicación Web	59
Figura 3.20 Página Inicio del modelo de presentación	60
Figura 3.21 Página Gestionar Encuesta del modelo de presentación	60
Figura 3.22 Página Encuesta del modelo de presentación	61
Figura 3.23 Página Generar Modelos del modelo de presentación	62
Figura 3.24 Página Resultados EPM del modelo de presentación	63
Figura 3.25 Área alternativa Parámetros	63
Figura 3.26 Área alternativa Resultados_EPM	64
Figura 3.27 Diagrama de actividad del caso de uso Responder Encuesta	65
Figura 3.28 Diagrama de actividad del caso de uso Consultar Encuesta	66
Figura 3.29 Diagrama de actividad del caso de uso Eliminar Encuesta	67
Figura 3.30 Diagrama de actividad del caso de uso Generar Modelo	67
Figura 3.31 Diagrama de actividad del caso de uso Interpretar Resultados de Minería	68
Figura 3.32 Proceso de descubrimiento del conocimiento	68
Figura 3.33 Conjunto de datos de vista_minable	72
Figura 3.34 Conjunto de datos de matriz_binaria	72
Figura 4.1 Página de inicio	96
Figura 4.2 Crear Encuesta	96
Figura 4.3 Formulario de Datos	96
Figura 4.4 Verificación de datos	97
Figura 4.5 Datos del centro de estudio	97
Figura 4.6 Formulario de Encuesta	98

Figura 4.7 Respuesta Clasificadas	99
Figura 4.8 Funciones del especialista	99
Figura 4.9 Modelo de iEPMiner, conjunto D, clase Fmr_sol_aut	100
Figura 4.10 Modelo de iEPMiner, conjunto D, clase Med_aut	101
Figura 4.11 Modelo de iEPMiner, conjunto D clase, Per_sol_aut	101
Figura 4.12 Modelo iEPMiner, conjunto D, clase No_hosp	101
Figura 4.13 Modelo iEPMiner, conjunto D, clase Rechazo_fam	102
Figura 4.14 Modelo iEPMiner, conjunto D, clase Mcc_no_aut	102
Figura 4.15 Modelo iEPMiner, conjunto D, clase Mcc_aut	102
Figura 4.16 Modelo DGCP-Tree, conjunto D, clase Fmr_sol_aut	103
Figura 4.17 Modelo DGCP-Tree, conjunto D, clase Med_aut	103
Figura 4.18 Modulo DGCP-Tree, conjunto D, clase Per_sol_aut	103
Figura 4.19 Modelo DGCP-Tree, conjunto D, clase No_hosp	103
Figura 4.20 Modulo DGCP-Tree, conjunto D, clase Rechazo_fam	104
Figura 4.21 Modelo DGCP-Tree, conjunto D, clase Mcc_no_aut	104
Figura 4.22 Modulo DGCP-Tree, conjunto D, clase Mcc_aut	105
Figura 4.23 Clase 1 y 2	105
Figura 4.24 Modelo iEPMiner, conjunto C	105
Figura 4.25 Modelo DGCP-Tree, conjunto C	106
Figura 4.26 Gestionar Encuesta	106
Figura 4.27 Filtrar encuesta	106
Figura 4.28 Consultar encuesta.	107
Figura 4.29 Eliminar encuesta	107
Figura 4.30 Cerrar Sesión	108
Figura 4.31 Resultados	108
Figura 4.32 Reglas con iEPMiner y conjunto C	109
Figura 4.33 Reglas con DGCP-Tree y conjunto C	109
Figura 4.34 Reglas con iEPMiner, conjunto D y la clase Med_aut	110
Figura 4.35 Reglas con DGCP-Tree , conjunto D y la clase Per_sol_aut	110
Figura 1 Último grado de estudios y Área a la que pertenece	117
Figura 2 Años de práctica y número de participación u observación de autopsias	117
Figura 3 Hallazgos encontrados en las autopsias	118
Figura 4 Principal causa de rechazo por parte del familiar para realizar una autopsia	119
Figura 5 Causas por las que no se realizan suficientes autopsias en su hospital	120
Figura 6 Solicitud de autopsia	120
Figura 7 Solicitud de autopsias por parte del médico	121
Figura 8 Solicitud de estudio de autopsias	121

Resumen

La alarmante disminución en la realización de autopsias tuvo como consecuencia el interés del director del área de patología del Hospital Regional Río Blanco (H. R. R. B.) por analizar cuáles son las razones, motivos o circunstancias por las que cada vez se realizan menos autopsias en su hospital. Por lo tanto, se desarrolló un sistema para el análisis de encuestas usando minería de datos en un proyecto anterior [1]. El sistema permite aplicar técnicas predictivas, específicamente la clasificación, así como descriptivas, minería de reglas de asociación. Ahora, se quiere analizar los resultados de las encuestas en otro hospital y aplicar otras tareas descriptivas de la minería de datos, como la caracterización y la discriminación, para comparar los resultados obtenidos en el H. R. R. B. con los de otro hospital. La importancia de este análisis radica en que será posible determinar si las causas por las que ya no se realizan autopsias en el H. R. R. B. son generales o son diferentes en otro hospital; esto permitirá que el área de patología de los hospitales tome acciones que logren incrementar el número de autopsias realizadas.

En este proyecto se agregó un módulo al sistema, el cual se desarrolló con base en la metodología UWE (*UML-based Web Engineering*, Ingeniería Web basada en UML), y utilizando el lenguaje de programación Java, el *framework* JavaServer Faces y el Sistema Gestor de Bases de Datos PostgreSQL; además, se siguió el proceso KDD (*Knowledge Discovery from Data*, Descubrimiento de Conocimiento a partir de los Datos) para la minería de datos y se usó la plataforma de aprendizaje automático EPM_Algorithms para generar los modelos descriptivos.

Abstract

The alarming decrease in the performance of autopsies resulted in the interest of the director of the pathology area of the Regional Hospital of Río Blanco (H.R.R.B.) to analyze the reasons, motives, or circumstances for which fewer autopsies are performed in his hospital. Therefore, a system for survey analysis using data mining was developed in a previous project [1]. The system allows applying predictive techniques, specifically classification, as well as descriptive, association rule mining. Now, we want to analyze the results of the surveys in another hospital and apply other descriptive tasks of data mining, such as characterization and discrimination, to compare the results obtained in the H. R. R. B. with those of another hospital. The importance of this analysis lies in the fact that it will be possible to determine if the reasons for which autopsies are no longer performed in the H. R. R. B. are general or are different in another hospital; this will allow the pathology area of hospitals to take actions that will increase the number of autopsies performed.

In this project a module was added to the system, which was developed based on the UWE methodology (UML-based Web Engineering), and using the Java programming language, the JavaServer Faces framework and the PostgreSQL Database Management System; in addition, the KDD (Knowledge Discovery from Data) process for data mining was followed and the machine learning platform EPM_Algorithms was used to generate the descriptive models.

Introducción

El Hospital Regional de Río Blanco (H. R. R. B.) es uno de los centros médicos más importantes de la región. Actualmente brinda atención médica a más de 735 mil pacientes de los 57 municipios más cercanos como son Río Blanco, Orizaba, Córdoba, Coscomatepec, Zongolica, Nogales y Ciudad Mendoza, por mencionar algunos, e incluso de otros estados como Oaxaca y Puebla. Debido a la reducción considerable del número de autopsias realizadas en el H. R. R. B., el departamento de Patología del hospital tuvo la necesidad de determinar las razones por las que ya casi no se realiza este estudio, por tal motivo, en un proyecto anterior se desarrolló un sistema para el análisis de encuestas sobre la disminución de autopsias en el H. R. R. B. utilizando minería de datos. El sistema utiliza minería de reglas de asociación y redes Bayesianas para analizar las encuestas. Las técnicas de minería de datos usadas permitieron determinar que las razones principales por las que no se realizan autopsias de acuerdo a los médicos del H. R. R. B. son: 1) Enfermedad de base conocida, 2) Negativa de los familiares, 3) Desinterés, 4) Factores sociales, 5) Temor a la demanda, 6) Cuestiones culturales, 7) Cuestiones religiosas, 8) Falta de indicación, 9) Falta de recursos humanos y 10) Falta de recursos materiales. En este proyecto se aplicaron técnicas descriptivas de la minería de datos, específicamente la minería de patrones emergentes, para comparar los motivos por los que no se practican autopsias en el H. R. R. B. con los resultados de otro hospital.

Este documento contiene cinco capítulos. En el primero se presenta el marco teórico, así como la situación tecnológica, económica y operativa del H. R. R. B., el planteamiento del problema, los objetivos y la justificación. En el segundo capítulo se describen algunos trabajos relacionados con este proyecto y se muestra un análisis comparativo. El tercer capítulo trata sobre la aplicación de la metodología, en donde se encuentra el diseño de la aplicación, arquitectura, análisis de requerimientos, desarrollo de los modelos de minería de datos utilizando KDD, entre otros puntos, en el capítulo cuatro se muestran los resultados, se presenta el sistema y caso de estudio, así como la evaluación de resultados y finalmente en el último capítulo se encuentran las conclusiones y recomendaciones.

Capítulo 1. Antecedentes

En este capítulo se presentan los conceptos básicos de este proyecto, el planteamiento del problema, los objetivos y la justificación.

Marco Teórico

En esta sección se presentan los conceptos más importantes de este trabajo de investigación.

1.1.1 Minería de datos

La minería de datos es el proceso de descubrir automáticamente información útil en grandes repositorios de datos. Las técnicas de minería de datos se implementan para examinar grandes conjuntos de datos con el fin de encontrar patrones nuevos y útiles que de otro modo permanecerían desconocidos. También brindan la capacidad de predecir el resultado de una observación futura, como la cantidad que un cliente gastará en una tienda en línea o en una tienda física. La minería de datos solo es una fase o etapa del proceso de KDD [2].

1.1.2 Tareas de la minería de datos

La clasificación inicial de las tareas de minería de datos se distingue entre predictivas, en las que las variables se clasifican en dependientes e independientes (similares a las técnicas del análisis de la dependencia o métodos explicativos del análisis multivariante), descriptivas, en las que todas las variables tienen inicialmente el mismo estatus (similares a las técnicas del análisis de la interdependencia o métodos descriptivos del análisis multivariante) y técnicas auxiliares [3].

1.1.2.1 Tareas descriptivas

Las tareas descriptivas son características generales de las propiedades de los datos, no se asigna ningún papel predeterminado a las variables. No se supone la existencia de variables dependientes ni independientes y tampoco de un modelo previo para los datos. Los modelos se crean automáticamente partiendo del reconocimiento de patrones, aquí se incluyen las técnicas de *clustering* y segmentación, las técnicas de asociación y dependencia, las técnicas de análisis

exploratorio de datos y las técnicas de reducción de dimensión y de escalamiento multidimensional [2].

Caracterización de datos

Es un resumen de las características generales de las clases de datos objetivo. Los datos correspondientes a las clases especificadas por el usuario normalmente se recolectan por medio de una consulta.

Hay varios métodos para realizar el resumen y la caracterización de datos de manera efectiva. Algunos son resúmenes de datos simples basados en medidas estadísticas y gráficos, es posible usar la operación *roll-up* de OLAP (*On-Line Analytical Processing*, Procesamiento Analítico en Línea) basada en cubos de datos para realizar resúmenes de datos controlados por el usuario a través de una dimensión específica [2].

Discriminación de datos

Han, Kamber y Pei [2] definen esta tarea como la comparación de la clase/concepto bajo estudio con otras clases/conceptos (clases contrastantes). Las clases objetivo y las contrarias se especifican por el usuario, y los objetos de datos correspondientes se recuperan de la base de datos por medio de consultas.

Las descripciones de discriminación incluyen medidas comparativas que ayudan a distinguir entre las clases objetivo y las clases contrarias, se expresan en forma de reglas y se refieren como reglas de discriminación.

Minería de reglas de asociación

La minería de reglas de asociación es un ejemplo donde el uso de restricciones y medidas de interés garantizan la integridad de la minería. La minería de reglas de asociación consiste en encontrar primero conjuntos de elementos frecuentes (conjuntos de elementos, como A y B, que satisfacen un umbral de soporte mínimo o porcentaje de las tuplas relevantes para la tarea), a partir de las cuales se generan reglas de asociación sólidas en forma de $A \Rightarrow B$.

Estas reglas también satisfacen un umbral de confianza mínimo (una probabilidad preespecificada de satisfacer B bajo la condición de que A se cumpla). Las asociaciones se analizan más a fondo para descubrir reglas de correlación, que transmiten correlaciones estadísticas entre los conjuntos de elementos A y B.

En general, la minería de reglas de asociación es un proceso de dos pasos [2]:

1. Encontrar todos los conjuntos de elementos frecuentes: Por definición, cada uno de estos conjuntos de elementos se producirá al menos con la misma frecuencia que un recuento de soporte mínimo predeterminado, *min_sup* (*minimum support*).
2. Generar reglas de asociación sólidas a partir de los conjuntos de elementos frecuentes: Por definición, estas reglas satisfacen el soporte mínimo y la confianza mínima.

Agrupamiento

El agrupamiento analiza objetos de datos sin consultar una etiqueta de clase conocida. En muchos casos, las etiquetas de clase simplemente no existen al inicio. El agrupamiento se usa para generar etiquetas de clase para un grupo de datos. Los objetos se agrupan con base en el principio de maximizar la similitud dentro de las clases y minimizar la similitud entre las clases. Esto es, se forman cúmulos de objetos, de tal forma que los objetos en un cúmulo tengan una similitud alta, pero sean muy diferentes a los objetos de otros cúmulos. Cada cúmulo que se forma se ve como una clase de objetos, del que se derivan reglas. El agrupamiento también facilita la formación de taxonomías, esto es, la organización de las observaciones en una jerarquía de clases que agrupa juntos a los eventos similares [2].

1.1.2.2 Tareas predictivas

Las tareas predictivas consisten en inferir algo para clasificar/predecir. Especifican el modelo para los datos con base en un conocimiento teórico previo, el modelo figurado para los datos debe contrastarse después del proceso de minería de datos antes de aceptarlo como válido, la aplicación de todo modelo supera las fases de identificación objetiva, estimación, diagnosis y predicción, en algunas ocasiones el modelo se adquiere con la combinación del conocimiento alcanzado antes y después de la minería de datos y también se comprueba antes de aprobar que sea válido [4].

Clasificación

La clasificación es el proceso de encontrar un modelo o función que describe y distingue clases o conceptos de datos. El modelo se origina con base en el análisis de un conjunto de datos de entrenamiento, este modelo se usa para predecir la clase de objetos cuya etiqueta de clase es desconocida. Es posible representar el modelo derivado en varias formas, como reglas de clasificación (por ejemplo, reglas IF-THEN), árboles de decisión, fórmulas matemáticas o redes neuronales [2].

Tanto los árboles de decisión, como las redes neuronales y el análisis discriminante son a su vez técnicas de clasificación capaces de extraer perfiles de comportamiento o clases, siendo el objetivo construir un modelo que permita clasificar cualquier dato nuevo. Los árboles de decisión permiten clasificar datos en grupos basados en los valores de las variables. Una red neuronal, cuando se usa para la clasificación, es normalmente una colección de unidades de procesamiento parecidas a las neuronas con conexiones con peso entre las unidades. Hay muchos otros métodos para construir modelos de clasificación, como clasificación Bayesiana ingenua, máquinas de vectores de soporte y clasificación de k -vecinos más cercanos [4].

Regresión

Mientras que la clasificación predice etiquetas categóricas (discretas, desordenadas), la regresión también conocida como predicción, predice una o más variables continuas en función del resto de los atributos. Se usa para predecir valores de datos numéricos ausentes o no disponibles en vez de etiquetas de clase. El término predicción se refiere tanto a predicción numérica como a predicción de etiquetas de clase. El análisis de regresión es una metodología estadística que con frecuencia se usa para predicción numérica, aunque también existen otros métodos. La predicción también incluye la identificación de las tendencias de distribución con base en los datos disponibles [2].

1.1.2.3 Descubrimiento supervisado de reglas descriptivas

El concepto de descubrimiento supervisado de reglas descriptivas agrupa un conjunto de técnicas de minería de datos cuyo objetivo es describir datos con respecto a una propiedad de interés. Entre las técnicas dentro de este concepto se encuentran el descubrimiento de subgrupos, los patrones emergentes y los conjuntos de contraste.

Patrones emergentes

Los EPs (*Emerging Patterns*, Patrones Emergentes) se definieron por Dong y Li [5] como conjuntos de elementos cuyo soporte (frecuencia) aumenta significativamente de un conjunto de datos (D1) a otro (D2) para descubrir tendencias en los datos, en el tiempo o la diferenciación

entre las características. Los EPs permiten capturar las tendencias emergentes en las bases de datos con estampas de tiempo, o contrastes entre las clases de datos.

La EPM (*Emerging Pattern Mining*, Minería de Patrones Emergentes) se considera un problema NP-*Hard* con respecto al número de variables. Para aproximar este problema eficientemente, los autores intentan reducir el número de patrones extraídos. Ellos tratan de tomar solo aquellos patrones que describen algunas relaciones específicas e interesantes entre las variables. Esto produce un conjunto de patrones más simple y más interesante que debería ser tan preciso como sea posible. Estos tipos de patrones permiten la descripción de un problema en una forma fácil y directa.

Los tipos de EPs más utilizados en la literatura son:

- **JEPs (*Jumping Emerging Patterns*, Patrones Emergentes de Saltos).** Estos son EPs con una tasa de crecimiento igual al infinito, es decir, el JEP cubre ejemplos para una sola clase. Ellos tienen un gran carácter diferenciador entre las clases.
- **MinEPs (*Minimal Emerging Patterns*, Patrones Emergentes Mínimos).** Un MinEP se define como un EP cuyos subpatrones no son EPs. Ellos son interesantes para la descripción porque, en general, ellos contienen un pequeño número de variables.
- **MaxEPs (*Maximal Emerging Patterns*, Patrones Emergentes Máximos).** Este tipo de patrón es el opuesto al mínimo. Un MaxEP es un EP cuyos super patrones no son EPs. Esto produce patrones muy específicos, que son muy precisos.
- **eJEPs (*Essential Jumping Emerging Patterns*, Patrones Emergentes de Saltos Esenciales).** También se conocen como SJEPs (*Strong Jumping Emerging Patterns*, Patrones Emergentes de Saltos Fuertes). Este tipo de patrón es la intersección del conjunto de los MinEPs con el conjunto de los JEPs.
- **NEPs (*Noise-Tolerant Emerging Patterns*, Patrones Emergentes Tolerantes al Ruido).** También conocidos como CEPs (*Constrained Emerging Patterns*, Patrones Emergentes Restringidos). Este tipo es muy cercano a los JEPs pero permiten obtener patrones precisos en ambientes con ruido.
- **Chi EPs (*Chi Emerging Patterns*, Patrones Emergentes Chi).** Son similares a los NEPs pero introducen una prueba estadística χ^2 con el fin de mejorar la capacidad descriptiva del patrón.

Conjuntos de contraste

Bay y Pazzani definieron la minería de conjuntos de contraste en 1999 como una tarea de análisis de datos para comprender las diferencias entre varios grupos de contraste. Los conjuntos de contraste son conjunciones de pares atributo-valor cuya distribución generalmente difiere entre grupos.

La minería de conjuntos de contraste es una de las principales áreas de investigación en minería supervisada de patrones descriptivos, también conocida como descubrimiento supervisado de reglas descriptivas debido a que los resultados generalmente se dan en forma de reglas. El contraste se realiza en conjuntos de datos que satisfacen condiciones definidas estáticamente, con el objetivo de descubrir diferencias y similitudes entre conjuntos de datos o subconjuntos de datos dados [6].

Descubrimiento de subgrupos

Es la técnica central del proceso de minería de subgrupos, y fue primero formalizada por Klösgen y Wrobel. El descubrimiento de subgrupos tiene una gran relevancia práctica en todos los ámbitos de la ciencia de los negocios.

El descubrimiento de subgrupos es adecuado para encontrar relaciones entre una variable objetivo (dependiente) y un conjunto de variables independientes. El descubrimiento de subgrupos se utiliza para la exploración y la inducción descriptiva y los subgrupos se aplican por ejemplo, para el apoyo a la toma de decisiones [7].

1.1.3 Autopsia

La autopsia es un procedimiento médico que emplea la disección, con el fin de obtener información anatómica sobre la causa, naturaleza, extensión y complicaciones de la enfermedad que sufrió en vida el sujeto autopsiado. La autopsia se considera como la última posibilidad de consulta médica, ya que la interpretación de los hallazgos *post mortem* muestran el estatus del proceso patológico en el momento de la muerte [8].

1.2 Situación tecnológica, económica y operativa de la empresa

El Hospital Regional de Río Blanco (H. R. R. B.) es uno de los centros médicos más importantes de la región. Actualmente brinda atención médica a más de 735 mil pacientes pertenecientes a

los 56 municipios que habitan en la zona centro del Estado de Veracruz, abarcando: la Sierra Madre de Huatusco, la Sierra de Zongolica, el Valle de Acultzingo, el Valle de Córdoba, el Valle de Maltrata, el Valle de Orizaba, y parte de la Población de los Estados de Puebla y Oaxaca.

El H.R.R.B son un equipo de profesionales de la salud pública, especializados en proporcionar servicios médicos de segundo nivel con calidad y sencillez, en la ciudad de Río Blanco, del Estado de Veracruz.

Dentro de sus logros está contar con un premio nacional de calidad, ser un hospital certificado y recertificado y contar con un gran capital humano quien da, diariamente, su mejor esfuerzo en la búsqueda continua de mejorar los procesos de atención con enfoque al usuario e impacto social a la otorgación de cuidados paliativos y tanatológicos.

Su oferta de productos y servicios se caracteriza por ser parte esencial en el Sistema de Salud Regional, contar con Innovación Tecnológica, establecer la Mejora Continua en procesos de Atención, y participar en la formación de especialistas e Investigadores en el Campo de las Ciencias Médicas.

1.3 Planteamiento del problema

La práctica de la autopsia contribuye de gran manera al conocimiento y aprendizaje médico, al igual que es de gran ayuda para familiares de la persona fallecida porque en ocasiones el saber las causas de muerte del fallecido ayuda en su duelo y en el caso de las enfermedades hereditarias, los familiares se beneficiarán al saber si son propensos a sufrir la misma enfermedad y así se logre dar una solución pronta para evitar la aparición de complicaciones y evitar la muerte. También ayuda en el campo académico o de investigación al ser capaz de descubrir nuevas enfermedades, ya que hay enfermedades que solo se diagnostican con certeza mediante una autopsia.

Últimamente se observó la disminución de autopsias clínicas en el Hospital Regional de Río Blanco (H.R.R.B), al saber la importancia y los beneficios de su realización, la disminución de esta práctica es una situación crítica y alarmante, esto tuvo como consecuencia el interés del

director del área de patología del H. R. R. B. por analizar cuáles son las razones, motivos o circunstancias por las que cada vez se realizan menos autopsias en su hospital.

En un trabajo anterior [1] se desarrolló un sistema para el análisis de encuestas aplicadas al personal médico del H. R. R. B sobre la disminución de autopsias. El sistema utiliza minería de reglas de asociación y redes Bayesianas para analizar las encuestas. Las técnicas de minería de datos usadas permitieron determinar que las razones principales por las que no se realizan autopsias de acuerdo a los médicos del H. R. R. B. son: 1) Enfermedad de base conocida; 2) Negativa de los familiares; 3) Desinterés; 4) Factores sociales; 5) Temor a la demanda; 6) Cuestiones culturales; 7) Cuestiones religiosas; 8) Falta de indicación; 9) Falta de recursos humanos, y 10) Falta de recursos materiales. En este proyecto se aplicó la encuesta a médicos de otro hospital y se compararon los resultados con los encontrados en el H. R. R. B. por lo que se aplicaron tareas descriptivas de la minería de datos, específicamente la minería de patrones emergentes, para encontrar las diferencias entre los dos conjuntos de datos.

1.4 Objetivo general y específicos

Objetivo general

Determinar las causas de la disminución de autopsias en hospitales mexicanos usando tareas descriptivas de la minería de datos, para que el H.R.R.B. establezca estrategias que incrementen esta práctica médica.

Objetivos específicos

1. Analizar el estado del arte de las técnicas de caracterización y discriminación para seleccionar las más adecuadas para resolver el problema.
2. Determinar las tecnologías que se utilizarán para la implementación de las técnicas descriptivas por medio de un análisis comparativo.
3. Generar los nuevos conjuntos de datos con los resultados de las encuestas aplicadas al personal médico.
4. Aplicar técnicas de caracterización y discriminación a los conjuntos de datos utilizando las tecnologías seleccionadas para establecer los mejores parámetros.

5. Desarrollar el módulo del sistema para la caracterización y discriminación, con base en la metodología seleccionada, que permita comparar los resultados encontrados en el H.R.R.B. con los de otros hospitales.
6. Validar el nuevo módulo desarrollado, usando un caso de estudio, para lograr la aprobación del jefe de la unidad de anatomía patológica del H.R.R.B.

1.5 Justificación

En años recientes, la descripción de conjuntos de datos se convirtió en un área interesante debido a la fácil adquisición de conocimiento por los expertos.

Dada la gran cantidad de datos almacenados en las bases de datos, es útil ser capaz de describir conceptos en términos concisos y breves en niveles de abstracción generalizados. Permitir que los conjuntos de datos se generalicen en múltiples niveles de abstracción facilita a los usuarios el análisis del comportamiento general de los datos.

El principal beneficio que se obtiene con el desarrollo de este proyecto es que utilizando técnicas de caracterización y discriminación, específicamente la minería de patrones emergentes, se comparan los motivos por los que no se realizan autopsias en otro hospital, para determinar si son los mismos que en el H. R. R. B., para que el personal del departamento de Patología tome medidas que permitan incrementar el número de autopsias realizadas en el hospital.

Capítulo 2. Estado de la práctica

En este capítulo se presenta una recopilación y análisis de trabajos relacionados con este proyecto de tesis.

2.1 Trabajos relacionados

En [9] se propuso una solución al problema de la detección de anomalías en los datos categóricos en un entorno semi-supervisado.

Las principales contribuciones son:

1. Diseño de un marco de trabajo de detección de anomalías para datos categóricos basados en un enfoque de aprendizaje a distancia.
2. Incorporación del algoritmo de aprendizaje a distancia en diferentes estrategias de clasificación y muestra de que el enfoque devuelve buenos candidatos atípicos para cada una de las cuatro estrategias de clasificación propuestas.
3. Comparación del método con los métodos de detección de valores atípicos semi-supervisados de vanguardia.
4. Demostración de que el método no es simplemente un método de trabajo, sino que también proporciona información explicativa sobre los datos.

Esta propuesta produjo un modelo discriminativo que se interpreta fácilmente y se utiliza para la exploración de los datos. Se utilizaron estrategias heurísticas para la selección de un conjunto representativo de K instancias del conjunto de datos (D) para el cálculo de la puntuación de valor atípico.

Como resultado se tiene un nuevo enfoque para la detección de anomalías para datos categóricos, que es capaz de modelar datos categóricos usando algoritmos basados en distancia y obtiene muy buenos resultados.

Los cambios en la preferencia de los viajeros influyen en el desempeño de las empresas hoteleras, creando así la necesidad de identificar y atender las demandas de sus huéspedes. La mayoría de los estudios existentes se centran en los atributos de la demanda actual y no en los emergentes, por lo tanto, a los gerentes de hoteles les resulta difícil tomar decisiones apropiadas en respuesta a los cambios en las preocupaciones de los viajeros.

Li et al. [10] adoptaron una técnica de EPM (*Emerging Pattern Mining*, Minería de Patrones Emergentes) para identificar características emergentes de hospedaje de interés para los viajeros internacionales.

Sus contribuciones son:

- Apuntar a llenar el vacío de investigación actual mediante la introducción de la técnica EPM para establecer características emergentes de hospedaje. El término "características de hospedaje" incluye cualquier entidad o concepto que se refiera a los viajeros al revisar un hotel.
- Identificación de las funciones emergentes que actualmente reciben más atención de los viajeros internacionales.
- Construcción de un conjunto de perfiles de usuario para ayudar a los gerentes de hoteles a mejorar las funciones disponibles en sus propiedades.
- Identificación de características específicas de las revisiones en línea.

Se utilizó EPM, la cual captura las tendencias emergentes en bases de datos con estampas de tiempo o contrastes agudos entre conjuntos de datos o grupos. Al utilizar EPM, los investigadores son capaces de identificar las características emergentes de los hoteles.

Los datos se derivaron de 118,000 registros de revisiones en línea. Los métodos y los hallazgos ayudan a los gerentes de hoteles a entender los intereses de los viajeros, permitiendo comprender mejor los rápidos cambios en las preferencias turísticas, por lo tanto, el método y los hallazgos de este estudio son potencialmente valiosos para los gerentes de hoteles que desean obtener información sobre las preocupaciones de los viajeros y encontrar formas de adaptarse a los rápidos cambios en el mercado turístico, al igual que se utiliza para identificar problemas de interés para los viajeros en otros contextos turísticos, como aerolíneas, restaurantes u otras atracciones.

Vimieiro y Moscato [11] propusieron un nuevo algoritmo de divide y vencerás que permite calcular de manera eficiente los patrones emergentes disyuntivos (conjuntos de características que son muy frecuentes entre muestras de una clase y muy raras entre todas las otras muestras) en entornos paralelos y distribuidos. Los autores contribuyeron al área de la bioinformática y la minería de datos al proporcionar otra alternativa útil para identificar patrones que distinguen muestras con diferentes etiquetas de clase, como las de los estudios de casos y controles.

El enfoque se basó en hipergrafos, transversales mínimas y DEP (Patrones Disyuntivos Emergentes), usaron muestras en lugar de características para encontrar patrones y compararon su propuesta con algunos algoritmos, los cuales fueron DEPMiner, DFS (*Depth First Search*, Búsqueda en Profundidad), RS (*Reverse Search*, Búsqueda en Reversa), kCNF (*k Conjunctive Normal Form*, Forma Normal Conjuntiva con k literales) y LogicRegression (Regresión Lógica).

Los resultados mostraron que el enfoque es más eficiente que la solución de vanguardia disponible en la literatura, y demostró ser mucho más eficiente tanto en términos de memoria como en tiempo de computación.

En [12] se presentó una estructura de árbol de patrones de contraste (DGCP-*tree*) que crece dinámicamente para almacenar patrones crecientes y sus matrices de códigos de ruta con conteos de un bit, que son del árbol de compresión de cadena de bits construido.

Dado que es un gran desafío descubrir patrones emergentes de saltos fuertes (SJEPs) a partir de un conjunto de datos de alta dimensión debido al enorme espacio de patrones, se desarrolló un método de extracción de SJEPs basado en el árbol DGCP (*Dynamically Growing Contrast Pattern*, Patrón de Contraste de Crecimiento Dinámico).

Con el fin de reducir el espacio de búsqueda de patrones, se introdujo un nuevo método de poda de patrones, que reduce drásticamente los patrones emergentes de saltos (JEPs) no mínimos durante el proceso de minería.

En comparación con el conocido método de árbol de CP (Patrón de Contraste), los resultados mostraron que el método propuesto es sustancialmente más rápido, capaz de manejar conjuntos de datos de dimensiones más altas y podar más JEPs no mínimos, entonces, se concluyó que los clasificadores SJEP construidos por el árbol DGCP lograrán una mayor precisión.

Más que un problema, el objetivo principal de García-Vico et al. [13] fue demostrar la capacidad descriptiva de los patrones extraídos por los algoritmos de patrones emergentes (EP), según el marco SDRD (Descubrimiento de Reglas Descriptivas basado en Aprendizaje Supervisado).

Se presentó un estudio completo de EPM (Minería de Patrones Emergentes) bajo el marco SDRD, también se mostró una taxonomía de los algoritmos presentes en la literatura y los tipos

de EP existentes. Además, se analizaron los Chi-EP (*Chi Emerging Patterns*, Patrones Emergentes Chi) y se resaltaron las estrategias de poda basadas en una prueba de χ^2 .

Comparar varias medidas de calidad para evaluar patrones de contraste es una tarea compleja, por lo que García-Borroto et al. [14] propusieron un método para comparar medidas de calidad, utilizando un conjunto de patrones minados y una colección de objetos que no se utilizan para la minería. Los patrones de contraste se extraen de una colección diversa de árboles de decisión generados por un tipo de bosque aleatorio (*Random Forest*), este procedimiento es capaz de extraer un subconjunto de los patrones de contraste más relevantes, que se usan para construir un clasificador preciso. Con respecto a otros mineros, este algoritmo tiene las ventajas de que no requiere la discretización *a priori* de las características numéricas y que maneja bases de datos más grandes.

En la mayoría de las aplicaciones de medidas de calidad, lo importante es la relación entre los valores de calidad, en lugar de los valores mismos por lo que se utilizó el coeficiente de correlación de rango de Spearmans, este coeficiente mide la relación entre el orden de los valores, devuelve valores cercanos a uno para las correlaciones positivas fuertes, cerca de cero para las variables no correlacionadas y cerca de -1 para las correlaciones negativas fuertes. Dado que las correlaciones obtenidas no fueron altas, se realizó un estudio de meta-aprendizaje para encontrar la forma de combinar sinérgicamente las medidas de calidad.

Los resultados de este método ayudarán a los investigadores a crear nuevas medidas de calidad o encontrar nuevas combinaciones de medidas de calidad para crear sistemas de clasificación más comprensibles.

En [15] se intentó comprender e identificar la evolución de los contrastes a lo largo de períodos definidos, uniendo las áreas de CSM (*Contrast Set Mining*, Conjunto de Minería de Contraste) y TDM (*Temporal Data Mining*, Minería de Datos Temporales). La principal contribución de este trabajo es una propuesta para representar la diferencia entre grupos en una base de datos temporal. Para lograr el objetivo de contrastar de manera oportuna, se propuso un conjunto de patrones temporales.

Magalhães y Azevedo [15] tuvieron como objetivo llevar el concepto de patrón de discriminación a un entorno temporal para comprobar cómo evolucionaron las diferencias entre grupos a lo largo del tiempo. Este conjunto de patrones permitirá detectar y representar situaciones de interés que marcan un cambio significativo en el comportamiento de contraste. Potencialmente, esta se considera información muy valiosa para el usuario final.

El descubrimiento de JEPs (Patrones Emergentes de Saltos) en un espacio grande de patrones suele ser una tarea difícil y que requiere mucho tiempo debido a su comportamiento exponencial. El algoritmo propuesto para el descubrimiento de JEPs se basa en algoritmo genético (GA) para descubrir JEPs en un gran espacio de patrones combinándolos para encontrar JEPs de alta calidad a partir de conjuntos de datos para mejorar el rendimiento del sistema de clasificación.

El método presentado en [16] explora un conjunto de JEPs de alta calidad desde el espacio de búsqueda de patrones, a diferencia de otros métodos en la literatura que calculan un conjunto completo de JEPs. Los resultados experimentales mostraron que los Genetic-JEPs propuestos son efectivos y precisos para la clasificación de una variedad de conjuntos de datos y, en general, logran una mayor precisión que otros clasificadores estándar.

Los clasificadores más comprensibles se basan en patrones de contraste, que se extraen con precisión de los árboles de decisión. Sin embargo, la diversidad de los árboles debe asegurarse para extraer una colección representativa de patrones.

García-Borroto et al. [17] compararon los procedimientos de generación de diversidad para los árboles de decisión utilizados para extraer patrones de contraste. El hallazgo principal es que *Random Forest* y *Bagging* son los mejores métodos para extraer patrones de contraste, ya que obtienen más patrones de mayor calidad.

Para estimar la calidad de un subconjunto de patrones, se utilizó el clasificador CAEP (*Classification by Aggregating Emerging Patterns*, Clasificación por medio de Agregación de Patrones Emergentes), con una validación cruzada de cinco pliegues como método de muestreo para todas las estimaciones, los resultados estadísticos se mostraron usando diagramas de CD (*Critical Difference*, Diferencia Crítica), la calidad de los subconjuntos de patrones se evaluó por medio de la precisión (proporción de objetos clasificados correctamente entre el total de

objetos) y la abstención (proporción de objetos que no concuerda con ningún patrón, por lo que ninguna predicción precisa se realiza).

Este resultado ayuda a los futuros investigadores a mejorar los procedimientos reales de generación de diversidad mediante el uso de información de tipo de característica.

Las grandes bases de datos de observación longitudinal brindan la oportunidad de extraer nuevos conocimientos de una manera rentable. Desafortunadamente, la capacidad de estas bases de datos para la inferencia causal es limitada debido a la forma pasiva en que se recopilan los datos, lo que da como resultado diversas formas de sesgo.

Reps et al. [18] presentaron una nueva metodología con el propósito de identificar factores de riesgo que aumentan la probabilidad de que los pacientes experimenten el raro efecto secundario conocido de la insuficiencia renal después de ingerir aminosalicilatos.

Las tecnologías utilizadas fueron:

1. Los algoritmos de minería de reglas de asociación y regresión logística.
2. La metodología de identificación del factor de riesgo candidato, la cual consta de cuatro pasos: Dividir la base de datos, Calcular la proporción de soporte, Aplicar regresión logística y Ordenación.
3. SQL (Structured Query Language, lenguaje de consulta estructurada) para administrar los datos y R para realizar el análisis.
4. El paquete arules se utilizó para identificar los conjuntos de elementos frecuentes.
5. El sistema de código READ.
6. La base de datos THIN (The Health Improvement Network, La Red de Mejora de la Salud).
7. El código de Fórmula Nacional Británica (BNF).

Los resultados de [18] mostraron que la metodología fue capaz de identificar los factores de riesgo investigados previamente, como la prescripción de diuréticos, y destacó que los pacientes con un riesgo de insuficiencia renal más alto que el promedio son incluso más susceptibles a experimentarlo como un efecto secundario después de ingerir aminosalicilatos.

Debido a que la identificación de las características estructurales asociadas con la actividad toxicológica de compuestos químicos tarda mucho tiempo y, a menudo, requiere una aportación significativa de los expertos de dominio, en [19], se describió un método de minería de patrones emergentes para la identificación automatizada de características estructurales de activación en conjuntos de datos de toxicidad que está diseñado para ayudar a acelerar el proceso de desarrollo de alertas, también se describió cómo los desarrolladores de conocimiento usan patrones emergentes para mejorar la especificidad y la sensibilidad de un sistema experto existente.

Se aplicó el algoritmo de minería del árbol de patrones de contraste (*CP-Tree*) para generar un conjunto de patrones emergentes de descriptores de fragmentos estructurales, este usa varias heurísticas para restringir los patrones a aquellos que tienen más probabilidades de ser de interés. El método de Sherhod et al. [19] no pretendió generar alertas toxicológicas definitivas de forma totalmente automática, sino que proporcionó una herramienta para que los expertos exploren los datos para desarrollar alertas. Se demostró que es eficaz para identificar características tóxicas comunes y clases reconocibles de sustancias tóxicas.

Una de las tareas habituales de los gerentes es formular diversas estrategias de *marketing* para enfrentar los entornos de mercado competitivos y cambiantes, por lo tanto, identificar las tendencias de ventas de productos para desarrollar una mejor estrategia es un tema importante. También es un problema que molesta a los gerentes porque las preferencias de los clientes a menudo cambian y capturarlas no es un proceso fácil.

En [20], se integraron los conceptos de ciclo de vida del producto y patrones emergentes para definir cuatro patrones emergentes, a los que se hace referencia como patrones de crecimiento (*Growth patterns*), patrones de incremento rápido (*Rapid-Rise patterns*), patrones de declive (*Decline patterns*) y patrones de hundimiento rápido (*Rapid-Sink patterns*), a fin de identificar tendencias de ventas interesantes.

Los resultados experimentales demostraron la efectividad del algoritmo de minería propuesto. Con la asistencia de los cuatro patrones interesantes y valiosos, los gerentes comprendieron tendencias interesantes de las ventas de productos y luego ajustaron sus estrategias de *marketing* para aumentar el beneficio de sus productos.

Los científicos de datos, con acceso a datos de rápido crecimiento y poder de cómputo, buscan constantemente algoritmos con mayor poder de detección para descubrir conocimiento "novedoso". Pero la mayoría de las veces, sus algoritmos les dan demasiados resultados que son altamente especulativos o simplemente que confirman lo que los expertos en dominios ya saben.

Nguyen et al. [21] propusieron una nueva técnica de ECSM (*Exceptional Contrast Set Mining*, Minería de Conjuntos de Contraste Excepcionales), la cual es una mejora efectiva de la minería de conjuntos de contraste, su novedad radica en los siguientes aspectos:

1. Reconocer el alto nivel de redundancia en la minería de conjuntos de contraste tradicional y remover estos patrones redundantes por medio de métodos estadísticos.
2. Distinción de conjuntos de contrastes afirmativos frente a excepcionales.
3. Recopilación precisa de un conjunto de contraste negativo para el descubrimiento de excepciones.

El algoritmo se aplicó con éxito a varios proyectos de consultoría analítica, en particular, durante un análisis de un registro estatal de cáncer, descubrió una sorprendente diferencia regional en la detección del cáncer de mama.

La evaluación en una amplia gama de problemas sugiere que el ECSM constituye una solución efectiva para descubrir contrastes relevantes, sucintos e informativos en grandes conjuntos de datos. Debido a la simplicidad del algoritmo, es posible volver a implementarlo fácilmente y, si es necesario, se personaliza para adaptarse a aplicaciones analíticas de datos similares.

La minería web es un área de investigación popular en minería de datos, con el uso extensivo de Internet, es esencial conocer las páginas web favoritas de los usuarios y agrupar a los usuarios web para comprender los patrones estructurales de su comportamiento de uso.

En [22] se propuso un enfoque eficiente para agrupar usuarios web con base en los intereses de los usuarios encontrados en los datos del registro web. Los autores consideraron técnicas de minería de patrones emergentes. En la tarea de minería, los patrones emergentes de cada usuario web se usan para definir los intereses que son frecuentemente accedidos por un usuario y escasamente accedidos por otros. Al encontrar los patrones emergentes para todos los usuarios

web, es posible descartar páginas web no esenciales (ruido) para cada usuario web y agrupar a los usuarios web de acuerdo a la generación de páginas web típicas.

Su enfoque consistió en generar páginas web grandes a partir de los datos del registro web, después examinar y transformar el conjunto de datos en SPLGs (*Simple Page-Linked Graphs*, Grafos Simples de Páginas Enlazadas) y luego generar patrones emergentes en estos SPLGs. Los autores agruparon a los usuarios web con base en los patrones emergentes, y finalmente etiquetaron los grupos con las páginas web típicas.

La autopsia es una práctica muy importante para la medicina, es el único estudio que permite identificar la verdadera causa de muerte del difunto, el estudio de la evolución de las enfermedades, la determinación de la efectividad de los tratamientos tradicionales, el descubrimiento de nuevas enfermedades, y mucho más. Sin embargo, en un hospital del estado de Veracruz, este método está prácticamente en desuso, por lo cual el departamento de patología investigó las posibles causas.

El objetivo de Rubio-Delgado et al. [23] fue analizar las posibles causas de la reducción de las autopsias en el sistema hospitalario de “Servicios de Salud de Veracruz” por medio de la extracción de reglas de asociación y redes Bayesianas a partir de los datos que pertenecen a las opiniones médicas acerca de esta práctica. Para la exploración y extracción de conocimientos, se utilizaron algoritmos como Apriori, FPGrowth, PredictiveApriori, Tertius, J48, NaiveBayes, MultilayerPerceptron y BayesNet, todos ellos proporcionados por la API de Weka, para generar modelos de minería y presentar los nuevos conocimientos en lenguaje natural.

El uso de técnicas de minería de reglas de asociación y las redes Bayesianas permitió realizar un análisis de la situación problemática y encontrar las correlaciones entre los atributos categóricos del conjunto de datos, que conformaron la información obtenida del personal médico, todo esto a través de una aplicación web desarrollada especialmente para el caso. La aplicación Web proporcionó una explicación en lenguaje natural de los resultados de los modelos de minería, de esta manera, el especialista fue capaz de evaluar subjetivamente, en función de su experiencia y conocimiento, la información extraída por los modelos.

2.2 Análisis Comparativo

A continuación, la Tabla 2.1 muestra una comparación de los trabajos descritos en la sección anterior, para que se observen las diferencias y las similitudes entre ellos de una forma más clara.

Tabla 2.1 Análisis comparativo de trabajos relacionados.

Artículo	Problema	Contribución	Tecnologías	Resultado	Estado
Ienco et al. [9]	Detección de anomalías en los datos categóricos con un entorno semi-supervisado.	Modelo discriminativo que es fácilmente interpretado y utilizado para la exploración de los datos.	Estrategias heurísticas MinDTK (<i>Minimum Distance Top-k</i>), MaxDTK (<i>Maximum Distance Top-k</i>), RandK (<i>Random k</i>), CentralK (<i>Central k</i>). DILCA (<i>Distance Learning for Categorical Attributes</i> , Aprendizaje de Distancia para Atributos Categóricos). Nube de palabras.	SAnDCat, un nuevo enfoque para la detección semi-supervisada de anomalías para datos categóricos, que es capaz de modelar datos categóricos usando algoritmos basados en distancia y obtiene muy buenos resultados.	Finalizado
Li et al. [10]	Los cambios en la preferencia influyen en el desempeño de las empresas hoteleras, creando así la necesidad de identificar y atender las demandas de sus	Técnica de minería de patrones emergentes de interés para los viajeros internacionales.	Minería de Patrones Emergentes (EPM). GATE (<i>General Architecture for Text Engineering</i> , Arquitectura General	El método y los hallazgos de este estudio son potencialmente valiosos para los gerentes de hoteles que desean obtener información sobre	Finalizado

Artículo	Problema	Contribución	Tecnologías	Resultado	Estado
	huéspedes. La mayoría de los estudios existentes se centran en los atributos de la demanda actual y no en los emergentes, por lo tanto, a los gerentes de hoteles les resulta difícil tomar decisiones apropiadas en respuesta a los cambios en las preocupaciones de los viajeros.		para Ingeniería de Texto). Visual Web Ripper.	las preocupaciones de los viajeros y encontrar formas de adaptarse a los rápidos cambios en el mercado turístico, al igual que se utiliza para identificar problemas de interés para los viajeros en otros contextos turísticos, como aerolíneas, restaurantes u otras atracciones.	
Vimieiro y Moscato [11]	Minería de patrones disyuntivos emergentes en conjuntos biomédicos de alta dimensión.	Nuevo algoritmo de divide y vencerás que permite calcular de manera eficiente los patrones emergentes disyuntivos en entornos paralelos y distribuidos.	EPM Hipergrafos, transversales mínimas y DEP. Python, C, C++.	Los resultados mostraron que el enfoque es más eficiente que la solución de vanguardia disponible en la literatura, y demostró ser mucho más eficiente tanto en términos de memoria como de	Finalizado

Artículo	Problema	Contribución	Tecnologías	Resultado	Estado
				tiempo de computación.	
Liu et al. [12]	Gran desafío al descubrir patrones emergentes de saltos fuertes (SJEPs) a partir de un conjunto de datos de alta dimensión debido al enorme espacio de patrones.	Con el fin de reducir el espacio de búsqueda de patrones, se introdujo un nuevo método de poda de patrones, que reduce drásticamente los patrones emergentes de saltos (JEP) no mínimos durante el proceso de minería.	EPM DGCP-tree, CP-tree, BSC-tree (<i>bit string compression tree</i> , árbol de compresión de cadenas de bits) y SJEPs. Java Weka	Los resultados mostraron que el método propuesto es sustancialmente más rápido, capaz de manejar conjuntos de datos de dimensiones más altas y podar más JEP no mínimos, entonces, los clasificadores SJEP construidos por el árbol de DGCP (Patrón de Contraste Creciente Dinámico) lograrán una mayor precisión.	Finalizado
García-Vico et al. [13]	Más que un problema el objetivo principal es demostrar la	Se presentó un estudio completo de EPM bajo el marco	EPM SDRD	Los resultados de este trabajo permitieron	Finalizado

Artículo	Problema	Contribución	Tecnologías	Resultado	Estado
	capacidad descriptiva de los patrones extraídos por los algoritmos de patrones emergentes (EP), según el marco SDRD (Descubrimiento Supervisado de Reglas Descriptivas).	SDRD, también se mostró una taxonomía de los algoritmos presentes en la literatura y los tipos de EP existentes.	Paquete scmamp (<i>statistical comparison of multiple algorithms on multiple problems</i> , comparación estadística de múltiples algoritmos en varios problemas). Software estadístico R.	determinar la idoneidad de los diferentes enfoques presentados a lo largo de la literatura y determinar nuevas líneas de investigación para crear nuevas propuestas centradas en la descripción, finalmente, se detallan las tendencias y perspectivas de EPM.	
García-Borroto et al. [14]	Los patrones de contraste, que se encuentran en el núcleo de los clasificadores más comprensibles, se evalúan frecuentemente mediante medidas de calidad. Dado que	Método para comparar medidas de calidad, utilizando un conjunto de patrones minados y una colección de objetos que no se	Minería de Conjuntos de Contraste (CSM), <i>Random Forest</i> y el coeficiente de correlación de rango de <i>Spearman</i> s.	Los resultados de este documento ayudarán a los investigadores a crear nuevas medidas de calidad o encontrar nuevas combinaciones de medidas de calidad para crear sistemas	Finalizado

Artículo	Problema	Contribución	Tecnologías	Resultado	Estado
	existen muchas medidas de calidad diferentes, deben compararse para seleccionar las más adecuadas para cada aplicación.	utilizan para la minería.		de clasificación más comprensibles.	
Magalhães y Azevedo [15]	Comprender las diferencias subyacentes entre grupos o clases en ciertos contextos es de mayor importancia. La minería de conjuntos de contraste se basa en descubrir patrones significativos al contrastar dos o más grupos.	Propuesta para representar la diferencia entre grupos en una base de datos temporal. Para lograr el objetivo de contrastar de manera oportuna, se propuso un conjunto de patrones temporales.	CSM y TDM. RCS (<i>Rules for Contrast Sets</i> , Reglas para Conjuntos de Constraste) CAREN (<i>Class Project Association Rule Engine</i> , Proyecto de Clase de un Motor de Reglas de Asociación).	Este conjunto de patrones permitirá detectar y representar situaciones de interés que marcan un cambio significativo en el comportamiento de contraste. Potencialmente, esta es información muy valiosa para el usuario final.	Finalizado
ul Ain y ur Rehman [16]	El descubrimiento de JEPs (Patrones Emergentes de Saltos) en un espacio de patrón grande suele ser una tarea difícil y que requiere	Método para descubrir JEPs usando el Algoritmo Genético (GA) para superar las deficiencias de	EPM GA, JEPs. Microsoft Visual C# .Net Framework 4.0. Microsoft Visual Studio 2015.	Los resultados experimentales mostraron que el método propuesto es efectivo y preciso para la clasificación de	Finalizado

Artículo	Problema	Contribución	Tecnologías	Resultado	Estado
	mucho tiempo debido a su comportamiento exponencial.	enfoques relacionados previos.	Weka	una variedad de conjuntos de datos y, en general, logró una mayor precisión que otros clasificadores estándar.	
García-Borroto et al. [17]	Los clasificadores más comprensibles se basan en patrones de contraste, que se extraen con precisión de los árboles de decisión. Sin embargo, la diversidad de árboles debe asegurarse para extraer una colección de patrones representativa.	El hallazgo principal es que <i>Random Forest</i> y <i>Bagging</i> son los mejores métodos para extraer patrones de contraste, ya que extraen más patrones de mayor calidad.	Patrones de contraste. <i>Random Forest</i> , <i>Bagging</i> , <i>Random Subspaces</i> , <i>Randomized C 4.5</i> , <i>LCMine (Logical Complex Mine</i> , Minería Lógica Compleja), clasificador CAEP, diagramas de CD.	Este análisis ayudará a los futuros investigadores a mejorar los procedimientos reales mediante el uso de información de tipo de característica.	Mejoras a futuro.

Artículo	Problema	Contribución	Tecnologías	Resultado	Estado
Reps et al. [18]	Las grandes bases de datos de observación longitudinal brindan la oportunidad de extraer nuevos conocimientos de una manera rentable. Desafortunadamente, la capacidad de estas bases de datos para ser utilizadas para la inferencia causal es limitada debido a la forma pasiva en que se recopilan los datos, lo que da como resultado diversas formas de sesgo.	Nueva metodología con el propósito de identificar factores de riesgo que aumentan la probabilidad de que los pacientes experimenten el raro efecto secundario conocido de la insuficiencia renal después de ingerir aminosalicilatos.	CSM Algoritmos de minería de reglas de asociación y regresión logística. Metodología de identificación del factor de riesgo candidato. SQL para administrar los datos y R para realizar el análisis. Paquete arules se utilizó para identificar los conjuntos de elementos frecuentes. El sistema de código <i>READ</i> . La base de datos THIN (<i>The Health Improvement Network</i> , La Red de Mejora de la Salud). El código de Fórmula Nacional Británica (BNF).	Los resultados muestran que la metodología fue capaz de identificar los factores de riesgo investigados previamente, como la prescripción de diuréticos, y destacó que los pacientes con un riesgo de insuficiencia renal más alto que el promedio son incluso más susceptibles a experimentarlo como un efecto secundario después de ingerir aminosalicilatos.	Finalizado

Artículo	Problema	Contribución	Tecnologías	Resultado	Estado
Sherhod et al. [19]	La identificación de las características estructurales asociadas con la actividad toxicológica de compuestos químicos se lleva mucho tiempo y, a menudo, requiere una aportación significativa de los expertos en dominio.	Método de minería de patrones emergentes que ayuda a los desarrolladores de bases de conocimiento a identificar características estructurales para mejorar el rendimiento de los sistemas expertos para la predicción de toxicidad.	EPM. Algoritmo de árbol de patrones de contraste (CP-tree). ChemAxon software. KNIME	Proporcionó una herramienta para que los expertos exploren los datos para desarrollar alertas. Se demostró que es eficaz para identificar características tóxicas comunes y clases reconocibles de sustancias tóxicas.	Mejoras a futuro.
Weng y Huang [20]	Una de las tareas habituales de los gerentes es formular diversas estrategias de <i>marketing</i> para enfrentar los entornos de mercado competitivos y cambiantes, por lo tanto, identificar las tendencias de ventas de productos para desarrollar una mejor estrategia es un tema	Integración de los conceptos de ciclo de vida del producto y patrones emergentes para definir cuatro patrones emergentes, llamados patrones de crecimiento (<i>Growth patterns</i>), patrones de incremento rápido (<i>Rapid-Rise</i>	EPM <i>Growth pattern, Rapid-Rise pattern, Decline pattern, and RapidSink pattern.</i>	Con la asistencia de los cuatro patrones interesantes y valiosos, los gerentes comprendieron tendencias interesantes de las ventas de productos y ajustaron sus estrategias de <i>marketing</i> para	Mejoras a futuro.

Artículo	Problema	Contribución	Tecnologías	Resultado	Estado
	importante. También es un problema que molesta a los gerentes porque las preferencias de los clientes a menudo cambian y capturarlas no es un proceso fácil.	<i>patterns</i>), patrones de declive (<i>Decline patterns</i>) y patrones de hundimiento rápido (<i>Rapid-Sink patterns</i>), para identificar tendencias de ventas interesantes.		aumentar el beneficio de sus productos.	
Nguyen et al. [21]	Los científicos de datos, con acceso a datos de rápido crecimiento y poder de cómputo, buscan constantemente algoritmos con mayor poder de detección para descubrir conocimiento "novedoso". Pero la mayoría de las veces, sus algoritmos les dan demasiados resultados que son altamente especulativos o simplemente que confirman lo que los	Propusieron una nueva técnica de Minería de Conjuntos de Contraste Excepcionales (ECSM), la cual es una mejora efectiva de la minería de conjuntos de contraste.	CSM. C# (.Net Framework 4.5.50938)	Constituye una solución efectiva para descubrir contrastes relevantes, sucintos e informativos en grandes conjuntos de datos. Debido a la simplicidad del algoritmo, es posible volver a implementarlo fácilmente y, si es necesario, se personaliza para adaptarse a aplicaciones analíticas de datos similares.	Mejoras a futuro.

Artículo	Problema	Contribución	Tecnologías	Resultado	Estado
	expertos en dominios ya saben.				
Yu et al. [22]	La minería web es un área de investigación popular en minería de datos, con el uso extensivo de Internet, es esencial conocer las páginas web favoritas de los usuarios y agrupar a los usuarios web para comprender los patrones estructurales de su comportamiento de uso.	Con el objetivo de agrupar usuarios web con base en sus intereses encontrados en datos de registros web, se propuso un enfoque eficiente considerando las técnicas de patrones emergentes.	EPM SPLGs <i>k</i> -means Etiquetaron los grupos considerando TF-IDF (<i>Term-Frequency-Inverse Document Frequency</i> , Frecuencia de Términos, Frecuencia de Documentos Inversa).	Se identificaron las páginas web favoritas de cada usuario por medio de la eliminación de ruido debido a páginas populares, y del agrupamiento de usuarios web de acuerdo a los patrones emergentes generados.	Mejoras a futuro.
Rubio Delgado et al. [23]	La disminución importante en el número de autopsias en los hospitales de todo el mundo se ha planteado preguntas sobre los motivos de este fenómeno.	El uso de técnicas de minería de reglas de asociación y las redes Bayesianas permitió realizar un análisis de la situación problemática y encontrar las correlaciones entre	Minería de reglas de asociación (Apriori, FPGrowth, PredictiveApriori, Tertius). Redes bayesianas (BayesNet).	El sistema proporcionó una explicación en lenguaje natural de los resultados de los modelos de minería, de modo que el patólogo los comprenda, de esta manera, el	Mejoras a futuro.

Artículo	Problema	Contribución	Tecnologías	Resultado	Estado
		los atributos categóricos del conjunto de datos, que conformaron la información obtenida del personal médico, todo esto a través de una aplicación web.	Clasificación (J48, NaiveBayes, MultilayerPerceptron). Weka	especialista evaluó los resultados analizando subjetivamente, en función de su experiencia y conocimiento, la información extraída por los modelos.	

Como se observa en la Tabla 2.1, las tareas descriptivas de minería de datos se utilizan para resolver problemas en varias áreas, por ejemplo, [9] produjo un modelo discriminativo que se interpreta y utiliza fácilmente para la exploración de los datos. En [14], se evaluaron medidas de calidad para patrones de contraste.

Mientras que la Minería de Patrones Emergentes (EPM) se utilizó en [10] para identificar las preferencias de hospedaje de los viajeros; en [11], se abordó el tema de minería de patrones emergentes disyuntivos en conjuntos de datos biomédicos de alta dimensionalidad; de igual forma, un nuevo enfoque de EPM de saltos fuertes basado en el árbol de compresión de cadenas de bits (*BSC-tree*) a partir de conjuntos de datos de alta dimensionalidad se propuso en [12]; un estudio completo de EPM se presentó en [13]; en [16], los autores presentaron un método para descubrir patrones emergentes de saltos usando el algoritmo genético; en [19], se aplicó la EPM para descubrir conocimiento toxicológico de los compuestos químicos; en [20], para identificar tendencias de ventas de productos, y en [22], como una técnica de Minería Web.

Por otro lado, la minería de conjuntos de contraste (CSM) se usó en [15] para entender e identificar la evolución de contrastes a lo largo de periodos definidos, uniendo las áreas de CSM y Minería de Datos Temporales (TDM); en [14], debido a que se analizaron procedimientos para generar diversidad para la CSM; en [18], para identificar factores de riesgo de que los pacientes sufran insuficiencia renal por la prescripción de aminosalicilatos y en [21], ya que se propuso una nueva técnica CSM, llamada ECSM (Minería de Conjuntos de Contraste Excepcionales).

Finalmente, en [23], el uso de técnicas de minería de reglas de asociación y las redes Bayesianas permitió analizar encuestas sobre la disminución de autopsias y determinar cuáles son las principales razones por las que los médicos consideran que cada vez se realizan menos autopsias en el H.R.R.B, todo esto a través de una aplicación web desarrollada especialmente para el caso. En este proyecto se extendió el trabajo realizado en [23], para esto se aplicaron las encuestas a médicos de otro hospital y se utilizaron técnicas descriptivas, específicamente EPM, para la comparación de los resultados encontrados en el H. R. R. B. con los del otro hospital.

2.3 Propuesta de solución

De acuerdo a la problemática y a los objetivos que se encuentran en el capítulo uno (Secciones 1.3 y 1.4), se necesitó recolectar y procesar la información obtenida de las encuestas aplicadas a los médicos del H.R.R.B y otro hospital, para esto se emplearon tareas descriptivas de la minería de datos, específicamente EPM, por este motivo se agregó un módulo al sistema para el análisis de encuestas sobre la disminución de autopsias usando minería de datos, el cual es una aplicación web y se busca que sea capaz de analizar los resultados de las encuestas en otro hospital.

El sistema permite acceder a la aplicación desde un dispositivo cliente para ingresar y consultar encuestas, ahora también genera modelos descriptivos de minería de datos haciendo uso del marco de trabajo EPM_Algorithms y los guarda físicamente en el servidor que la hospeda.

El marco de trabajo permite cargar datos y crear modelos de aprendizaje, para esto se utiliza la biblioteca del marco de trabajo para consultar los modelos e interpretarlos y a través de la aplicación web mostrar los resultados de manera gráfica y entendible para los usuarios finales.

La tarea de EPM ayuda a conseguir patrones de coincidencia que permiten caracterizar los datos dentro de una clase y también comparar los dos conjuntos de datos, el del H.R.R.B y del otro hospital.

Para resolver esta problemática se eligieron las siguientes herramientas como las más adecuadas para el desarrollo de este proyecto, teniendo como herramienta de minería de datos a EPM_Algorithms, ya que cuenta con los algoritmos más importantes de minería de patrones emergentes; como marco de trabajo JSF debido a que está basado en el patrón MVC (Modelo Vista Controlador), por su facilidad de desarrollo, su flexibilidad al permitir crear componentes propios, por ser gratuito y por la facilidad de crear aplicaciones web con lenguaje puro de Java; por consiguiente se trabajó con el lenguaje de programación Java; como IDE se eligió NetBeans debido a que no necesita estar descargando tantos *plugins* y durante su instalación ya trae el servidor en donde se ejecuta JSF, el cual es Apache Tomcat, sin necesidad de estar configurando tantas cosas; como sistema gestor de bases de datos (SGBD) se utilizó PostgreSQL puesto que cuenta con una gran documentación para su uso y configuración en comparación con otros SGBD, es un software libre y de código abierto, cumple con ACID (atomicidad, consistencia, aislamiento, durabilidad) lo que asegura que ningún dato se pierda o transmita a otro lugar en

el sistema en caso de una falla, se utiliza ampliamente en sistemas grandes donde la velocidad de lectura y escritura son cruciales y los datos necesitan validarse, y en cuestión de seguridad cuenta con roles y roles heredados para mantener los permisos; la metodología para el desarrollo de la aplicación es UWE, ya que provee una serie de pasos concretos y se basa en el proceso unificado y UML para el desarrollo de aplicaciones web, cubriendo la navegación, la presentación, los procesos de negocios y los aspectos de adaptación; finalmente como metodología de minería de datos se usó KDD puesto que los modelos se construyen de manera rápida y es generalmente más utilizado en la academia en comparación con CRISP-DM que se usa más en la industria [24].

Capítulo 3. Aplicación de la metodología

Tomando en cuenta la problemática y la solución presentada en el capítulo uno (Sección 1.3), se requiere hacer la comparación entre el conjunto de datos del H.R.R.B con el conjunto de datos de otros hospitales, mediante técnicas de minería de datos, específicamente minería de patrones emergentes.

En este estudio se propone utilizar la aplicación web realizada en una tesis anterior [1] e implementar nuevos ajustes y tareas de MD (Minería de Datos) y presentar los resultados generados por los algoritmos de EPM (Minería de Patrones Emergentes).

En la Figura 3.1 se presenta el esquema de solución, en el cual a través de un cliente el médico ingresa a la aplicación web, contesta y consulta la encuesta, la aplicación permite generar modelos de minería de datos mediante el uso de la API de Weka y EPM_Algorithms y guardarlos físicamente en el servidor.

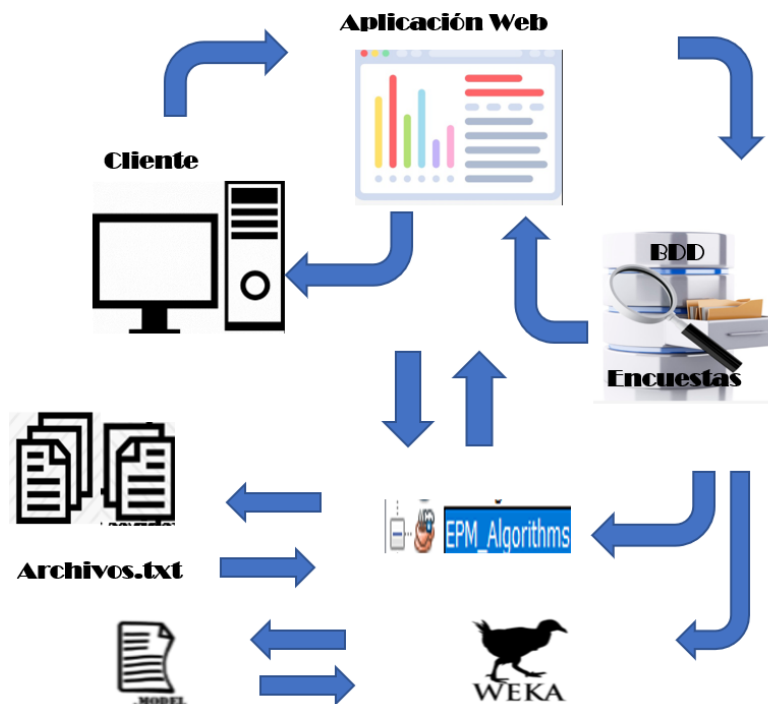


Figura 3.1 Esquema de solución

EPM_Algorithms permite consultar los archivos para interpretarlos y mediante la aplicación, mostrar los resultados de manera gráfica y comprensible para el usuario final.

3.1 Análisis de requerimientos

Durante esta fase se adquieren, reúnen y especifican las características funcionales y no funcionales que se deben cumplir en la aplicación de acuerdo con las necesidades del cliente.

Como primer paso se presentan en la Tabla 3.1 los diferentes actores de la aplicación y la descripción de las actividades que realiza cada uno.

Tabla 3.1 Actores de la aplicación Web

Actores	Descripción
Administrador	Este usuario se autentica en el sistema y tiene permisos para insertar, modificar, consultar y eliminar usuarios, consultar los resultados y cerrar sesión.
Especialista	Este usuario se autentica en el sistema y tiene permisos para consultar y eliminar encuestas, generar modelos, consultar resultados y cerrar sesión.
Encuestado	Este usuario solo es capaz de responder encuestas y consultar resultados.

Después de identificar los actores y funciones que realiza cada uno en la aplicación, se muestra una representación mediante un diagrama de casos de uso en la Figura 3.2

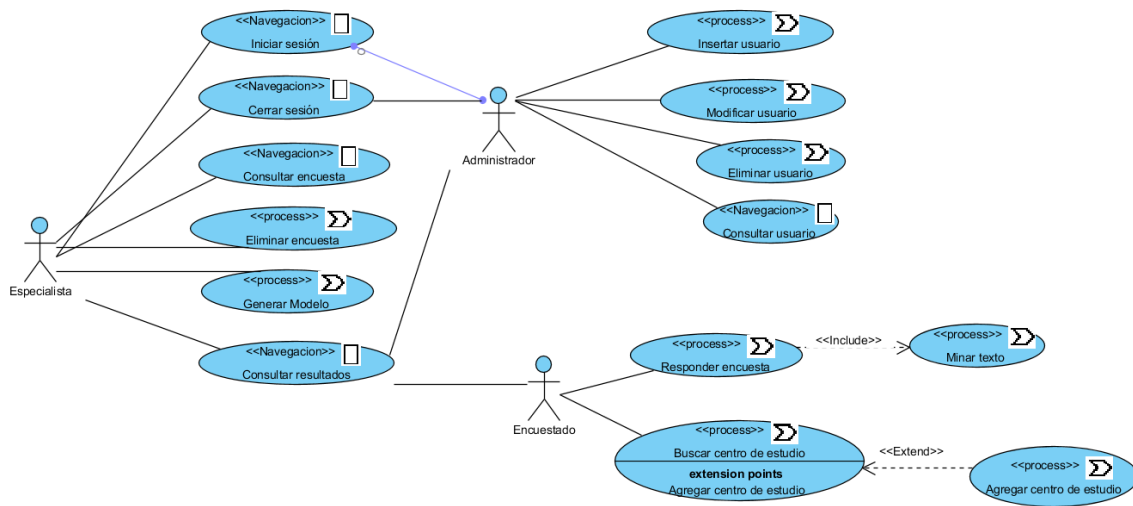


Figura 3.2 Diagrama de casos de uso de la aplicación web

Los actores *Especialista* y *Administrador* se deben autenticar en el sistema para realizar las operaciones correspondientes a cada uno y por consiguiente finalizan la sesión en el momento que lo deseen.

El *Administrador* es el encargado de gestionar los usuarios registrados en la aplicación, por lo que tiene permisos para insertar, modificar, eliminar y consultar a los usuarios.

El *Especialista* es el encargado de gestionar la información de las encuestas, es capaz de consultar y eliminar encuestas, así como generar los modelos y consultar los resultados de las encuestas minadas.

El *Encuestado* no se autentica en la aplicación, ya que es un usuario anónimo, una de sus actividades es buscar si se encuentra el centro de estudio donde realizó sus estudios de medicina general o su especialidad, en caso de no encontrarlo, entonces, lo agrega, su tarea principal es responder la encuesta, proceso que invoca a otro llamado *Minar Texto*, este es el encargado de clasificar las respuestas dadas por los encuestados en las preguntas abiertas.

Todos los actores son capaces de consultar los resultados.

Mediante el diagrama de casos de uso descrito se representa la interacción de los actores con la aplicación, permitiendo identificar los requerimientos funcionales, cada caso de uso se

representa a continuación por un diagrama de *workflow* para verificar el flujo y las actividades que cada uno realiza.

El caso de uso *Iniciar sesión* se representa en la Figura 3.3 mediante un diagrama de actividades, indica que la primera acción muestra en pantalla un formulario mediante el cual el usuario proporciona sus datos de acceso para que el sistema prosiga a iniciar sesión si las credenciales son correctas, termina su proceso con la actividad de iniciar sesión.

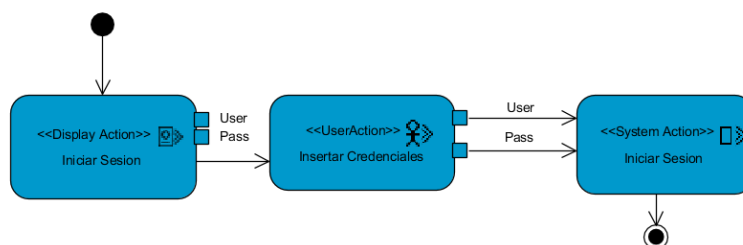


Figura 3.3 Diagrama de actividad del caso de uso Iniciar Sesión

El diagrama de actividad del caso de uso *Cerrar Sesión*, se muestra en la Figura 3.4, como primera acción inicia mostrando en pantalla un formulario en donde el usuario selecciona la opción de cerrar sesión y así la actividad termina finalizando la sesión actual.

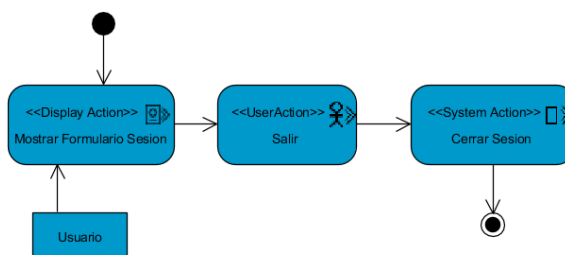


Figura 3.4 Diagrama de actividad del caso de uso Cerrar Sesión.

En la Figura 3.5 se representa el diagrama de actividad del caso de uso *Responder Encuesta*, la primera acción es mostrar en pantalla un formulario, el cual el usuario llenará respondiendo cada una de las preguntas de la encuesta, para que el sistema posteriormente guarde la nueva encuesta y lance el caso de uso *Minar Texto*. Cuando la encuesta queda registrada en la aplicación y fue objeto de la actividad de minería de texto, el caso de uso finaliza.

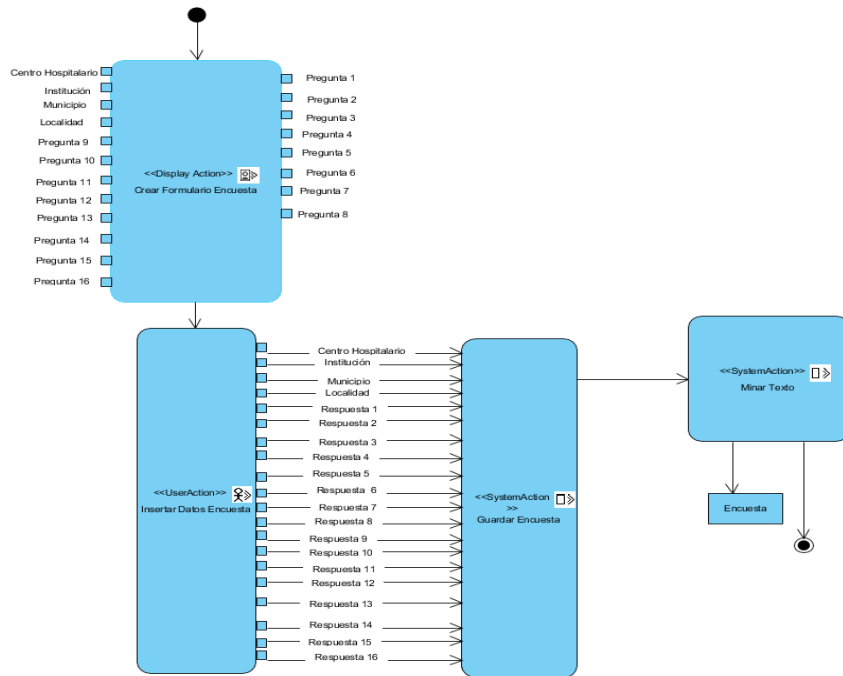


Figura 3.5 Diagrama de actividad del caso de uso Responder Encuesta.

El caso de uso *Consultar encuesta*, se representa en la Figura 3.6 mediante un diagrama de actividades, inicia mostrando en pantalla la lista de las encuestas y un formulario mediante el cual el usuario proporciona el número de control que desea consultar, entonces el sistema procede a proporcionar los datos de la encuesta y así finaliza la actividad.

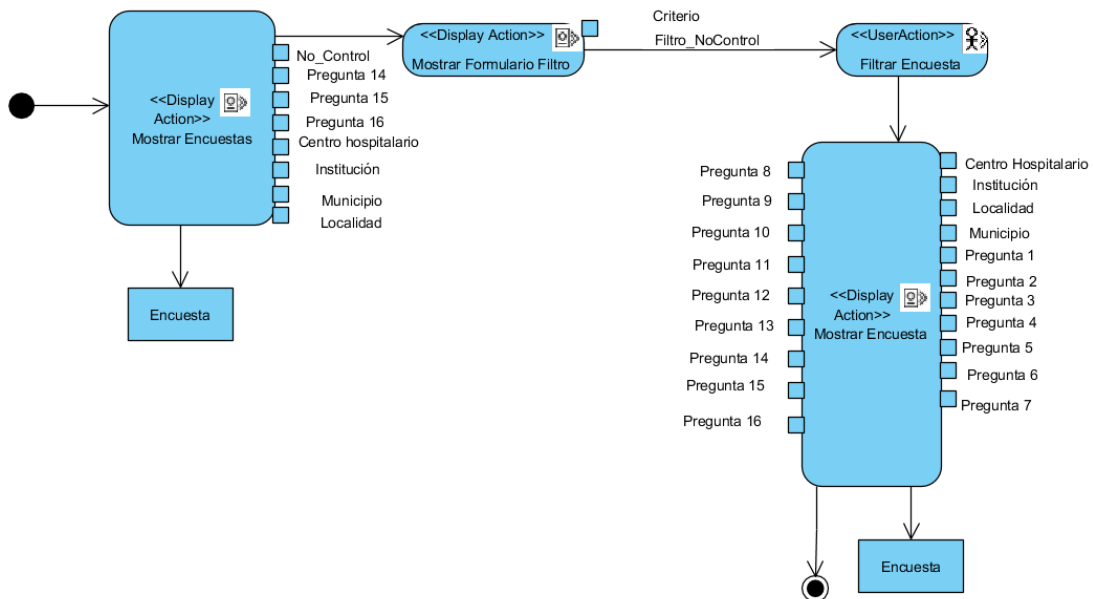


Figura 3.6 Diagrama de actividad del caso de uso Consultar Encuesta

En la Figura 3.7 se representa el diagrama de actividades del caso de uso *Eliminar Encuesta*, el sistema inicia la acción mostrando en pantalla el listado de las encuestas y un formulario en el cual el usuario proporciona el número de control de la encuesta que desea eliminar, el sistema responde con la muestra de los datos, el usuario procede a indicar que desea eliminar la encuesta por consiguiente el sistema elimina la encuesta y finaliza la actividad.

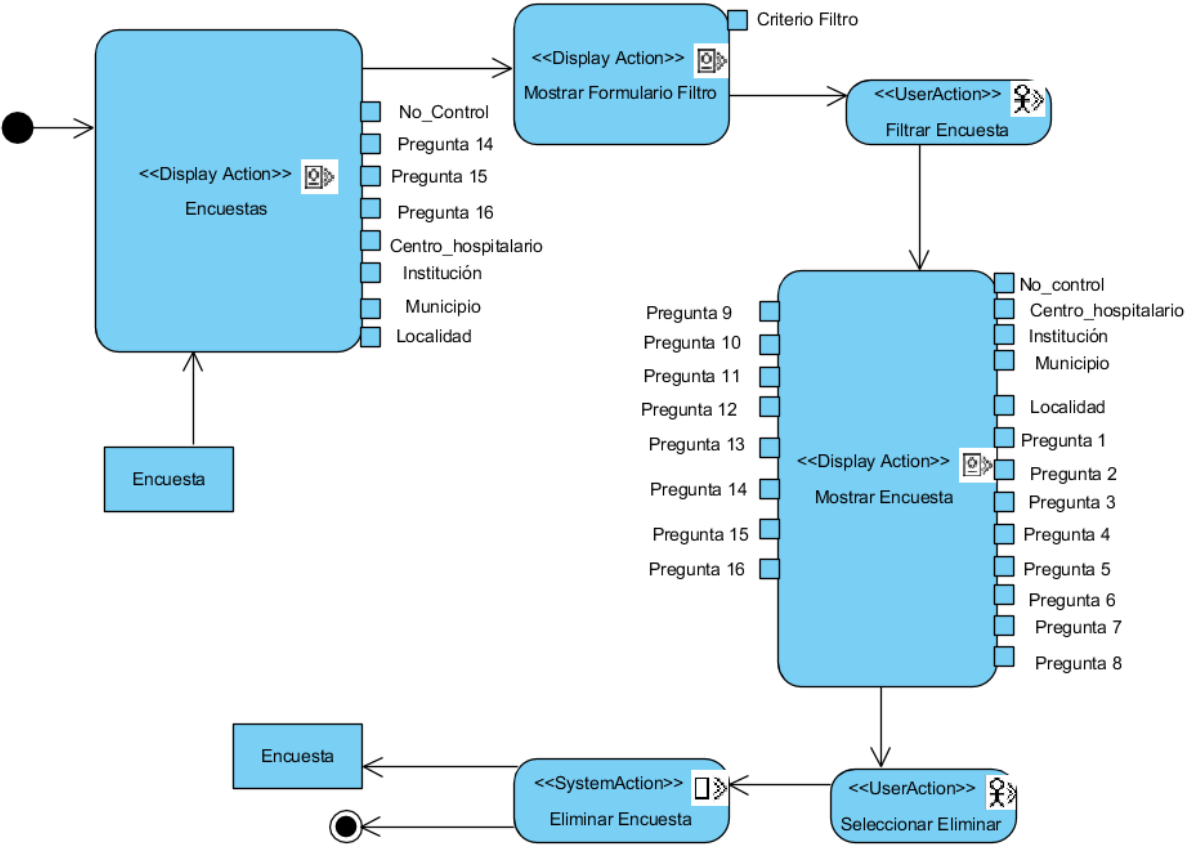


Figura 3.7 Diagrama de actividad del caso de uso *Eliminar Encuesta*

El caso de uso *Generar Modelo* se representa en el diagrama de actividades de la Figura 3.8, inicia mostrando en pantalla los resultados del modelo existente, el usuario selecciona la opción de *Generar Modelo*, por lo que, el sistema procede a generar el nuevo modelo y finaliza la actividad.

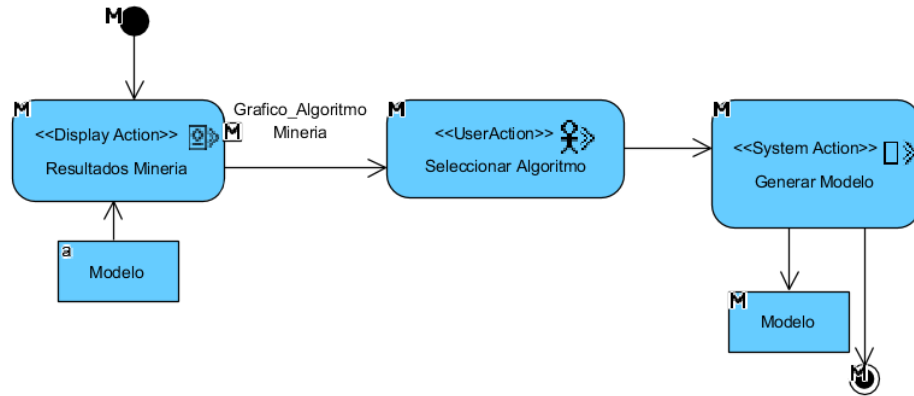


Figura 3.8 Diagrama de actividad del caso de uso Generar Modelo

En la Figura 3.9 se representa el caso de uso *Interpretar Resultados de Minería* mediante un diagrama de actividad, el sistema inicia haciendo la interpretación del modelo y finaliza una vez que la pantalla muestra los resultados de minería.

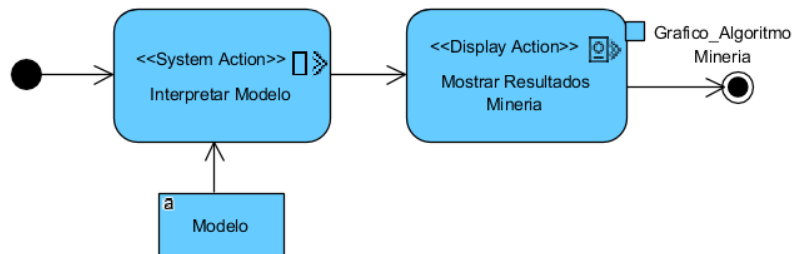


Figura 3.9 Diagrama de actividad del caso de uso Interpretar Resultados de Minería

En la Figura 3.10 se representa el caso de uso de *Buscar Centro de Estudio y Agregar Centro de Estudio* mediante un diagrama de actividades, el usuario realiza la acción de buscar su centro de estudio, si se encuentra ahí, finaliza la actividad, en caso de que no se encuentre, entonces prosigue a llenar el formulario y finaliza la actividad.

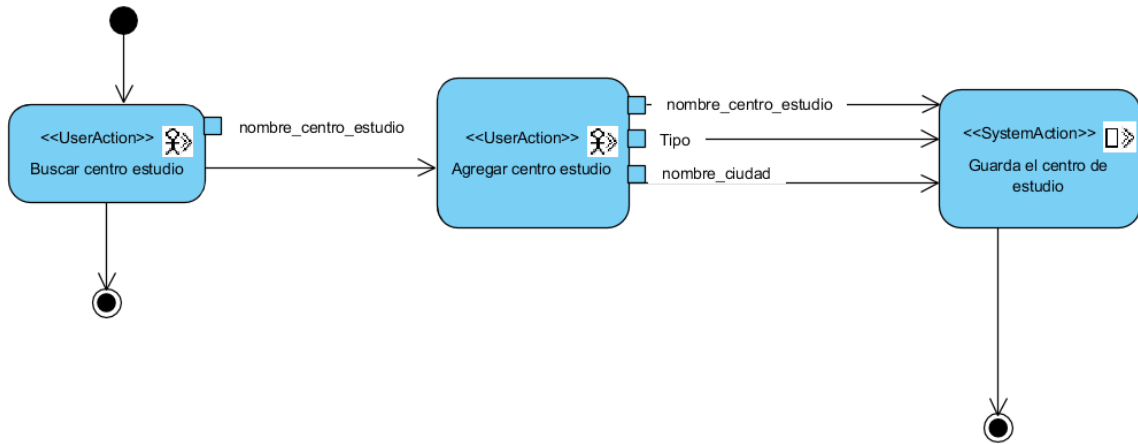


Figura 3.10 Diagrama de actividad del caso de uso *Buscar centro estudio* y *Agregar centro estudio*

El diagrama de actividad del caso de uso *Insertar Usuario* se muestra en la Figura 3.11, inicia mostrando en la pantalla el formulario en el cual el usuario autorizado proporciona los datos necesarios, para que el sistema prosiga a guardar el nuevo usuario y finalice la actividad.

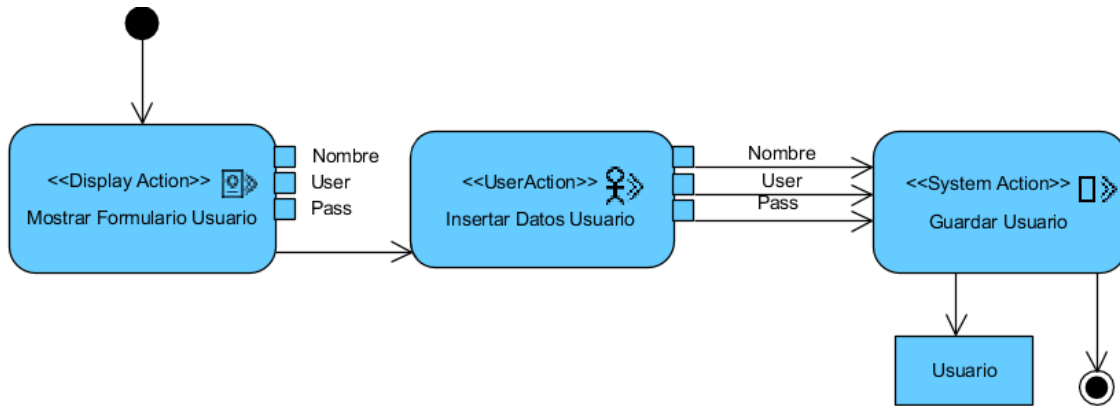


Figura 3.11 Diagrama de actividad del caso de uso *Insertar Usuario*.

En la Figura 3.12 se muestra el diagrama de actividad del caso de uso *Consultar Usuario*, inicia mostrando en pantalla la lista de usuarios registrados y un formulario en el cual el usuario autorizado mediante el identificador del mismo filtra los resultados, a lo que el sistema responde mostrando la información del usuario consultado y finaliza la actividad.

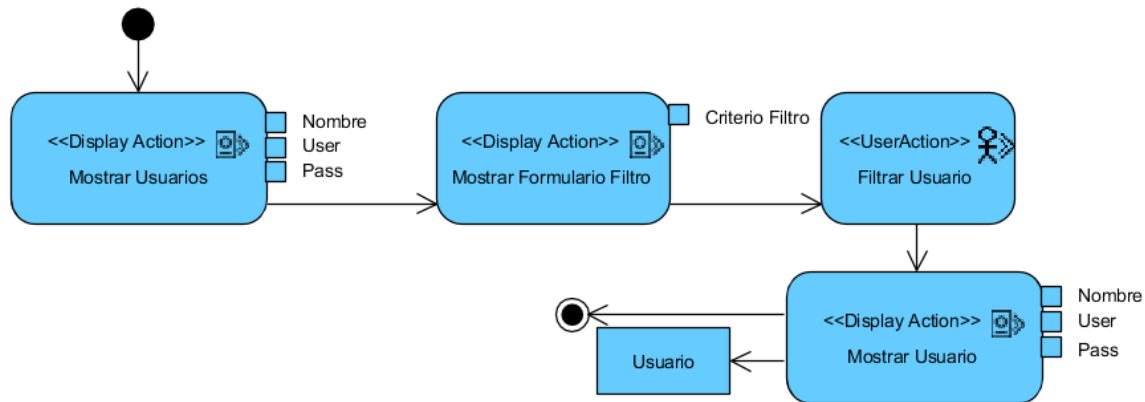


Figura 3.12 Diagrama de actividad del caso de uso Consultar Usuario.

El diagrama de actividad que se encuentra en la Figura 3.13 representa al caso de uso *Actualizar Usuario*, inicia mostrando en pantalla la lista de los usuarios registrados en la aplicación y un formulario en el cual el usuario autorizado proporciona el identificador de el o los usuarios que desea actualizar, el sistema prosigue a mostrar los datos para que el usuario realice los cambios y actualice la información, una vez realizado, el sistema guarda la modificación y finaliza la actividad.

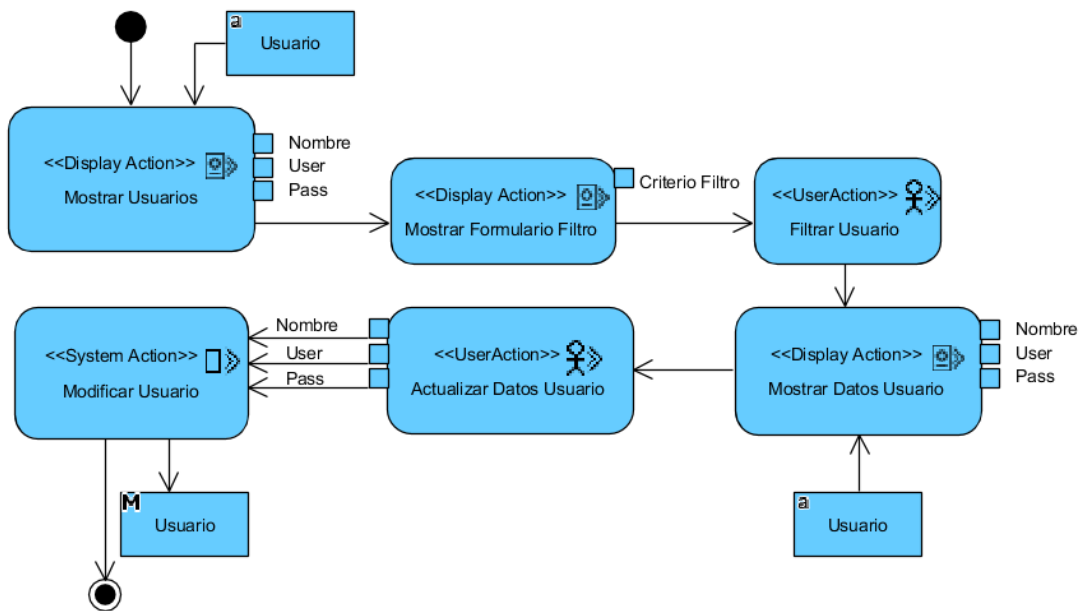


Figura 3.13 Diagrama de actividad del caso de uso Actualizar Usuario

El caso de uso *Eliminar Usuario* se representa en el diagrama de actividades que se encuentra en la Figura 3.14, inicia mostrando en pantalla la lista de usuarios registrados y un formulario en el cual el usuario autorizado proporciona el identificador del usuario que desea eliminar, el

sistema procede a mostrar los datos del mismo y el usuario selecciona la opción de eliminar, el sistema responde realizando la acción y así finaliza la actividad.

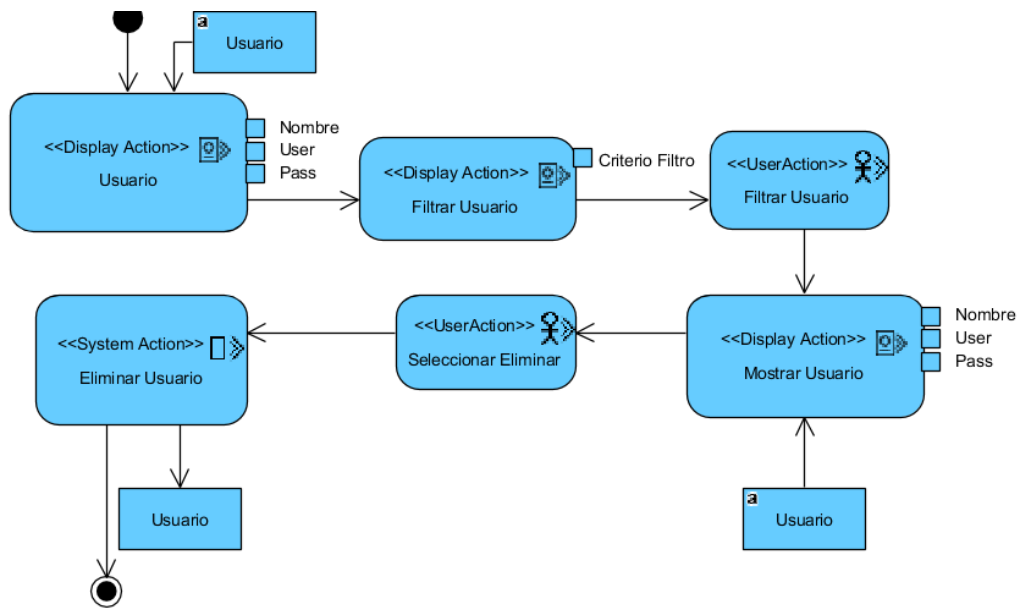


Figura 3.14 Diagrama de actividad del caso de uso Eliminar Usuario.

3.2 Diseño de la aplicación

En esta fase se desarrolló el diseño de solución que permitirá satisfacer los requisitos marcados en la primera fase de análisis, de manera que el cliente conozca su comportamiento, además de estimar tiempos, costos, detectar errores y realizar correcciones a tiempo. La primera actividad es la definición de la arquitectura del sistema para identificar las necesidades a cubrir. En la Figura 3.15 se representa la arquitectura propuesta para la aplicación Web.

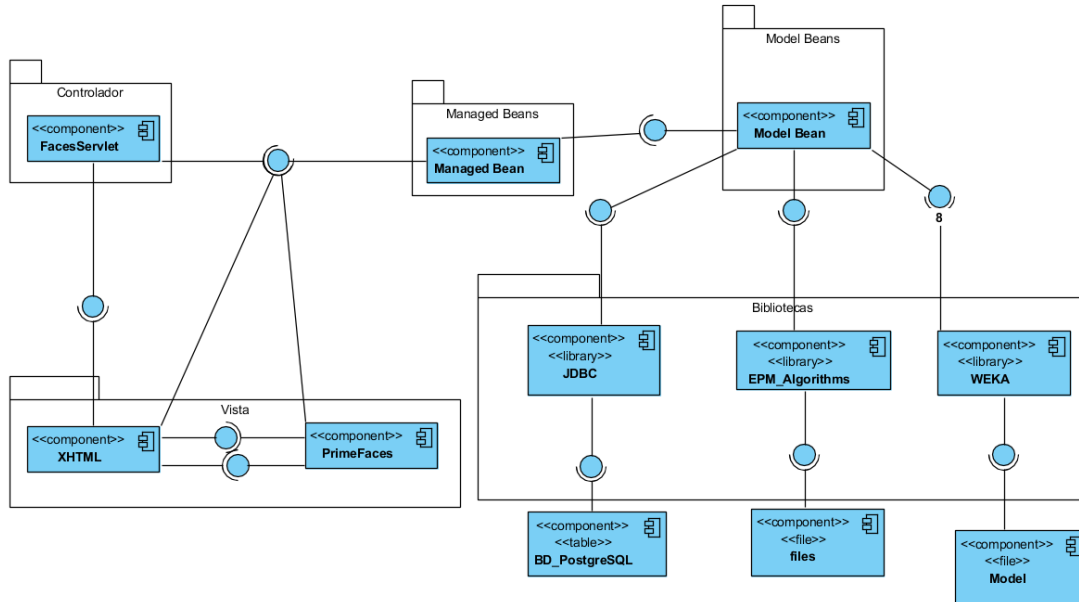


Figura 3.15 Arquitectura propuesta

3.2.1 Arquitectura

Los patrones arquitectónicos ayudan a expresar una estructura de organización para un software, ofrecen además soluciones a problemas de arquitectura y dan una descripción de los elementos y el tipo de relación que existe entre ellos.

Algunos de los beneficios que genera la aplicación de patrones arquitectónicos es la reducción en tiempos de desarrollo, mayor robustez, calidad y facilita el mantenimiento.

La arquitectura propuesta para la aplicación web es Modelo-Vista-Controlador, debido a que permite hacer modificaciones independientes en cada una de las partes sin afectar a la otra.

Patrón MVC

En este patrón de diseño existe un sistema central el cual es el controlador (gestiona las entradas y salidas del sistema), el modelo que busca datos y/o información requerida y finalmente una interfaz que muestra los resultados que solicita el usuario.

Modelo:

Este componente se encarga de gestionar, manipular y actualizar los datos. Si se utiliza una base de datos aquí es donde se realizan las consultas, búsquedas, filtros y actualizaciones, agrupa la lógica de la aplicación, la cual está representada en los *beans* administrados, con acceso a los componentes de la interfaz y con capacidad de pasar la información de los *beans*

de modelo, estos últimos representan las clases importantes del dominio, usando la API (*Application Programming Interface*) de Weka y EPM_Algorithms para generar los modelos de minería de datos de donde se extrae el nuevo conocimiento y el JDBC (*Conectividad a bases de datos de Java*) para controlar el acceso a la base de datos y finalmente lograr manipular la información.

Controlador:

Este componente se encarga de gestionar las instrucciones que se reciben, atenderlas y procesarlas, aquí está contenido el *servlet* de JSF (*JavaServer Faces*), por medio de él se comunican el modelo y la vista, seleccionando la vista adecuada para representarlo.

Vista:

Este componente se encarga de mostrarle al usuario final las pantallas, ventanas, páginas y formularios, mediante archivos XHTML (*eXtensible HyperText Markup Language*), para este propósito se utilizan etiquetas propias de JSF y componentes de la biblioteca *PrimeFaces*.

La arquitectura del software instrumentado, se presentó mediante el Lenguaje de Modelado Unificado (UML) por ser estándar, lo cual facilita el entendimiento y lo hace coherente con el lenguaje que propone UWE, la metodología que se eligió para el desarrollo de la aplicación.

3.2.2 Modelo conceptual

Este modelo también es conocido como modelo de dominio, ayuda a representar un problema de manera gráfica, en este caso representa de manera conceptual el dominio de la aplicación a partir de los requisitos, propone diagramas como: conceptual, lógico y físico de la base de datos, como se observa en las Figuras 3.16, 3.17 y 3.18.

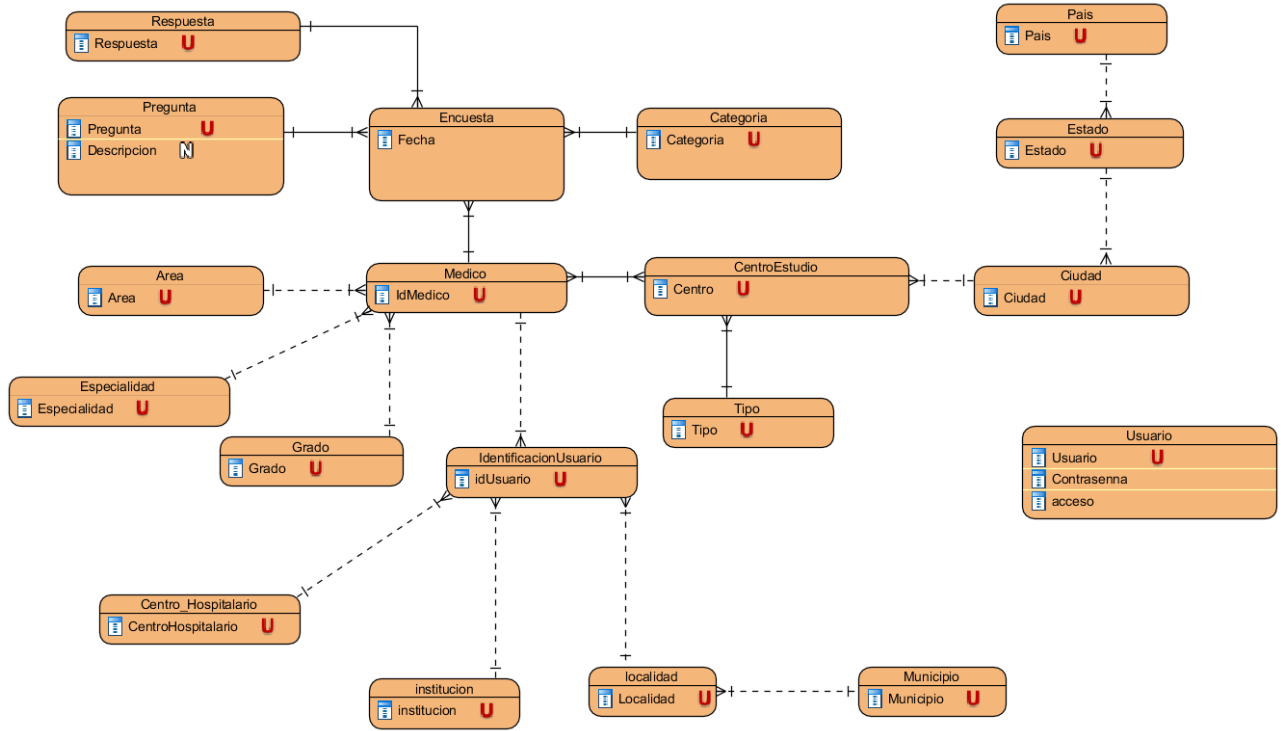


Figura 3.16 Diagrama conceptual de la aplicación

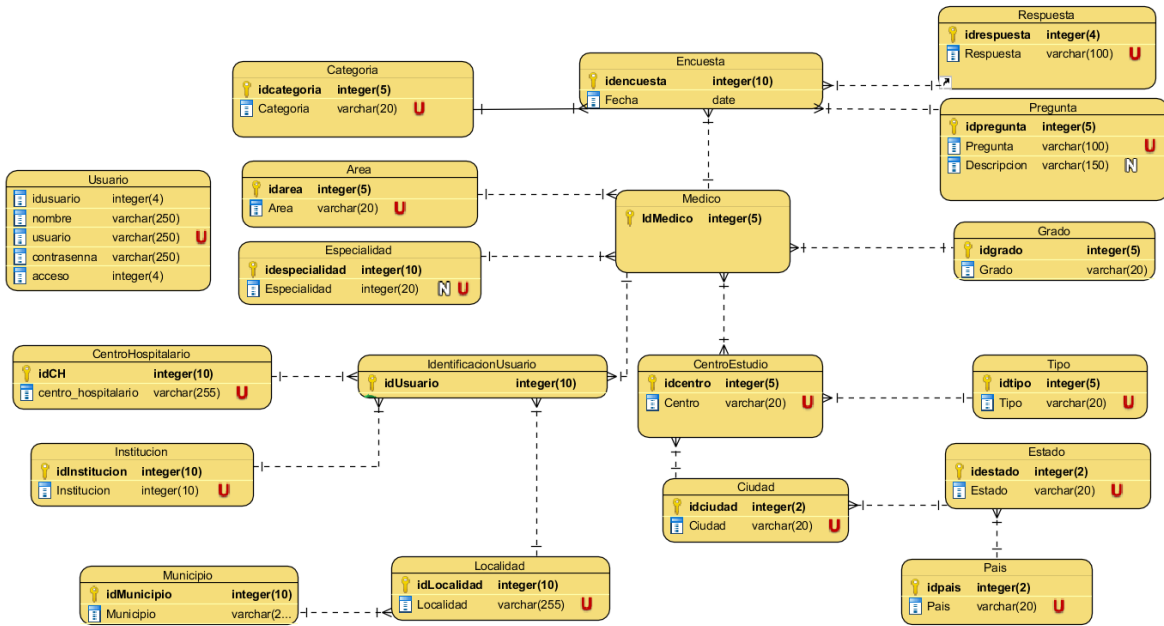


Figura 3.17 Diagrama lógico de la aplicación

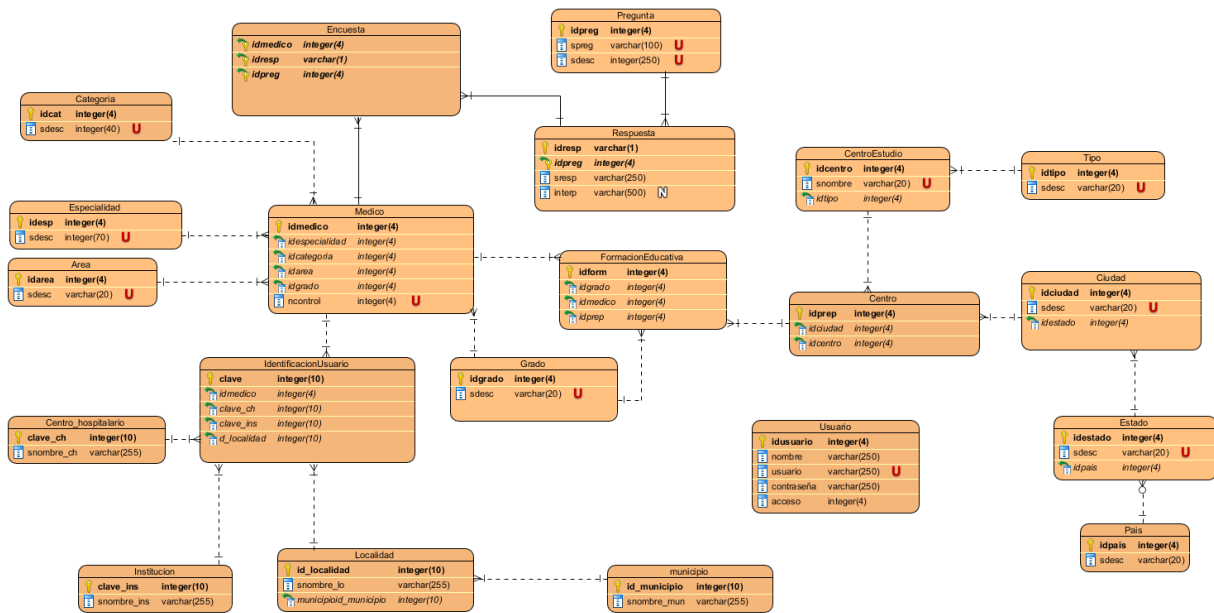


Figura 3.18 Diagrama físico de la aplicación

Tablas de la base datos

Categoría: Almacena la información referente a la forma en la que el encuestado accedió a responder la encuesta, por ejemplo, si fue mediante una "Invitación interna" o "Invitación accionada fuera del servicio".

Especialidad: Guarda el valor que permitirá indicar a qué campo pertenece un encuestado. Para este caso de estudio el valor que hace referencia al sector médico es "Medicina".

Área: Registra la información que señala en calidad de qué presta servicio el encuestado al hospital, por ejemplo, si es "Adscrito", "Residente" o "Interno".

Grado: Guarda el valor que permitirá indicar el nivel de especialización dentro del campo médico que tiene el encuestado, como "Práctica Universitaria", "Medicina General" o "Especialidad".

Centro Estudio: Contiene los nombres de centros que están involucrados en la enseñanza de medicina. Algunos ejemplos son: "Facultad Medicina", "Universidad Veracruzana", "Hospital Regional de Río Blanco", entre otros.

Tipo: La información contenida en esta tabla responde a la clasificación de los centros de preparación, como son "Escuela" u "Hospital".

Ciudad, Estado, País: Estas tablas y sus relaciones permiten darles una ubicación geográfica a los centros educativos. Por ejemplo, es posible conocer si en un momento determinado se hace referencia a la "Universidad Veracruzana" del estado "Veracruz" y ciudad "Mendoza", o del estado de "Veracruz" y ciudad "Xalapa".

Pregunta: La tabla registra cada una de las preguntas que conforman la encuesta. Como ejemplo se mencionan las preguntas "Años de práctica" y "¿Quién considera que debe solicitar la autopsia?".

Respuesta: La tabla respuesta contiene todas las respuestas posibles para la encuesta y establece la relación de éstas con su pregunta correspondiente. Ejemplo de ello son casos como "Menos de 5" y "Más de 20", que son posibles respuestas de la pregunta "Años de práctica". La relación se establece mediante la llave foránea *idpreg* en la tabla Respuesta, que responde a la llave primaria *idpreg* de la tabla Pregunta.

Centro: De acuerdo a las relaciones establecidas con las tablas Centro Estudio y Ciudad se logra asignar a cada centro de preparación médica su ubicación geográfica.

Médico: Es la tabla que almacena las encuestas y relaciona el número de control de cada una con su categoría, especialidad, grado y área.

Formación Educativa: En esta tabla se establecen relaciones con las tablas Médico, Grado y Centro; de esta manera se registra la información de los centros donde se formó cada encuestado.

Encuesta: Las relaciones establecidas con las tablas Médico, Pregunta y Respuesta permiten registrar las respuestas para cada una de las preguntas por cada encuestado.

Clasificado: En esta tabla persisten las respuestas de las preguntas abiertas con su clasificación correspondiente, resultado del proceso de minería para la clasificación de texto.

Interpretación: Con el objetivo de dar una explicación lógica de los resultados que generan los algoritmos de minería aplicados en esta investigación, esta tabla guarda el significado y la descripción para cada atributo contenido en los conjuntos de datos analizados.

Matriz_binaria y Vista_minable: Estas tablas representan a toda la información de las encuestas insertadas en la base de datos y se crean automáticamente por funciones SQL.

Menú: La estructura para el menú de la aplicación está contenida en esta tabla y brinda la posibilidad de personalizarse de acuerdo a los diferentes usuarios.

IdentificacionUsuario: Esta tabla contiene la información del usuario, como el centro hospitalario, institución, localidad y médico, por lo que está relacionada con varias tablas, esta tabla ayuda a saber de qué centro contestaron la encuesta.

Municipio: Proporciona los nombres de los municipios.

Localidad: Proporciona los nombres de las localidades y a qué municipio pertenecen.

CentroHospitalario: Almacena el nombre de todos los centros hospitalarios.

Institución: Guarda el nombre de todas las instituciones.

3.2.3 Modelo de navegación

Este modelo define la estructura de *links* y nodos de la aplicación o WebApp, muestra cómo se realiza la navegación utilizando elementos de acceso como índices, consultas y vistas guiadas, se representa mediante un diagrama de clases estereotipado, el cual abarca todos los caminos posibles para la navegación de los usuarios.

En la Figura 3.19 se muestra el diagrama de navegación, se muestran todos los caminos en el cual el usuario se mueve de un lugar a otro dentro de la aplicación. Partiendo de la página de inicio, desde el Menú Principal el usuario elige ir a consultar los resultados, insertar una encuesta (desde ahí tiene la opción de insertar un centro de estudio) o iniciar sesión, si decide seleccionar el inicio de sesión tiene distintas opciones, si es un usuario administrativo, navega hasta la página de Usuarios, ahí es capaz de insertar un nuevo usuario, ver información al seleccionar uno específico, modificar y eliminar.

Si inicia sesión como especialista navega hasta la página de Encuestas, en donde ve la lista de las encuestas, elige una en específico, desde ese punto navega hacia la página en donde se

mostrará la información completa de la encuesta o tiene la opción de eliminar encuesta, otra opción que tiene es la de generar los modelos de minería.

Ambos usuarios, administrador y especialista, tienen la opción de cerrar sesión.

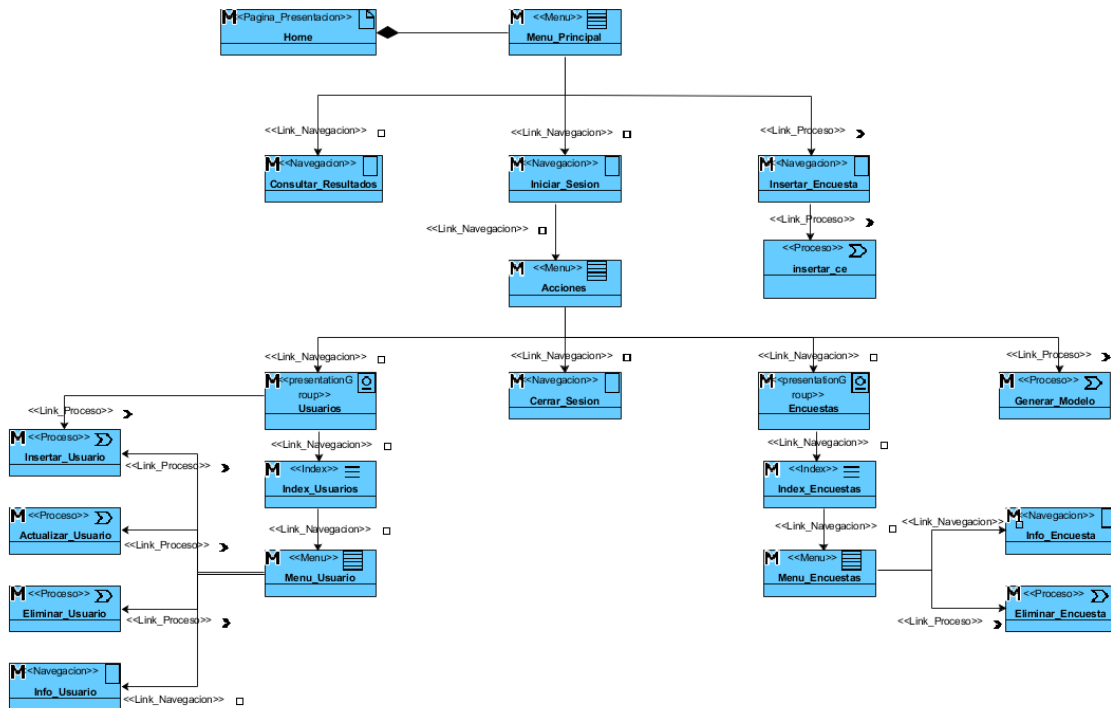


Figura 3.19 Diagrama de navegación de la aplicación Web

3.2.4 Modelo de presentación

En el modelo de presentación cada una de las clases que se muestran en el modelo de navegación, se diseñan como *storyboard*, se lleva a cabo el modelado del diseño de las propiedades y elementos de la página web. La WebApp tendrá seis páginas, las cuales se describen a continuación.

En la Figura 3.20 se muestra la página de *Inicio*, da la bienvenida mediante un texto a los usuarios que acceden a la aplicación, está compuesta por un menú principal que despliega opciones de acuerdo a los permisos de cada usuario y un formulario en el cual los usuarios registrados abren sesión y cierran sesión.

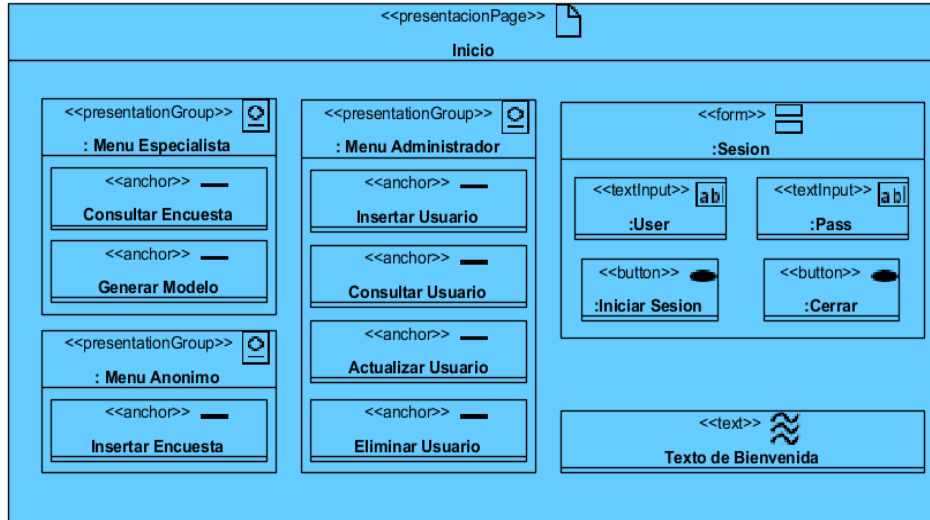


Figura 3.20 Página Inicio del modelo de presentación

La página *Gestionar Encuesta* se representa en la Figura 3.21, permite solo al especialista realizar acciones, se muestra una tabla con algunos datos de las encuestas, las opciones de seleccionar y eliminar, vinculadas a cada una de ellas. Cuenta con un formulario *Filtrar Encuesta* en el cual se debe especificar cuál se desea gestionar, si el especialista desea ver la información de alguna de las encuestas, se elige la operación *Seleccionar* y esta hace una llamada a la página *Encuesta*, en la alternativa de *Mensajes* se indicarán los errores de validación o de confirmación.

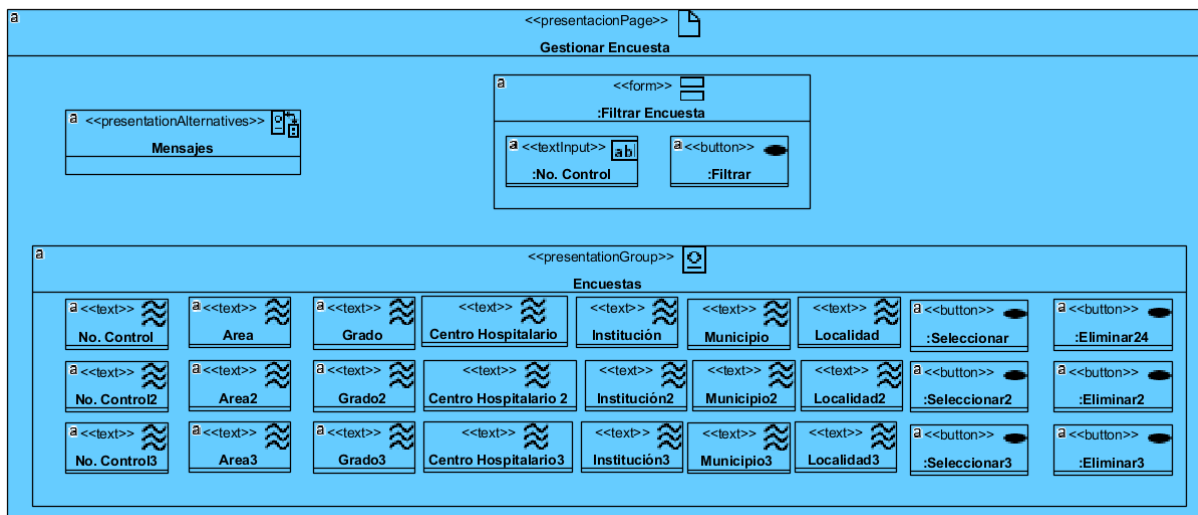


Figura 3.21 Página Gestionar Encuesta del modelo de presentación

En la Figura 3.22 se representa la página *Encuesta*, proporciona un formulario para responder una nueva encuesta o visualizarla, de acuerdo a la operación que haya hecho la llamada a esta página. Los mensajes de error y validación en esta página se manejan igual que las anteriores.

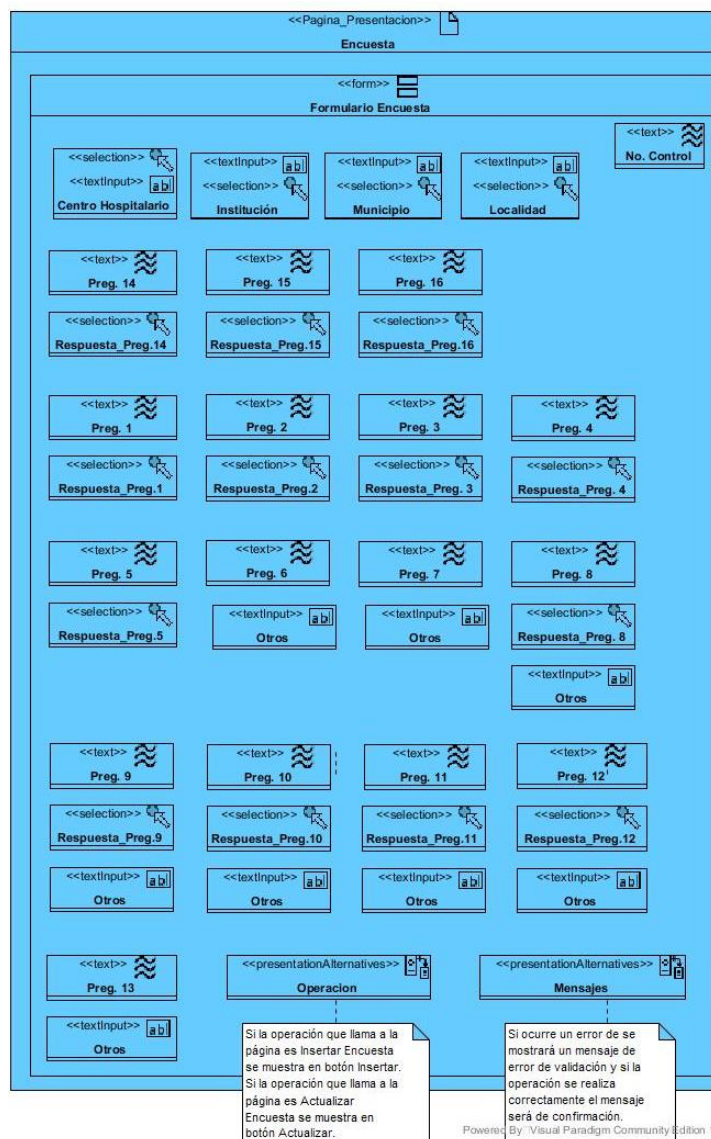


Figura 3.22 Página Encuesta del modelo de presentación

El especialista es capaz de generar los modelos desde la página *Generar Modelos*, donde selecciona el conjunto de datos que va a minar y el algoritmo que quiere aplicar especificando sus parámetros. Al generar el modelo, se muestra a un lado el resultado, incluyendo el número de regla, la regla y la interpretación de cada regla, como se ve en la Figura 3.23.

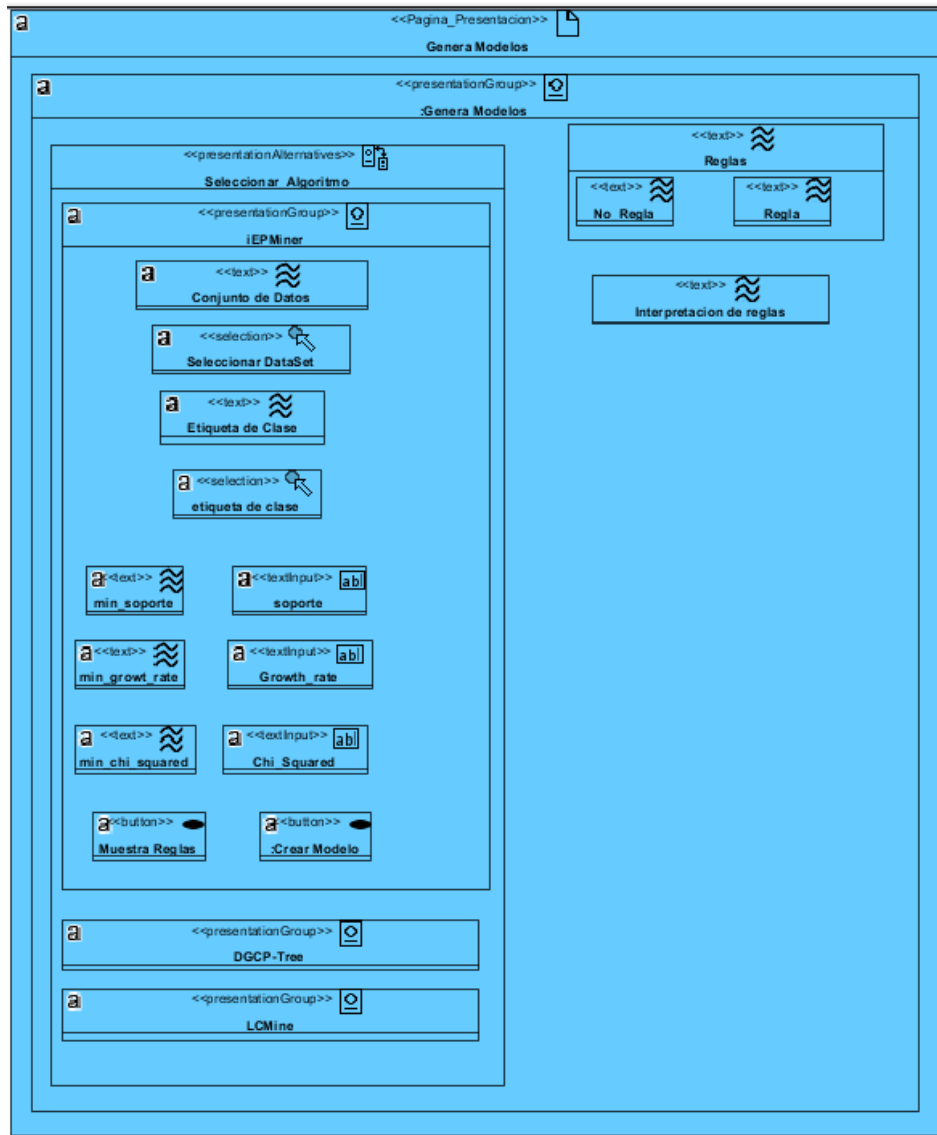


Figura 3.23 Página Generar Modelos del modelo de presentación

En la página *Resultados*, el especialista y encuestado ven la interpretación de los modelos generados, después de seleccionar el algoritmo y el conjunto de datos, como se muestra en la Figura 3.24.

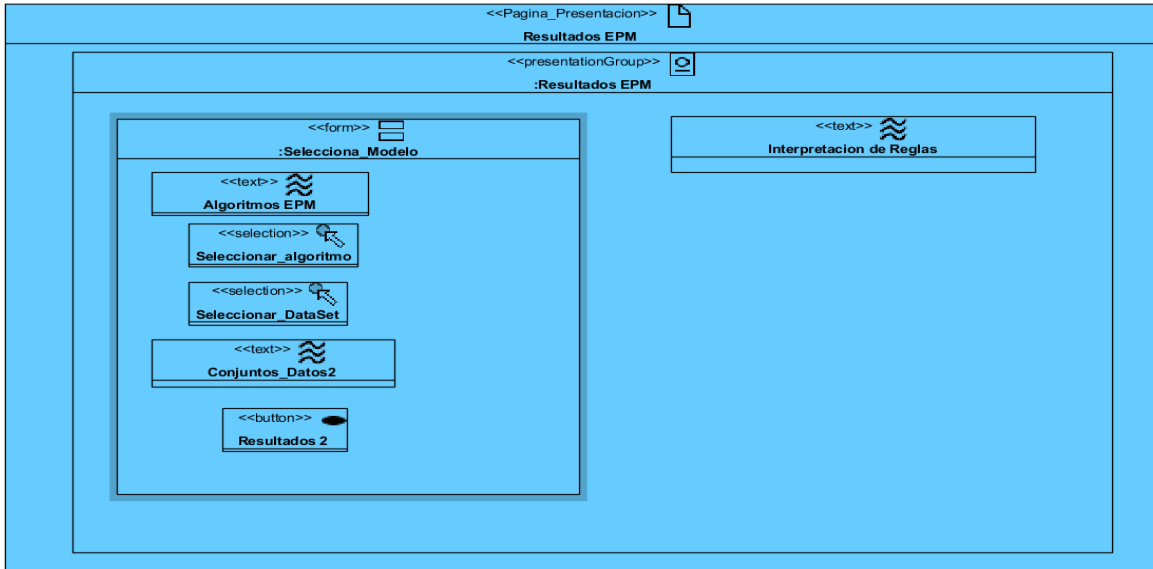


Figura 3.24 Página Resultados EPM del modelo de presentación

En la Figura 3.25 se muestra los parámetros correspondientes a cada algoritmo.

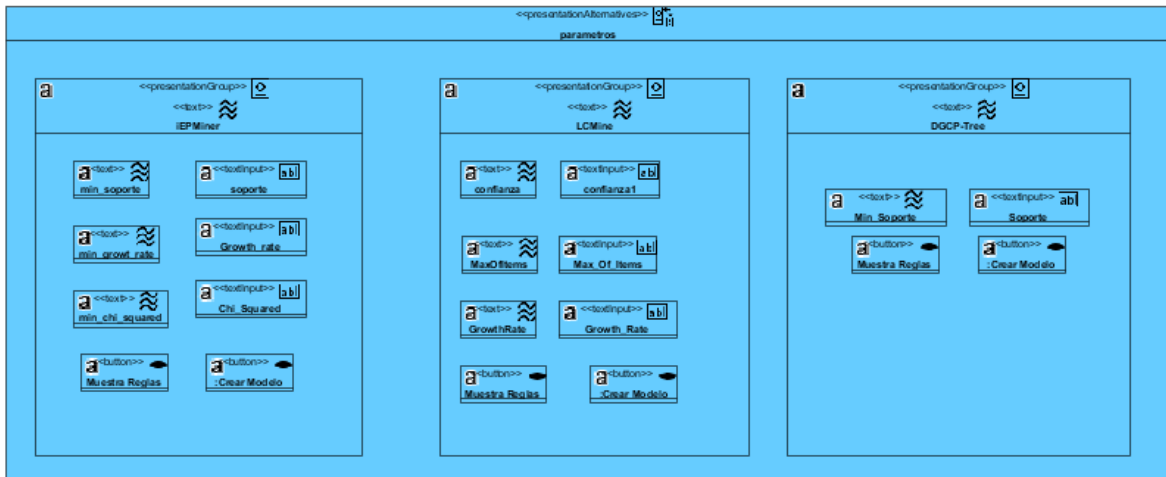


Figura 3.25 Área alternativa Parámetros

La página *Resultados_EPM* muestra la interpretación de cada regla para cada algoritmo, como se ve en la Figura 3.26.

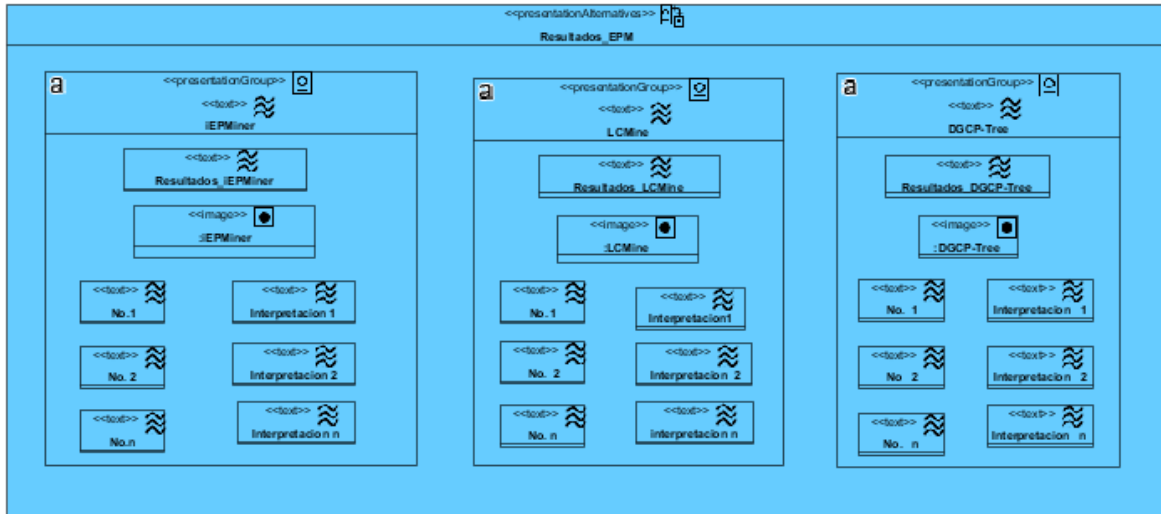


Figura 3.26 Área alternativa Resultados_EPM

3.2.5 Modelo de procesos

El modelo de proceso es representado como un diagrama de actividades en donde se detallan los casos de uso, también se describe como un método de diagrama de flujo el cual modela los pasos de un proceso. A continuación, se presentan los diagramas de cada caso de uso que se especificaron inicialmente.

En el proceso del caso de uso *Responder Encuesta*, como vista inicial se muestra un formulario en el cual el usuario tiene que responder las preguntas de la encuesta, una vez contestado todo correctamente, el sistema pasa a validar los datos, si los datos son incorrectos el sistema envía un mensaje de error, por lo que se mostrará nuevamente el formulario para corregir la información, cuando el sistema valida los datos y son correctos, procede a guardar la nueva encuesta y lanza el caso de uso *Minar Texto*, el cual es el encargado de clasificar las respuestas de las preguntas abiertas insertadas por el usuario, como se observa en la Figura 3.27.

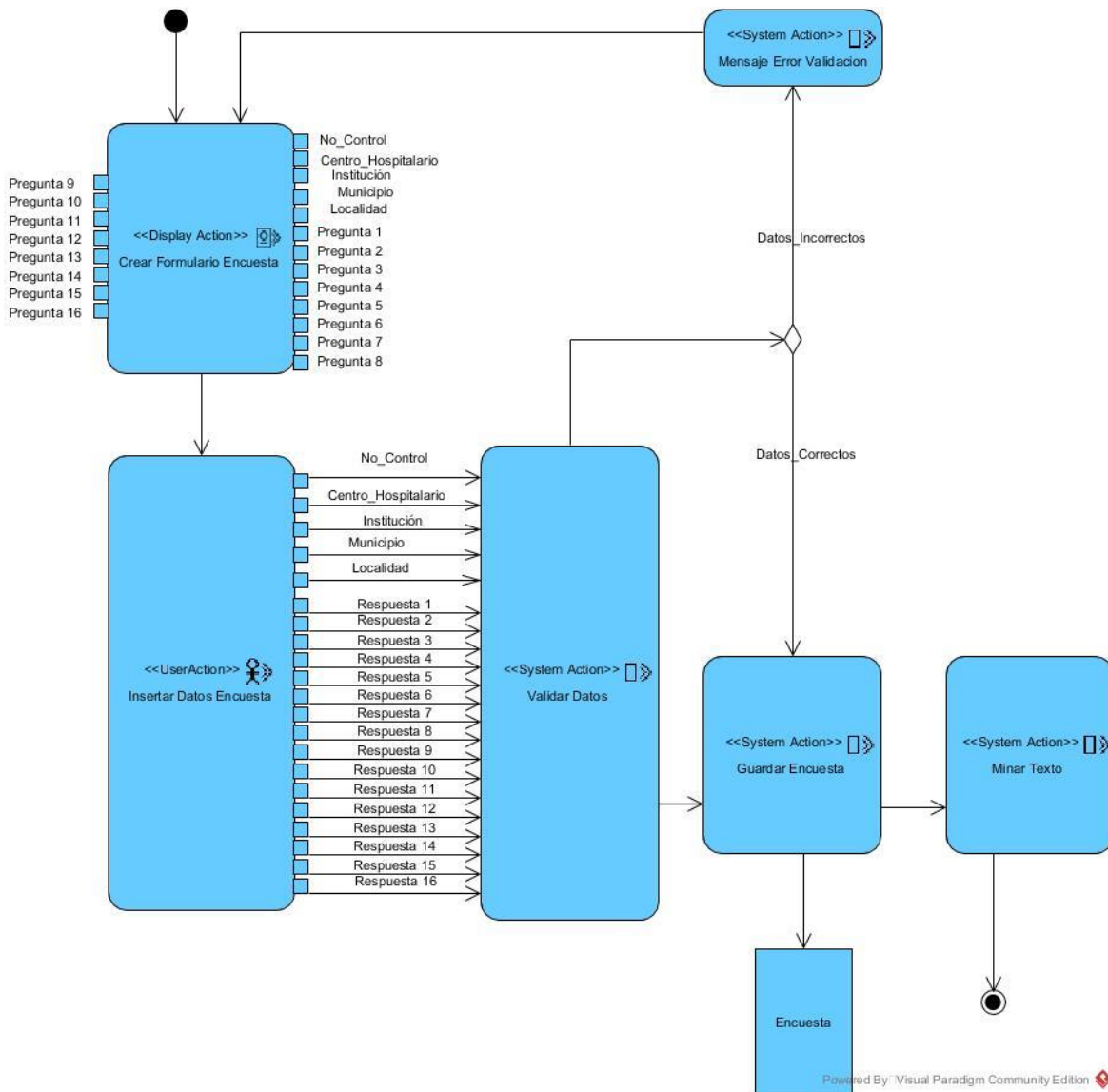


Figura 3.27 Diagrama de actividad del caso de uso Responder Encuesta

En la Figura 3.28 se muestra el proceso para el caso de uso *Consultar Encuesta*, inicia mostrando el listado de las encuestas y un formulario en donde el usuario ingresa el número de control de la encuesta que quiere filtrar, si no se obtienen resultados con el valor ingresado por el usuario, el sistema presenta un mensaje indicando que la encuesta no existe, el usuario debe aceptar y regresa nuevamente al listado en donde se encuentran las encuestas, si el filtro encuentra el valor especificado, entonces el sistema muestra los datos de la encuesta.

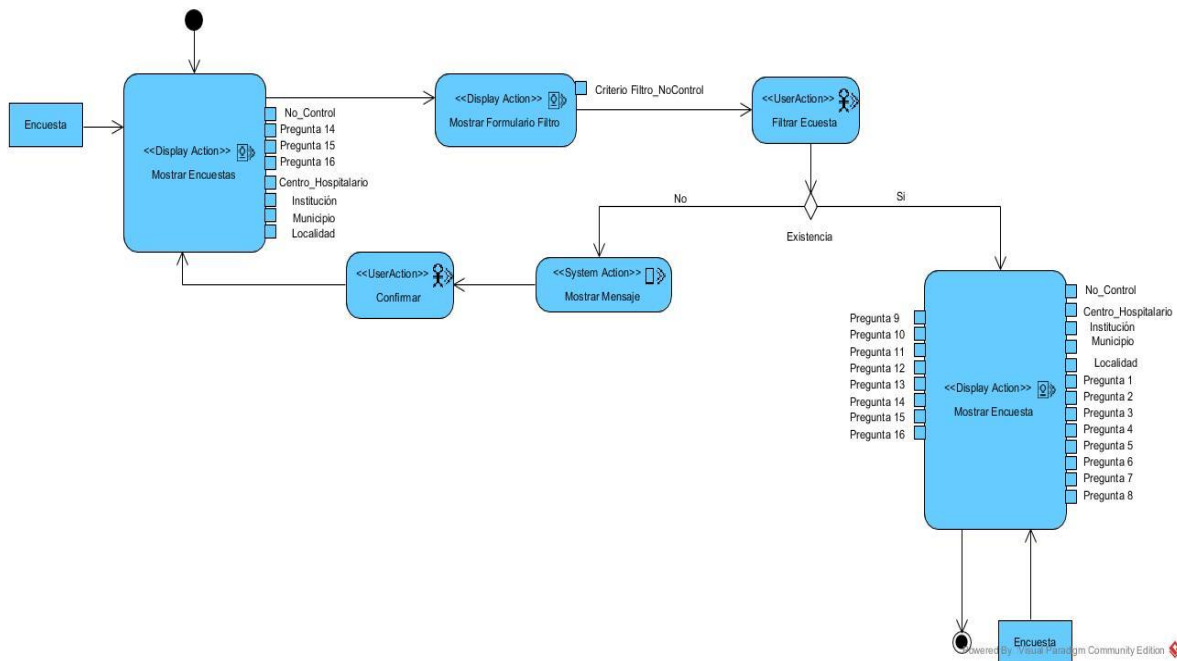


Figura 3.28 Diagrama de actividad del caso de uso Consultar Encuesta

El proceso de *Eliminar Encuesta* inicia mostrando en pantalla el listado de las encuestas, mediante un formulario el usuario filtra información proporcionando el número de control de la encuesta que desea eliminar, si no existe, el sistema lanza un mensaje en donde indica que la encuesta no existe, el usuario acepta y regresa el listado de las encuestas, cuando el sistema encuentra la encuesta presenta su información, en ese momento el usuario selecciona la opción de eliminar y el sistema muestra un mensaje en donde el usuario debe confirmar para eliminar preceder con la operación, si el usuario cancela la operación, entonces el sistema le vuelve a mostrar la lista de encuestas y en caso contrario, elimina la encuesta seleccionada por el usuario, como se muestra en la Figura 3.29.

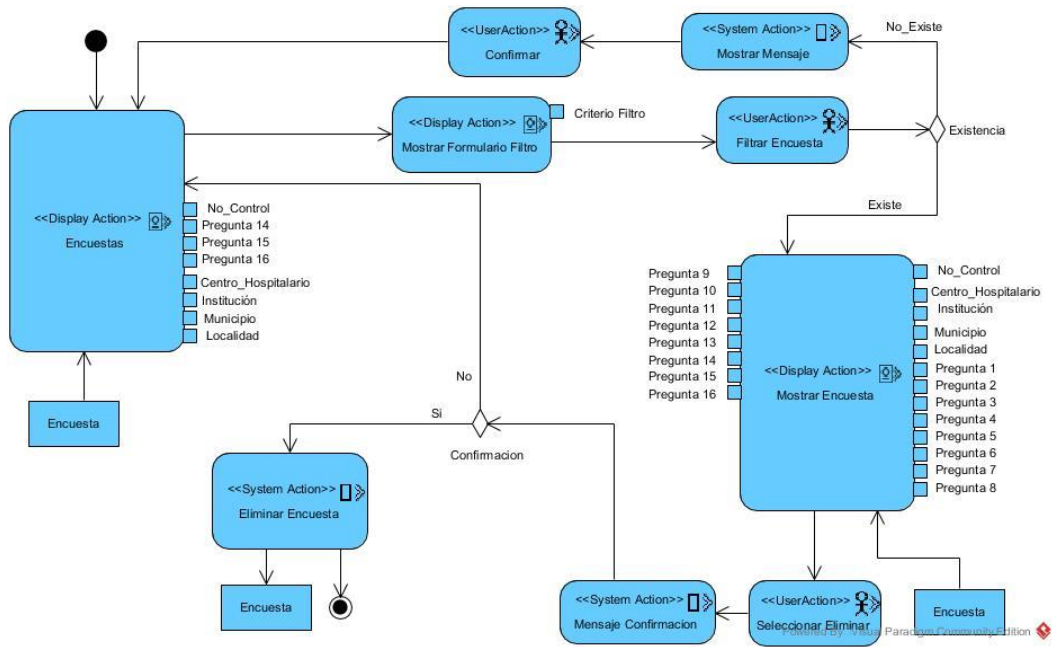


Figura 3.29 Diagrama de actividad del caso de uso Eliminar Encuesta

El proceso *Generar Modelo* inicia mostrando al especialista los resultados del modelo existente, el especialista selecciona el algoritmo de minería de patrones emergentes que desea aplicar y el sistema procede a generar el modelo correspondiente, en caso que haya un fallo el sistema muestra un mensaje de error y regresará al escenario donde comienza el caso de uso, si la operación es exitosa, el modelo se guarda físicamente en la aplicación, como se ve en la Figura 3.30.

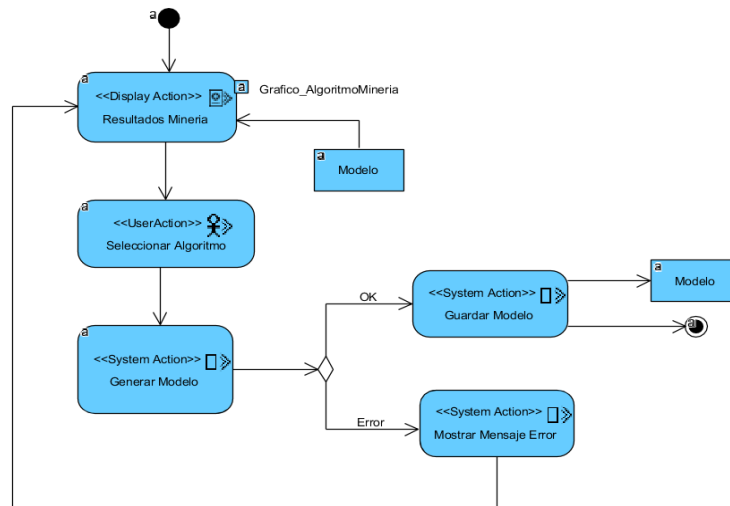


Figura 3.30 Diagrama de actividad del caso de uso Generar Modelo

En la Figura 3.31 se indica el proceso para la interpretación de los resultados de los modelos, en donde inicia cuando el sistema comienza la interpretación del modelo, si hay un error, el sistema notifica al usuario mediante un mensaje y regresa a interpretar el modelo nuevamente, si la interpretación fue exitosa, muestra los resultados.

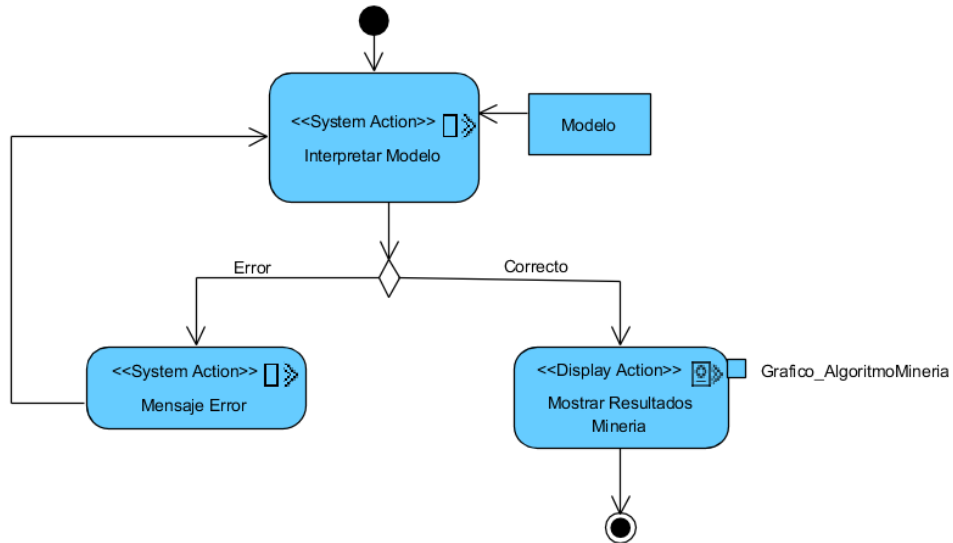


Figura 3.31 Diagrama de actividad del caso de uso Interpretar Resultados de Minería

3.3. Desarrollo de los modelos de minería de datos utilizando KDD

Se utilizó el proceso guiado por la metodología KDD como se muestra en la Figura 3.32, en este apartado se describe la aplicación de la minería de datos, con el objetivo de encontrar conocimiento útil para identificar los motivos del porqué cada vez se realizan menos autopsias en el H.R.R.B.

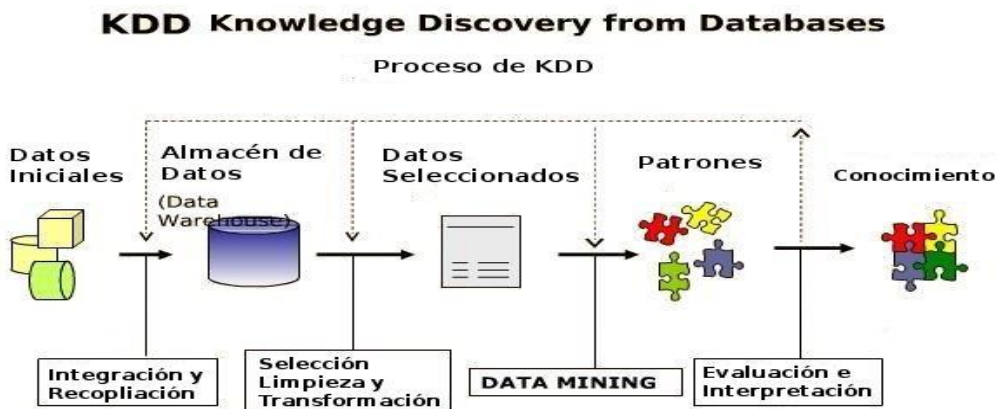


Figura 3.32 Proceso de descubrimiento del conocimiento

3.3.1 Selección de Datos

Al identificar la problemática que aborda esta investigación, se observó que no existía una fuente de información que permitiera de forma directa investigar y especificar los motivos por los que no se solicitan las autopsias en el Hospital de Río Blanco y compararlas con las de otro hospital, para saber si son los mismos motivos causas y/o circunstancias que en el H.R.R.B.

Por esta razón, se consideró utilizar una encuesta como técnica de investigación social cuyo objetivo es indagar sobre la opinión, actitudes y/o comportamientos de los médicos ante la realización de autopsias, al igual que los valores, creencias y/o motivos que los caracterizan.

La encuesta contiene un total de dieciséis preguntas, tres de tipo abiertas y trece de tipo cerradas, las cuales cinco incluyen una sección para especificar otras opciones consideradas por los encuestados, también cuenta con cuatro datos extras los cuales son: centro hospitalario, institución, municipio y localidad.

Se presenta un resumen a continuación de la información identificada de acuerdo al instrumento:

- 27 categorías relacionadas con factores que los médicos consideran negativos para la realización de autopsias y 26 más para los factores positivos.
- Nueve motivos por los que el familiar no solicita el estudio y ocho posibles razones por las que no se realizan suficientes autopsias en el hospital.
- Respecto a la opinión de los médicos sobre el procedimiento para solicitar una autopsia, se consideraron 14 métodos eficientes y seis opciones sobre personal adecuado para solicitarla.
- Las respuestas de comentarios generales que dieron los médicos se consideraron en 25 categorías.
- Tres posibles respuestas para el área y el grado del médico y cinco para cada una de las tres preguntas relacionadas con la opinión de los médicos sobre los hallazgos encontrados en las autopsias.

La Tabla 3.2 muestra un resumen de la encuesta aplicada y la cantidad de categorías generadas por respuesta.

Tabla 3.2 Resumen de la encuesta aplicada al personal médico

Aspectos	Preguntas	Tipo de Pregunta	Categorías generadas
Información del médico	Centro Hospitalario	Abierta	4
	Institución	Abierta	4
	Municipio	Abierta	2
	Localidad	Abierta	3
Formación del médico	Área	Cerrada	3
	Grado	Cerrada	3
	Centro Formación Medicina General	Cerrada	49
	Centro Formación Especialidad	Cerrada	49
Experiencia del médico	Años de práctica	Cerrada	5
	Casos de autopsias en los que ha participado	Cerrada	5
Opinión del médico sobre los hallazgos de autopsias	Originan discrepancias con los diagnósticos clínico	Cerrada	5
	Originan casos de demandas	Cerrada	5
	Originan casos de arbitraje	Cerrada	5
Opinión del encuestado ante la solicitud de autopsias	Motivos para que el médico la solicite	Abierta	26
	Motivos para que el médico no la solicite	Abierta	27
	Motivos para que el familiar no la solicite	Cerrada	8
	Motivos por los que en el hospital no se realizan suficientes autopsias	Cerrada	9
Opinión del médico sobre el procedimiento para solicitar una autopsia	Personal adecuado para solicitar autopsias	Cerrada	7
	Método para solicitar autopsia por parte del médico	Cerrada	11
	Método eficiente para solicitar autopsia	Cerrada	11
Aspecto General	Comentarios	Abierta	25

3.3.2 Pre-procesamiento de Datos

Las respuestas de los médicos se transformaron en dos representaciones adecuadas: *matriz_binaria* y *vista_minable*, para aplicar técnicas de minería de datos, como se observa en la Figura 3.33. Estas estructuras se crearon mediante funciones SQL (*Structured Query Language*), Lenguaje de Consulta Estructurado), teniendo así dos conjuntos de datos distintos a partir de los mismos datos, C (*matriz-binaria*) y D (*vista-minable*).

También se crearon los conjuntos de datos *mcc_aut*, *mcc_no_aut* y *com_sug_op* que se utilizarán para la clasificación de las respuestas de los encuestados de las preguntas abiertas como son: motivos para solicitar autopsias, motivos para no solicitarlas y comentarios, respectivamente. Estos conjuntos contienen las respuestas de los encuestados a las preguntas anteriormente mencionadas y se consideraron como valores para el atributo de clase las categorías seleccionadas a priori de conjunto con el experto.

Vista_minable: En esta tabla, las filas significan las combinaciones de respuestas para las encuestas y las columnas representan las respuestas.

El valor de cada columna responde a la intersección que se lee como un par <pregunta, respuesta> (Figura 3.33).

A continuación, se muestran las características de los conjuntos de datos C y D en la Tabla 3.3.

Tabla 3.3 Características de los conjuntos de datos C y D

Características	Conjunto C	Conjunto D
Atributos	166	18
Instancias	93	7859
Tipo de datos	Nominal-Binarios-Asimétricos	Nominal
Descripción	Matriz binaria <respuesta, valor>	Matriz representada por <pregunta, respuesta>
Valores faltantes	Sí	No
Valores fuera de rango	No	No
Valores inconsistentes	No	No

3.3.3 Minería de Datos

En la fase de transformación se analizaron las características de los conjuntos de datos y se consideró que no era necesario realizar alguna transformación, por lo que se procedió a la fase de minería de datos.

De acuerdo a los datos y al objetivo principal de este proyecto se consideraron dos tareas de la minería (caracterización y discriminación) para atacar el problema. Se aplicó la minería de patrones emergentes para encontrar las diferencias entre clases (H.R.R.B y otro hospital).

3.3.3.1 Análisis de minería de patrones emergentes

Se analizaron cada uno de los algoritmos que propone el *framework EPM_Algorithms* y se consideró la posibilidad de su aplicación en los conjuntos C y D. A continuación, se muestran los aspectos de cada uno de ellos.

EPM-Framework [25] es un marco de trabajo públicamente disponible y escrito en Java que permite ejecutar algoritmos de EPM en varios conjuntos de datos. Algunos algoritmos de EPM implementados en este marco de trabajo son **BCEP** (*Bayesian Classification based on Emerging Patterns*, Clasificación Bayesiana basada en Patrones Emergentes) [26], **DeEPS** (*Decision-making by Emerging Patterns*, Toma de Decisiones por medio de Patrones Emergentes) [27], **iEPMiner** (*Interesting Emerging Pattern Miner*, Minero de Patrones Emergentes Interesantes) [28], **Tree-based JEP-C** (*Tree-based Jumping Emerging Pattern Classifier*, Clasificador de Patrones Emergentes de Saltos basado en Árbol) [29], **Top-k**

minimal SJEPs (*Strong Jumping Emerging Patterns*, Patrones Emergentes de Saltos Fuertes) [30], **SJEP-C** (*Strong Jumping Emerging Pattern Classifier*, Clasificador de Patrones Emergentes de Saltos Fuertes) [31], **DGCP-Tree** (*Dynamically Growing Contrast Pattern Tree*, Árbol de Patrones de Contraste que Crecen Dinámicamente) [12], **EvAEP** (*Evolutionary Algorithm for Extracting Emerging Patterns*, Algoritmo Evolutivo para Extraer Patrones Emergentes) [32], **CEPMine** (*Crisp Emerging Pattern Mining*, Minería de Patrones Emergentes Definidos) [33], **LCMine** (*Logical Complex Miner*, Minero de Lógica Compleja) [34] y **EP-Random Forest** [12].

BCEP [13]. Este método utiliza la estructura del Árbol de patrones (PTree) para extraer el conjunto completo de SJEPs. El árbol se construye ordenando los selectores por GR (Tasa de crecimiento). La estrategia de minería se realiza mediante un primer SJEP de búsqueda transversal profunda. El árbol se poda mediante un umbral de soporte mínimo o cuando se encuentra un JEP, para obtener solo SJEPs. Después de eso, el método mantiene solo los SJEPs más importantes con una estrategia de filtrado similar al procedimiento de competencia de *tokens*. En este procedimiento, solo se mantienen los SJEPs que cubren ejemplos que aún no están cubiertos por patrones mejor clasificados.

DeEPS [13]. Este método se basa en el aprendizaje diferido, es decir, los datos de entrenamiento son el modelo en sí. Para cada nueva instancia de prueba, los selectores que no pertenecen a la instancia de prueba se eliminan del modelo.

Top-k minimal SJEPs. En [30] se consideró el problema de extraer los patrones emergentes de saltos (JEPs) mínimos *top-k*, es decir, los k JEPs mínimos (un EP es mínimo cuando sus subpatrones no son EPs, los EPs mínimos son interesantes para la descripción porque, en general, contienen un pequeño número de variables [13]) más admitidos en cada clase de decisión, por lo que se propuso el algoritmo Top-k minimal SJEPs que sigue un enfoque basado en un árbol de patrones de contraste (*CP-Tree*) para encontrar JEPs fuertes (los EPs son fuertes cuando son EPs de saltos y mínimos, los EPs fuertes son fáciles de entender y tienen un alto poder predictivo [13]). Se poda el espacio de búsqueda con un soporte mínimo, cuyo valor se aumenta gradualmente a medida que se descubren nuevos JEPs mínimos. Para permitir esto, se verifica que cada nuevo JEP sea mínimo en el momento en que se encuentra en lugar de al final del

proceso. Este algoritmo logra ahorros significativos tanto en tiempo como en el número de patrones examinados, cuando se solicita un pequeño número de JEPs altamente compatibles.

iEPMiner [28]. Es un algoritmo eficiente para la minería de patrones emergentes interesantes basados en una estructura de datos de árbol, donde la prueba de *chi-Squared* se utiliza como heurística para podar el espacio de búsqueda, esto hace que *iEPMiner* sean más rápido. La heurística siempre descubre el 90% de EPs más interesantes, que son suficientes para construir clasificadores de alta precisión en muchas aplicaciones de la vida real.

Tree-based JEP-C [29]. Este es un algoritmo eficiente para minar JEPs que tiene dos características: usa una estructura de datos basada en árbol para almacenar los datos y el desarrollo del algoritmo de minería opera directamente en los datos contenidos en los árboles. Esto logra ganancias de desempeño considerables en comparación con otros enfoques.

LCMine [34]. Es un algoritmo eficiente para encontrar regularidades discriminatorias en una muestra de entrenamiento con datos mixtos e incompletos para la clasificación supervisada. Este algoritmo se basa en una inducción de árbol de decisión diversa y una etapa de filtrado posterior al procesamiento, que permite encontrar un conjunto reducido de propiedades discriminatorias de alta calidad para cada clase. *LCMine* tiene tres características distintivas: 1) No aplica una discretización *a priori* en las características numéricas, a diferencia de la mayoría de los algoritmos para extraer regularidades discriminatorias; 2) Utiliza una representación extendida para las regularidades, y 3) Utiliza una nueva estrategia de filtrado para eliminar las regularidades redundantes.

DGCP-Tree [12]. Es un método para minar SJEPs con base en un *DGCP-Tree* (*Dynamically Growing Contrast Pattern Tree*, Árbol de Patrones de Contraste que Crece Dinámicamente). Un *DGCP-Tree* almacena parcialmente patrones y sus arreglos de códigos de rutas en ambas clases. Este método es eficiente y capaz de manejar conjuntos de datos de altas dimensiones.

EvAEP [32]. Su principal objetivo es la extracción de un número variable de reglas o patrones que describan información con respecto a una variable objetivo. Considerando que es posible que una variable objetivo tenga distintos valores, el algoritmo es capaz de obtener patrones para todos los valores de esa variable objetivo al ejecutarse una vez para cada valor.

CEPMiner [33]. Un algoritmo novedoso para encontrar patrones emergentes, que no aplica la discretización global de atributos numéricos. CEPMiner extrae patrones usando un procedimiento especial, de una colección de árboles de decisión C4.5. Para encontrar una colección representativa de patrones.

Se evaluaron los 11 algoritmos considerando cada atributo como etiqueta de clase, tomando en cuenta las siguientes métricas de evaluación:

- Velocidad. Se refiere al tiempo en segundos que tardó la ejecución de cada algoritmo.
- #Patrones con confianza > 0.6 . EPM-Framework permite obtener un archivo con los patrones que superan un valor de 0.6 en confianza en los datos de prueba, la confianza se define como la precisión de la capacidad predictiva de un patrón para la clase positiva.
- #Patrones. Número de patrones.
- FPR. Tasa de Falsos Positivos, mide el porcentaje de ejemplos incorrectamente cubiertos con respecto a la cantidad total de ejemplos negativos, se busca minimizar este valor.
- GR. Tasa de crecimiento, esta es la medida que define un patrón emergente, calcula la proporción entre el soporte del patrón en la clase positiva y el soporte en la negativa. Se interpreta como el poder discriminativo de un patrón.
- WRAcc. *Weighted Relative Accuracy*, Precisión Relativa Pesada, mide el intercambio entre la ganancia de generalidad y de precisión. Un valor alto se interpreta como un alto balance entre la generalidad del patrón y su confianza.
- TPR. Tasa de Verdaderos Positivos, mide el porcentaje de ejemplos correctamente cubiertos con respecto al número total de ejemplos positivos.

Se estudiaron cada uno de los algoritmos que propone EPM_Framework y se consideró la posibilidad de su aplicación en los conjuntos C y D (ver Tabla 3.4).

Tabla 3.4 Aplicación de los algoritmos en los conjuntos de datos

Algoritmos	Conjunto de datos	
	C	D
DeEPS	✗	✓
LCMine	✗	✓
CEPMine	✗	✓
EP-Random Forest	✗	✓
EvAEP	✓	✓
Tree-based JEP-C	✗	✓
BCEP	✗	✓
iEPMiner	✓	✓
SJEP-C	✗	✓
Top-k minimal	✗	✓
DGCP-Tree	✓	✓

A continuación, se describen los resultados del experimento realizado para determinar la eficiencia de estos algoritmos en cada conjunto de datos, cada prueba se ejecutó 10 veces para estimar la velocidad promedio de cada algoritmo.

Conjunto D

Conjunto de datos: Encuestas3.dat

Número de atributos: 17

Número de tuplas: 7859

Clase: *Med_aut* (¿Cuál o cuáles serían las causas por parte del médico para solicitar una autopsia?), en la Tabla 3.5 me muestran las causas por parte del médico para solicitar la autopsia.

Tabla 3.5 Clase *Med_aut*

20a	Interés
20b	Error médico
20c	Medicamentos mal recetados
20d	Negligencia médica
20e	Diagnóstico erróneo
20f	Orientación diagnóstica
20g	Educación Médica Continua
20h	Problema Médico Legal
20i	Acoso por delincuencia organizada
20j	Enseñanza

Algoritmos aplicados

En la Tabla 3.6 se muestra la comparación de los algoritmos tomando en cuenta la velocidad, número de reglas > 0.6 y el número de reglas.

Tabla 3.6 Comparación de algoritmos de la clase Med_aut

Algoritmo	Tiempo (segundos)	Número de reglas con confianza>0.6	Número de reglas
BCEP	-	-	-
DeEPS	30.939	1447	2599
iEPMiner	1.694	153	153
Tree-based JEP-C	5.072	780	987
Top-k minimal SJEPs	-	26	44
SJEP-C	0.496	3220	3410
DGCP-Tree	31.84	13	13
EvAEP	577.664	11	252
CEPMine	0.187	0	1
LCMine	3.633	5	133
EP-Random Forest	-	-	-

Cómo se observa, el algoritmo con mayor velocidad es CEPMine y en cuanto a número de reglas totales y con una confianza mayor a 0.6 el mejor es SJEP-C.

En la Tabla 3.7 se tomaron en cuenta solo las reglas con confianza mayor a 0.6.

Tabla 3.7 Comparación considerando la confianza

Algoritmo	WRAcc	Conf	GR	TPR	FPR	#Patrones
BCEP	-	-	-	-	-	-
DeEPS	-	-	-	-	-	-
iEPMiner	0.341048	0.879229	1	0.057011	0.004244	153
Tree-based JEP-C	0.378884	0.991582	1	0.004368	0.000116	780
Top-k minimal SJEPs	0.323508	0.947884	1	0.047411	0.002824	26
SJEP-C	0.317957	0.958203	1	0.022509	0.000691	3220
DGCP-Tree	0.335128	1	1	0.061223	0	13
EvAEP	0.312357	0.813195	1	0.082291	0.014714	11
CEPMine	0	0	0	0	0	0
LCMine	0.398855	1	1	0.02891	0	5
EP-Random Forest	-	-	-	-	-	-

Como se observa en la Tabla 3.7 el que obtuvo mejores resultados fue DGCP-Tree tomando en cuenta la confianza y la tasa de falsos positivos, LCMine también arrojó buenos resultados, sin embargo, en el número de patrones solo obtuvo 5 con confianza > 0.6.

La Tabla 3.8 muestra cuatro EPs obtenidos con el algoritmo DGCP-Tree, para la clase 20e, que en este caso es Diagnóstico erróneo.

Tabla 3.8 EPs obtenidos con DGCP-Tree para Med_aut

Regla	Descripción
IF Mcc_no_aut = 16u THEN 20e	Si los encuestados piensan que una causa o motivo para no solicitar autopsia es el olvido en la práctica de este estudio, entonces estiman como motivo para solicitar autopsia un diagnóstico erróneo
IF Categoria = c2 AND Esc_esp = c7 THEN 20e	Si los encuestados respondieron la encuesta por una invitación interna y estudiaron su especialidad en Nachón/CEM, entonces estiman como motivo para solicitar autopsia un diagnóstico erróneo
IF Anios_prac = 3b AND Esc_esp = c7 AND Mcc_no_aut = 16b THEN 20e	Si los encuestados tienen entre 5 y 10 años de práctica y estudiaron su especialidad en Nachón/CEM y piensan que una causa o motivo para no solicitar autopsia es el temor a la demanda, entonces estiman como motivo para solicitar autopsia un diagnóstico erróneo
IF Hall_disc = 7b AND Hall_dem = 11a AND Esc_esp = c7 THEN 20e	Si los encuestados están de acuerdo en que los hallazgos de las autopsias son discrepantes con los diagnósticos clínicos y están totalmente de acuerdo en que los hallazgos de las autopsias dan origen a casos de demanda y estudiaron medicina general en Nachón/CEM, entonces estiman como motivo para solicitar autopsia un diagnóstico erróneo

Estos patrones además de tener un alto valor de confianza (1 en los cuatro casos), son mínimos, por lo tanto, tienen un pequeño número de variables (tres máximos), lo cual los hace interesantes para la descripción. También son máximos, es decir, sus super-patrones no son EPs, por lo que son muy precisos.

Conjunto de datos: Encuestas4.dat

Número de atributos: 17

Número de tuplas: 7859

Clase *Per_sol_aut* (¿Quién considera que debe solicitar la autopsia?), en la Tabla 3.9 se muestran las opciones de quienes consideran que debe solicitar la autopsia.

Tabla 3.9 Calse *Per_sol_aut*

19a	Médico
19b	Enfermera
19c	Trabajadora Social
19d	La Institución
19e	Familia
19f	Criminología Forense

Algoritmos aplicados

En la Tabla 3.10 se muestra la comparación de los algoritmos tomando en cuenta la velocidad, número de reglas > 0.6 y el número de reglas.

Tabla 3.10 Comparación de algoritmos de la clase *Per_sol_aut*

Algoritmo	Tiempo (segundos)	Número de reglas con confianza>0.6	Número de reglas
BCEP	-	-	-
DeEPS	90.387	7770	14717
iEPMiner	3.011	1372	1372
Tree-based JEP-C	12.441	6095	6252
Top-k minimal SJEPs	-	15	33
SJEP-C	0.416	6080	6168
DGCP-Tree	24.377	555	555
EvAEP	533.896	6	102
CEPMine	0.207	1	1
LCMine	4.143	29	174
EP-Random Forest	-	-	-

Como se observa en la Tabla 3.10 el algoritmo más rápido fue CEPMine, sin embargo, el que obtuvo mayor cantidad de reglas y más patrones con confianza >0.6 fue DeEPS.

La Tabla 3.11 contiene las reglas que superan un valor de 0.6 en confianza.

Tabla 3.11 Comparación considerando la confianza

Algoritmo	WRAcc	Conf	GR	TPR	FPR	#Patrones
BCEP	-	-	-	-	-	-
DeEPS	-	-	-	-	-	7770
iEPMiner	0.803973	0.969796	0.978134	0.06148	0.014806	1372
Tree-based JEP-C	0.789396	0.975062	0.952912	0.011015	0.004787	6095
Top-k minimal SJEPs	0.625772	0.973973	1	0.082081	0.000662	15
SJEP-C	0.638512	0.988028	0.989967	0.02674	0.002015	6080
DGCP-Tree	0.801249	1	1	0.033425	0	555
EvAEP	0.48246	0.940951	1	0.375788	0.195751	6
CEPMine	0.801918	1	1	0.036681	0	1
LCMine	0.725448	0.970076	1	0.020295	0.000277	29
EP-Random Forest	-	-	-	-	-	-

Como se observa en la Tabla 3.11 el que obtuvo mejores resultados fue DGCP-Tree tomando en cuenta la confianza, tasa de crecimiento y la tasa de falsos positivos, iEPMiner también arrojó buenos resultados, sin embargo, solo lo hizo en dos de las métricas evaluadas.

La Tabla 3.12 muestra cuatro EPs obtenidos con el algoritmo DGCP-Tree, para la clase 19a, que en este caso es que el médico solicite la autopsia.

Tabla 3.12 EPs obtenidos con DGCP-Tree para Per_sol_aut

Regla	Descripción
IF Anios_prac = 3c THEN 19a	Si los encuestados tienen entre 11 y 15 años de práctica, entonces alegan que es el médico el personal adecuado para ordenar autopsia
IF Esc_med_gral = c12 AND Categoria = c2 THEN 19^a	Si los encuestados estudiaron medicina general en Universidad Villa Rica y respondieron la encuesta por una invitación interna entonces alegan que es el médico el personal adecuado para ordenar autopsia
IF Ult_grado = g2 AND Mcc_no_aut = 16e THEN 19a	Si los encuestados son médicos generales y piensan que una causa o motivo para no solicitar autopsia es el desinterés, entonces alegan que es el médico el personal adecuado para ordenar autopsia
IF Casos = 4e AND Rechazo_fam = 17d THEN 19a	Si los encuestados han participado en más de 20 casos de autopsias y creen que la causa de rechazo por parte del familiar es por no haber solicitado la autorización en forma adecuada, entonces alegan que es el médico el personal adecuado para ordenar autopsia

Estos patrones además de tener un alto valor de confianza (1 en todos los casos), son mínimos (solo con una o dos variables), por lo tanto, son interesantes para la descripción. También son máximos, por lo que son muy precisos.

Conjunto de datos: Encuestas5.dat

Número de atributos: 17

Número de tuplas: 7859

Clase *No_hosp* (¿Cuáles considera que son las causas por las que no se realizan suficientes autopsias en su hospital?), en la Tabla 3.13 me muestran las causas por parte del médico para no solicitar una autopsia.

Tabla 3.13 Clase *No_hosp*

18a	Falta de recursos humanos
18b	Falta de material
18c	Falta de recursos financieros
18d	No se solicitan
18e	Desconocimiento de su utilidad didáctica
18f	Se considera que los nuevos procedimientos diagnósticos y otros estudios especializados eliminan la necesidad de su práctica
18g	Desinterés del área de patología
18h	Negación de los familiares

Algoritmos aplicados

En la Tabla 3.14 se muestra la comparación de los algoritmos tomando en cuenta la velocidad, número de reglas > 0.6 y el número de reglas.

Tabla 3.14 Comparación de algoritmos de la clase *No_hosp*

Algoritmo	Tiempo (segundos)	Número de reglas con confianza>0.6	Número de reglas
BCEP	-	-	-
DeEPS	32.255	994	2648
iEPMiner	1.673	63	63
Tree-based JEP-C	10.681	741	973
Top-k minimal SJEPs	-	29	55
SJEP-C	0.48	6386	8383
DGCP-Tree	23.034	1	1

Algoritmo	Tiempo (segundos)	Número de reglas con confianza>0.6	Número de reglas
EvAEP	2309.225	13	906
CEPMine	0.073	1	1
LCMine	3.979	14	61
EP-Random Forest	-	-	-

Como se observa el algoritmo con mayor velocidad es CEPMine, sin embargo, solo encontró una regla, el algoritmo con mayor número de reglas es SJEP-C, al igual que obtuvo más reglas con confianza > 0.6 .

La Tabla 3.15 contiene las reglas que superan un valor de 0.6 en confianza.

Tabla 3.15 Comparación considerando la confianza

Algoritmo	WRAcc	Conf	GR	TPR	FPR	#Patrones
BCEP	-	-	-	-	-	-
DeEPS	-	-	-	-	-	994
iEPMiner	0.323591	0.842503	1	0.122709	0.003927	63
Tree-based JEP-C	0.214377	0.991587	1	0.011537	0.000022	741
Top-k minimal SJEPs	0.246066	0.892024	1	0.070265	0.002394	29
SJEP-C	0.147447	0.952364	1	0.03378	0.000441	6386
DGCP-Tree	1	0.282624	0.015269	0.5	0.086016	1
EvAEP	0.193326	0.971435	1	0.029427	0.000586	13
CEPMine	0.33775	1	1	0.00921	0	1
LCMine	0.226062	1	1	0.008986	0	14
EP-Random Forest	-	-	-	-	-	-

En la Tabla 3.15 el algoritmo que obtuvo mejores resultados fue iEPMiner tomando en cuenta la precisión relativa pesada, tasa de crecimiento y la tasa de verdaderos positivos.

La Tabla 3.16 muestra cuatro EPs obtenidos con el algoritmo iEPMiner, para las clases 18c, y 18d, que se refieren a la falta de recursos financieros y a que no se solicitan las autopsias.

Tabla 3.16 EPs obtenidos con iEPMiner para No_hosp

Regla	Descripción
IF Esc_med_gral = c18 THEN 18c	Si los encuestados estudiaron medicina general en Universidad Juárez Autónoma, entonces consideran que no se realizan suficientes autopsias en el hospital por falta de recursos financieros
IF Hall_dem = 11c AND Hall_disc = 7a AND Per_sol_aut = 19a THEN 18c	Si los encuestados consideran incierto que los hallazgos de las autopsias dan origen a casos de demanda y están totalmente de acuerdo en que los hallazgos de las autopsias son discrepantes con los diagnósticos clínicos y alegan que es el médico el personal adecuado para ordenar autopsia, entonces consideran que no se realizan suficientes autopsias en el hospital por falta de recursos financieros
IF Casos = 4a AND Esc_esp = c39 AND Hall_arb = 8a THEN 18c	Si los encuestados no han participado en ningún estudio de autopsia y no especifican donde hicieron su especialidad y están totalmente de acuerdo en que los hallazgos de las autopsias dan origen a casos de arbitraje, entonces consideran que no se realizan suficientes autopsias en el hospital por falta de recursos financieros
IF Hall_arb = 8b AND Hall_disc = 7a AND Ult_grado = g1 THEN 18d	Si los encuestados están de acuerdo en que los hallazgos de las autopsias dan origen a casos de arbitraje y están totalmente de acuerdo en que los hallazgos de las autopsias son discrepantes con los diagnósticos clínicos y son especialistas entonces consideran que no se realizan suficientes autopsias en el hospital debido a que no se solicitan

Estos patrones además de tener un alto valor de confianza (1, 0.759494, 1 y 1, respectivamente), son mínimos (solo con una o tres variables), por lo tanto, interesantes para la descripción. También son máximos, por lo que son muy precisos.

Conjunto de datos: Encuestas6.dat

Número de atributos: 17

Número de tuplas: 7859

Clase *Rechazo_fam* (¿Cuál considera que es la principal causa de rechazo por parte del familiar para realizar una autopsia?), en la Tabla 3.17 me muestran las causas de rechazo por parte del familiar.

Tabla 3.17 Clase Rechazo_fam

17a	Motivos religiosos
17b	Motivos políticos
17c	Motivos morales
17d	No haber solicitado la autorización en forma adecuada
17e	Inconformidad de los familiares con la atención prestada al paciente
17f	Deficiente comunicación del médico con el paciente y sus familiares
17g	Circunstancial

Algoritmos aplicados

En la Tabla 3.18 se muestra la comparación de los algoritmos tomando en cuenta la velocidad, número de reglas > 0.6 y el número de reglas.

Tabla 3.18 Comparación de algoritmos de la clase Rechazo_fam

Algoritmo	Tiempo (segundos)	Número de reglas con confianza>0.6	Número de reglas
BCEP	-	-	-
DeEPS	86.877	1423	3468
iEPMiner	2.458	405	405
Tree-based JEP-C	9.582	1060	1288
Top-k minimal SJEPs	-	39	44
SJEP-C	0.274	3752	4944
DGCP-Tree	31.039	33	33
EvAEP	208.164	8	33
CEPMine	0.427	0	1
LCMine	9.57	7	82
EP-Random Forest	-	-	-

Como se observa el algoritmo mejor evaluado es SJEP-C, tanto en tiempo, número de reglas con confianza > 0.6 y número de reglas, sin embargo, de las 4944 reglas encontradas, solo 3752 cumplen con el umbral de confianza.

La Tabla 3.19 contiene las reglas que superan un valor de 0.6 en confianza.

Tabla 3.19 Comparación considerando confianza

Algoritmo	WRAcc	Conf	GR	TPR	FPR	#Patrones
BCEP	-	-	-	-	-	-
DeEPS	-	-	-	-	-	-
iEPMiner	0.404964	0.883026	1	0.052229	0.004826	405
Tree-based JEP-C	0.303683	0.977619	1	0.00741	0.000179	1060
Top-k minimal SJEPs	0.279102	0.930873	1	0.043807	0.002146	39
SJEP-C	0.259232	0.964749	1	0.019708	0.000455	3752
DGCP-Tree	0.415639	1	1	0.062182	0	33
EvAEP	0.246579	0.807921	1	0.045838	0.007494	8
CEPMine	-	-	-	-	-	-
LCMine	0.413573	1	1	0.104592	0	7
EP-Random Forest	-	-	-	-	-	-

En la Tabla 3.19 el algoritmo que obtuvo mejores resultados fue DGCP-Tree tomando en cuenta la precisión relativa pesada, confianza, tasa de crecimiento, la tasa de verdaderos positivos y tasa de falsos positivos.

La Tabla 3.20 muestra cuatro EPs obtenidos con el algoritmo DGCP-Tree, para la clase 17f, que en este caso es Deficiente comunicación del médico con el paciente y sus familiares.

Tabla 3.20 EPs obtenidos con DGCP-Tree para Rechazo_fam

Regla	Descripción
IF Esc_esp = c30 THEN 17f	Si los médicos estudiaron su especialidad en San Luis Potosí, entonces creen que la causa de rechazo por parte del familiar es por deficiente comunicación del médico con el paciente y sus familiares
IF Hall_disc = 7b AND Area = a2 AND Hall_arb = 8d THEN 17f	Si los médicos están de acuerdo en que los hallazgos de las autopsias son discrepantes con los diagnósticos clínicos y son residentes y están en desacuerdo en que los hallazgos de las autopsias dan origen a casos de arbitraje, entonces creen que la causa de rechazo por parte del familiar es por deficiente comunicación del médico con el paciente y sus familiares
IF No_hosp = 18f AND Categoria = c2 AND Ult_grado = g2 THEN 17f	Si los médicos consideran que no se realizan suficientes autopsias en el hospital porque se considera que los nuevos procedimientos diagnósticos y otros estudios especializados eliminan la necesidad de su práctica y respondieron la encuesta por una invitación interna y son médicos generales, entonces creen que la causa de rechazo por parte del familiar es por

Regla	Descripción
	deficiente comunicación del médico con el paciente y sus familiares
IF Categoria = c2 AND Fmr_sol_aut = 21d AND Hall_arb = 8d THEN 17f	Si los médicos respondieron la encuesta por una invitación interna y están de acuerdo en que el método eficiente para solicitar autopsia es que el médico tratante sea quién la solicite y están en desacuerdo en que los hallazgos de las autopsias dan origen a casos de arbitraje, entonces creen que la causa de rechazo por parte del familiar es por deficiente comunicación del médico con el paciente y sus familiares

Estos patrones además de tener un alto valor de confianza (1 en todos los casos), son mínimos (solo con una o tres variables), por lo tanto, son interesantes para la descripción. También son máximos, por lo que son muy precisos.

Conjunto de datos: Encuestas7.dat

Número de atributos: 17

Número de tuplas: 7859

Clase *Mcc_no_aut* (¿Cuáles son los motivos, las causas y las circunstancias para que no solicite un estudio de autopsia?) en la Tabla 3.21 me muestran las causas para no solicitar un estudio de autopsia.

Tabla 3.21 Clase *Mcc_no_aut*

16a	Que no exista el servicio
16b	Temor a la demanda
16c	Negativa de los familiares
16d	Desconocimiento de la práctica de autopsias
16e	Desinterés
16f	Enfermedad de base conocida
16g	Trámites administrativos
16h	Falta de indicación
16i	Religiosos
16j	Culturales
16k	Legales
16l	Cuando el cuerpo se encuentra en descomposición
16m	De la condición del finado
16n	Falta de recursos humanos
16ñ	Falta de recursos materiales
16o	Falta de recursos financieros
16p	Que se haga sin fines de enseñanza

16q	Sociales
16r	Que el servicio de patología no la realice
16s	Políticos
16t	Falta de comunicación médico-paciente
16u	Olvido en la práctica de las autopsias
16z	No especificado

Algoritmos aplicados

En la Tabla 3.22 se muestra la comparación de los algoritmos tomando en cuenta la velocidad, número de reglas > 0.6 y el número de reglas.

Tabla 3.22 Comparación de algoritmos de la clase Mcc_no_aut

Algoritmo	Tiempo (segundos)	Número de reglas con confianza>0.6	Número de reglas
BCEP	-	-	-
DeEPS	88.178	8384	15187
iEPMiner	6.952	1537	1537
Tree-based JEP-C	26.844	3505	4336
Top-k minimal SJEPs	-	116	143
SJEP-C	1.291	127469	142268
DGCP-Tree	72.254	113	113
EvAEP	1294.807	28	510
CEPMine	0.597	0	1
LCMine	6.106	17	294
EP-Random Forest	-	-	-

Como se observa el algoritmo más rápido es CEPMine, sin embargo, solo encontró una regla, en el caso del número de reglas con confianza > 0.6 y numero de reglas el mejor es SJEP-C.

La Tabla 3.23 contiene las reglas que superan un valor de 0.6 en confianza.

Tabla 3.23 Comparación considerando confianza

Algoritmo	WRAcc	Conf	GR	TPR	FPR	#Patrones
BCEP	-	-	-	-	-	-
DeEPS	-	-	-	-	-	-
iEPMiner	0.282754	0.906411	1	0.110848	0.003477	1537
Tree-based JEP-C	0.194691	0.979787	1	0.022894	0.000194	3505
Top-k minimal SJEPs	0.445805	0.984146	1	0.391532	0.000235	116
SJEP-C	0.155368	0.98098	1	0.03538	0.000169	127469

Algoritmo	WRAcc	Conf	GR	TPR	FPR	#Patrones
DGCP-Tree	0.265033	1	1	0.116288	0	113
EvAEP	0.328496	0.882785	1	0.249992	0.006159	28
CEPMine	0	0	0	0	0	0
LCMine	0.449692	0.994901	1	0.348205	0.000403	17
EP-Random Forest	-	-	-	-	-	-

Como se observa en la tabla 3.23 el algoritmo con mejores resultados es DGCP-Tree, pues de las seis métricas tomadas en cuenta para la evaluación, cumple con tres, las cuales son la confianza, tasa de crecimiento y tasa de falsos positivos.

La Tabla 3.24 muestran cuatro EPs obtenidos con el algoritmo DGCP-Tree, para cuatro distintos valores de clase 16c, 16e, 16f y 16z, que en este caso se refieren a Desinterés, Negativa de los familiares, Enfermedad de base conocida y No especificado.

Tabla 3.24 EPs obtenidos con DGCP-Tree para Mcc_no_aut

Regla	Descripción
IF Esc_esp = c40 THEN 16c	Si los encuestados estudiaron su especialidad en UNAM, entonces piensan que una causa o motivo para no solicitar autopsia es la negativa de los familiares
IF Esc_esp = c15 THEN 16e	Si los encuestados estudiaron su especialidad en Instituto Nacional de Pediatría, entonces piensan que una causa o motivo para no solicitar autopsia es el desinterés
Anios_prac = 3c AND Esc_esp = c7 THEN 16z	Si los encuestados tienen entre 11 y 15 años de práctica y estudiaron su especialidad en Nachón/CEM, entonces no especifican motivos para solicitar autopsia
IF Rechazo_fam = 17d AND Esc_med_gral = c12 THEN 16f	Si los encuestados creen que la causa de rechazo por parte del familiar es por no haber solicitado la autorización en forma adecuada y estudiaron medicina general en Universidad Villa Rica, entonces piensan que una causa o motivo para no solicitar autopsia es por tratarse de una enfermedad de base conocida

Estos patrones además de tener un alto valor de confianza (1 en todos los casos), son mínimos (solo con una o dos variables), por lo tanto, interesantes para la descripción. También son máximos, por lo que son muy precisos.

Encuestas8.dat

Número de atributos: 17

Número de tuplas: 7859

Clase *Mcc_aut* (¿Cuáles son los motivos, las causas y las circunstancias para que solicite una autopsia?), en la Tabla 3.25 me muestran las causas para que se solicite un estudio de autopsia.

Tabla 3.25 Clase *Mcc_aut*

14a	Establecer diagnósticos definitivos
14b	Temas relacionados con medicina forense
14c	Objetivos Pedagógicos
14d	Aplicabilidad del estudio
14e	Falta de correlación entre los datos clínicos y los de laboratorio
14f	Proporcionar información esencial a familiares en casos de enfermedades infecciosas o congénitas
14g	Aplicación correcta o no correcta del tratamiento empleado
14h	Problemas de atención médica
14i	Diagnóstico difícil
14z	No especificado

Algoritmos aplicados

En la Tabla 3.26 se muestra la comparación de los algoritmos tomando en cuenta la velocidad, número de reglas > 0.6 y el número de reglas.

Tabla 3.26 Comparación de algoritmos de la clase

Algoritmo	Tiempo (segundos)	Número de reglas con confianza>0.6	Número de reglas
BCEP	-	-	-
DeEPS	31.953	2415	5731
iEPMiner	2.012	236	236
Tree-based JEP-C	6.377	977	1237
Top-k minimal SJEPs	-	67	77
SJEP-C	0.121	103518	120663
DGCP-Tree	30.721	18	18
EvAEP	964.369	20	369
CEPMine	0.181	0	1
LCMine	3.355	14	126
EP-Random Forest	-	-	-

En la tabla 3.26 se observa que el algoritmo que cumplió con las tres métricas de evaluación establecidas es SJEP-C.

La Tabla 3.27 muestra la comparación considerando los patrones que cumplen con la confianza mínima de 0.6.

Tabla 3.27 Comparación considerando confianza

Algoritmo	WRAcc	Conf	GR	TPR	FPR	#Patrones
BCEP	-	-	-	-	-	-
DeEPS	-	-	-	-	-	-
iEPMiner	0.272784	0.887847	1	0.1962	0.002721	236
Tree-based JEP-C	0.140966	0.982238	1	0.028774	0.000123	977
Top-k minimal SJEPs	0.249333	0.951433	1	0.12724	0.001048	67
SJEP-C	0.078656	0.992585	1	0.033478	0.000037	103518
DGCP-Tree	0.331188	1	1	0.286137	0	18
EvAEP	0.197475	0.916829	1	0.099568	0.001703	20
CEPMine	0	0	0	0	0	0
LCMine	0.15379	1	1	0.103265	0	14
EP-Random Forest	-	-	-	-	-	-

El algoritmo que obtuvo el mejor resultado es DGCP-Tree, ya que cumplió con cinco métricas, la cuales son precisión relativa pesada, confianza, tasa de crecimiento, tasa de verdaderos positivos y tasa de falsos positivos.

La Tabla 3.28 muestra cuatro EPs obtenidos con el algoritmo DGCP-Tree, para tres distintos valores de clase 14b, 14d y 14z, que son Temas relacionados con medicina forense, Aplicabilidad del estudio y No especificado.

Tabla 3.28 EPs obtenidos con DGCP-Tree para Mcc_aut

Regla	Descripción
IF Area = a3 AND Esc_med_gral = c17 THEN 14b	Si el área a la que pertenece el medico es igual a adscrito y la escuela de estudios de medicina general es la Universidad Benito Juárez, entonces ellos plantean como motivo para solicitar autopsia cuando existe relación con temas relacionados con medicina forense
IF Esc_esp = c8 THEN 14d	Si la escuela/ hospital en donde los médicos estudiaron la especialidad es la Alta Especialidad en IMSS,

Regla	Descripción
	entonces plantean como motivo para solicitar autopsia la aplicabilidad del estudio
IF Anios_prac = 3e AND Mcc_no_aut = 16z THEN 14z	Si los médicos tienen más de 20 años de práctica y no especifican motivos para solicitar autopsia, entonces no especifican motivos para solicitar autopsia
IF Fmr_sol_aut = 21d AND Esc_med_gral = c17 THEN 14b	Si los médicos consideran que la forma, método o recurso que funciona en forma eficiente para la solicitud de un estudio de autopsia es que el médico tratante sea quién la solicite y la escuela donde estudiaron medicina general es la Universidad Benito Juárez, entonces ellos plantean como motivo para solicitar autopsia cuando existe relación con temas relacionados con medicina forense

Estos patrones además de tener un alto valor de confianza (1 en todos los casos), son mínimos (solo con una o dos variables), por lo tanto, son interesantes para la descripción. También son máximos, por lo que son muy precisos.

Finalmente, se obtuvieron los algoritmos con mejores resultados para cada atributo de clase, como se muestra en la Tabla 3.29.

Tabla 3.29 Los mejores algoritmos para cada atributo de clase

Clase	Descripción	Mejor algoritmo
Fmr_sol_aut	Forma, método o recurso que funciona en forma eficiente para la solicitud de un estudio de autopsia.	iEPMiner
Med_aut	Causas por parte del médico para solicitar una autopsia	DGCP-Tree
Per_sol_aut	Persona que debe solicitar la autopsia	DGCP-Tree
No_hosp	Causas por las que no se realizan suficientes autopsias en el hospital	iEPMiner
Rechazo_fam	Principal causa de rechazo por parte del familiar para realizar una autopsia	DGCP-Tree
Mcc_no_aut	Motivos, causas y circunstancias para que no se solicite una autopsia	DGCP-Tree
Mcc_aut	Motivos, causas y circunstancias para que se solicite una autopsia	DGCP-Tree

En el conjunto de datos D los algoritmos seleccionados son DGCP-Tree e iEPMiner, ya que obtuvieron los mejores resultados para este conjunto.

Conjunto C

Para el conjunto C, solo tres algoritmos fueron los que obtuvieron resultados, a continuación, se muestra en la Tabla 3.30 los resultados.

Tabla 3.30 Comparación para la clase del hospital {1,2}

Algoritmo	Tiempo (segundos)	Número de reglas con confianza>0.6	Número de reglas
iEPMiner	0.364	537	537
DGCP-Tree	0.875	125	125
EvAEP	4.142	2	2

La Tabla 3.31 muestra la comparación considerando los patrones que cumplen con la confianza mínima de 0.6.

Tabla 3.31 Comparación considerando confianza

Algoritmo	WRAcc	Conf	GR	TPR	FPR	#Patrones
iEPMiner	0.490523	0.980158	0.955307	0.17735	0.019126	537
DGCP-Tree	0.920201	1	1	0.104027	0	125
EvAEP	0.566441	1	1	0.077311	0	2

Finalmente se observa que solo iEPMiner y DGCP-Tree, son los dos algoritmos con mejores resultados, tomando en cuenta las métricas de evaluación, se descartó EvAEP, ya que solo encontró dos patrones.

La Tabla 3.32 muestra cuatro EPs obtenidos con el algoritmo iEPMiner.

Tabla 3.32 EPs obtenidos con iEPMiner

Regla	Descripción
IF P10R06 = S THEN 2	Si los médicos alegan que es criminología forense quien debe ordenar la autopsia, entonces son del segundo hospital
IF P08R01 = S AND P11R01 = S THEN 1	Si los médicos creen que la causa de rechazo por parte del familiar es por motivos religiosos y estiman como motivo para solicitar autopsia el interés, entonces son del H.R.R.B.
IF P04R03 = S AND P11R02 = S AND P11R03 = S THEN 1	Si los médicos consideran incierto que los hallazgos de las autopsias dan origen a casos de arbitraje y estiman como motivo para solicitar autopsia un error médico y estiman como motivo para solicitar autopsia el consumo de medicamentos mal recetados, entonces son del H.R.R.B.

IF P11R01 = S AND P11R02 = S AND P17R27 = S THEN 2	Si los médicos estiman como motivo para solicitar autopsia el interés y estiman como motivo para solicitar autopsia un error médico y estudiaron su especialidad en UMAC 14 Adolfo Ruiz Cortines, entonces son del segundo hospital
---	---

La Tabla 3.33 muestra tres EPs obtenidos con el algoritmo DGCP-Tree.

Tabla 3.33 EPs obtenidos con DGCP-Tree

Regla	Descripción
IF P04R05 = S THEN 1	Si los médicos están totalmente en desacuerdo en que los hallazgos de las autopsias dan origen a casos de arbitraje, entonces son del H.R.R.B.
IF P05R02 = S THEN 1	Si los médicos están de acuerdo en que los hallazgos de las autopsias dan origen a casos de demanda, entonces son del H.R.R.B.
IF P16R20 = S THEN 2	Si los médicos estudiaron medicina general en Instituto Nacional de Nutrición, entonces son del segundo hospital

De los 11 algoritmos evaluados para la matriz binaria, solo dos dieron buenos resultados los cuales son DGCP-Tree y iEPMiner, en el caso de la vista minable también estos fueron los mejores, por lo que serán los algoritmos que se implementarán en el módulo nuevo del sistema.

Capítulo 4. Resultados

Como parte del proyecto se agregó un módulo de EPM a la aplicación Web desde la cual los médicos responden la encuesta, la cual proporciona información, que se procesa de manera automatizada y mantiene los datos actualizados.

El módulo de EPM proporciona un conjunto de operaciones de minería de patrones emergentes utilizando el *framework* EPM_Algorithms, el cual permite analizar los datos y extraer de ellos conocimiento útil e interesante. La información generada por los algoritmos es el resultado de las correlaciones entre las variables de interés expuestas en la encuesta y la probabilidad de que determinados patrones ocurran a partir del comportamiento de los datos entrenados.

El conocimiento obtenido es de gran importancia para el especialista del área de patología, porque de esta manera valida los resultados de las encuestas de manera objetiva mediante algoritmos de EPM, además que va de la mano su evaluación subjetiva sustentada por su experiencia y conocimiento de su área.

El nuevo módulo de la aplicación, permite analizar los datos mediante técnicas de discriminación y caracterización, el especialista elige entre dos algoritmos, los cuales son: iEPMiner y DGCP-tree, al igual que tiene la opción de seleccionar las distintas clases y el conjunto de datos del cual quiere extraer los patrones para generar el modelo de datos y guardarlo.

Los resultados que se muestran a los usuarios que visitan la aplicación son los extraídos de las reglas generadas, los cuales se observan en lenguaje natural eligiendo el algoritmo, el conjunto de datos y la clase, esto les permite analizar y estudiar el conocimiento generado sin requerir de la intervención de un experto en minería de datos.

En este capítulo se muestra y describe el funcionamiento de cada una de las opciones y operaciones de la aplicación. Con base en el caso de estudio del H.R.R.B. y el segundo hospital, así se observa cómo desde la aplicación un especialista es capaz de generar y guardar los modelos para que después muestre los resultados de los modelos de minería a todos los usuarios de la aplicación los cuales son el especialista y anónimo.

4.1 Presentación del sistema y caso de estudio

Se accede al sistema a través de su página de bienvenida, esta página se presenta a cualquier usuario, solo muestra información referente a la investigación (Figura 4.1).



Figura 4.1 Página de inicio

El usuario anónimo es capaz de contestar una encuesta, debe seleccionar “Encuesta” en la sesión del menú “Crear” (ver Figura 4.2).



Figura 4.2 Crear Encuesta

Una vez haciendo eso, el sistema muestra la página para rellenar el formulario, como se observa en la Figura 4.3.

Hospital Regional de Río Blanco

Inicio Sesión Gestionar Consultar Crear

Inicio Sesión Gestionar Consultar Crear

Inicio Sesión Gestionar Consultar Crear

FORMATO 1 ÁREA MÉDICA

Respetuosamente me dirijo a usted para solicitarle unos minutos de su tiempo para participar en una encuesta sobre la práctica de autopsias. El motivo de la encuesta, es el de establecer la Política de Autopsias en el Hospital Regional de Río Blanco. La información proporcionada en este documento será utilizada para la elaboración de un documento para su publicación y difusión.

SERVICIOS DE SALUD DE VERACRUZ

Datos

Centro Hospitalario: Hospital Regional Río Blanco

Institución: Río Blanco

Municipio: Río Blanco

Localidad: Río Blanco

Nota: Verificar

Numero de encuestas

Ingreso centro Estudio: [Campo]

Enviar

Siguiente

Figura 4.3 Formulario de Datos

En un inicio el encuestado debe llenar los campos de la ventana emergente “Datos”, esto ayudará a saber cuántas encuestas se tienen registradas de cada centro hospitalario, el médico, antes de iniciar la encuesta, debe verificar si el centro de estudio al que pertenece se encuentra en el sistema, como se muestra en la Figura 4.4.

Presentación Formación Experiencia Hallazgos Solicitud Procedimiento Comentarios

FORMATO 1
ÁREA MÉDICA

SERVICIOS DE SALUD DE VERACRUZ
HOSPITAL REGIONAL RIO BLANCO
SERVICIOS DE NEUROLOGÍA

Numero de encuestas
HRRB/2.0

Respetuosamente me dirijo a usted para solicitarle unos minutos de su tiempo para hacerle algunas preguntas, que tienen que ver con la práctica de autopsias.
El motivo de la encuesta, es el de establecer la Política de Autopsias en el Hospital Regional Río Blanco, así como la elaboración de un documento para su publicación y difusión.
Nota: Verificar si se encuentra el centro de estudio al que pertenece

Nota: Verificar si se encuentra el centro de estudio al que pertenece

Ingresar centro Estudio:

- Alta Especialidad en IMSS - Desconocido - Monterrey
- Centro de Especialidades Médicas - Desconocido - Veracruz
- Centro Médico de Occidente - Desconocido - Guadalajara
- Desconocido - Desconocido - Desconocido
- Facultad Medicina - Xalapa - Veracruz
- Facultad Medicina - Desconocido - Desconocido
- General Manuel Gea González - Desconocido - Desconocido
- Hospital de la Mujer - Xalapa - Veracruz

→ Siguiete

Figura 4.4 Verificación de datos

En caso de que se encuentre el centro de estudio, el médico dará clic en el botón que dice **siguiete**, si no se encuentra, entonces llena los campos solicitados (ver Figura 4.5).

Presentación Formación Experiencia Hallazgos Solicitud Procedimiento Comentarios

FORMATO 1
ÁREA MÉDICA

SERVICIOS DE SALUD DE VERACRUZ
HOSPITAL REGIONAL RIO BLANCO
SERVICIOS DE NEUROLOGÍA

Numero de encuestas
HRRB/2.0

Respetuosamente me dirijo a usted para solicitarle unos minutos de su tiempo para hacerle algunas preguntas, que tienen que ver con la práctica de autopsias.
El motivo de la encuesta, es el de establecer la Política de Autopsias en el Hospital Regional Río Blanco, así como la elaboración de un documento para su publicación y difusión.
Nota: Verificar si se encuentra el centro de estudio al que pertenece

Nota: Verificar si se encuentra el centro de estudio al que pertenece

Ingresar centro Estudio: Alta Especialidad en IMSS - Desconocido - Monterrey

Tipo: Escuela Ciudad: Xalapa

Enviar

→ Siguiete

Figura 4.5 Datos del centro de estudio

Una vez hecho el proceso anterior, el médico continúa a llenar el formulario de la encuesta, la cual cuenta con seis apartados, los cuales son: Formación, Experiencia, Hallazgos, Solicitud, Procedimiento y Comentarios (ver Figura 4.6).

The screenshot shows a web interface for a survey. At the top, there is a navigation bar with links for 'Inicio', 'Sesion', 'Gestionar', and 'Consultar'. Below this is a horizontal menu with tabs for 'Presentación', 'Formación' (highlighted), 'Experiencia', 'Hallazgos', 'Solicitud', 'Procedimiento', and 'Comentarios'. The main content area is titled 'Formación del médico' and contains four numbered sections:

- 1** Área a la que perteneces: *
 Interno
 Residente
 Adscrito
- 2** Último grado de estudios: *
 Especialidad
 Medicina General
 Práctica Universitaria
- 3** ¿Dónde efectuaste tus estudios de medicina general?
Alta Especialidad en IMSS - Desconocido - Monterrey
- 4** Escuela/Hospital donde estudiaste la especialidad que ejerces:
Alta Especialidad en IMSS - Desconocido - Monterrey

At the bottom of the form, there are two buttons: '← Anterior' on the left and '→ Siguiente' on the right.

Figura 4.6 Formulario de Encuesta

Cuando el usuario contesta todas las preguntas y envía el formulario, el sistema inicia el proceso de clasificación de texto para las preguntas abiertas, una vez que finaliza, se guarda la encuesta. El resultado de clasificación de las preguntas abiertas se encuentra en la tabla “*clasificado*” de la base de datos. Cabe mencionar que algunas funciones se realizaron en la tesis anterior, sin embargo, se reutilizaron para la realización del módulo nuevo (ver Figura 4.7).

id_medico	id_prog	texto	clasificado
integer	integer	character varying	character varying
393	04	22	
394	04	22	j
395	04	22	c
396	05	14	a
397	05	14	i
398	05	16	f
399	05	22	a
400	06	14	a
401	06	16	f
402	07	14	e
403	07	16	i
404	07	22	k
405	08	14	a
406	08	16	f
407	08	22	c
408	09	14	a
409	09	16	j
410	09	17	z
411	09	22	k
412	90	14	a
413	90	16	f
414	90	22	c
415	91	14	a
416	91	16	c
417	91	22	c
418	92	14	a
419	92	16	j
420	92	22	g
421	93	14	a
422	93	16	f
423	93	17	z
424	93	22	e

Figura 4.7 Respuesta Clasificadas

El especialista forma un rol muy importante, pues es el que es capaz de iniciar sesión, actualizar los *datasets*, crear los modelos, consultar y eliminar encuestas, para que él y el usuario anónimo consulten los resultados (Figura 4.8).

- Inicio
- Sesión**
- Iniciar
- Cerrar
- Eliminar
- Encuesta
- Consultar**
- Encuestas
- Reglas asociación
- Minería de patrones emergentes
- Redes bayesianas
- Crear**
- Reglas asociación
- Redes bayesianas
- Minería de patrones emergentes
- Conjuntos de datos

Páginas: (1 de 1) 1 10

No. Control	Grado	Área	Categoría	Especialidad	Centro Hospitalario	Institución	Municipio	Localidad	Operaciones
88	Medicina General	Adscrito	Espontáneo	Medicina	Jurisdicción Santiana	Secretaría de Salud	Orizaba	Orizaba	🔍 ✖
89	Especialidad	Adscrito	Espontáneo	Medicina	Hospital Regional Río Blanco	Secretaría de Salud de Veracruz	Río Blanco	Río Blanco	🔍 ✖
90	Especialidad	Adscrito	Espontáneo	Medicina	Hospital Regional Río Blanco	Secretaría de Salud	Río Blanco	Río Blanco	🔍 ✖
91	Especialidad	Adscrito	Espontáneo	Medicina	Hospital Regional Río Blanco	Secretaría de Salud de Veracruz	Río Blanco	Río Blanco	🔍 ✖
92	Especialidad	Residente	Espontáneo	Medicina	Hospital Regional Río Blanco	Secretaría de Salud de Veracruz	Río Blanco	Río Blanco	🔍 ✖
93	Especialidad	Adscrito	Espontáneo	Medicina	IMSS HGZ No 1, Orizaba	IMSS	Orizaba	Orizaba	🔍 ✖
94	Medicina General	Residente	Espontáneo	Medicina	Hospital Regional Río Blanco	Secretaría de Salud de Veracruz	Río Blanco	Río Blanco	🔍 ✖

Páginas: (1 de 1) 1 10

Datos de los encuestados

Nuevas encuestas: 0
Actualizar datasets

Figura 4.8 Funciones del especialista

Para el nuevo conjunto de datos es necesario actualizar el *dataset*, una vez actualizado se modifican las cantidades en las gráficas y el sistema actualiza los conjuntos de datos “C y D”, quedando adecuados para generar los modelos de minería.

Desde la sección de “*Crear*” el especialista selecciona “*Minería de patrones emergentes*”, el sistema muestra la página donde el especialista configura cada uno de los algoritmos de acuerdo a su interés, se proponen los valores que resultaron eficientes en esta investigación, cuando el especialista selecciona uno de los algoritmos, él es capaz de seleccionar la clase, la cual es una pregunta de la encuesta para que a partir de esta, genere el modelo, se consideran como clase a los siguientes atributos:

- *Fmr_sol_aut*: Forma, método o recurso que funciona en forma eficiente para la solicitud de un estudio de autopsia.
- *Med_aut*: Causas por parte del médico para solicitar una autopsia.
- *Per_sol_aut*: Persona que debe solicitar la autopsia.
- *No_hosp*: Causas por las que no se realizan suficientes autopsias en el hospital.
- *Rechazo_fam*: Principal causa de rechazo por parte del familiar para realizar una autopsia.
- *Mcc_no_aut*: Motivos, causas y circunstancias para que no se solicite una autopsia.
- *Mcc_aut*: Motivos, causas y circunstancias para que se solicite una autopsia.

En las Figuras 4.9 a 4.22 se muestran los patrones emergentes creados a partir del conjunto de datos D.

No.	Regla
1	IF Esc_esp = c43 AND No_hosp = 18e THEN 21a
2	IF Esc_esp = c43 AND No_hosp = 18d THEN 21a
3	IF Esc_esp = c43 AND Med_aut = 20e THEN 21a
4	IF Esc_esp = c43 AND Med_aut = 20a THEN 21a
5	IF Esc_esp = c43 AND Esc_med_gral = c1 THEN 21a
6	IF Area = a2 AND Hall_dem = 11a THEN 21a
7	IF Anios_prac = 3e AND Esc_esp = c43 THEN 21a
8	IF Mcc_no_aut = 16c AND Uit_grado = g8 THEN 21b
9	IF Esc_med_gral = c12 AND Uit_grado = g8 THEN 21b
10	IF Categoria = c2 AND Esc_med_gral = c12 THEN 21b
11	IF Area = a1 AND Mcc_no_aut = 16c THEN 21b
12	IF Area = a1 AND Esc_med_gral = c12 THEN 21b
13	IF Hall_arb = 8c AND Hall_dem = 11b THEN 21c
14	IF Anios_prac = 3c AND Casos = 4b THEN 21c
15	IF Esc_esp = c43 AND Esc_med_gral = c1 AND Med_aut = 20e THEN 21a
16	IF Esc_esp = c43 AND Esc_med_gral = c1 AND Hall_disc = 7a THEN 21a

Figura 4.9 Modelo de iEPMiner, conjunto D, clase *Fmr_sol_aut*

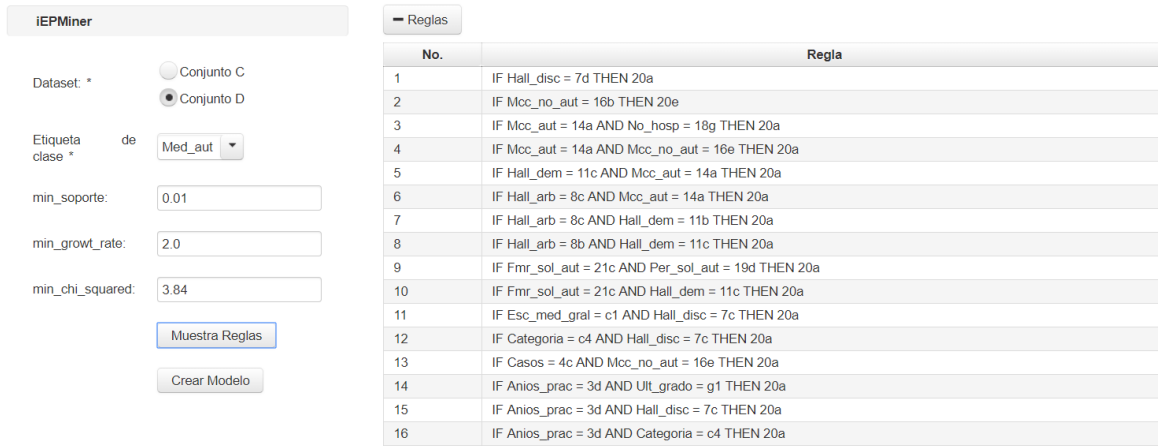


Figura 4.10 Modelo de iEPMiner, conjunto D, clase Med_aut

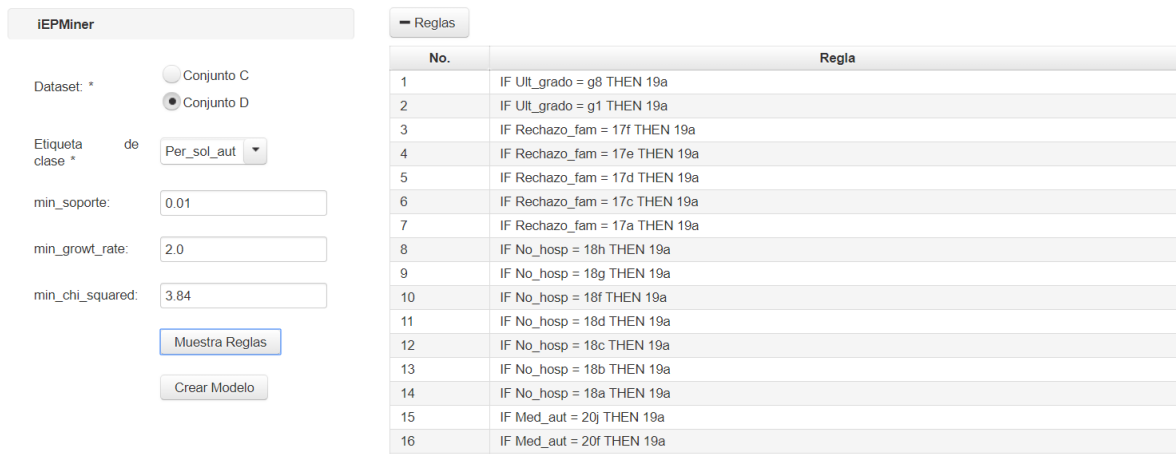


Figura 4.11 Modelo de iEPMiner, conjunto D clase, Per_sol_aut

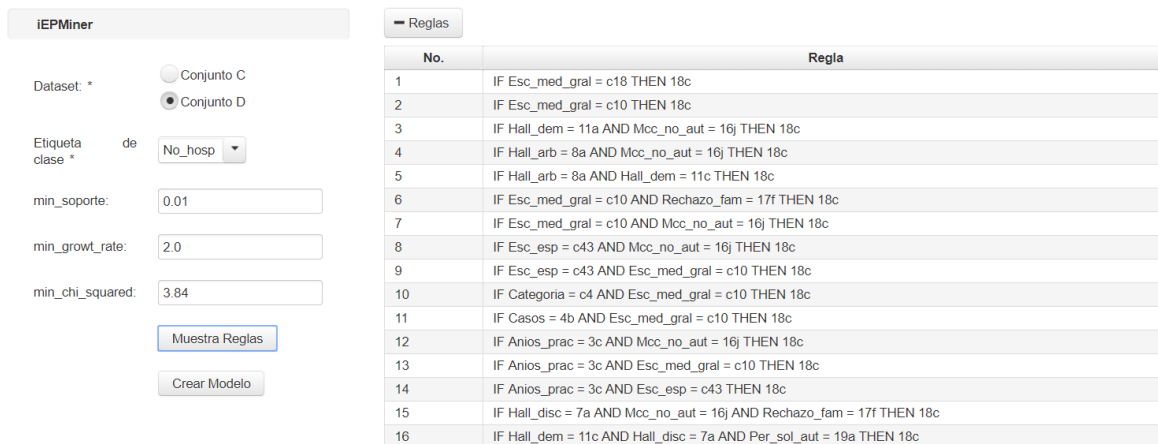


Figura 4.12 Modelo iEPMiner, conjunto D, clase No_hosp

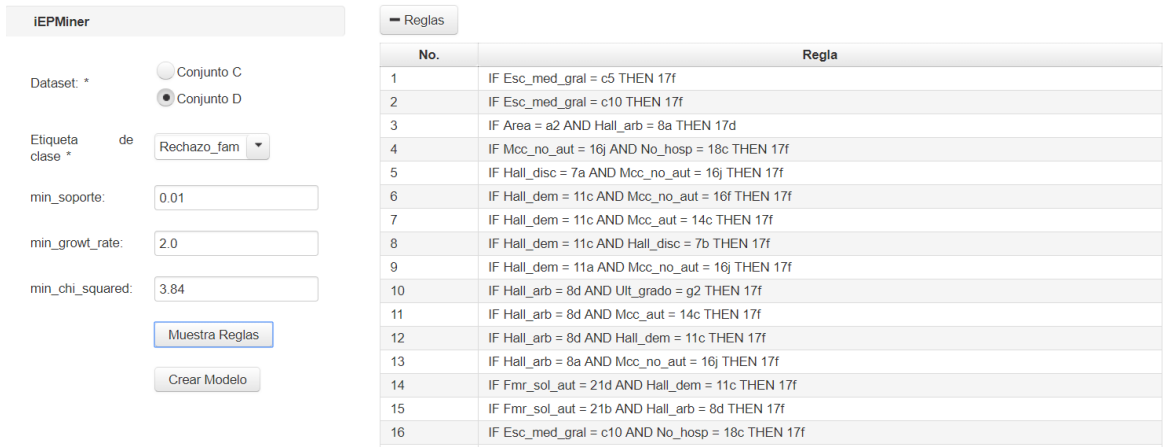


Figura 4.13 Modelo iEPMiner, conjunto D, clase Rechazo_fam

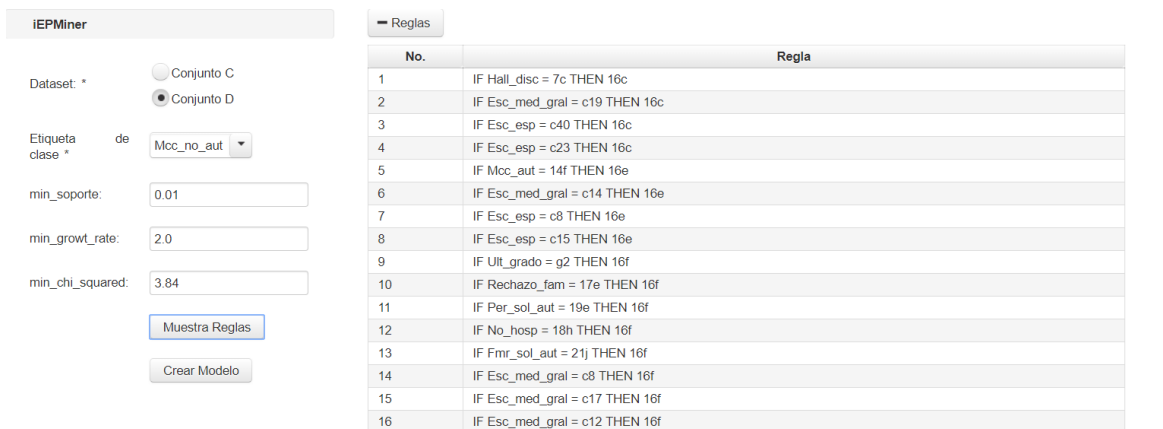


Figura 4.14 Modelo iEPMiner, conjunto D, clase Mcc_no_aut

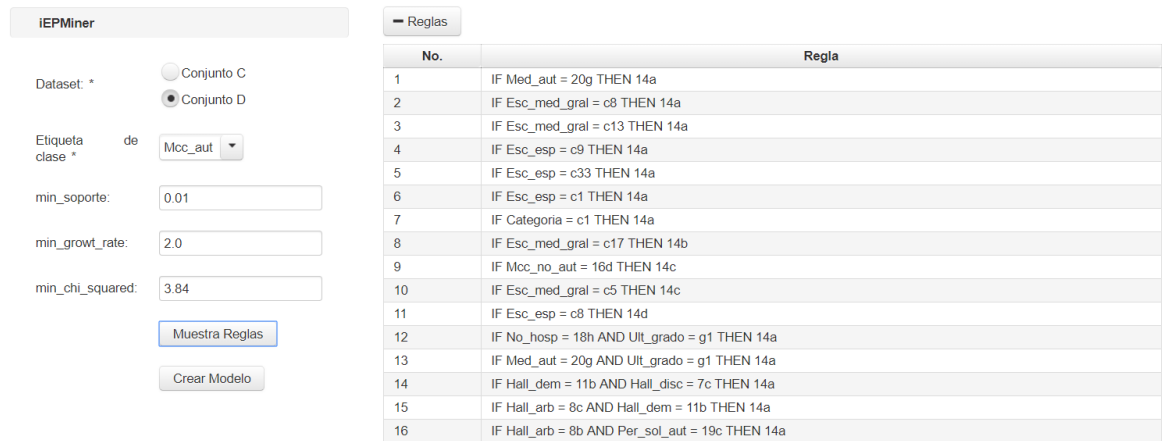


Figura 4.15 Modelo iEPMiner, conjunto D, clase Mcc_aut

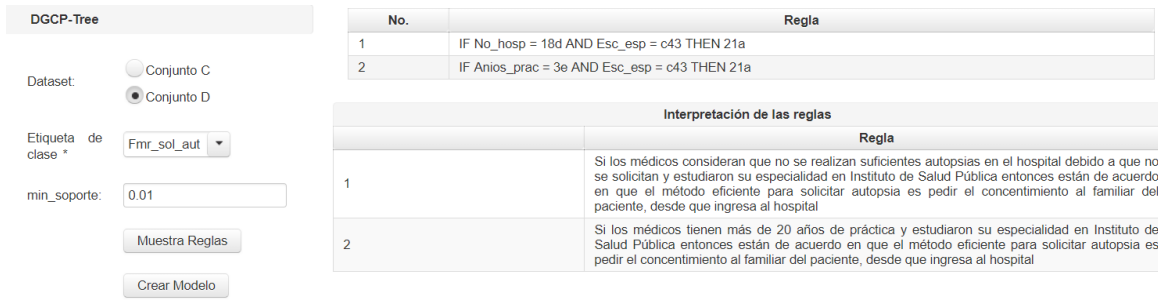


Figura 4.16 Modelo DGCP-Tree, conjunto D, clase Fmr_sol_aut

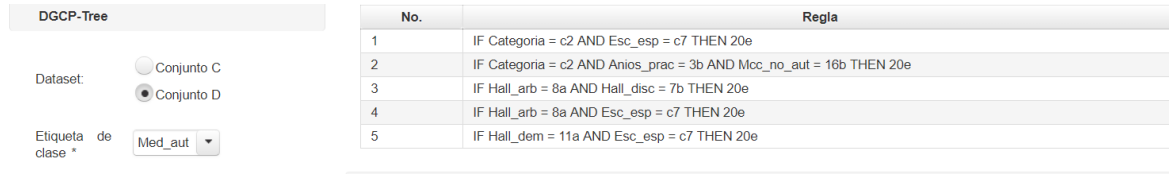


Figura 4.17 Modelo DGCP-Tree, conjunto D, clase Med_aut

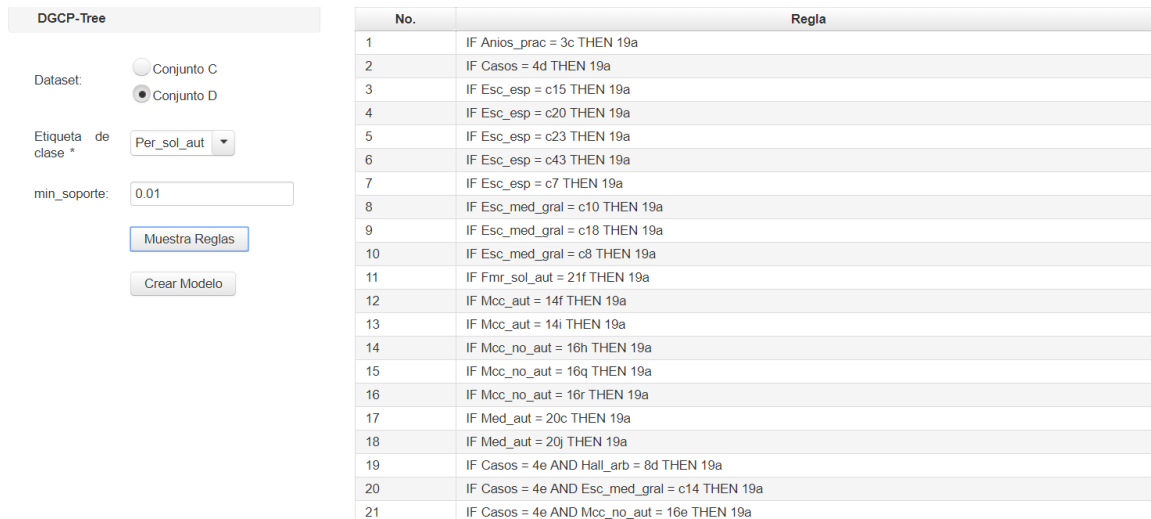


Figura 4.18 Modulo DGCP-Tree, conjunto D, clase Per_sol_aut

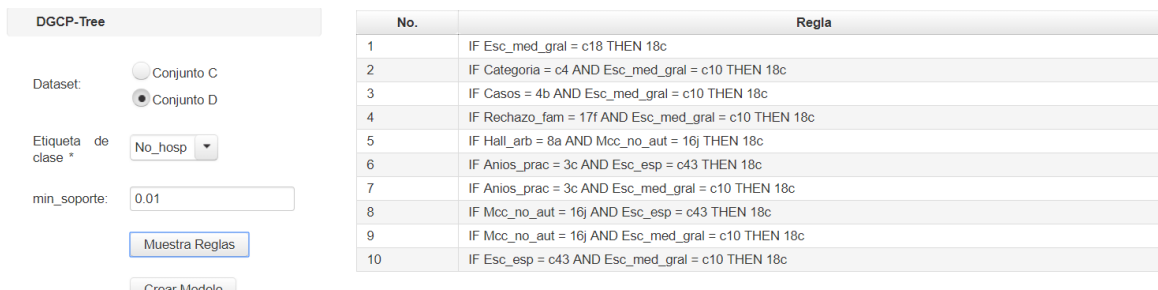


Figura 4.19 Modelo DGCP-Tree, conjunto D, clase No_hosp

DGCP-Tree	
Dataset:	<input type="radio"/> Conjunto C <input checked="" type="radio"/> Conjunto D
Etiqueta de clase *	Rechazo_fam
min_soporte:	0.01
	<input type="button" value="Muestra Reglas"/>
	<input type="button" value="Crear Modelo"/>

No.	Regla
1	IF Esc_med_gral = c5 THEN 17f
2	IF Anios_prac = 3a AND Hall_arb = 8d THEN 17f
3	IF Categoria = c2 AND Area = a2 AND Hall_arb = 8d THEN 17f
4	IF Categoria = c2 AND Hall_dem = 11c AND Hall_arb = 8d THEN 17f
5	IF Hall_disc = 7b AND Hall_dem = 11c AND Hall_arb = 8d THEN 17f
6	IF Esc_esp = c39 AND Hall_arb = 8d THEN 17f
7	IF Categoria = c4 AND Esc_med_gral = c10 THEN 17f
8	IF Hall_arb = 8a AND Mcc_no_aut = 16j THEN 17f
9	IF Mcc_aut = 14c AND Hall_dem = 11c AND Hall_arb = 8d THEN 17f
10	IF No_hosp = 18c AND Esc_esp = c43 THEN 17f
11	IF No_hosp = 18c AND Esc_med_gral = c10 THEN 17f
12	IF No_hosp = 18c AND Mcc_no_aut = 16j AND Casos = 4b THEN 17f
13	IF No_hosp = 18c AND Mcc_no_aut = 16j AND Anios_prac = 3c THEN 17f
14	IF No_hosp = 18c AND Casos = 4b AND Anios_prac = 3c THEN 17f
15	IF Mcc_no_aut = 16j AND Esc_esp = c43 THEN 17f
16	IF Mcc_no_aut = 16j AND Esc_med_gral = c10 THEN 17f
17	IF Casos = 4b AND Esc_med_gral = c10 THEN 17f
18	IF Anios_prac = 3c AND Esc_esp = c43 THEN 17f
19	IF Anios_prac = 3c AND Esc_med_gral = c10 THEN 17f
20	IF Esc_esp = c43 AND Esc_med_gral = c10 THEN 17f

Figura 4.20 Modulo DGCP-Tree, conjunto D, clase Rechazo_fam

DGCP-Tree	
Dataset:	<input type="radio"/> Conjunto C <input checked="" type="radio"/> Conjunto D
Etiqueta de clase *	Mcc_no_aut
min_soporte:	0.01
	<input type="button" value="Muestra Reglas"/>
	<input type="button" value="Crear Modelo"/>

No.	Regla
1	IF Hall_arb = 8a AND Esc_esp = c7 THEN 16b
2	IF Hall_dem = 11a AND Esc_esp = c7 THEN 16b
3	IF Categoria = c2 AND Esc_esp = c7 THEN 16b
4	IF Esc_esp = c23 THEN 16c
5	IF Anios_prac = 3b AND Esc_esp = c40 THEN 16c
6	IF Anios_prac = 3b AND Esc_med_gral = c19 THEN 16c
7	IF Area = a3 AND Esc_med_gral = c19 THEN 16c
8	IF Hall_dem = 11b AND Anios_prac = 3d AND Hall_disc = 7c THEN 16c
9	IF Esc_esp = c6 AND Anios_prac = 3d THEN 16c
10	IF Esc_esp = c6 AND Hall_disc = 7c THEN 16c
11	IF Anios_prac = 3e AND Med_aut = 20c THEN 16c
12	IF Anios_prac = 3e AND Mcc_aut = 14e AND Casos = 4c THEN 16c
13	IF Anios_prac = 3e AND Med_aut = 20b AND Casos = 4c THEN 16c
14	IF Ult_grado = g1 AND Med_aut = 20c THEN 16c
15	IF Ult_grado = g1 AND Esc_med_gral = c19 THEN 16c
16	IF No_hosp = 18f AND Hall_arb = 8a THEN 16c
17	IF No_hosp = 18f AND Hall_dem = 11a AND Casos = 4c THEN 16c
18	IF Categoria = c2 AND Esc_esp = c40 THEN 16c
19	IF Categoria = c2 AND Esc_med_gral = c19 THEN 16c
20	IF Casos = 4b AND Esc_esp = c40 THEN 16c
21	IF Mcc_aut = 14e AND Casos = 4c THEN 16c

Figura 4.21 Modelo DGCP-Tree, conjunto D, clase Mcc_no_aut

No.	Regla
1	IF p10_1 = S THEN 1
2	IF p10_2 = S THEN 1
3	IF p10_3 = S THEN 1
4	IF p10_4 = S THEN 1
5	IF p10_5 = S THEN 1
6	IF p10_6 = S THEN 1
7	IF p11_1 = S THEN 1
8	IF p11_10 = S THEN 1
9	IF p11_2 = S THEN 1
10	IF p11_3 = S THEN 1
11	IF p11_5 = S THEN 1
12	IF p11_6 = S THEN 1
13	IF p11_7 = S THEN 1
14	IF p11_9 = S THEN 1
15	IF p12_1 = S THEN 1
16	IF p12_2 = S THEN 1
17	IF p12_3 = S THEN 1
18	IF p12_5 = S THEN 1
19	IF p12_7 = S THEN 1
20	IF p13_10 = S THEN 1
21	IF p13_11 = S THEN 1

Figura 4.25 Modelo DGCP-Tree, conjunto C

Como se mencionó anteriormente el especialista es capaz de seleccionar “Encuesta” desde el menú, es capaz de ver y eliminar encuestas, a las cuales se le agregaron cuatro campos más, los cuales son: Centro hospitalario, Institución, Municipio y Localidad (Figura 4.26).

Páginas: (1 de 1)									
No. Control	Grado	Área	Categoría	Especialidad	Centro Hospitalario	Institución	Municipio	Localidad	Operaciones
88	Medicina General	Adscrito	Espontáneo	Medicina	Jurisdicción Sanitaria	Secretaría de Salud	Orizaba	Orizaba	
89	Especialidad	Adscrito	Espontáneo	Medicina	Hospital Regional Río Blanco	Secretaría de Salud de Veracruz	Río Blanco	Río Blanco	
90	Especialidad	Adscrito	Espontáneo	Medicina	Hospital Regional Río Blanco	Secretaría de Salud	Río Blanco	Río Blanco	
91	Especialidad	Adscrito	Espontáneo	Medicina	Hospital Regional Río Blanco	Secretaría de Salud de Veracruz	Río Blanco	Río Blanco	
92	Especialidad	Residente	Espontáneo	Medicina	Hospital Regional Río Blanco	Secretaría de Salud de Veracruz	Río Blanco	Río Blanco	
93	Especialidad	Adscrito	Espontáneo	Medicina	IMSS, HGZ No 1, Orizaba	IMSS	Orizaba	Orizaba	
94	Medicina General	Residente	Espontáneo	Medicina	Hospital Regional Río Blanco	Secretaría de Salud de Veracruz	Río Blanco	Río Blanco	

Figura 4.26 Gestionar Encuesta

El especialista cuenta con un filtro que facilita la ubicación de la encuesta requerida mediante el número de control, el cual es un valor único (Ver Figura 4.27).

No. Control:

Páginas: (1 de 1)									
No. Control	Grado	Área	Categoría	Especialidad	Centro Hospitalario	Institución	Municipio	Localidad	Operaciones
93	Especialidad	Adscrito	Espontáneo	Medicina	IMSS, HGZ No 1, Orizaba	IMSS	Orizaba	Orizaba	

Figura 4.27 Filtrar encuesta

El especialista también tiene la opción “Operaciones” en donde es posible consultar las respuestas de la encuesta seleccionada o eliminar directamente, como se muestra en las Figuras 4.28 y 4.29.

Figura 4.28 Consultar encuesta.

No. Control	Grado	Área	Categoría	Especialidad	Operaciones
88	Especialidad	Adscrito	Espontáneo	Medicina	Eliminar

Figura 4.29 Eliminar encuesta

El especialista permanece con su sesión iniciada hasta que decida cerrarla, una vez cerrada la sesión, el sistema regresa a la página de bienvenida (Figura 4.30).

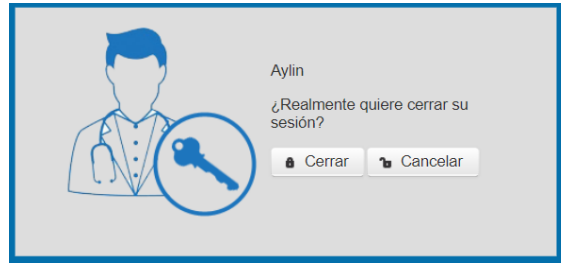


Figura 4.30 Cerrar Sesión

Cualquier usuario es capaz de consultar los resultados, para esto accede desde el apartado “Consultar” y seleccionar “Minería de patrones emergentes”, aquí el usuario elige el algoritmo, dataset y la etiqueta de clase, una vez seleccionado lo requerido, consulta las reglas interpretadas en lenguaje natural, para que cualquier persona las entienda, sin necesidad de contar con algún especialista de Minería de Datos (Ver Figuras 4.31 a 4.35).

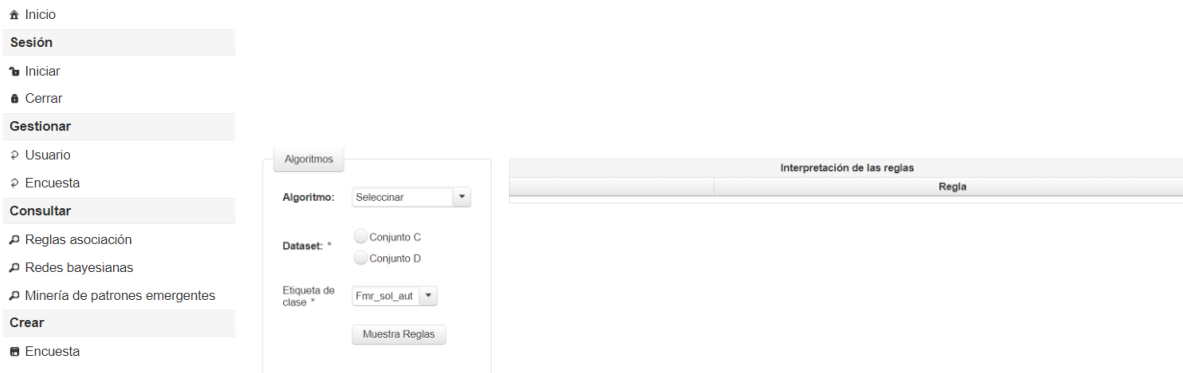


Figura 4.31 Resultados

Interpretación de las reglas	
	Regla
1	Si los médicos consideran que no se realizan suficientes autopsias en el hospital debido a la negación de los familiares entonces son del hospital 1
2	Si los médicos consideran que no se realizan suficientes autopsias en el hospital por desinterés del área de patología entonces son del hospital 1
3	Si los médicos consideran que no se realizan suficientes autopsias en el hospital porque se considera que los nuevos procedimientos diagnósticos y otros estudios especializados eliminan la necesidad de su práctica entonces son del hospital 1
4	Si los médicos consideran que no se realizan suficientes autopsias en el hospital por desconocimiento de su utilidad didáctica entonces son del hospital 1
5	Si los médicos consideran que no se realizan suficientes autopsias en el hospital debido a que no se solicitan entonces son del hospital 1
6	Si los médicos consideran que no se realizan suficientes autopsias en el hospital por falta de recursos financieros entonces son del hospital 1
7	Si los médicos consideran que no se realizan suficientes autopsias en el hospital por falta de material entonces son del hospital 1
8	Si los médicos consideran que no se realizan suficientes autopsias en el hospital por falta de recursos humanos entonces son del hospital 1
9	Si los médicos creen que la causa de rechazo por parte del familiar es circunstancial entonces son del hospital 1
10	Si los médicos creen que la causa de rechazo por parte del familiar es por deficiente comunicación del médico con el paciente y sus familiares entonces son del hospital 1
11	Si los médicos creen que la causa de rechazo por parte del familiar es por inconformidad de los familiares con la atención prestada al paciente entonces son del hospital 1
12	Si los médicos creen que la causa de rechazo por parte del familiar es por no haber solicitado la autorización en forma adecuada entonces son del hospital 1
13	Si los médicos creen que la causa de rechazo por parte del familiar es por motivos morales entonces son del hospital 1

Figura 4.32 Reglas con iEPMIner y conjunto C

Interpretación de las reglas	
	Regla
1	Si los médicos consideran que no se realizan suficientes autopsias en el hospital debido a la negación de los familiares entonces son del hospital 1
2	Si los médicos consideran que no se realizan suficientes autopsias en el hospital por desinterés del área de patología entonces son del hospital 1
3	Si los médicos consideran que no se realizan suficientes autopsias en el hospital porque se considera que los nuevos procedimientos diagnósticos y otros estudios especializados eliminan la necesidad de su práctica entonces son del hospital 1
4	Si los médicos consideran que no se realizan suficientes autopsias en el hospital por desconocimiento de su utilidad didáctica entonces son del hospital 1
5	Si los médicos consideran que no se realizan suficientes autopsias en el hospital debido a que no se solicitan entonces son del hospital 1
6	Si los médicos consideran que no se realizan suficientes autopsias en el hospital por falta de recursos financieros entonces son del hospital 1
7	Si los médicos consideran que no se realizan suficientes autopsias en el hospital por falta de material entonces son del hospital 1
8	Si los médicos consideran que no se realizan suficientes autopsias en el hospital por falta de recursos humanos entonces son del hospital 1
9	Si los médicos creen que la causa de rechazo por parte del familiar es circunstancial entonces son del hospital 1
10	Si los médicos creen que la causa de rechazo por parte del familiar es por deficiente comunicación del médico con el paciente y sus familiares entonces son del hospital 1
11	Si los médicos creen que la causa de rechazo por parte del familiar es por inconformidad de los familiares con la atención prestada al paciente entonces son del hospital 1
12	Si los médicos creen que la causa de rechazo por parte del familiar es por no haber solicitado la autorización en forma adecuada entonces son del hospital 1
13	Si los médicos creen que la causa de rechazo por parte del familiar es por motivos morales entonces son del hospital 1

Figura 4.33 Reglas con DGCP-Tree y conjunto C



Figura 4.34 Reglas con iEPMiner, conjunto D y la clase Med_aut



Figura 4.35 Reglas con DGCP-Tree, conjunto D y la clase Per_sol_aut

4.2 Evaluación de los resultados

Ya que se tienen los modelos se analiza el conocimiento extraído de los datos (las respuestas de las encuestas). Esta actividad es realizada por el o los expertos en el área de patología que está siendo objeto de investigación, por esta razón fue necesario que el sistema proporcionara una interpretación entendible de las reglas para apoyar al experto en el proceso de evaluación. De esta manera el especialista evalúa los resultados analizando de manera subjetiva la información obtenida por los modelos con base en sus conocimiento y experiencia.

Capítulo 5. Conclusiones y recomendaciones

5.1. Conclusiones

Dado el alcance que presentan los estudios de autopsias en la práctica médica y tomando en cuenta la problemática planteada en un inicio, en esta investigación se busca identificar los elementos que intervienen en dicha práctica empleando técnicas de minería de datos, la cual es una de las herramientas más útiles para el manejo de grandes volúmenes de datos, la minería hace posible la extracción de conocimiento oculto en los datos.

La finalidad de este proyecto fue determinar las posibles causas de la disminución en la realización de autopsias medicas en el H.R.R.B. y en otro hospital, para saber si son las mismas causas, razones y/o circunstancias del porqué cada vez se realiza menos esta práctica en el H.R.R.B. Esto a través de un módulo de minería de patrones emergentes, el cual permite encontrar diferencias entre los conjuntos de los dos hospitales, todo el análisis se realizó a partir de las opiniones médicas, logrando con esto un aprendizaje automático, ya que el sistema capaz de aprender o predecir conocimiento a partir de los conjuntos de datos.

El módulo de minería de patrones emergentes hizo posible el análisis profundo de los datos, para obtener reglas significativas a través de una aplicación Web, el sistema crea el modelo de aprendizaje y generar patrones emergentes, los cuales son las reglas obtenidas, que se interpretan en lenguaje natural, para que sean entendidas por cualquier personal médico.

En la evaluación se observó que el algoritmo que obtuvo mayor número de reglas totales fue iEPMiner, ya que tuvo un total de 4865 reglas en comparación con DGCP-Tree el cual solo obtuvo 848, esto para el conjunto D, para el conjunto C, también ganó iEPMiner, ya que tiene un total de 262 EP (Patrones Emergentes) en comparación con DGCP-Tree que obtuvo 111, sin embargo, ambos cumplen en su totalidad con la confianza > 0.6 , por lo que los resultados son altamente confiables.

5.2. Recomendaciones

Como recomendaciones se propone trabajar en la implementación de otro módulo que permita la comparación entre diversos conjuntos de datos obtenidos de distintos hospitales, usando otras tareas de la Minería de Datos, como parte de la investigación que se realizó, se observó que la disminución de autopsias médicas no solo es un problema de H.R.R.B, si no, de muchos hospitales en todo el mundo, es un tema bastante interesante e importante que debe ser atendido,

por lo que es necesario saber cuáles son las causas de su disminución en todos los hospitales, para tener resultados mucho más precisos y se logre hacer algo al respecto a nivel mundial y no solo local.

Productos Académicos



Ingrid Aylin Ríos Méndez, Lisbeth Rodríguez Mazahua, María Antonieta Abud Figueroa, Giner Alor Hernández, José Antonio Palet Guzmán.

“Análisis de Encuestas sobre la Disminución de Autopsias en un Hospital Mexicano usando Minería de Patrones Emergentes”, Congreso de Asociación Nacional de Instituciones de Educación en Tecnologías de la Información (ANIEI) 2019.

Avances en la Transformación Digital hacia la industria 4.0 (ISBN 978-607-538-551-8) (pág. 327-333)

Estado: Publicado.



Lisbeth Rodríguez-Mazahua, Ingrid Aylin Ríos-Méndez, José Antonio Palet Guzmán, Jair Cervantes, Giner Alor-Hernández, Isaac Machorro-Cano.

“Analysis of Medical Opinions about the Decrease of Autopsies in a Mexican Hospital Using Emerging Pattern Mining”, Journal of Universal Computer Science.

Estado: Enviado.



Ingrid Aylin Ríos-Méndez, Lisbeth Rodríguez-Mazahua, José Antonio Palet Guzmán, Isaac Machorro-Cano, Silvestre Gustavo Peláez-Camarena, Celia Romero-Torres, Hilarión Muñoz-Contreras.

“Discovering Emerging Patterns from Medical Opinions about the Decrease of Autopsies Performed in a Mexican Hospital” 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), Hong Kong, Hong Kong, 2020, pp. 798-803, doi: 10.1109/CASE48305.2020.9216751.

Estado: Publicado.

Referencias

- [1] E. Rubio Delgado, «Determinación del diagnóstico situacional de las autopsias en el H.R.R.B aplicando algoritmos de aprendizaje automático para las tareas de minería de datos», Orizaba, Veracruz, México, 2017.
- [2] J. Han, M. Kamber, y J. Pei, *Data Mining Concepts and Techniques*, Third edition. Morgan Kaufmann Publishers, jun. 2011.
- [3] C. P. López, *Minería de datos: técnicas y herramientas*. Thomson Paraninfo, 2007.
- [4] P.-N. Tan, M. Steinbach, A. Karpatne, y V. Kumar, *Introduction to Data Mining*, Second edition. Pearson, ene.2018.
- [5] G. Dong y J. Li, «Efficient Mining of Emerging Patterns: Discovering Trends and Differences» In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 43-52, 1999.
- [6] S. Ventura y J. M. Luna, *Supervised Descriptive Pattern Mining*. Springer International Publishing, 2018.
- [7] M. Atzmüller, *Knowledge-intensive Subgroup Mining: Techniques for Automatic and Interactive Discovery*. Aka Akademische Verlagsgesellschaft, 2007.
- [8] M. Sánchez-Mena, «EL CLINICO Y LA MUERTE.», vol. 8, n.º 3, p. 26, 1997.
- [9] D. Ienco, R. G. Pensa, y R. Meo, «A Semisupervised Approach to the Detection and Characterization of Outliers in Categorical Data», *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, n.º 5, pp. 1017-1029, may 2017.
- [10] G. Li, R. Law, H. Q. Vu, J. Rong, y X. (Roy) Zhao, «Identifying emerging hotel preferences using Emerging Pattern Mining technique», *Tour. Manag.*, vol. 46, pp. 311-321, feb. 2015.
- [11] R. Vimieiro y P. Moscato, «A new method for mining disjunctive emerging patterns in high-dimensional datasets using hypergraphs», *Inf. Syst.*, vol. 40, pp. 1-10, mar. 2014.
- [12] Q. Liu, P. Shi, Z. Hu, y Y. Zhang, «A novel approach of mining strong jumping emerging patterns based on BSC-tree», *Int. J. Syst. Sci.*, vol. 45, n.º 3, pp. 598-615, mar. 2014.
- [13] A. M. García-Vico, C. J. Carmona, D. Martín, M. García-Borroto, y M. J. del Jesus, «An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects: Emerging pattern mining», *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, n.º 1, p. e1231, ene. 2018.
- [14] M. García-Borroto, O. Loyola-González, J. Fco. Martínez-Trinidad, y J. A. Carrasco-Ochoa, «Evaluation of quality measures for contrast patterns by using unseen objects», *Expert Syst. Appl.*, vol. 83, pp. 104-113, oct. 2017.
- [15] A. Magalhães y P. J. Azevedo, «Contrast set mining in temporal databases», *Expert Syst.*, vol. 32, n.º 3, pp. 435-443, jun. 2015.
- [16] S. Q. ul Ain y S. ur Rehman, «Discovery of Jumping Emerging Patterns Using Genetic Algorithm», *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 16, n.º 4, pp. 1-6, abr. 2018.
- [17] M. García-Borroto, J. Fco. Martínez-Trinidad, y J. A. Carrasco-Ochoa, «Finding the best diversity generation procedures for mining contrast patterns», *Expert Syst. Appl.*, vol. 42, n.º 11, pp. 4859-4866, jul. 2015.
- [18] J. Reps, Z. Guo, H. Zhu, y U. Aickelin, «Identifying Candidate Risk Factors for Prescription Drug Side Effects Using Causal Contrast Set Mining», en *Health Information Science*, 2015, pp. 45-55.

- [19] R. Sherhod, P. N. Judson, T. Hanser, J. D. Vessey, S. J. Webb, y V. J. Gillet, «Emerging Pattern Mining To Aid Toxicological Knowledge Discovery», *J. Chem. Inf. Model.*, vol. 54, n.º 7, pp. 1864-1879, jul. 2014.
- [20] C.-H. Weng y T. Huang Cheng-Kui, «Observation of sales trends by mining emerging patterns in dynamic markets», *Appl. Intell.*, vol. 48, n.º 11, pp. 4515-4529, nov. 2018.
- [21] D. Nguyen, W. Luo, D. Phung, y S. Venkatesh, «Exceptional Contrast Set Mining: Moving Beyond the Deluge of the Obvious», en *AI 2016: Advances in Artificial Intelligence*, vol. 9992, B. H. Kang y Q. Bai, Eds. Cham: Springer International Publishing, 2016, pp. 455-468.
- [22] X. Yu, M. Li, K. Kim, J. Chung, y K. Ryu, «Emerging Pattern-Based Clustering of Web Users Utilizing a Simple Page-Linked Graph», *Sustainability*, vol. 8, n.º 3, p. 239, mar. 2016.
- [23] E. Rubio Delgado *et al.*, «Analysis of Medical Opinions about the Nonrealization of Autopsies in a Mexican Hospital Using Association Rules and Bayesian Networks», *Sci. Program.*, vol. 2018, pp. 1-21, 2018.
- [24] «Minería de datos sobre ontologías», 2011. [En línea]. Disponible en: <http://rtdibermatica.com/?tag=kdd>.
- [25] SIMIDAT/epm-framework, <http://github.com/SIMIDAT/epm-framework>.
- [26] H. Fan y K. Ramamohanarao, A Bayesian Approach to Use Emerging Pattern for Classification, En: Proc. of the 14th Australasian database conference, Vol. 17, pp. 39-48, 2003.
- [27] J. Li, G. Dong, K. Ramamohanarao y L. Wong, DeEPS: A New Instance-Based Lazy Discovery and Classification System, Machine Learning, Vol. 54, pp. 99-124, 2004.
- [28] H. Fan y K. Ramamohanarao, Efficiently Mining Interesting Emerging Patterns, En: Dong et al. (Eds.), WAIM 2003, LNCS 2762, pp. 189-201, Springer-Verlag, 2003.
- [29] J. Bailey, T. Manoukian y K. Ramamohanarao, Fast Algorithms for Mining Emerging Patterns, En: T. Elomaa (Eds.), PKDD, LNAI 2431, pp. 39-50, Springer-Verlag, 2002.
- [30] P. Terlecki y K. Walczak, Efficient Discovery of Top-K Minimal Jumping Emerging Patterns, En: C.-C. Chan et al. (Eds.), RSCTC 2008, LNAI 5306, pp. 438-447, SpringerVerlag 2008.
- [31] H. Fan y K. Ramamohanarao, Fast Discovery and the Generalization of Strong Jumping Emerging Patterns for Building Compact and Accurate Classifiers, IEE Transactions on Knowledge and Data Engineering, Vol. 18, No. 6, 2006.
- [32] A. M. García-Vico, J. Montes, J. Aguilera, C. J. Carmona y M. J. del Jesus, Analysing Concentrating Photovoltaics Technology Through the Use of Emerging Pattern Mining, En: M. Graña et al. (Eds.), International Joint Conference SOCO'16-CISIS'16-ICEUTE'16, Advances in Intelligent Systems and Computing 527, Springer International Publishing AG 2017.
- [33] M. García-Borroto, J. F. Martínez-Trinidad y J. A. Carrasco-Ochoa, A New Emerging Pattern Mining Algorithm and Its Application in Supervised Classification, En: M. J. Zaki et al. (Eds.) PAKDD 2010, Part I, LNAI 6118, pp. 150-157, Springer-Verlag, 2010.

[34] M. García-Borroto, J. F. Martínez-Trinidad, J. A. Carrasco-Ochoa, M. A. Medina-Pérez y J. Ruiz-Shulcloper, LCMine: An efficient algorithm for mining discriminative regularities and its application in supervised classification, *Pattern Recognition*, Vol. 43, pp. 3025-3034, 2010.

Anexos

Gráficas de las encuestas.

En la Figura 1 se muestran las gráficas del total de encuestas (Total-93), Último grado de estudios (Especialista-48, Médico general-29, Estudiante de medicina-16) y el área a la que pertenece cada médico (Interno-16, Residente-22, Adscrito-55).

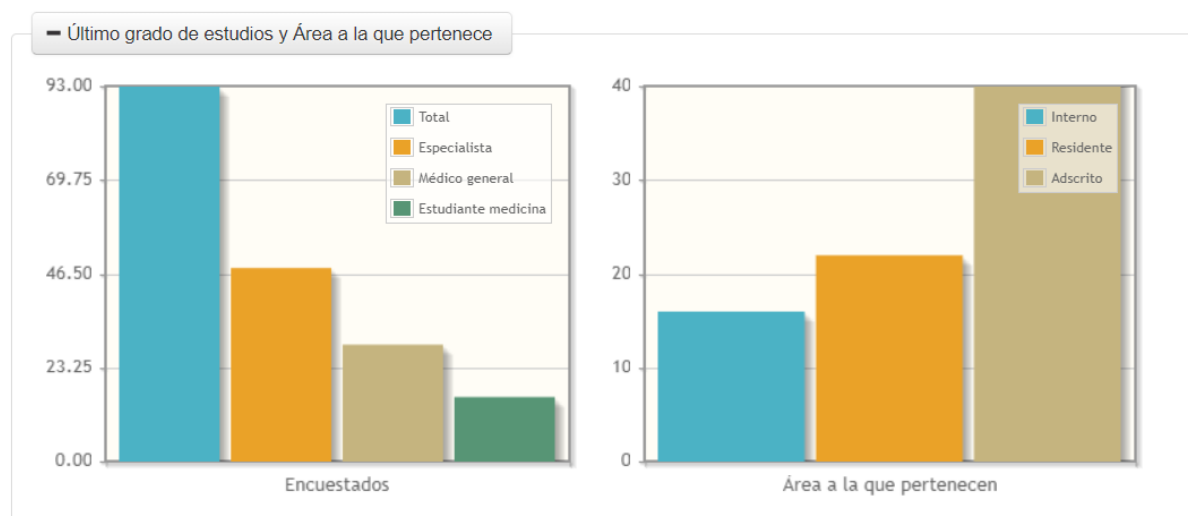


Figura 1 Último grado de estudios y Área a la que pertenece

En la Figura 2 se observan las gráficas de los años de práctica (Menos de 5 - 31, 5 a 10 - 22, de 11 a 15 - 6, 16 a 20 -7, Más de 20 - 27) y número de participación u observación de autopsias (Ningún caso-30, Menos de 5 casos-40, Entre 6 y 10 casos-13, Entre 11 y 20 casos-2, Más de 20 casos- 8).

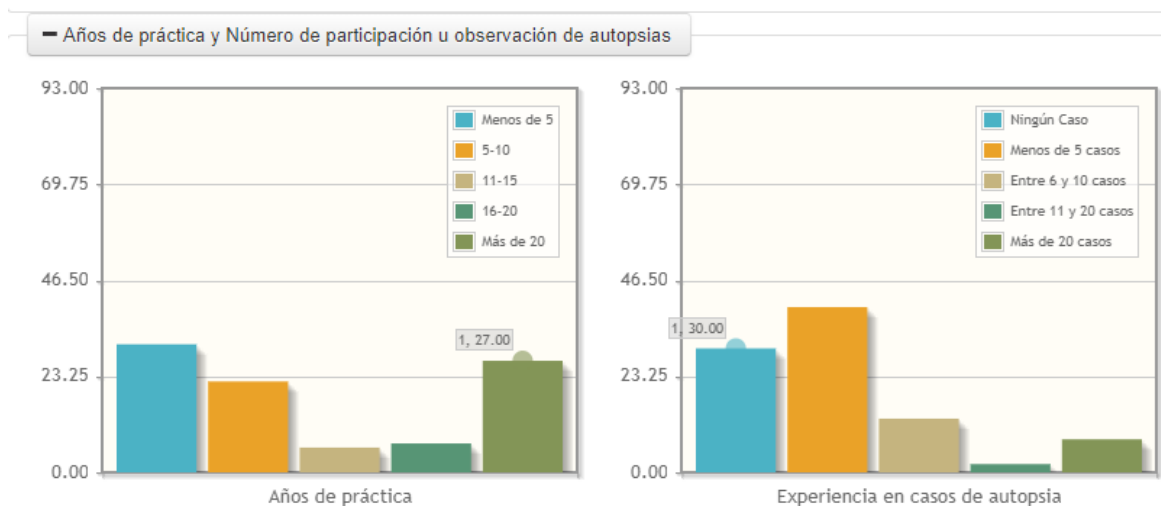


Figura 2 Años de práctica y número de participación u observación de autopsias

En la Figura 3 se muestran las gráficas de los hallazgos encontrados en las autopsias como: Los hallazgos encontrados en las autopsias pueden ser discrepantes con los diagnósticos clínicos (Totalmente de acuerdo-37, De acuerdo-32, Incierto- 16, En desacuerdo-7, Totalmente desacuerdo-1).

Los hallazgos encontrados en las autopsias dan origen a casos de arbitraje (Totalmente de acuerdo-19, De acuerdo-52, Incierto-13, En desacuerdo-6, Totalmente en desacuerdo-3).

Los hallazgos encontrados en las autopsias dan origen a casos de demanda (Totalmente de acuerdo-18, De acuerdo-40, Incierto-18, En desacuerdo-14, Totalmente desacuerdo-3).

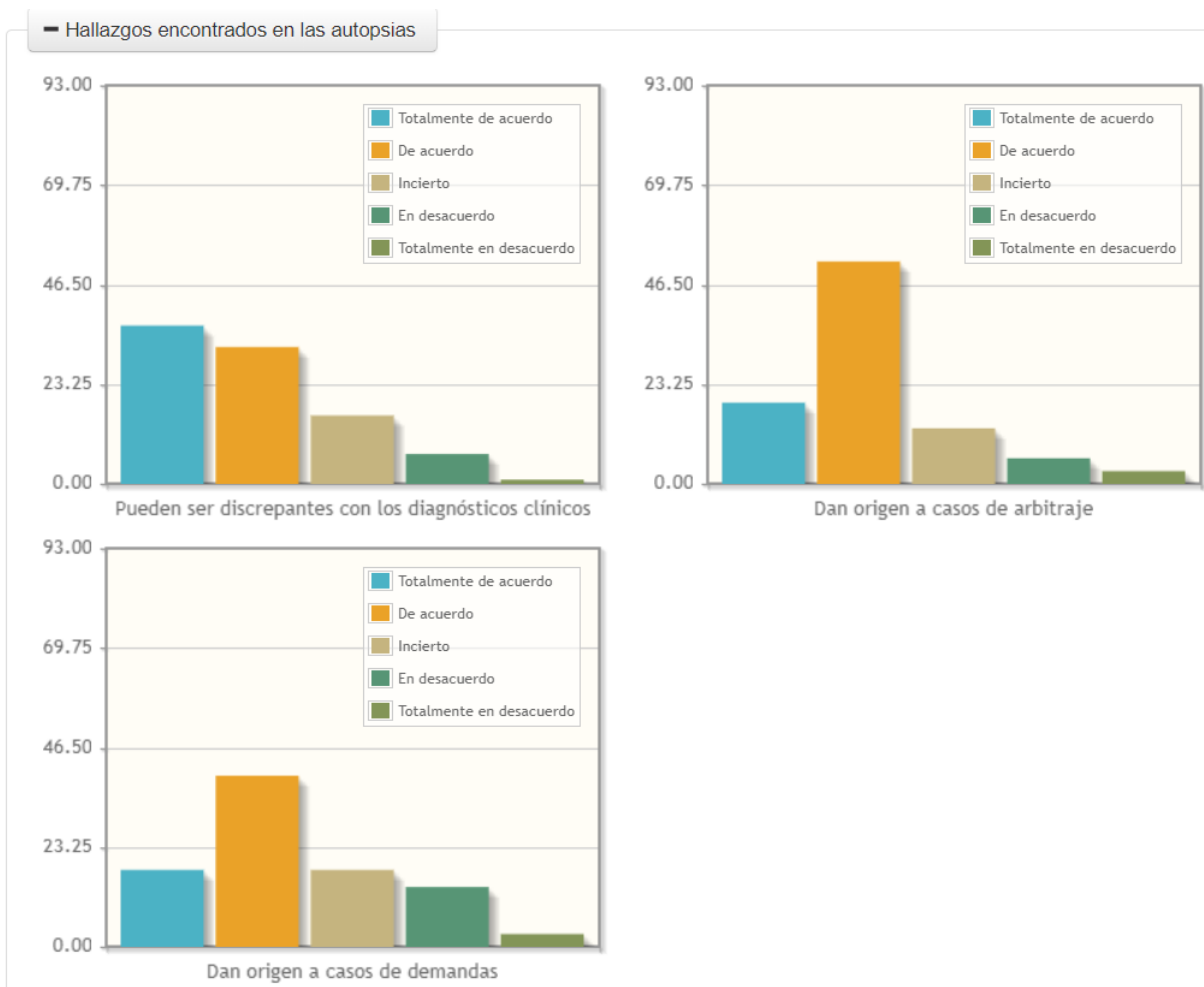


Figura 3 Hallazgos encontrados en las autopsias

La Figura 4 muestra la gráfica de la principal causa de rechazo por parte del familiar para realizar una autopsia:

Motivos religiosos-48

Motivos políticos-3

Motivos morales-36

No haber solicitado la autorización en forma adecuada -27

Inconformidad de los familiares con la atención prestada al paciente-5

Deficiente comunicación del médico con el paciente y sus familiares-44

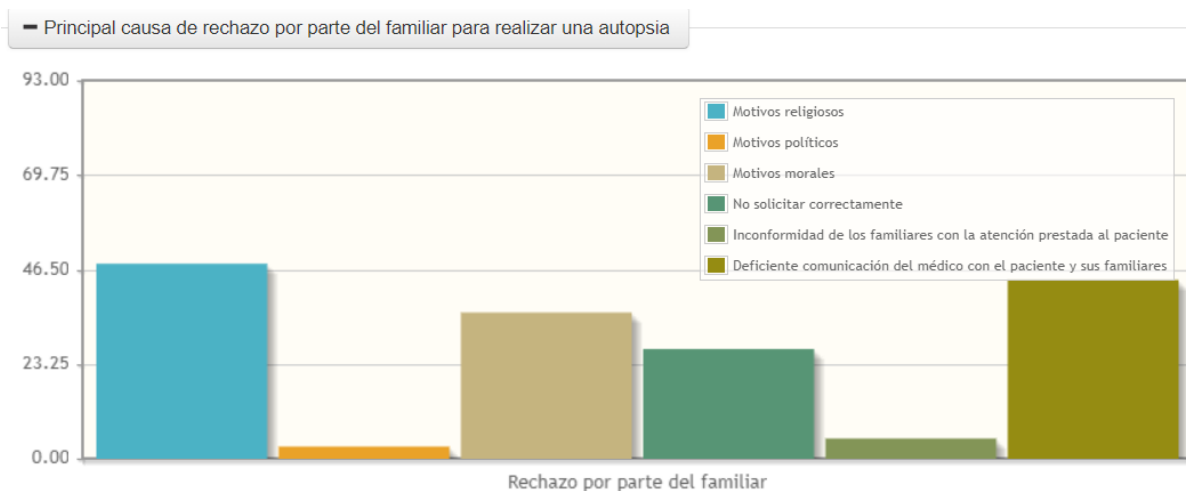


Figura 4 Principal causa de rechazo por parte del familiar para realizar una autopsia

En la Figura 5 se presentan los resultados que los médicos consideran que son las causas por las que no se realizan suficientes autopsias en su hospital:

Falta de recursos humanos-18

Falta de material-16

Falta de recursos financieros-23

Desconocimiento de su utilidad didáctica-35

Se considera que los nuevos procedimientos diagnósticos y otros estudios especializados eliminan la necesidad de su práctica-10

Desinterés del área de patología-14

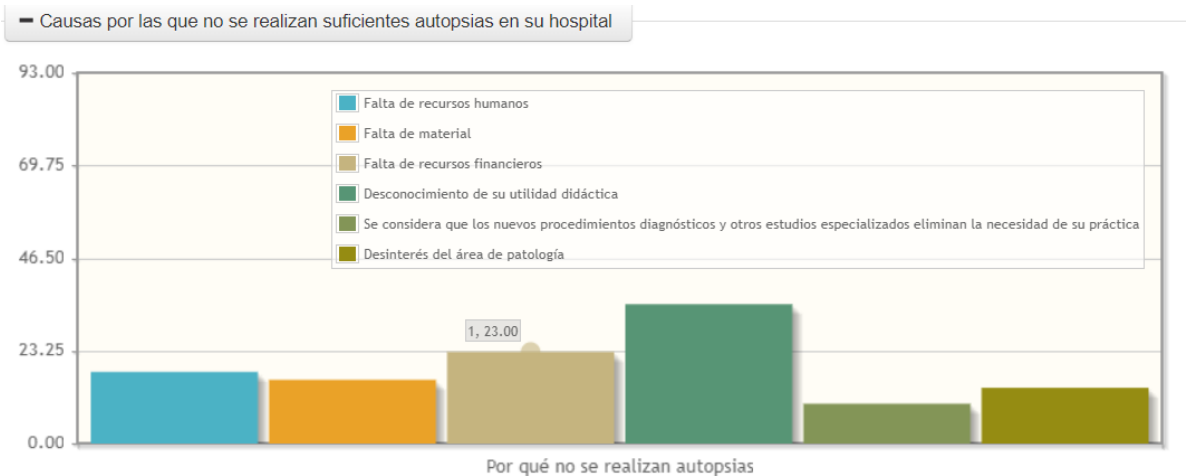


Figura 5 Causas por las que no se realizan suficientes autopsias en su hospital

En la Figura 6 se muestra la gráfica de quiénes consideran que debe solicitar la autopsia:

- Médico - 83
- Enfermera - 1
- Trabajadora Social - 5
- La Institución - 20
- Familiar - 11
- Criminología Forense - 10

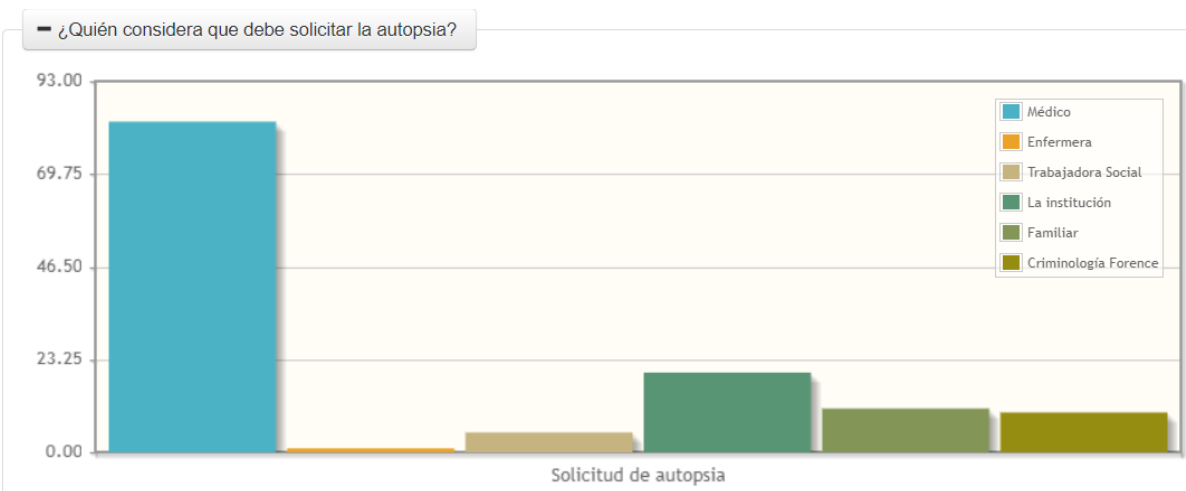


Figura 6 Solicitud de autopsia

En la Figura 7 se muestran cuál o cuáles serían las causas por parte del médico para solicitar una autopsia, Interés-73, Error médico-7, Medicamentos mal recetados-3, Negligencia médica-10, Diagnóstico erróneo-44, Orientación diagnóstica-17, Educación Médica Continua-5, Problema Médico Legal-4, Acoso por delincuencia organizada-1, Enseñanza-2.

¿Cuál o cuáles serían las causas por parte del médico para solicitar una autopsia?

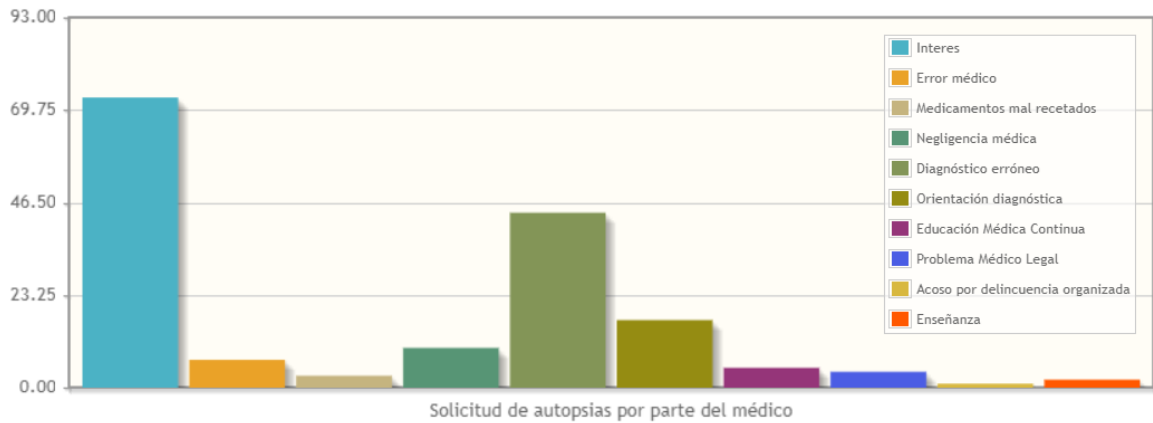


Figura 7 Solicitud de autopsias por parte del médico

En la Figura 8 se visualiza cuál sería la forma, método o recurso que pudiera funcionar en forma eficiente para la solicitud de un estudio de autopsia.

Pedir el consentimiento desde que ingresa al hospital al hospital-34.

Solicitar el consentimiento inmediatamente después de la defunción-37.

Que la institución a nivel regional o estatal reglamente la práctica de autopsias de forma obligatoria-48.

El médico tratante sea quien la solicite-20.

A través de otra área-10.

Comité para determinar-3.

A través del área de trabajo social-4.

¿Cuál sería la forma, método o recurso que pudiera funcionar en forma eficiente para la solicitud de un estudio de autopsia?

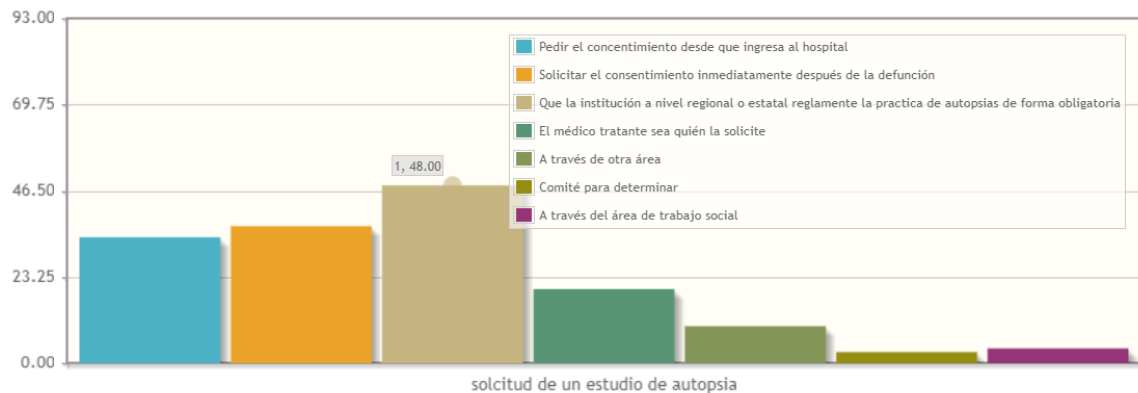


Figura 8 Solicitud de estudio de autopsias