



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE
MÉXICO

FACULTAD DE INGENIERÍA
MAESTRÍA EN CIENCIAS DE LA INGENIERÍA

MINADO DE REGLAS DE ASOCIACIÓN
APLICADO AL ANÁLISIS DE RIESGO
SÍSMICO. CASO DE ESTUDIO: SISMOS
DEL 07 Y 19 DE SEPTIEMBRE DE 2017

TESIS

QUE PARA OBTENER EL TÍTULO DE:
Maestra en Ciencias de la Ingeniería

PRESENTA:
Graciela García Rueda

Directores de Tesis:
Dra. Rosa María Valdovinos Rosas
Dr. Jesús Valdés González



Toluca, México, Noviembre 2020

Resumen

Los sismos son eventos que no se pueden predecir. Sin embargo, cuando ocurren, se muestran consecuencias devastadoras en diferentes ámbitos como: económico, social y estructural, entre otros. En México el 7 y 19 de Septiembre de 2017 se presentaron dos sismos de magnitud 8.2 y 7.1 respectivamente, los cuales provocaron fuertes daños en diferentes sectores. Un sector que llamó la atención de la comunidad científica y la sociedad en general es el sector educativo, en el que varias escuelas sufrieron colapsos de edificios provocando la muerte de niños y adultos. Esto último es lo que justifica la realización de estudios que permitan la detección del nivel vulnerabilidad de los edificios en determinadas zonas, y de este modo minimizar los daños derivados de la ocurrencia de un sismo.

El análisis de riesgo sísmico es aquel que proporciona los elementos para tomar decisiones técnicas como planificación de tipos de construcciones mas seguras y económicas tales como la destinación de fondos necesarios para la reconstrucción, mantenimiento de edificios, ante un sismo.

En esta tesis se hace uso de técnicas de minería de datos (Reglas de Asociación), para realizar el análisis de riesgo sísmico, utilizando los datos públicos recopilados por el FONDEN (Fondo para la Atención de Emergencias). En específico, se analizaron las escuelas que reportaron daños a nivel nacional, producto de los sismos del 7 y 19 de septiembre del 2017.

Las reglas de asociación obtenidas dieron a conocer las características de las escuelas que mas daños sufrieron durante los sismos, así como los costos aproximados que se utilizarán para las reparaciones, con el objetivo de obtener un resumen de la evolución de los costos por cada tipo de escuela, en base a la distancia epicentral.

Contenido

	Página
Índice de figuras	ix
Índice de tablas	xi
1. Introducción	1
1.1. Objetivo	4
1.2. Hipótesis	4
1.3. Estructura de la Tesis	5
2. Marco teórico y Estado del Arte	7
2.1. Minería de Datos	7
2.2. Técnicas Descriptivas de MD	8
2.3. Reglas de Asociación	8
2.3.1. Conjunto de datos transaccional	9
2.3.2. RA	11
2.3.3. Tipos de RA	11
2.3.3.1. RA multinivel	11
2.3.3.2. RA secuenciales	12
2.3.4. Algoritmos para obtención de RA	12
2.3.4.1. Algoritmo Apriori	12
2.3.4.2. Algoritmo PSO	13
2.3.4.3. Algoritmo PSO-GES	15
2.3.5. Métricas de calidad para RA	16
2.4. Análisis de Riesgo Sísmico	17
2.5. Estado del Arte	18
2.5.1. Análisis de riesgo sísmico con técnicas decriptivas	18
2.5.2. Análisis de riesgo sísmico con técnicas predictivas	25

CONTENIDO

3. Marco Metodológico	29
3.1. Selección	30
3.2. Pre-procesado	31
3.3. Transformación	33
3.3.1. Unificación del tipo de dato	33
3.3.2. Complementación de datos	34
3.3.3. Categorización	36
3.3.4. Discretización del conjunto de datos	37
3.3.5. Generación de datos transaccionales	39
3.4. Minería de Datos	39
3.5. Interpretación y evaluación	41
4. Resultados y Discusiones	43
4.1. Resultados	43
4.1.1. Patrones Frecuentes	43
4.1.2. Reglas de Asociación	45
4.1.2.1. Sismos del 7 de Septiembre de 2017	45
4.1.2.2. Sismos del 19 de Septiembre de 2017	47
4.2. Discusión	51
5. Conclusiones y Trabajo futuro	53
Bibliografía	57
Anexos	63
A. Formato de dictamen A	65
B. Formato de dictamen B	67
C. Formato de dictamen C	69

Índice de figuras

	Página
2.1. Vector Horizontal de Ítems[1].	9
2.2. Lista Horizontal de Ítems[1].	9
2.3. Vector Vertical de Ítems[2].	10
2.4. Lista Vertical de Ítems[2].	10
3.1. Fases de la metodología KDD [3].	29
3.2. Repositorio de las bases de datos del Fondo de Transparencia Pre- supuestaria	30
3.3. Muestra de conjunto de datos iniciales [4].	32
3.4. Regionalización sísmica de la República Mexicana [5]	37
4.1. Variación de los costos individuales promedio de reparación para el sismo del 7 de septiembre de 2017 en función del tipo de cons- trucción y la distancia epicentral.	47
4.2. Variación de los costos individuales promedio de reparación pa- ra el sismo del 19 de septiembre de 2017 en función del tipo de construcción y la distancia epicentral.	50

Índice de tablas

	Página
2.1. Ejemplo de base de datos transaccional	11
2.2. Resumen Estado del Arte	28
3.1. Ejemplo de datos iniciales	33
3.2. Ejemplo de datos finales	34
3.3. Categorías detalles del daño	38
3.4. Conjunto de datos discretizado	38
3.5. Conjunto de datos transaccional	39
4.1. Patrones frecuentes generados por el algoritmo FP-Growth. Sismo del 7 de septiembre de 2017.	44
4.2. Patrones frecuentes generados por el algoritmo FP-Growth. Sismo del 19 de septiembre de 2017.	45
4.3. Reglas de Asociación para el sismo del 7 de septiembre de 2017.	46
4.4. Reglas de Asociación para el sismo del 19 de septiembre de 2017	48

Capítulo 1

Introducción

Los sismos de origen tectónico se definen como la vibración de la corteza terrestre debida al deslizamiento súbito de las placas tectónicas, o al desplazamiento de las fallas geológicas que se encuentran dentro de las placas. Los desplazamientos y/o deslizamientos de las placas se deben a la liberación de presiones a las cuales están sometidas estas placas. Esta liberación súbita de energía es la responsable de los sismos [6].

Los movimientos sísmicos registrados a lo largo de la historia han evidenciado que uno de los factores que más genera pérdidas materiales y humanas es el derrumbe de las construcciones. Materiales de deficiente calidad, incumplimiento de normas sismo resistentes y mano de obra no calificada influyen en el derrumbe de las construcciones [7]. Ante esto, el estudio de la vulnerabilidad sísmica determina el nivel de afectación de una zona o de una estructura ante los movimientos sísmicos que pueden acontecer en el sitio. El análisis y diseño sísmico de las construcciones se lleva a cabo haciendo un diagnóstico del desempeño y comportamiento que éstas tendrán ante la ocurrencia de los sismos que las afecten durante su vida útil.

Comúnmente el análisis estructural se realiza por métodos probabilistas y deterministas. Por un lado los métodos probabilísticos se encargan de realizar la modelación estadística de la ocurrencia temporal de sismos, de sus tamaños y la atenuación del movimiento del suelo dada la ocurrencia de éstos, con el objeto de determinar la probabilidad de excedencia de determinados niveles del movimiento del suelo. Por otro lado, los métodos deterministas buscan establecer los parámetros del movimiento del terreno más desfavorables, al considerar la ocurrencia del mayor sismo del que se tenga conocimiento [8].

Una de las áreas que apoyan el diagnóstico y el análisis del desempeño de las estructuras, es la Minería de Datos (MD). La MD, constituye un campo en las ciencias de la computación que utiliza técnicas para buscar sentido a la explosión

1. INTRODUCCIÓN

de información que actualmente puede ser almacenada mediante el descubrimiento de conocimiento, ayudando con esto a la toma de decisiones [9].

De acuerdo al tipo y estructura de los datos recabados después de la ocurrencia de los sismos, las técnicas de MD que más se han utilizado para su análisis son las conocidas como minado de reglas de asociación (RA). El minado de RA representa una técnica para descubrir asociaciones o correlaciones interesantes a partir de un conjunto de datos [10].

Un ejemplo de su aplicación en el estudio relacionado con los sismos, es el caso de Martínez-Álvarez et al. [11] quienes aplican técnicas descriptivas para obtener Reglas de Asociación Cuantitativas (QAR) y la regresión (Algoritmo M5P) como técnica predictiva de la ocurrencia de sismos bajo determinadas circunstancias.

Por su parte, Nivedhitha U. et al. [12] hacen uso de técnicas descriptivas de MD, donde, utilizando algoritmos de clustering y el algoritmo Apriori-TFT generan Reglas de asociación con el objetivo de anticipar la ocurrencia de sismos y tomar decisiones al respecto.

En otro estudio realizado por Galán Montaña F. [13] se utilizan dos técnicas de MD, obteniendo Reglas de Asociación para la etapa predictiva y el algoritmo del vecino más cercano (KNN), J48 y redes neuronales artificiales para la etapa de la predicción, con la intención de encontrar patrones y realizar modelos del comportamiento de las series temporales que comprenden la ocurrencia de sismos de magnitud media.

En los últimos 17 años, datos del Centro Nacional de Prevención de Desastres (CENAPRED) de acuerdo a los datos se tienen registrados 28 sismos de magnitud media (5 a 6.9 Richter) y alta (7 o más Richter), ocupando el lugar 20 entre los fenómenos que ocurren con mayor frecuencia en el país. No obstante, cuando se calculan los daños económicos que producen, su posición cambia drásticamente al quinto lugar. Del año 2000 a 2017 las pérdidas económicas por los sismos en México fueron de 62 mil 139 millones de pesos (70 % de este costo contempla los sismos ocurridos el 7 y 19 de septiembre de 2017) [14].

El 7 de Septiembre de 2017 se registró un sismo con magnitud 8.2 en Oaxaca, causando daños en 107 municipios, 57,621 viviendas, 1,988 escuelas, 102 inmuebles culturales y 104 edificios públicos resultaron dañados [15]. A pocos días de lo ocurrido, el 19 de Septiembre de 2017 un sismo de magnitud 7.1 grados con epicentro en los límites de Puebla y Morelos afectó la zona centro del país, principalmente. Resultando afectadas un gran número de construcciones de la ciudad de

México, Puebla, Morelos, Estado de México, Guerrero, e incluso Oaxaca y Chiapas. Se estima que el costo de reparación de las más de 150,000 viviendas que resultaron dañadas en estos estados fué cercano a los 38,150 millones de pesos [16].

En términos generales, el Sector Educativo tuvo deterioros considerables derivado de los dos sismos, registrando 12,931 escuelas con daños, de las cuales 577 requerirán una reconstrucción total, 1,847 una reconstrucción parcial y el resto con daños menores, de las cuales el costo de reparación se estima será de 13,650 millones de pesos [17].

Otras estructuras con afectación por el sismo del 19 de Septiembre, son los inmuebles históricos y de valor cultural, como la zona arqueológica de Chiapa de Corzo, Zócalo de la Ciudad de México, el Museo Nacional de Arte, entre otros, cuya reparación tiene un costo estimado de 8000 millones de pesos [18].

Sin embargo no solo se tuvieron daños estructurales sino que también hubo personas heridas y fallecidas. Para obtener el costo aproximado de muertes se consideraron parámetros como: el promedio de salario mensual de acuerdo al IMSS que es de \$10,582.56, edad de jubilación de 65 años, promedio de vida de 75 años y número de muertes de los sismos del 7 y 19 de septiembre del 2017.

El sismo del 7 de septiembre provocó la muerte de 102 personas, mientras que el del 19 de septiembre de 369 personas. De las cuales para este último sismo, 228 muertes se presentaron en la Ciudad de México, correspondiendo 200 a personas adultas y 28 a niños [19]. De acuerdo a cálculos actuariales aproximados, se estima que el costo estimado de las muertes provocadas por ambos sismos fué cercano a \$2,236,179,588.00.

Derivado de lo anterior es indispensable evaluar detalladamente la seguridad estructural de los inmuebles, a través de pruebas de campo, de laboratorio y cálculos para establecer técnicamente las condiciones de habitabilidad y riesgo de los inmuebles respecto de los parámetros técnicos y conforme a las normas aplicables [20]. Si bien ya nada se puede hacer respecto a lo ocurrido, si es importante determinar y encontrar una explicación a los daños sufridos, como producto de estos sismos.

Esto último permite realizar un análisis de riesgo sísmico para un conjunto de datos el cual debe contener, las características estructurales más importantes a considerar en el análisis del riesgo sísmico como son [21]: número de niveles del edificio, tipo de material de construcción (Mampostería, Concreto, Acero), año de construcción, irregularidades, entre otros. No obstante, debido a la naturaleza

1. INTRODUCCIÓN

del problema, la recolección de datos es compleja y escasa.

En base a lo anterior, en esta tesis se presenta el análisis de riesgo sísmico aplicando MD, en el cual se realiza una clasificación de los daños de acuerdo a ciertas características de interés desde el punto de vista estructural de los sismos del 7 y 19 de Septiembre, utilizando los datos proporcionados por el Fondo de Transparencia Presupuestaria¹.

1.1. Objetivo

Realizar el minado de reglas de asociación en un conjunto de datos de infraestructuras afectadas del sector educativo, para los sismos del 7 y 19 de Septiembre de 2017, con la intención de obtener los patrones descriptivos, de los daños ocasionados por el sismo en función de las características estructurales de construcciones afectadas. Para ello se establecieron los siguientes objetivos particulares:

- Obtener los conjuntos de datos de inmuebles afectados derivados de los sismos del 7 y 19 de Septiembre de 2017 en los estados de la República Mexicana que sufrieron afectaciones.
- Realizar el preprocesado del conjunto de datos existente para ser utilizados con los algoritmos de minado de RA más utilizados.
- Obtener los patrones frecuentes y las reglas de asociación para el análisis de riesgo sísmico.
- Validar la calidad de las reglas de asociación que aporten información para la toma de decisiones.
- Interpretación de las reglas de asociación por un experto en el área de estructuras.

1.2. Hipótesis

Con el minado de reglas de asociación de datos de las construcciones del sector educativo afectadas por los sismos del 7 y 19 de Septiembre de 2017 se podrá conocer el comportamiento general del daño producido por un sismo en función de las características estructurales, con un grado de confianza superior a 75 %.

¹Fuerza México, https://www.transparenciapresupuestaria.gob.mx/es/PTP/fuerzamexico_datosabiertos

1.3. Estructura de la Tesis

La Tesis consta de cinco capítulos, el primer capítulo contextualiza el problema de estudio, mientras que el marco teórico y estado del arte se desarrollan en el capítulo 2. La metodología a seguir se describe en el capítulo 3, en tanto que en el capítulo 4 se presentan los resultados obtenidos. Por último se presentan las conclusiones y trabajo futuro en el capítulo 5.

De forma adicional se incluyen una serie de anexos en los que se muestran algunos formatos de dictámenes y la estructura del conjunto de datos utilizado.

Marco teórico y Estado del Arte

En este capítulo se incluye el sustento teórico de la investigación, en la que se encuentran temas como la definición de la MD, y las RA, clasificación y métricas de calidad de las RA, algoritmos utilizados para el minado de RA. Por último, se presentan algunos trabajos en lo que se realiza el minado de RA con datos sísmicos.

2.1. Minería de Datos

La minería de datos es el proceso de descubrimiento automático de información útil en grandes repositorios de datos [3]. Para ello, la MD prepara, sondea y explota los datos para sacar la información oculta en ellos, de modo que un pequeño hallazgo ó relación que se descubre, puede ser un hecho de alto impacto [22]. Hoy en día la utilización de las técnicas de MD se han realizado en diferentes áreas, tales como [23]: Aplicaciones financieras y banca, educación, procesos industriales, medicina, biología, telecomunicaciones, análisis de mercado, entre otros.

La MD cuenta con dos enfoques de aplicación, el descriptivo que se encarga de identificar patrones que expliquen o generen un resumen de los datos y el predictivo que permite estimar valores de interés, a partir de otras variables de un conjunto de datos. En este trabajo se abordan las técnicas descriptivas derivado que para el área estructural la información de dos sismos es muy escasa y por lo cual no se puede aplicar una predicción por dicho motivo, sin embargo si se puede hacer un análisis descriptivo para entender el comportamiento sísmico en las escuelas.

2.2. Técnicas Descriptivas de MD

Las técnicas descriptivas exploran las propiedades de los datos, proporcionando información sobre las relaciones existentes en los mismos. Los algoritmos utilizados para este fin detallan el comportamiento de los datos de forma que sea interpretable por un usuario experto. En su funcionamiento no suponen la existencia de variables dependientes ni independientes, ni conocimiento a partir de la distribución de los datos y se enfocan al descubrimiento del conocimiento. Algunas tareas que dan lugar a modelos descriptivos son [24]:

- **Agrupamiento (clustering):** Al agrupamiento también se le suele llamar segmentación, ya que parte o segmenta los datos en grupos, de tal manera que los patrones asignados al mismo grupo sean similares, o lo mas similares entre sí y lo mas diferente a los patrones de otros grupos.

Para la realización de un agrupamiento existen distintos algoritmos los cuales pueden clasificarse atendiendo a varios criterios como [25]:

- Forma en que procesan los objetos.
 - Pertenencia de los objetos a los grupos formados.
 - Forma en que se organizan o se relacionan los grupos obtenidos.
 - Mecanismo en que se basan para agrupar los objetos.
- **Correlaciones y factorizaciones:** Su objetivo es analizar la relevancia de los atributos que generan el patrón, detectar atributos redundantes o dependencias entre atributos, con el objeto de seleccionar un subconjunto representativo.
 - **Reglas de asociación:** Son una manera muy popular de expresar patrones de datos en una base de datos. Estos patrones pueden servir para conocer el comportamiento general del problema que genera una base de datos, y de esta manera, se tenga más información que pueda asistir en la toma de decisiones.

En esta tesis se realizó el minado de RA para encontrar relaciones entre distancia epicentral, tipos de edificios y costos por lo que en la siguiente sección se aborda con mas a profundidad sobre RA.

2.3. Reglas de Asociación

Las Reglas de Asociación (RA) representan una técnica para descubrir asociaciones o correlaciones interesantes a partir de un conjunto de datos, comúnmente transaccionales.

2.3.1. Conjunto de datos transaccional

Conceptualmente, un conjunto de datos transaccional es una matriz de dos dimensiones donde las filas representan las transacciones y las columnas representan los *ítems* o atributos que conforman la transacción. Esta matriz se puede implementar de las siguientes cuatro formas [2]:

1. Vector Horizontal de *Ítems* (*VHI*): La base de datos se organiza como un conjunto de filas donde cada una de ellas almacena un identificador de transacción (*Tid*) y un vector de bits de 1 y 0 para representar cada uno de los *ítems*, su presencia o ausencia, respectivamente, en la transacción (Figura 2.1).

Tid	Ítems				
	1	2	3	4	5 ...
1	1	0	1	0	0
2	0	1	0	0	0
3	0	1	1	0	0
4	1	0	1	0	1

Figura 2.1: Vector Horizontal de *Ítems*[1].

2. Lista Horizontal de *Ítems* (*LHI*): Esto es similar al *VIH*, excepto que cada fila almacena una lista ordenada de identificadores de *ítems* (*Tid*) presentes en cada transacción (Figura 2.2).

Tid	Ítems
1	1 3 7 9
2	2 8 15
3	2 3 7 8 11
4	1 3 5 10

Figura 2.2: Lista Horizontal de *Ítems*[1].

2. MARCO TEÓRICO Y ESTADO DEL ARTE

3. Vector Vertical de Identificadores de Transacciones (*VTV*): La base de datos está organizada como un conjunto de columnas, donde cada una de ellas almacena un *Tids* y un vector de bits de 1 y 0 para representar la presencia o ausencia, respectivamente, del *ítem* en el conjunto de transacciones (Figura 2.3).

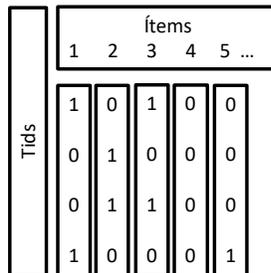


Figura 2.3: Vector Vertical de Ítems[2].

4. Lista Vertical de Identificadores de Transacciones (*LVTid*): Esto es similar a *VTV*, excepto que cada columna almacena una lista ordenada de solo los *Tids* de las transacciones en la que aparece el *ítem* (Figura 2.4).

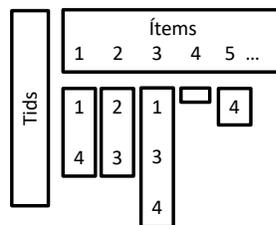


Figura 2.4: Lista Vertical de Ítems[2].

Para esta investigación se utilizó un conjunto de datos transaccional de lista horizontal de *ítems* como se muestra en el ejemplo de la Tabla 2.1 [10], debido a que de esta manera se reduce la dimensionalidad de las columnas dejando solo los atributos que cuenten con presencia en cada transacción.

Transacciones	
Tid	Artículos
1	Zona sísmica B, Daño acabados, Magnitud 8.1
2	Zona sísmica B, Daño estructural, Magnitud 7.2
3	Muros de mampostería, Daño estructural, Magnitud 7.2
4	Zona sísmica B, Daño acabados, Daño estructural, Magnitud 7.2
5	Zona sísmica B, Daño acabados, Magnitud 8.2, Magnitud 7.2

Tabla 2.1: Ejemplo de base de datos transaccional

2.3.2. RA

La definición de una RA plantea que sea $I = i_1, i_2, \dots, i_n$ un conjunto de n atributos llamados *ítems* y $D = t_1, t_2, \dots, t_m$ un conjunto de transacciones almacenadas en un conjunto de datos. Cada transacción en D tiene un *Tid* (identificador único) y contiene un subconjunto de *ítems* de I (conjunto de *ítems* al cual se le denomina *itemset*) [26]. Por lo que una RA se puede definir como una implicación de la forma X entonces Y ($X \Rightarrow Y$), donde $X, Y \subset I$ y $X \cap Y = \emptyset$, entiendo con ello que toda transacción que satisface X también satisface a Y [27].

Por ejemplo, considerando el conjunto de datos de la Tabla 2.1, se tienen las posibles asociaciones entre los datos sísmicos:

Daño estructural \Rightarrow *Magnitud 7.2*

Daño acabados \wedge *Zona sísmica B* \Rightarrow *Magnitud 8.1*

Magnitud 7.2 \wedge *Zona sísmica B* \Rightarrow *Magnitud 8.1* \wedge *Daño acabados*

2.3.3. Tipos de RA

Las reglas de asociación se pueden clasificar dependiendo la manera en que los datos se presenten, las cuales se muestran a continuación.

2.3.3.1. RA multinivel

Las RA multinivel, son aquellas que involucran *ítems* con diferentes niveles de abstracción. Para este tipo de Reglas se utilizan varios niveles de conceptos jerárquicos para expresar las relaciones, por lo que se debe proporcionar, además de los datos, una jerarquía de conceptos. El concepto de jerarquía define una secuencia de diagramas, iniciando en un conjunto de conceptos de bajo nivel hasta llegar a conceptos de alto nivel; en otras palabras, se inicia con conceptos específicos en el nivel más bajo, y se va subiendo en jerarquía a conceptos más generales

[23]. Para la generación de reglas de asociación multinivel existen principalmente dos métodos [28]:

- Utilizando el mismo soporte mínimo para todos los niveles (soporte uniforme). Este soporte es empleado cuando se realiza minería en cada nivel de abstracción, simplificando de esta manera el procedimiento de búsqueda. Sin embargo, el método tiene algunas dificultades: es improbable que los ítems en un nivel de abstracción inferior ocurran tan frecuentemente como aquellos que están en un nivel de abstracción superior. Si el mínimo soporte establecido es muy alto podrían perderse muchas asociaciones significativas en niveles inferiores de abstracción; si es muy bajo podrían generarse muchas asociaciones poco interesantes en niveles de abstracción más altos.
- Utilizando soportes mínimos más reducidos en niveles de abstracción inferiores. Cada nivel de abstracción tiene su propio soporte mínimo, por lo cual el nivel más bajo tendrá el soporte más pequeño de toda la jerarquía.

2.3.3.2. RA secuenciales

Expresan patrones de comportamiento secuencial, es decir que tiende a producirse en un orden previsible [29]. Este tipo de información es de importancia crucial en áreas de aplicación tales como el análisis de navegación sobre páginas web.

El aprendizaje de una RA secuencial se basa en encontrar secuencias más comunes. Una secuencia se define formalmente como una lista de conjuntos de *ítems*. Uno de los algoritmos más populares para el minado de reglas de asociación secuenciales es el algoritmo Apriori All que cuenta con cinco fases [23]: Ordenación, selección de conjuntos de ítems, transformación y renombramiento, construcción de secuencias frecuentes y selección de secuencias mixtas.

2.3.4. Algoritmos para obtención de RA

Existen diferentes algoritmos que permiten la extracción de RA, dentro de la literatura se pueden encontrar los siguientes.

2.3.4.1. Algoritmo Apriori

El algoritmo Apriori (2.1) se basa en la búsqueda de los conjuntos de *ítems* con determinada cobertura. En este algoritmo, C_k contiene los conjuntos de *ítems* candidatos a ser frecuentes de tamaño i y L_1 los conjuntos frecuentes de *ítems*

de tamaño $i - 1$.

Primero, son obtenidos y almacenados en C_k todos los *ítems* con sus respectivos valores de soporte (Algoritmo 2.1, línea 1). Para cada elemento de C_k se verifica si el soporte es mayor o igual que el mínimo soporte establecido (Algoritmo 2.1, línea 2), si es mayor se va agregar a L_1 (Algoritmo 2.1, línea 6). Posteriormente en cada iteración del algoritmo se generan los conjuntos de *ítems* candidatos a frecuentes de tamaño $i + 1$ combinando los conjuntos de L_i hasta que los *itemset* generados no cumplan con la condición de mínimo soporte (Algoritmo 2.1, línea 5). Lo que permite ir eliminando posibles combinaciones. Aquellas que no cumplan con los requerimientos de soporte no entrarán en el análisis, esto se aprovecha en la construcción de candidatos, para no considerar todas las opciones [23].

Algunas de las ventajas del algoritmo Apriori es que es eficiente para grandes volúmenes de datos, es uno de los mas usados, sin embargo sus desventajas es que consume una gran cantidad de recurso(memoria).

Algoritmo 2.1 *Apriori*

Entrada: Conjunto de datos T

Mínimo Soporte $minSupport$

Salida: Retorna los *itemset* frecuentes L

$L_1 = \{itemsfrecuentes\}$

- 1: **para** $k \leftarrow 2$ a $L_{k-1} \neq 0$ **hacer**
 - 2: $C_k = GenerarCandidatosdeL_{k-1}$
 - 3: **para todo** $t \in T$ **hacer**
 - 4: Incrementar las cantidades de todos los candidatos en C_k si están presentes en t
 - 5: $L_k =$ candidatos en C_k con soporte mayor que el $minSupport$
 - 6: **fin para**
 - 7: **fin para**
 - 8: **devolver** $\cup_k L_k$
-

2.3.4.2. Algoritmo PSO

El algoritmo PSO (Particle Swarm Optimization, PSO), es una metaheurística poblacional basada en la simulación de modelos sociales simples e inspirada en el comportamiento social del vuelo de las bandadas de aves o el movimiento de los bancos de peces. En PSO, se utiliza una población de tamaño fijo. Cada partícula de la población, es una solución candidata al problema y sus movimientos se encuentran acotados al espacio de búsqueda. Este espacio se encuentra

2. MARCO TEÓRICO Y ESTADO DEL ARTE

definido de antemano y no se permite que las partículas se desplacen fuera de él. El Algoritmo de búsqueda PSO 2.2, cada partícula, se inicializa su posición (X_i) y su velocidad (V_i) de manera aleatoria.

Una vez que la población es inicializada los individuos comienzan a moverse por el espacio de búsqueda por medio de un proceso iterativo. Con la nueva posición del individuo, se calcula y actualiza su $fitness(fitness_{X_i})$ (el $fitness$ se define sobre la representación genética y mide la calidad de la solución representada). Además, si el nuevo $fitness$ del individuo es el mejor encontrado hasta el momento, se actualizan los valores de mejor posición $PBest_i$ y $fitnessPBest_i$ (Algoritmo 2.2, línea 6). En cada iteración el vector de velocidad de la partícula se actualiza tomando en cuenta su experiencia y su entorno (Algoritmo 2.2, línea 3), es decir que toma la mejor solución encontrada [30].

Algunas de las ventajas del algoritmo PSO son la facilidad de implementación, tiene pocos parámetros de ajuste y es eficiente en la búsqueda global, y entre sus desventajas es que tiende a tener una rápida y prematura convergencia en los puntos medios óptimos.

Algoritmo 2.2 PSO

Entrada: Número de Partículas N

Número de Iteraciones T

Salida: Devuelve las mejores soluciones encontradas $GBest$

$POP = CrearPoblacion(N)$

```
1: mientras  $t < T$  hacer
2:   para  $i = 1$  a  $size(POP)$  hacer
3:      $V_i^{(t+1)} = w * V_i^t + c_1 * rand_1 * (PBest_i - P_i^t) + c_2 * rand_2 * (GBest - P_i^t)$ 
4:      $V_i^{(t+1)} = P_i^t + V_i^{(t+1)}$ 
5:     si  $Fitness(P_1) > Fitness(PBest_i)$  entonces
6:        $PBest_i \leftarrow P_i$ 
7:     fin si
8:     si  $Fitness(P_1) > Fitness(GBest_i)$  entonces
9:        $GBest_i \leftarrow P_i$ 
10:    fin si
11:  fin para
12: fin mientras
13: devolver  $GBest$ 
```

2.3.4.3. Algoritmo PSO-GES

El algoritmo PSO-GES (2.3), realiza el minado de RA para la identificación automática de RA que permite predecir las posibles relaciones entre dos o más *items* en conjuntos de datos transaccionales, sin generar patrones frecuentes. Para ello este algoritmo esta basado en PSO (*Particle Swarm Optimization*) [30] con una estrategia de búsqueda guiada con el objetivo de generar RAs de mayor calidad.

Algoritmo 2.3 Algoritmo PSO-GES

Entrada: Número de partículas N
 Número de iteraciones T
 Número de RAs que se desean encontrar M
 Dataset D
 Número de particiones K
 Mínimo Soporte $MinSup$
 Mínima Confianza $MinCof$
 Factor de Inercia w
 Constantes de aceleración c_1 and c_2

Salida: Devuelve las M mejores reglas de asociación encontradas $GBests$

- 1: $Table \leftarrow CrearTabla(D, K)$
- 2: $POP \leftarrow CrearPoblacion(N)$
- 3: **mientras** $t \leq T$ **hacer**
- 4: **para** $i \leftarrow 0$ a $size(POP)$ **hacer**
- 5: **para** $j \leftarrow 0$ a $size(P_i^t)$ **hacer**
- 6: $v_i^{(t+1)} \leftarrow w * v_{ij}^t + c_1 * rand_1 * (PBest_{ij}^t - P_{ij}^t) + c_2 * rand_2 * (GBest_{ij}^t - P_{ij}^t)$
- 7: $P_{ij}^{t+1} \leftarrow ActualizarPosicion(P_{ij}^{t+1}, j, v_{ij}^{t+1}, Table, MinSup, MinCof)$
- 8: **fin para**
- 9: **si** $Fitness(P_i^{t+1}) > Fitness(PBest_i)$ **entonces**
- 10: $PBest_i \leftarrow P_i^{t+1}$
- 11: **fin si**
- 12: **si** $Fitness(P_i^{t+1}) > Fitness(GBest_i)$ **entonces**
- 13: $GBest_i \leftarrow P_i^{t+1}$
- 14: **fin si**
- 15: **fin para**
- 16: **fin mientras**
- 17: **devolver** $GBests$

El algoritmo PSO-GES tiene como parámetros de entrada el número de partículas (N), el cual evolucionará durante varias iteraciones (T), requiere el número de (M) reglas que se desean encontrar, así como los umbrales de soporte ($MinSup$) y confianza ($MinCof$), el conjunto de datos (D) a minar y el número de particiones (K) en que se dividirá el conjunto de datos. En PSO-GES, cada partícula (solución candidata al problema de estudio) está compuesta por 3 vectores y 2 valores de *fitness*.

El primer paso para PSO-GES es la generación de una Tabla de Resumen (Algoritmo 2.3, línea 1) con un número de filas igual a K , a partir del conjunto de datos D . Luego se genera aleatoriamente la población inicial con un

tamaño N (Algoritmo 2.3, línea 2). Durante un número de iteraciones T (Algoritmo 2.3, línea 3), para cada partícula P_i^t de la población de tamaño N (Algoritmo 2.3, línea 4), se actualiza su posición P_i^{t+1} (Algoritmo 2.3, línea 7) y en seguida se actualiza su *fitness*. Si el nuevo *fitness* es el mejor encontrado hasta el momento por la partícula P_i (Algoritmo 2.3, línea 8), tanto $PBest_i$ como su *fitness* ($fitnessPBest_i$) se actualizan (Algoritmo 2.3, línea 9). De igual manera, si el nuevo *fitness* es la mejor solución global encontrada hasta el momento por la partícula P_i (Algoritmo 2.3, línea 10), tanto $GBest_i$ como su *fitness* ($fitnessGBest_i$) se actualizan (Algoritmo 2.3, línea 11) [31].

2.3.5. Métricas de calidad para RA

La calidad de las RA se evalúa mediante una serie de índices que expresan el grado de incertidumbre de las mismas, por lo que en la literatura se encuentran varias métricas algunas de las más utilizadas son las siguientes [32]:

- **Soporte o Cobertura (*Support*):** En un conjunto de transacciones T , es la fracción de transacciones de T que contienen los *ítems* de $X \cup Y$, es decir la probabilidad (P) con que aparece un conjunto de *ítems* involucrados en una serie de transacciones. Por lo que el soporte esta definido por la ecuación 2.1.

$$soporte(X \Rightarrow Y) = P(X \cup Y) \quad (2.1)$$

Considerando los datos de la tabla 2.1, se obtienen la RA {cerveza \Rightarrow pañales} donde si se tiene pañales y cerveza en la transacción 2, 3 y 4, se tiene 3 veces de las 5 transacciones, por lo que el soporte sería de la siguiente manera:

$$soporte(\{Magnitud 7.2 \Rightarrow Daño estructural\}) = \frac{3}{5} = 0.6 \text{ (60\%)}$$

- **Confianza (*Confidence*):** Fracción de las transacciones en las que aparece X que también incluyen a Y ; esto es, la confianza mide con qué frecuencia aparece Y en las transacciones que incluyen X . El soporte es calculado mediante la ecuación 2.2.

$$confianza(X \Rightarrow Y) = \frac{soporte(X \cup Y)}{soporte(X)} \quad (2.2)$$

Siguiendo con los datos de la Tabla 2.1, y la RA {cerveza \Rightarrow pañales}, para obtener la confianza se obtiene el soporte de $X \cup Y$ (el cual se obtuvo anteriormente) y el soporte de Y donde tenemos cerveza en la transacción 2,

3, 4 y 5 por lo cual esto aparece 4 veces de las 5 transacciones, por lo que la confianza sería de la siguiente manera:

$$\text{confianza}(\{\text{Daño estructural} \Rightarrow \text{Magnitud 7.2}\}) = \frac{3/5}{4/5} = 0.75 \text{ (75 \%)}$$

- **Lift:** Es una medida utilizada para evaluar el grado de dependencia de los términos de una regla. El lift representa en qué grado Y tiende a ser frecuente cuando X ocurre, o viceversa y puede ser obtenida con la ecuación 2.3.

$$\text{lift}(X \Rightarrow Y) = \frac{\text{confianza}(X \Rightarrow Y)}{\text{soporte}(Y)} \quad (2.3)$$

Siguiendo con los datos de la tabla 2.1, y la RA {cerveza \Rightarrow pañales}, para obtener el lift se obtiene la confianza de $X \cup Y$ y el soporte de Y los cuales se han calculado anteriormente y el lift sería de la siguiente manera:

$$\text{lift}(\{\text{Daño estructural} \Rightarrow \text{Magnitud 7.2}\}) = \frac{3/4}{4/5} = 0.93 \text{ (93 \%)}$$

2.4. Análisis de Riesgo Sísmico

Los riesgos están íntimamente ligados a las actividades humanas. Una de las definiciones más aceptadas del riesgo R , se expresa como el producto de tres factores: el valor de los bienes expuestos C , la vulnerabilidad V y la probabilidad P de que ocurra un hecho potencialmente dañino para lo expuesto. El riesgo se expresa como [33]:

$$R = CVP \quad (2.4)$$

donde:

C es el valor de los bienes expuestos, tales como vidas humanas, edificios, carreteras, puertos, tuberías, etc.

V representa la vulnerabilidad, la cual es una medida que describe el nivel de afectación de un bien expuesto a la ocurrencia de un fenómeno perturbador.

P es a probabilidad de que ocurra un hecho potencialmente dañino, es decir, es la medida del peligro que enfrenta el bien expuesto.

El riesgo sísmico es la expresión de la peligrosidad sísmica en términos de pérdidas ocasionadas por las acciones sobre las construcciones, es decir, las pérdidas económicas debidas al movimiento del suelo. En base a lo anterior, un análisis de riesgo sísmico es aquel que proporciona los elementos para tomar decisiones técnicas (planeación al momento de realizar las construcciones), económicas (administración de fondos financieros para sismos). Por lo cual un análisis de riesgo sísmico es una herramienta que tratada adecuadamente puede ayudar a cuantificar la influencia del sismo en la toma de decisiones.

La gran ventaja de realizar un análisis de riesgo sísmico es poder seleccionar un nivel apropiado de movimiento del suelo con fines de diseño, que corresponda a un riesgo sísmico aceptable. Por lo que un estudio detallado del análisis sísmico debe incluir la incertidumbre en los aspectos fundamentales del análisis, con la consideración de las múltiples hipótesis que puedan darse, estos estudios paramétricos, de la influencia en el valor final de las posibles variaciones de los distintos parámetros puede proporcionar estimaciones aproximadas del riesgo, ya que incluye probabilidades subjetivas difíciles de cuantificar [34].

2.5. Estado del Arte

En la literatura existen diferentes trabajos donde se han utilizado técnicas de MD para analizar sismos y poder tomar medidas preventivas que minimicen las pérdidas que estos eventos puedan generar. En el estado del arte es posible encontrar trabajos en los que se han utilizado técnicas tanto descriptivas, como predictivas de MD. En las siguientes secciones se presenta una revisión literaria considerando ambos enfoques.

2.5.1. Análisis de riesgo sísmico con técnicas descriptivas

Dentro de los trabajos que han utilizado técnicas descriptivas para el análisis de riesgo sísmico, el primer estudio citado es el realizado por Martínez-Álvarez, 2011 [11], en el que analizó la ocurrencia de sismos, haciendo uso de dos técnicas de minería de datos: reglas de asociación cuantitativas (QAR) y regresión. Para ello se utilizó un conjunto de datos con información de 873 sismos obtenidos del catálogo del Instituto Geográfico Nacional de España¹, los cuales contienen la localización y magnitud del sismo, para demostrar que la variación junto con el tiempo transcurrido es útil a la hora de modelar sismos de diferentes magnitudes

¹IGN, <http://www.ign.es/web/ign/portal>

y dichos patrones pueden ser utilizados para futuras decisiones.

En su funcionamiento, el método QAR inicia ordenando el conjunto de datos de acuerdo al consecuente de la regla, en el cual se consideró únicamente un único atributo, magnitud del sismo. Posteriormente, para la obtención de patrones frecuentes utilizaron el algoritmo MP5 (extensión del algoritmo M5 proporcionado por Quinlan J. R.) el cual añade técnicas de valores perdidos y transformación de características de valores discretos y binarios.

En su proceso utilizó un árbol de decisión convencional con regresiones lineales en cada uno de los nodos; el cual se obtiene mediante un algoritmo de inducción clásico, y la predicción se obtiene al maximizar la reducción de la varianza en lugar de maximizar la ganancia de la información. Como resultado se obtuvieron las reglas de asociación que contienen los patrones frecuentes que preceden a la ocurrencia de un sismo, forzando a que la magnitud del sismo actual, sea el único atributo del consecuente.

Derivado del trabajo de Martínez-Álvarez [11], Florido-Navarro, 2013 [35] lo utilizó como base de una nueva metodología para descubrir patrones en series temporales de origen sísmico y su posterior aplicación en la predicción de sismos. Para ello obtuvo los datos del catálogo del Servicio Nacional de Sismología de Chile¹, en el que cada sismo está representado por: magnitud, fecha de ocurrencia y su valor *b-value* (se determinó teniendo en cuenta los 50 eventos anteriores al sismo estudiado).

Una vez obteniendo el conjunto de datos, estos son agrupados en orden cronológico para así encontrar patrones en los datos que preceden a la aparición de sismos con magnitud mayor o igual a 4.5. Obtenidos los patrones frecuentes, con el algoritmo *K-medias* se realizó el agrupamiento de los datos y en los análisis posteriores cada muestra es considerada exclusivamente por la etiqueta asignada por el grupo al que pertenece, con la intención de buscar secuencias específicas de etiquetas como patrones precursores de sismos con magnitud mayor o igual a 4.5.

Otra propuesta es la realizada por Martínez-Ballesteros et al., 2009 [36] quienes tuvieron como objetivo el descubrimiento de reglas de asociación basadas en técnicas evolutivas, con el fin de obtener relaciones entre series de tiempo sísmicas (sismos ocurridos durante un determinado tiempo). Para ello, utilizaron un algoritmo genético (AG), el cual determina los intervalos que forman las reglas sin

¹SNSC, <http://www.sismologia.cl/>

2. MARCO TEÓRICO Y ESTADO DEL ARTE

discretizar los atributos y permitiendo la superposición de las regiones cubiertas por las reglas.

Los resultados obtenidos con el algoritmo genético, fueron reglas de asociación que muestran la variación de las series de tiempo para los sismos. Estas reglas fueron también de mejor calidad en comparación con el algoritmo Apriori, las obtenidas con el AG tuvieron una aptitud mayor al 70 %, en tanto que las obtenidas con el algoritmo Apriori, tuvieron una confianza del 50 %. Con estos resultados, los autores concluyen que utilizando el algoritmo genético, se puede tener una mejor descripción del comportamiento de las series temporales sísmicas en los sismos y, de este modo entender el comportamiento de estos.

Otro caso de estudio se realizó en México en el año 2011 por Mendoza, J. D. [37] quien hizo una caracterización de las actividades sísmicas de la República Mexicana. Para este fin se utilizó un conjunto de datos de 9,142 sismos obtenidos del catálogo de Servicios Sismológicos Nacionales ¹.

En su estudio Mendoza, J. D., aplicó el algoritmo de clasificación KnowledgeFlow, el cual permite ver en modo texto una estructura de árbol de decisión estándar con funciones de regresión logística en las hojas. De forma adicional utilizó el algoritmo no supervisado *K-medias*, del cual se obtuvieron 5 grupos de donde se extrajo la información de la latitud y longitud del sismo, ubicando el lugar con mayor magnitud en la República Mexicana en el estado de Guerrero, cerca de El Monteal, Chapuluapan y El Veladero.

Por otro lado, Pita Martín A., 2010 [38], utilizó una metaheurística para la extracción de reglas de asociación, aplicada a sismos. Para ello se recopilaron los datos del catálogo del Instituto Geográfico Nacional ², que contiene información sobre la localización y magnitud de sismos en España, obteniendo 873 datos de sismos entre los años 1981 y 2008.

La propuesta se basa en la obtención del *k-ésimo mayor* y *k-ésimo menor* elemento para cada una de las transacciones. Para ello el algoritmo se divide en tres partes:

1. Integración de un formulario donde se indica el número de atributos con los que se generarán las reglas de asociación, así como un rango permisible de consecuentes, con los que se generan las reglas.

¹SSN, <http://www.ssn.unam.mx/>

²IGN, <http://www.ign.es/web/ign/portal>

2. Generación aleatoria de un número de reglas a partir de los datos.
3. Medición de la calidad de la regla, guardando las reglas con mejor puntuación para el lift y las reglas con la mejor puntuación para la media.

Como resultado, con las reglas de asociación obtenidas concluyeron que antes de que ocurra un sismo de magnitud elevada hay grandes decrementos del valor del antecedente, por lo que a cambios incrementales se tendría una elevada probabilidad de que ocurra un sismo de magnitud 4.4.

Mamani-Coaquira et al., 2020 [39] realizaron un estudio que tiene como objetivo la búsqueda de reglas de asociación para predecir la ocurrencia de sismos y patrones que permitan descubrir fenómenos sísmicos.

Mamani-Coaquira et al., utilizaron datos de sismos y datos de placas tectónicas cercanas a América del Sur. El conjunto de datos fue obtenido del repositorio ANSS (Advanced National Seismic System) ¹, el cual contenía datos de sismos ocurridos del 01/01/2000 hasta el 31/12/2009 con magnitud mayor o igual a 3 según la escala de Richter, con un total de 177,227 registros. Las variables seleccionadas para el estudio fueron: fecha, latitud, longitud y magnitud.

Para la generación de reglas de asociación utilizaron el lenguaje R con las librerías: `arulesViz` y `arules`, las cuales permiten utilizar las funciones de patrones frecuentes y reglas de asociación mediante el algoritmo Apriori. Para probar la calidad de las reglas de asociación se determinaron las medidas de soporte y confianza con valor mayor o igual a 50 %.

Las reglas obtenidas por el algoritmo Apriori fueron 458, entre los años 2000 y 2009 con magnitudes de los sismos mayores o igual a 5 y con los parámetros de soporte mayor o igual a 0.50, confianza mayor o igual a 0.50 y lift mayor o igual a 1. Concluyendo que el patrón que ocurre frecuentemente es: que si ocurre un sismo de magnitud mayor o igual a 5 en la placa Sudamericana y dentro de las 94 horas vuelve a ocurrir con magnitud mayor o igual a 5 entonces significa que dentro de las 120 horas volverá a ocurrir otro sismo con magnitud mayor o igual a 5.

Aqdas Ikram y Usman Qamar, 2015 [40] realizó una investigación que tuvo el objetivo de implementar un sistema experto para predecir sismos a partir de datos históricos de 1972 a 2013.

¹<http://ncedc.org/anass/catalog-search.html>

2. MARCO TEÓRICO Y ESTADO DEL ARTE

La base de datos de sismos que utilizaron Aqdas Ikram y Usman Qamar, se encuentra en el catálogo del USGS ¹, teniendo atributos como: magnitud, origen (latitud y longitud) y profundidad, además de 777,245 registros de sismos. Una vez adquiridos los datos, se realizó la generación de reglas de asociación a partir de dos pasos: en el primer paso se crean las reglas mediante minado de patrones frecuentes utilizando el algoritmo Apriori y FP-Growth, mientras que el segundo paso se validan las reglas mediante el minado inferencial a través de la lógica de predicados.

Para la generación de las reglas de asociación, utilizaron FG-Growth en Rapid-Miner ©, las reglas generadas tuvieron un soporte mínimo de 0.3 y una confianza mínima de 0.5 y un aumento mayor que 1, lo cual dio como resultado 80 reglas. Una vez obtenidas las reglas de asociación, para obtener distintas interpretaciones lógicas de las reglas generadas, Aqdas Ikram y Usman Qamar aplicaron la lógica de predicados. Siguiendo los procedimientos de la lógica de predicados se asegura que el valor de verdad de las premisas se mantenga en la conclusión. En la lógica de predicados se aplicó el Conocimiento de Dominio (DK) que se refiere al conocimiento de la tierra y las geociencias.

Una vez terminada la interpretación de las reglas, estas fueron pasadas a un sistema experto, teniendo como resultado que el sistema pudiera predecir todos los sismos que realmente ocurrieron dentro de las 12 hrs como máximo.

Otro estudio es el realizado en 2016 por Nivedhitha U. et al., [12] que utilizan dos técnicas de MD, las Reglas de Asociación y la agrupación; para esto los datos utilizados fueron obtenidos del repositorio del Servicio Geológico de los Estados Unidos ² considerando aquellos con una profundidad de la tierra (donde se libera la energía en un sismo) mínima de 100 km dejando así un conjunto de 235 eventos sísmicos en el periodo de 1990 a 2013.

Para la etapa del agrupamiento Nivedhitha U. et al. utilizaron el algoritmo *K-medias*, de los cuales se extraen los patrones frecuentes y para la obtención de las reglas de asociación aplica el algoritmo Apriori-TFP. Una vez implementados dichos algoritmos se obtienen como resultados Reglas de Asociación que tienen un valor de confianza por encima del umbral mínimo del 70 %, para la realización de predicciones a largo plazo, las cuales puede ayudar tomar las precauciones necesarias, así como también ayudar a los sismólogos para un análisis de su comportamiento.

¹<http://earthquake.usgs.gov/earthquakes/>

²USGS, <https://www.usgs.gov/>

Lorenzo Diana et al., 2019 [41] quienes plantea el uso del aprendizaje de reglas de asociación (ARL) para atribuir clases tipológicas para evaluación sísmica a edificios. Para la investigación Lorenzo Diana et al. utilizaron un conjunto de datos del Registro Federal de Edificios y Viviendas (RegBL), con datos de 700 edificios de Basilea en Suiza, donde se recopilan cuatro atributos como: periodo de construcción, número de historias de sismos, superficie de la huella del edificio y la forma del techo. Estos atributos fueron considerados por cumplir dos criterios: tener influencia en el comportamiento estructural de los edificios y ser recuperables de las bases de datos de toda la ciudad.

Para la generación de las reglas de asociación, Lorenzo Diana et al., utilizaron el método ARL el cual determina relaciones entre variables en grandes bases de datos. Como primer paso para la aplicación del método ARL y para permitir un paso de validación posterior, el conjunto de aprendizaje se dividió en dos partes. La primera parte que fue un conjunto con el 30 % de los edificios seleccionados al azar que forman el conjunto de aprendizaje. El conjunto de formación se utilizó para definir correlaciones entre los atributos de los edificios y las clases tipológicas. El 70 % restante de los edificios seleccionados se utilizó para evaluar la precisión de la atribución de clase utilizando las correlaciones que se definen en el conjunto de formación.

Finalmente para el proceso de minería de datos que define el método ARL, el cual consiste en generar una matriz de distribución que contiene la probabilidad de que un edificio definido por una combinación de los atributos seleccionados, pertenezca a cada clase tipológica. Una vez que la matriz de distribución se deriva del conjunto de formación, la combinación de atributos permite asignar a cada edificio una clase tipológica.

Como resultado se obtuvieron reglas de asociación basadas en el método ARL, con varias combinaciones de los atributos: períodos de construcción, número de pisos, forma del techo y superficie de la huella. Los cuales permiten decir que las clases de construcción se pueden asignar a todos los edificios considerando solo tres atributos que están fácilmente disponibles en las bases de datos públicas (período de construcción, número de pisos, forma del techo). La reducción en términos de la demanda de tiempo para la preparación de escenarios confiables de vulnerabilidad sísmica a escala de ciudad es particularmente pronunciada y solo reduce ligeramente la precisión ($< 5\%$ de error).

Otra investigación es la realizada por Riedel Ismael et al., 2015 [42], quienes propiciaron una forma de estimación para la evaluación de la vulnerabilidad

2. MARCO TEÓRICO Y ESTADO DEL ARTE

sísmica en entornos urbanos, utilizando datos de construcciones, mediante técnicas de minería de datos. El conjunto de datos utilizado es de edificios existentes en Grenoble, con una clasificación de vulnerabilidad EMS98 (escala de intensidad microsísmica).

La primera técnica utilizada para la evaluación de la vulnerabilidad son las reglas de asociación, las cuales fueron obtenidas mediante el método ARL, las cuales ayudan a revelar las relaciones entre datos aparentemente no relacionados en una base de datos relacional u otro depósito de información. Mediante el método ARL, Riedel Ismael et al., realizaron un proxy de vulnerabilidad, teniendo como atributos en el antecedente la clase de vulnerabilidad y en el consecuente la información estructural (pisos y periodo de construcción). Teniendo como resultados 5 clases de vulnerabilidad en el conjunto de Grenoble, por lo que esto sugiere que predice más edificios de las clases más vulnerables y menos de las clases menos vulnerables.

Otro método utilizado para la estimación de vulnerabilidad fue el aprendizaje estadístico SVM, para etiquetar los edificios de acuerdo con el estándar EMS98 deseado para las clases de vulnerabilidad sísmica. Este método se divide en dos fases: la primera que es el aprendizaje, donde todo el conjunto de datos se divide en dos, en el cual el conjunto de entrenamiento se selecciona aleatoriamente y la segunda fase que es aplicación al conjunto de datos de Grenoble, donde se encuentra una distribución geolocalizada de clases de vulnerabilidad en los atributos de vulnerabilidad.

Otro trabajo que realizó la evaluación de la vulnerabilidad es el de Abdelheq Guettiche et al., 2017 [43] los cuales realizaron una evaluación de vulnerabilidad sísmica de la ciudad de Constantine (Argelia), utilizando los métodos Risk-UE y basados en minería de datos. El conjunto de datos utilizado consta con la información de 2,252 edificios de la ciudad de Constantine, teniendo características como: diseño y construcción de los edificios, número de pisos, año de construcción.

Para la realización de la evaluación de vulnerabilidad, Abdelheq Guettiche et al., primero usaron el método Risk-UE, el cual clasifica cada edificio en la tipología definida por los materiales y/o sistemas estructurales, una vez teniendo esta clasificación se realizó un proxy de vulnerabilidad sísmica, que es un método basado en minería de datos, este método es el aprendizaje de reglas de asociación (ARL).

Los resultados obtenidos indican que la calidad de una estimación alcanza una asíntota más allá de un tamaño de subconjunto de aprendizaje que representa el

30 % de los datos totales, por lo tanto, seleccionaron 2500 datos al azar, cada uno de los cuales representa el 30 % conjunto de datos completos. Por lo que se descubrieron relaciones “ocultas” entre los atributos elementales de los edificios y la vulnerabilidad sísmica y puede brindar una evaluación general del riesgo sísmico en un área urbana.

Yaohui Liu et al., 2019 [44] presentaron un enfoque integrado para una evaluación de la vulnerabilidad macrosísmica para Urumqi (China), para la esto se utilizaron métodos de minería de datos. La base de datos utilizada en su investigación se compone de dos partes: la base de datos Urumqi y los datos in situ del distrito de Tianshan. La base de datos de Urumqi se refiere a la base de datos de la fundación de emergencia de sismos de Urumqi, que comprende información social, económica y de población, además de un inventario de edificios y un mapa de la ciudad, así como accidentes geográficos naturales, ubicaciones de objetos clave, información del equipo de rescate, datos de comunicación de socorro y sismos planificación previa de datos. Mientras que la base de datos de un situ, contiene características como: el tipo de material, el número de pisos, el período de construcción y el tipo de techo.

Para la evaluación de la vulnerabilidad se utilizo el método ARL, simplificado a la base de datos in situ Urumqi. Las probabilidades condicionales entre las clases de vulnerabilidad EMS-98 y la información básica sobre los edificios se derivan para obtener el proxy de vulnerabilidad in situ. Por lo que Yaohui Liu et al., usando dos atributos (el período de construcción y el número de pisos) y el método ARL con 616 edificios de la base de datos in situ, obtuvieron el proxy de vulnerabilidad ARL, teniendo como resultado que los edificios construidos antes de 1960 con menos de dos pisos pertenecen a la Clase A. No se construyeron edificios antes de 1960 con tres o más pisos. Este fenómeno corresponde a la historia de la planificación y construcción urbanística de Urumqi.

Por lo que se han concluido que las ARL, para la extracción de datos se aplico con éxito a la base de datos de Urumqi y lograron precisiones y estabildades deseables para la estimación de la vulnerabilidad sísmica.

2.5.2. Análisis de riesgo sísmico con técnicas predictivas

No solo se han utilizado técnicas descriptivas de minería de datos, si no que también se han utilizado las ya mencionas técnicas predictivas para el análisis de los sismos, como lo hizo Morales-Esteban A. et al. [45] en 2010, donde utilizó técnicas de agrupamiento para obtener patrones que modelen el comportamiento

2. MARCO TEÓRICO Y ESTADO DEL ARTE

de los datos temporales sísmicos y puedan ayudar a predecir sismos de tamaño mediano a grande. El conjunto de datos que utilizaron fue del catálogo del Instituto Geográfico Español (SGI), el cual contiene la ubicación (latitud y longitud) y magnitud del sismo, adicional a estos parámetros, se calculó el $b - value$ de la ley de Gutenberg-Richter la cual es un importante parámetro, ya que refleja las propiedades tectónicas y geofísicas de las rocas y las variaciones de presión de los fluidos en la región en cuestión.

Mediante el uso del algoritmo $k-medias$ Morales-Esteban A. et al., realizaron la agrupación, donde obtuvieron que el número óptimo de agrupaciones es tres ya que el índice de precisión alcanza el valor máximo para las tres agrupaciones, en las áreas estudiadas. Los valores máximos que utilizaron fueron iguales a 0.69 y 0.75. Con referencia al $b - value$, los grupos 1, 2 y 3 se caracterizan como: el grupo 1 presenta una disminución del $b - value$, el grupo 2 un incremento cercano a cero, y finalmente el grupo 3 presenta un aumento del $b - value$. Por lo que en resumen, todos los sismos con magnitud mayor o igual a 4.5 se han clasificado en el grupo 1 y se caracterizan por un incremento del $b - value$ negativo.

Los resultados obtenidos por Morales-Esteban A. et al., para pronosticar sismos con magnitud mayor o igual a 4.5, proporcionan una precisión del 70-90 %, ya que todos los sismos medianos-grandes se han caracterizado por una disminución del $b - value$. Por lo que, este parámetro puede considerarse un precursor sísmico de los datos sísmicos españoles.

Galán Montaña F., 2013 [13] se enfocó en sismos de Japón utilizando un conjunto de datos de 41,553 registros divididos en 5 zonas de acuerdo a su latitud y longitud (Zona A 730 sismos, Zona B 651 sismos, Zona C 6904 sismos, Zona D 3260 sismos y Zona E 668 sismos). El objetivo del estudio se centró en encontrar patrones frecuentes y realizar modelos del comportamiento de las series temporales que comprenden la ocurrencia de sismos de magnitud elevada mayor a 4.5. Una vez que se extraen dichos patrones, se utilizan para describir el comportamiento del sistema, dependiendo de la zona del sismo y la magnitud del mismo.

Como resultados finales concluyen que los datos obtenidos en las Zonas C y D se pueden considerar satisfactorios ya que por tener la mayor cantidad de datos se consiguen patrones frecuentes de interés teniendo descripciones que se pueden utilizar para trabajos futuros en el área predictiva.

Otro estudio realizado es el de Ascencio-Cortes G. et al., 2015 [46], donde utilizaron datos de Tokio, con el objetivo de evaluar la precisión de las redes neuronales artificiales para predecir la magnitud de los sismos en dicho lugar. El

conjunto de datos que Ascencio-Cortes G. et al. han utilizado fue obtenido del del catálogo sísmico del Servicio Geológico de los Estados Unidos (USGS), teniendo características como fecha del evento, ubicación del sismo (latitud y longitud), y la magnitud, en base a estas características obtuvieron el $b - value$. Para la predicción de la magnitud de los sismos, los autores propusieron un algoritmo (EQP-ANN) que se basa en la utilización de un perceptrón multicapa con aprendizaje de backpropagation.

Los resultados obtenidos de esta investigación se consideraron satisfactorios ya que se han alcanzado valores superiores al 70 % de precisión, para la evaluación de su efectividad, se ha comparado con otros métodos conocidos (C4.5, KNN, NB), superando a todos ellos en términos de los parámetros de precisión.

De manera general la Tabla 2.2, muestra los aspectos mas destacados de los trabajos del estado del arte revisados en esta sección, resaltando los siguientes aspectos: objetivo del estudio, los datos y los algoritmos utilizados.

A modo de resumen se puede decir que, autores que han usado reglas de asociación como Martínez-Ballesteros, Mammani-Coaquira, Nivedhitha, han utilizado el algoritmo Apriori ya que es uno de los más empleados dentro de la literatura, aplicado para diferentes áreas. Sin embargo, debido a que no siempre es el que mejores resultados proporciona se han implementado otros algoritmos, tal es el caso de Martínez-Ballesteros quien utilizó un algoritmo genético.

Pese a que la utilización del algoritmo genético mostró mejores resultados, en la actualidad existen otros algoritmos que han demostrado mayor capacidad de inferencia, como el de colonia hormigas o abejas y los basados en enjambres de partículas, como el PSO-GES el cual es implementado en esta investigación.

Finalmente, aún con la considerable cantidad de estudios que utilizan técnicas de minería de datos realizados al rededor del mundo, para el análisis de riesgo sísmico aún tienen la limitante de no considerar todas las características para dicho análisis. Sin embargo, estos estudios significan en sí mismos un gran avance en el área de ingeniería sísmica.

2. MARCO TEÓRICO Y ESTADO DEL ARTE

Tabla 2.2: Resumen Estado del Arte

	Autor (es) y Año	Objetivo	Descripción datos	Algoritmos
Método descriptivo	Martínez-Álavarez 2012	Predice la ocurrencia de terremotos bajo determinadas circunstancias, haciendo uso de dos técnicas de MD: Reglas de Asociación Cuantitativas y Regresión.	latitud, longitud magnitud <i>b - value</i>	QAR MP5
	Martínez-Ballesteros 2009	Descubrimiento de reglas de asociación basadas en técnicas evolutivas con el fin de obtener relaciones entre series de tiempo correlacionadas.	latitud, longitud magnitud	Apriori, GA
	Mendoza J. D. 2011	Técnicas de MD y Weka para identificar, analizar y extraer información relevante y obtener patrones frecuentes con la finalidad de crear un mejor criterio para la toma de decisiones.	Fecha, lugar magnitud	KnowledgeFlow K-means,
	Pita Martín A. 2012	Encontrar patrones y relaciones dentro de los datos, permitiendo la creación de modelos en los que la representación del conocimiento esté basada en reglas de asociación.	latitud, longitud magnitud <i>b - value</i>	QAR pasado en k-ésimo mayor k-ésimo menor
	Mamani-Coaquira Y. 2020	Búsqueda de reglas de asociación para predecir la ocurrencia de sismos, patrones que permitan descubrir fenómenos sísmicos.	latitud, longitud coordenadas de placas tectónicas	Apriori librería arules en R
	Aqdas Ikram y Usman Qamar 2015	Implementar un sistema experto para predecir sismos a partir de datos anteriores. Esto se logra aplicando reglas de asociación en datos desde 1972 a 2013.	latitud, longitud magnitud profundidad	FG-Growth, Apriori
	Nivedhitha U. y Krishna Anand S. 2016	Generación de un modelo de predicción para anticipar sismos mediante la aplicación de reglas de asociación y agrupamiento en datos históricos de sismos.	magnitud profundidad fecha y tiempo	Apriori-TFP K-means
	Lorenzo Diana et al. 2019	Propone la aplicación del método ARL para atribuir clases tipológicas (para evaluación sísmica) a edificios.	año construcción superficie huella historia sismos forma techo	ARL
	Ridiel Ismael et al. 2015	Proponer una forma de estimación para la evaluación de la vulnerabilidad sísmica en entornos urbanos, utilizando datos de construcciones, mediante técnicas de minería de datos	pisos, periodo de construcción	ARL y SVM
	Guettiche A. et. al 2017	Realizan una evaluación de vulnerabilidad sísmica de la ciudad de Constantine (Argelia), utilizando los métodos Risk-UE y basados en minería de datos.	diseño construcción número de pisos año de construcción	ARL y Risk-UE
Yaohui Liu et al. 2019	presentaron un enfoque integrado para una evaluación de la vulnerabilidad macrosísmica para Urumqi (China), para la esto se utilizaron métodos de minería de datos.	tipo de material número de pisos año construcción el tipo de techo	ARL	
Método predictivo	Florido-Navarro 2013	Agrupación para clasificar las muestras en grupos y obtener los patrones frecuentes de los sismos de gran magnitud.	magnitud fecha <i>b - value</i>	K-means
	Morales-Esteban et al. 2010	Técnicas de agrupamiento para obtener patrones que modelen el comportamiento de los datos temporales sísmicos y puedan ayudar a predecir sismos de tamaño mediano a grande.	magnitud <i>b - value</i> fecha	K-means
	Montaño F. 2013	Encontrar patrones frecuentes y determinar el comportamiento de series temporales, para después predecir el comportamiento del sistema de modo tan preciso utilizando clasificadores.	latitud, longitud magnitud	PF-Growth, árbol de clasificación
	Asencio-Cortés G. et al. 2015	Evalúa la precisión de las redes neuronales artificiales para predecir la magnitud de los sismos en Tokio.	fecha, magnitud longitud, latitud	EQP-ANN, árboles de decisión, KNN

Llevar a cabo el proceso de MD requiere la aplicación de una metodología estructurada para la utilización eficiente de las técnicas y herramientas disponibles.

En la literatura se pueden encontrar las siguientes metodologías más utilizadas en MD como: KDD (*Knowledge Discovery in Databases*) [3], SEMMA (el acrónimo corresponde a las fases básicas del proceso: *Sample, Explore, Modify, Model, Assess*) [47], CRISP-DM (*Cross Industry Standard Process for Data Mining*) [48] y Catalyst, conocida como P3TQ (*Product, Place, Price, Time, Quantity*) [49]. No obstante, por la adaptación al problema de estudio en esta investigación se utilizó la metodología KDD.

La metodología KDD es un proceso iterativo ya que la salida de algunas de las fases puede hacer volver a pasos anteriores, por lo que la metodología define una secuencia de cinco fases: selección, preprocesado, transformación, data mining e interpretación y evaluación como se muestra en la Figura 3.1. En esta tesis las fases de la metodología se siguieron de la siguiente manera:

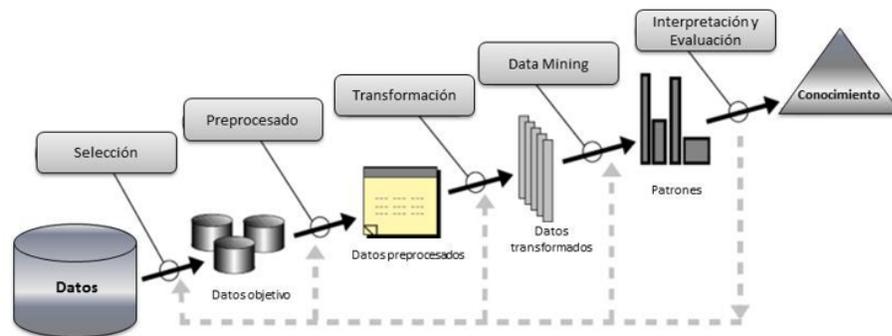


Figura 3.1: Fases de la metodología KDD [3].

3. MARCO METODOLÓGICO

3.1. Selección

En esta fase se hace la recopilación de los datos disponibles para el estudio con el fin de alcanzar los objetivos de la investigación.

El 7 y 19 de Septiembre del 2017 en la República Mexicana sucedieron dos sismos de magnitud 8.2 y 7.1 respectivamente, causando daños considerables en estructuras, por lo que después de lo ocurrido varios ingenieros civiles o arquitectos directores responsables de obra (ODR), provenientes de dependencias como el Instituto Nacional de la Infraestructura Física Educativa (INIFED), realizaron evaluaciones en los edificios afectado, obteniendo dictámenes que incluyen parte de la información necesaria para un análisis de riesgo sísmico. El concentrado de dictámenes se encuentran en una base de datos de acceso libre por medio de una página del Gobierno de la República Mexicana, la cual lleva por nombre Fondo de Transparencia Presupuestaria ¹.

La figura 3.2 es una impresión de pantalla en la que se muestra la información que el Fondo de Transparencia Presupuestaria pone a disposición del público en general para su consulta.

Tipo de información	Nombre	Tipo	Diccionario	Fuente	Estatus de actualización ⁹	Fecha de actualización
Entrega de apoyos	FONDEN (emergencias)	 		Secretaría de Hacienda y Crédito Público	Activo	22/10/2017
Entrega de apoyos	Apoyos internacionales	 		Secretaría de Relaciones Exteriores	Concluido	01/12/2017
Entrega de apoyos	Seguimiento a apoyos internacionales en especie	 		Entrega: Secretaría de Relaciones Exteriores Seguimiento: Secretaría de Gobernación	Concluido	26/04/2018
Entrega de apoyos	Comedores comunitarios	 		Secretaría de Desarrollo Social	Concluido	19/01/2018
Entrega de apoyos	Beneficiarios de comedores comunitarios	 		Secretaría de Desarrollo Social	Concluido	19/01/2018
Daños y reconstrucción	Patrimonio cultural e histórico dañado	 		Secretaría de Cultura	Activo	09/05/2018
Daños y reconstrucción	Reconstrucción de escuelas	 		Secretaría de Educación Pública	Activo	02/08/2018
Daños y reconstrucción	Apoyos a viviendas	 		Banco del Ahorro Nacional y Servicios Financieros	Activo	31/07/2018
Daños y reconstrucción	Préstamos para reconstrucción de viviendas	 		Instituto del Fondo Nacional de la Vivienda para los Trabajadores	Concluido	28/03/2018

Figura 3.2: Repositorio de las bases de datos del Fondo de Transparencia Presupuestaria

¹https://www.transparenciapresupuestaria.gob.mx/es/PTP/fuerzamexico_datosabiertos

La realización de un análisis de riesgo implica considerar variables como: número de pisos, tipo de construcción (mampostería, losa reticular, concreto, acero), fecha de construcción, refuerzo, desplome (con o sin hundimiento), irregularidad (planta, elevación), edificio en esquina, golpeteo (edificios de igual altura o distinta), daños previos y sobrepeso. Sin embargo la obtención de todas las variables antes mencionadas son difíciles de obtener, por lo que la información obtenida de la base de datos utilizada contiene información valiosa como tipo de construcción, detalle del daño, costos aproximados de reparación y tipo del daño, sin embargo esta información solo contienen algunos datos para el análisis de riesgo sísmico.

3.2. Pre-procesado

Para esta fase se realiza la valoración de cada una de las características que tiene el conjunto de datos, con el objetivo de obtener aquellas que son de utilidad, además de analizar cada uno de los atributos de cada característica.

El conjunto de datos de infraestructuras afectadas en el sector educativo obtenido, consta de 19,194 registros y 76 atributos tales como: fecha del evento, nombre de la escuela, dirección, nivel, matrícula, entidad federativa, municipio, localidad, latitud, longitud, tipo de daño, detalle del daño, fotos del daño, estatus de operación, url del dictamen, monto y costo de reparación y porcentajes de avance.

Al conjunto de datos se le realizó una validación de cada uno de los campos, para determinar las características relevantes a considerar en el análisis de riesgo sísmico. Esta validación se realizó con el apoyo de un experto en el área de estructuras quien determinó que de acuerdo a las características disponibles, las que proporcionan información relevante son: fecha de evento, matrícula, entidad federativa, latitud, longitud, tipo de daño, detalle del daño, costos totales, y url del dictamen. La determinación de los datos se realizó en base al enfoque de estudio que al experto le interesaba en esta investigación.

Dentro del conjunto de datos cada atributo está constituido por un formato y dimensión diferente como se puede observar en la imagen 3.3:

3. MARCO METODOLÓGICO

LOCALIDAD	LATITUD	LONGITUD	INFRAESTRU	NIVEL	GOBI	INSTANCIA	TIPO DAÑO	DETALLE_DA	FOTOS	DAÑO	ESTATUS	OPFECHA	VERIURL	DICTAM	COSTO	ORIC	MONTO	TOTIMONTO	EJE	DESC	OBRA	FECHA	INICI	
TECPÁTAN	17.134268	-93.307855	Información	Información	Información	2	Moderado	EDIF. "M": D	http://143.1:	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
SAN CRISTÓ	16.712909	-92.635684	SI	Estatal	PROTECCIO	2	Moderado	Muros divis	http://143.1:	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
TUKTLA GUTI	16.733122	-93.120502	SI	Estatal	INFIECH	3	Menor	Información	http://143.1:	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
MATZAM	16.79	-92.477865	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	14/10/2017	http://dgde	50000	50000	50000	1.-REPARACI	14/11/2017							
ABSALÓN CA	16.978333	-93.418611	SI	Estatal	PROTECCI	2	Moderado	Muros divis	http://143.1:	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
CHIKTONTIC	16.891506	-92.457908	SI	Estatal	PROTECCIO	2	Moderado	Información	No disponible	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
YASHANAL	16.82809	-92.45194	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	06/11/2017	http://dgde	50000	50000	50000	1.-REPARACI	09/11/2017							
OCOCH	16.838056	-92.475278	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	27/09/2017	http://dgde	50000	50000	50000	1.-REPARACI	09/11/2017							
EL PORVENI	16.855	-93.4	SI	Estatal	Gobierno Est	3	Menor	Información	http://dgde	En su inmueble	25/09/2017	http://dgde	25000	25000	25000	1.-REPARACI	18/01/2018							
ÁLVARO OBF	17.391111	-92.991111	Información	Información	Información	2	Moderado	EDIF. "A" "Y	http://143.1:	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
MIGUEL HID	17.268333	-93.448333	SI	Información	Información	2	Moderado	EDIF. "C": D	http://143.1:	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
EL COPAL	16.966111	-92.928056	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	27/09/2017	http://dgde	50000	50000	50000	1.-REPARACI	08/11/2017							
KOTOLTE	16.86444	-92.46111	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	09/11/2017	http://dgde	50000	50000	50000	1.-REPARACI	09/11/2017							
NARANJATIC	16.94944	-92.52917	SI	Estatal	Gobierno Est	3	Menor	1.-DAÑOS M	http://dgde	En su inmueble	14/09/2017	http://dgde	50000	50000	50000	1.-DAÑOS M	30/11/2017							
FRANCISCO	17.130694	-91.486431	SI	Estatal	Gobierno Est	3	Menor	1.-DAÑOS M	http://dgde	En su inmueble	15/10/2017	http://dgde	50000	50000	50000	1.-DAÑOS M	17/11/2017							
LAS MARAVI	17.134722	-92.556389	SI	Estatal	PROTECCIO	2	Moderado	Información	No disponible	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
CRUZCHÉN	16.829444	-92.471667	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	Información	http://dgde	50000	49955.84	49955.84	1.-REPARACI	09/11/2017							
PANTELHÓ	17.005692	-92.471234	Información	Información	Información	2	Moderado	Información	No disponible	En su inmueble	Información	Información	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
YOSHIB	16.846617	-92.444727	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	26/12/2017	http://dgde	75000	75000	75000	1.-REPARACI	13/12/2017							
CHIMIX DOS	17.014077	-92.525245	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	Información	http://dgde	50000	50000	50000	1.-REPARACI	16/11/2017							
MAIOSIK	16.89778	-92.49333	SI	Estatal	PROTECCIO	2	Moderado	Información	No disponible	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
YERBAUEN	16.933056	-92.865833	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	07/11/2017	http://dgde	50000	50000	50000	1.-REPARACI	07/11/2017							
TZ'QUIUILI	16.855278	-92.459722	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	04/12/2017	http://dgde	50000	50000	50000	1.-REPARACI	30/11/2017							
PUERTO RIC	16.61	-91.872778	Información	Información	Información	2	Moderado	EDIF. "A": D	http://143.1:	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
BAIXULUM	16.9	-92.60889	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	27/11/2017	http://dgde	50000	50000	50000	1.-REPARACI	07/11/2017							
CHENALHÓ	16.892434	-92.629841	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	Información	http://dgde	50000	50000	50000	1.-REPARACI	07/11/2017							
CHIQUIUITE	15.087134	-92.09355	Información	Información	Información	2	Moderado	Información	No disponible	En su inmueble	Información	Información	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
PAJALÓN A	16.829167	-92.572222	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	09/11/2017	http://dgde	50000	50000	50000	1.-REPARACI	08/11/2017							
VELASCO SU	17.074093	-93.746599	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	29/09/2017	http://dgde	25000	25000	25000	1.-REPARACI	15/11/2017							
PAVENCUL	15.201389	-92.179722	SI	Información	Información	2	Moderado	Información	No disponible	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
LLANO GRAN	17.032778	-92.981111	SI	Estatal	Gobierno Est	3	Menor	1.-REPARACI	http://dgde	En su inmueble	Información	http://dgde	50000	50000	50000	1.-REPARACI	07/11/2017							
CHANA	16.885	-92.4775	SI	Estatal	PROTECCIO	2	Moderado	Información	No disponible	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica
CHULIÁ	16.86611	-92.48667	SI	Estatal	PROTECCIO	2	Moderado	Información	No disponible	En su inmueble	Información	http://143.1:	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica	No aplica

Figura 3.3: Muestra de conjunto de datos iniciales [4].

Como puede verse, atributos como latitud y longitud son campos de tipo numérico, con una dimensión máxima de 10 caracteres, la fecha del evento el formato es de tipo fecha con una estructura de dd/mm/yyyy, en la cual solo incluyen fechas correspondientes al 7 y 19 de Septiembre del 2017, para el atributo de entidad el formato es una cadena de texto, el tipo de daño esta definido por un formato de cadena de texto en el cual solo se pueden esperar tres valores establecidos (menor, moderado y grave), el atributo url direcciona al sitio donde se encuentra la ubicación del dictamen el cual contiene la información estructural de los daños en los edificios educativos. Los formatos para descargar se encontraron en pdf, png, jpg, word (Ver anexo A, B, C).

Un atributo importante para el análisis de riesgo sísmico es el costo de los daños ocasionados por el sismo tales como: costo de reparación, costo de muertes o lesionados, costo de operatividad entre otros. No obstante dentro del conjunto de datos utilizado únicamente se encuentran costos aproximados de reparación, los cuales se encontraron en formato numérico y cadena de texto, estos se debe a que si existía alguna cantidad se colocaba el costo aproximado mientras que si no se genero ningún costo de reparación se colocaba la leyenda de “No aplica”, si el dato aún no estaba disponible se manejaba la leyenda de “Información en construcción”.

Por último, para el atributo de detalles del daño se encontró con un tipo de dato en cadena de texto en donde cada descripción del daño estaba separada por una coma o por un punto y coma, o bien todo escrito en un mismo párrafo. Sin

embargo, cada descripción fue un formato libre donde se encontraron distintas formas de referirse a algo, por ejemplo se encontraron enunciados como “Daños en pisos”, “fisuras en piso”, “grietas en pisos”.

3.3. Transformación

En esta fase se hace la unificación del tipo de datos a las características que aplican, además de completar los datos estructurales obtenidos de dictámenes, y la formación de categorías para cada característica del conjunto de datos a utilizar, la cual se explica más detalladamente a continuación.

Para la generación de RA el conjunto de datos debe tener un formato uniforme para todos los atributos, sin embargo como se explicó en la sección anterior cada uno de ellos está conformado por formatos y dimensiones diferentes. Además también se tuvo que recopilar información faltante que se encontraba en los dictámenes proporcionados a fin de obtener los datos estructurales importantes para la realización del análisis de riesgo sísmico. Por ello, se realizaron las siguientes tareas para obtener un conjunto de datos con formato uniforme.

3.3.1. Unificación del tipo de dato

La unificación del tipo de dato fue necesaria para poder manipular más fácilmente los datos, por lo que se realizó la unificación con los atributos: “monto total INIFED”, “monto total seguro”, “monto total FONDEN”, “monto total escuela”, “detalle del daño”. Para el caso del atributo del “costo total”, se tenía el registro de diferentes instituciones que se encargan de cubrir una cantidad de reconstrucción, por tal motivo la suma de cada institución daba el “costo total” de reparación. Sin embargo, para dicho atributo no todos los registros estaban en formato numérico sino que se tenían cadenas de texto por lo que para realizar la suma total de estos valores se tuvieron que unificar todos los datos a un formato numérico. La Tabla 3.1 muestra tres atributos de costo, donde tiene valores numéricos como 7500 y valores tipo texto como “No aplica”.

Tabla 3.1: Ejemplo de datos iniciales

Costo_total_FONDEN	Costo_total_INIFED	Costo_total_seguro
75000	75000	75000
No aplica	No aplica	No aplica
50000	50000	No aplica
69276.74	Información en construcción	Información en construcción

3. MARCO METODOLÓGICO

Una vez observado lo anterior se realizó la programación de un proceso que se encargó de leer todo el archivo y cambiar todos aquellos datos con la leyenda “No aplica” por un valor numérico 0, mientras que para la leyenda “Información en construcción” se asignó un valor numérico el cual representará que es información no disponible por el momento, pero que en cualquier momento puede ser actualizada debido a que el proceso sigue activo, por lo que se asignó un valor -1. Una vez terminado el proceso se obtuvo un resultado con los datos en formato numérico y la suma total de todos los costos como se observa en la Tabla 3.2.

Tabla 3.2: Ejemplo de datos finales

Costo_total.FONDEN	Costo_total.INIFED	Costo_total.seguro	Costos_totales
75000	75000	75000	225000
0	0	0	0
50000	50000	0	100000
69276.74	-1	-1	69274.74

Para los atributos de detalles del daño al ser un tipo de dato cadena de texto el tratamiento es diferente, el primer paso consistió en la separación de palabras para obtener las más comunes en las descripciones, quedando con aquellas palabras importantes como colapso, fisuras, grietas, paredes, pisos, columnas, entre otras. Terminado el paso anterior el siguiente paso fue la detección de las palabras con significados parecidos (piso, pisos) unificando y pasando esas palabras a un solo término para evitar la duplicidad de detalles.

3.3.2. Complementación de datos

Como se ha mencionado anteriormente para la realización del análisis de riesgos sísmico se requiere de características estructurales, las cuales se encontraban dentro de los dictámenes almacenados en un repositorio, por lo que sin estos datos estructurales el análisis de riesgo no se podría realizar. Para la obtención de datos estructurales se realizaron las siguientes etapas:

1. **Obtención de dictamen:** La primera etapa fue realizar la descarga de dictámenes de la URL proporcionada en el conjunto de datos, cada archivo podía estar en un formato pdf, jpg, png y doc.
2. **Extracción de características:** La segunda etapa consistió en obtener las características estructurales de cada uno de los dictámenes descargados ya que se obtuvieron más de un tipo de formato diferente. Por ejemplo, el Anexo A muestra un tipo de formato de dictamen, dentro de éste los detalles estructurales están dentro del recuadro rojo. Existen otros formatos

de dictámenes tales como se muestra en el Anexo B y C, en los cuales los detalles estructurales se encuentran en la descripción de los reportes o de manera general por categorías ya establecidas por alguna institución encargada de la recolección de los daños.

Identificado cada formato de dictamen se hizo uso de una herramienta en línea de uso libre llamada online OCR¹, la cual permitió realizar la conversión de la imagen con los detalles estructurales a un formato tipo texto (es decir transformar una imagen en un archivo “.txt”), el uso de esta herramienta se debió a que se tenía que pasar la información a un tipo texto para ser utilizada de manera más sencilla para el proceso de minado de reglas de asociación y patrones frecuentes.

3. **Integración:** La tercera etapa consistió en realizar la programación de una rutina que integrará los datos estructurales obtenidos por la herramienta antes mencionada al conjunto de datos completo.
4. **Validación de datos:** La última etapa consistió en la validación de los datos estructurales agregados, este paso fue requerido debido a que la calidad de algunos formatos descargados era poco visible por lo que la herramienta online OCR, no generó la información con los datos estructurales requeridos.

Por tal motivo si la recopilación de los datos estructurales no era la esperada derivado de la mala calidad de los documentos descargados, se realizó la obtención de los datos de forma manual, es decir que se abrió cada dictamen de manera individual y se agregaron los campos de forma directa al archivo, sin ningún proceso automático. Este proceso, se realizó para un 13% del total de los dictámenes descargados.

Otro dato importante es la distancia epicentral, para lo cual se tomaron los epicentros de los sismos reportados por el Servicio Sismológico Nacional-UNAM [50]. Para el sismo del 7 de Septiembre tuvo coordenadas de latitud 14.761 y longitud -98.72, mientras que para el sismo del 19 de Septiembre las coordenadas fueron de latitud 18.40 y longitud -98.7. Teniendo la información correspondiente de las coordenadas de las escuelas y las coordenadas de los epicentros, se utilizó la fórmula de Haversine (Ecuación 3.1) la cual se emplea para obtener la aproximación esférica de la distancia (d) entre dos puntos de la superficie terrestre (donde ϕ_2, ϕ_1 y γ_2, γ_1 representan las latitudes y longitudes y r es el radio terrestre (6371) [51]).

¹<https://www.onlineocr.net/es/>

3. MARCO METODOLÓGICO

$$d = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\phi_2 - \phi_1}{2} \right) + \cos(\phi_1) \cos(\phi_2) \sin^2 \left(\frac{\gamma_2 - \gamma_1}{2} \right)} \right) \quad (3.1)$$

Por ejemplo considerando la ubicación del sismo del 7 de Septiembre con latitud 14.761 y longitud -98.78 y tomando la ubicación de una escuela la cual tiene una latitud de 17.134268 y longitud -93.307855 se calculó la distancia epicentral como se muestra a continuación:

$$d = 2(6371) \arcsin \left(\sqrt{\sin^2 \left(\frac{17.134268 - 18.40}{2} \right) + \cos(18.40) \cos(17.134268) \sin^2 \left(\frac{(-93.307855) - (-98.78)}{2} \right)} \right) = 277.2484Km$$

La formula anterior fue utilizada para la obtención de la distancia epicentral de cada escuela del conjunto de datos utilizado.

Por último, se utilizó el municipio y de este modo se obtuvo la zona sísmica mediante la clasificación de municipios ¹, según el grado de peligro al que están expuestos.

3.3.3. Categorización

En este paso se obtuvieron las categorías para los atributos de fecha de evento, distancia epicentral, monto total del daño, detalle del daño, entidad, tipo de daño, zona sísmica. Para la entidad se tomaron 11 entidades las cuales fueron las que tuvieron daños de los sismos: Chiapas, Ciudad de México, Guerrero, Hidalgo, Estado de México, Tlaxcala, Morelos, Oaxaca, Puebla, Veracruz de Ignacio de la Llave y Michoacán de Ocampo.

Para la fecha de evento únicamente se tomaron dos categorías y esta dada por la fecha del evento que es 07/09/2017 y 19/09/2017. Las categorías para el tipo de daño se determinaron en menor, moderado y grave, estas categorías se tomaron en base a los daños que el sismo ocasiono.

Las categorías para la zona sísmica se han determinado por el manual de la CFE, donde la República Mexicana se divide en cuatro zonas sísmicas (Figura 3.4) [5]: Zona A es una zona donde no se tienen registros históricos de sismos, es decir que no se han reportado sismos en los últimos 80 años y no se esperan aceleraciones del suelo mayores a un 10%, Zona B y C son aquellas zonas intermedias, donde se registran sismos no tan frecuentemente es decir que son zonas afectadas por aceleraciones altas pero no mayores a un 70% de la aceleración

¹http://www.cenapred.unam.mx/es/Transparencia/FAQ/SISI/Anexo7.2/00_07_CI.RG_30052000.pdf

del suelo y Zona D la cual es una zona donde se han reportado grandes sismos históricos, esto es debido a que la ocurrencia de sismos es muy frecuente y las aceleraciones del suelo pueden sobrepasar el 70 % de la aceleración de la gravedad.

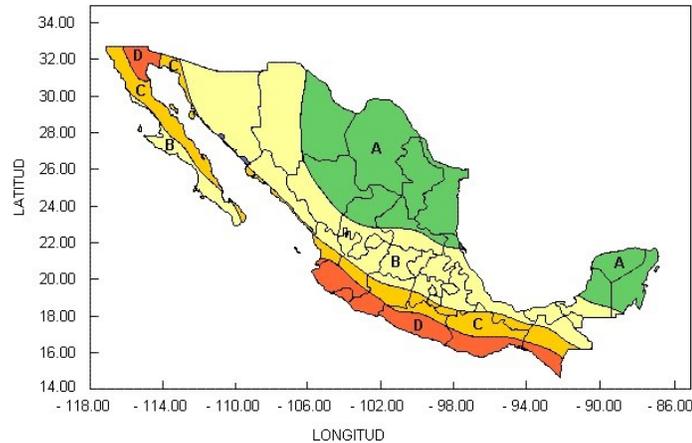


Figura 3.4: Regionalización sísmica de la República Mexicana [5]

La categorización para los detalles del daño se realizó por rubros específicos, para lo cual solo se obtuvieron 7 categorías, que engloban información descriptiva del detalle del daño como se puede observar en la tabla 3.3.

Por último, para la categorización de la distancia epicentral y los costos totales se obtuvieron sus respectivas tablas de frecuencia, obteniendo los rangos para cada uno de los rubros. Las categorías obtenidas para la distancia epicentral fue de 16, donde para cada intervalo de distancias es de 50 km, teniendo como mínima distancia 1 km y máxima distancia 800 km. Para los costos totales se obtuvieron 15 rangos donde cada intervalo es de 4080401 pesos teniendo un valor mínimo de \$0 y un valor máximo de \$57,125,601, estos rangos fueron considerados para que no todos los valores de costos de reconstrucción estuvieran concentrados dentro de un solo intervalo.

3.3.4. Discretización del conjunto de datos

Teniendo todas las categorías definidas, el siguiente paso fue la obtención de un conjunto de datos representado en formato binario a través de vectores horizontales donde 1 denota la presencia del *ítem* en la transacción, mientras que 0 denota la ausencia del *ítem*, tal como se muestra en el ejemplo de la Tabla 3.4.

Tomando en cuenta la tabla anterior se puede decir a modo de ejemplo que 1 corresponde a daños en instalaciones hidráulicas, 2 daños en instalaciones eléctri-

3. MARCO METODOLÓGICO

Tabla 3.3: Categorías detalles del daño

Instalación hidráulica y sanitaria	Instalación Eléctrica	Equipos	Mobiliario	Acabados	Estructuras	Exteriores
baños	alumbrado	aparatos	anaqueles	antipechos	azotea	fachada
bombas	apagadores	computadora	archivero	azulejo	bardas	impermeabilización
cisterna	eléctrica	cañón	banca	barandales	castillos	patio
drenaje	luminarias	copiadora	butacas	resane	colapso	jardín
hidráulicas	poste	bafe	sillas	restauración	columnas	gradas
hidrosanitarias	lámparas	bocina	escritorio	rotulación	cuarteaduras	banqueta
inodoro		equipo	gabinetes	cancelería	rupturas	perimetral
lavabos		televisión	librero	cerradura	tabique	
tinacos		grabadora	mesas	crystalería	demolición	
tubería		impresora	mobiliario	cubierta	derrumbe	
plomaría		micrófonos	pizarrones	desprendimiento	escaleras	
		pantalla		entortado	estructural	
		ventilador		enyesado	fractura	
		proyector		fisuras	techos	
				tablaroca	grietas	
				herrajes	infraestructura	
				herrería	trabes	
				lamina	losas	
				loseta	menores	
				marco	multipanel	
				ventanas	muros	
				vidrios	pasillos	
				pintura	pisos	
				plafones	rampa	
				puertas	reconstrucción	
				recubrimientos	reforzamiento	
				rejas	rehabilitación	
					reparación	

cas, 3 daños en equipos, 4 daños en mobiliario, 5 daños en acabados, 6 daños en estructuras y 7 daños en exteriores, $T1, T2, \dots, T7$, es un equivalente a cada transacción. Para la primera transacción se tiene que el edificio tuvo daños en equipos, mobiliario y daños estructurales, por lo que al hacer la discretización se coloca un 1 en los rubros 3, 4 y 6 ya que existe información correspondiente a

	1	2	3	4	5	6	7
$T1$	0	0	1	1	0	1	0
$T2$	1	1	0	0	0	0	1
$T3$	0	0	0	0	1	1	0
$T4$	1	1	1	1	0	1	1
$T5$	0	0	0	1	1	0	1
$T6$	0	1	0	1	1	0	1
$T7$	0	1	1	1	0	0	0

Tabla 3.4: Conjunto de datos discretizado

dicha categoría, mientras que para los rubros 1, 2, 5 y 7 se coloca 0 ya que no existen información de estas categorías.

El resultado de la discretización dio como resultado un conjunto de datos con una dimensión de 84 atributos, esto se debió a la categorización de cada atributo realizada anteriormente.

3.3.5. Generación de datos transaccionales

Tomando el conjunto de datos en formato discretizado, se obtuvo un conjunto de datos transaccional (obteniendo una transacción con tamaño máximo de características de 16 y el tamaño mínimo de características es de 9), el cual consta de obtener solo los *ítems* existentes (marcados con 1) en cada transacción. Tomando como referencia la Tabla 3.5 el conjunto transaccional final se representa en la Tabla 3.4. Para esto solo se consideran los valores correspondientes a 1, es decir para *T1* se toma el número de columna donde se tiene presencia el *ítem*; para este caso es la columna 3,4,6.

<i>T1</i>	3,4,6
<i>T2</i>	1,2,7
<i>T3</i>	5,6
<i>T4</i>	1,2,3,4,6,7
<i>T5</i>	4,5,7
<i>T6</i>	2,4,5,7
<i>T7</i>	2,3,4

Tabla 3.5: Conjunto de datos transaccional

3.4. Minería de Datos

Teniendo el conjunto de datos procesado, se realizó el minado de los patrones frecuentes y las RA. Para ello se utilizaron los algoritmos tales como el FP-Growth [52], Apriori [23] y el PSO-GES [31].

Los valores correspondientes al soporte se establecieron en base a los criterios estructurales, ya que para el análisis de riesgo sísmico, es indispensable que los datos estructurales sean frecuentes, de tal manera que se dividió el número aproximado de veces que se pretende salga una característica estructural entre el total de los datos (0.00002). Mientras que la confianza se ha establecido tomando

3. MARCO METODOLÓGICO

como referencia valores utilizados en el estado del arte ([11](75%), [36](70%), [37](75 %)).

■ Algoritmo FP-Growth

La obtención de patrones frecuentes fueron mediante el uso del algoritmo FP-Growth, incluido en el software Weka¹. Este algoritmo utiliza parámetros libres los cuales fueron los siguientes:

- Umbral de mínimo soporte igual a 0.00002
- Umbral de mínimo de confianza igual a 0.75
- Número de patrones frecuente igual a 50

■ Algoritmo Apriori

Para la obtención de las Reglas de Asociación con el algoritmo Apriori se uso la librería de Python llamada efficient-apriori. Este algoritmo utilizó un conjunto de datos transaccional y dos parámetros predeterminados los cuales fueron los siguientes:

- Umbral de mínimo soporte igual a 0.00002
- Umbral de mínimo de confianza igual a 0.75

■ PSO-GES

Para el algoritmo PSO-GES desarrollado en Visual Studio 2015, se ocupo el conjunto de datos discretizado, donde los parámetros utilizados son los especificados en [31]:

- Población de 20 partículas
- Las constantes c_1 y c_2 con valor 2
- Inercia w con valor de 1
- Número de iteraciones igual a 10 (se tomó por el punto de referencia encontrado en la literatura)

Mientras que los últimos tres han sido establecidos en base a fines óptimos para la investigación, por lo que los parámetros fueron los siguientes:

- Umbral de mínimo soporte igual a 0.0002
- Umbral de mínimo de confianza igual a 0.75
- Parámetro K (número de particiones) igual a 1 (es decir, una transacción por partición)

¹<https://www.cs.waikato.ac.nz/ml/weka/>

3.5. Interpretación y evaluación

Una vez finalizado el minado de reglas de asociación se obtuvieron una gran cantidad de RA, sin embargo para fines de análisis de riesgo sísmico se realizó el filtrado de las RA considerando solo las mejores reglas las cuales tienen un rango de confianza entre el 75 % y 100 %. Además, otras características que se han tomado son que las reglas tengan el costo total, los detalles estructurales (número de niveles, tipo de construcción, año de diseño), entre otros.

Las reglas que se obtuvieron de cada algoritmo fueron con parámetros similares, sin embargo los resultados han sido analizados, en conjunto con un experto del área de estructuras para así solo obtener las reglas con mayor interés para la realización del análisis de riesgo sísmico. No obstante la descripción de las RA obtenidas serán explicadas de manera detallada en el capítulo siguiente de resultados.

Resultados y Discusiones

En este capítulo se presenta en primer lugar los resultados obtenidos al utilizar el conjunto de datos de las escuelas afectadas por los sismos del 7 y 19 de Septiembre del 2017 en la República Mexicana. Al finalizar los resultados se presentan algunas discusiones

4.1. Resultados

4.1.1. Patrones Frecuentes

Los patrones frecuentes son aquellos que representan características o rasgos constantes y recurrentes en un conjunto de datos y ayudan a identificar un fenómeno o un problema, su obtención conforma la primera etapa en el proceso de generación de RA.

Para la obtención de patrones frecuentes en esta investigación se utilizó el algoritmo FP-Growth. Para el sismos del 7 de septiembre se obtuvieron 319 patrones frecuentes (Anexo PF_FP-GrowthA), mientras que para el sismo del 19 de Septiembre se 489 patrones frecuentes (Anexo PF_FP-GrowthB).

En la Tabla 4.1 se presentan a modo de ejemplo 5 de los patrones frecuentes para el sismo del 7 de septiembre del 2017, los cuales fueron elegidos ya que son algunos de los patrones frecuentes con mayores características estructurales de interés (descripción y tipo de daño, distancia epicentral, costos).

El primer patrón frecuente de la Tabla 4.1 identifica que las estructuras que resultaron más afectadas eran de dos niveles, la longitud de sus claros (transversal y longitudinal) era de 8 m y 3.08 m, fueron diseñadas en 1984, tenían acero y losa de concreto y estaban ubicadas en la zona sísmica C; estas estructuras tuvieron daños principalmente en acabados y daños estructurales, con un costo

4. RESULTADOS Y DISCUSIONES

Tabla 4.1: Patrones frecuentes generados por el algoritmo FP-Growth. Sismo del 7 de septiembre de 2017.

Patrones Frecuentes sismo 7 de septiembre de 2017
Nivel 2 claro transversal 8.00 m, claro longitudinal 3.80 m, año diseño 1984, acero y losa de concreto, daño en acabados, daño estructural, distancia epicentral 251-300 Km, zona C, daño menor, costo aproximado de reparación \$4,080,401-\$8,160,801
Claro transversal 6.00 m, claro longitudinal 6.00 m, año diseño 1986, muros de mampostería y losa de concreto, daño en acabados, daño estructural, distancia epicentral 201-250 Km, zona C, daño menor, costo aproximado de reparación \$0-\$12,241,202
Instalaciones eléctricas, nivel 1, claro transversal 6.00 m, claro longitudinal 6.00 m, año diseño 1986, muros de mampostería y losa de concreto, daño estructural, equipo, distancia epicentral 151-200 Km, zona C, daño menor, costo aproximado de reparación \$8,160,802-\$12,241,202
Nivel 2, claro transversal 8.00 m, claro longitudinal 3.06 m, año diseño 1970, Muros de mampostería y losa de concreto, daño en acabados, daño estructural, instalaciones eléctricas, distancia epicentral 151-200 Km, zona D, daño moderado, costo aproximado de reparación \$20,402,005-\$24,482,405
Instalaciones hidráulicas y sanitarias, nivel 1, claro transversal 6.00 m, claro longitudinal 6.00 m, año diseño 1984, acero y losa de concreto, daño en acabados, daño estructural, distancia epicentral 151-200 Km, zona D, daño menor, costo aproximado de reparación \$4,080,401-\$8,160,801

de reparación que oscila entre los \$4,080,401-\$8,160,801.

Sin embargo las características con mayor predominancia en los patrones frecuentes para el sismo del 7 de septiembre consiste en que las escuelas con un nivel, con claro (transversal y longitudinal) de 6.0 m y 5.80 m, con un diseño de 1970, tuvieron daños menores con un costo aproximado de reparación que va entre los \$0 a los \$12,241,202.

La Tabla 4.2 representa a modo ejemplo 5 de los patrones frecuentes para el sismo del 19 de septiembre del 2017, los cuales fueron elegidos derivado que son los que contienen mayores características estructurales de interés (descripción y tipo de daño, distancia epicentral, costos).

Analizando el primer patrón frecuente de la Tabla 4.2, se puede decir que las estructuras que resultaron más afectadas para este sismo eran de dos niveles, la longitud de sus claros (transversal y longitudinal) era de 6 m, fueron diseñadas en 1984, tenían muros de mampostería y techumbre ligera y estaban ubicadas en la zona sísmica C; estas escuelas presentaron daños principalmente en acabados y daños estructurales, con un costo de reparación que oscila entre los \$28,567,807-\$32,643,207.

Sin embargo las características con mayor predominancia en los patrones frecuentes para el sismo del 19 de septiembre consiste en que las escuelas con un nivel, con claro (transversal y longitudinal) de 6.00 m, con un diseño de 1984, tuvieron daños moderados con un costo aproximado de reparación que va entre los \$32,643,208 a los \$40,804,009.

Tabla 4.2: Patrones frecuentes generados por el algoritmo FP-Growth. Sismo del 19 de septiembre de 2017.

Patrones Frecuentes sismo 19 de septiembre de 2017
Instalaciones eléctricas, hidráulicas y sanitarias, nivel 1, claro transversal 6.00 m, claro longitudinal 6.00 m, año diseño 1984, Muros de mampostería y techumbre ligera, daño estructural, exteriores, distancia epicentral 151-200 Km, zona C, daño moderado, costo aproximado de reparación \$28,567,807-\$32,643,207
Nivel 1 claro transversal 6.00 m, claro longitudinal 6.00 m, año diseño 1984, acero y losa de concreto, daño en acabados, daño estructural y equipos, distancia epicentral 1-50 Km, zona C, daño menor, costo aproximado de reparación \$24,482,406-\$28,562,806
Daño estructura, daño acabados, daño mobiliario, daño en instalaciones eléctricas, claro transversal 8.00 m, claro longitudinal 3.06 m, año diseño 1986, muros de mampostería y losa de concreto, distancia epicentral 151-200 Km, zona B, daño moderado, costo aproximado de reparación \$36,723,609-\$40,804,009
Nivel 2, claro transversal 8.00 m, claro longitudinal 3.06 m, año diseño 1970, Muros de mampostería y techumbre ligera, daño en acabados, instalaciones eléctricas, distancia epicentral 351-400 Km, zona C, daño menor, costo aproximado de reparación \$8,160,802-\$12,241,202
Nivel 1, claro transversal 6.00 m, claro longitudinal 6.00 m, año diseño 1986, Muros de mampostería y losa de concreto, distancia epicentral 251-300 Km, zona C, daño moderado, daños estructurales, daños acabados, instalaciones eléctricas, costo aproximado de reparación \$32,643,208-\$36,723,608.

4.1.2. Reglas de Asociación

Atendiendo al Análisis de riesgo sísmico, el objetivo del minado de RA se orientó en ese sentido. Para ello, los algoritmos (Apriori y PSO-GES) se adaptaron en su ejecución, de tal manera que, adicional a los *ítems* que el algoritmo determinase, se debía cumplir la condición de que las RA resultantes incluyeran: distancia epicentral y tipo de construcción en el antecedente y tipo de daño, detalle del daño, costo total de reparación y zona sísmica en el consecuente. Considerando esta condición, en los resultados, el algoritmo PSO-GES obtuvo una mayor cantidad de reglas que el algoritmo Apriori (Anexo RA_apriori), en las siguientes secciones se realiza un análisis de éstas.

4.1.2.1. Sismos del 7 de Septiembre de 2017

La Tabla 4.3 muestra las RA obtenidas para el sismo del 7 de septiembre de 2017. Estas RA estas agrupadas considerando la distancia epicentral y el tipo de daño que se fue presentando, dichas reglas fueron generadas con el algoritmo PSO-GES.

Con las RA mostradas en la Tabla 4.3 se puede decir que las escuelas con menos variación de daños son las escuelas construidas con acero y losa de concreto y las escuelas de muros de mampostería y techumbre ligera ya que para ambos tipos de escuelas a una distancia epicentral de 101-300 Km los daños reportados fueron moderados para una zona sísmica C, donde los mayores daños presentados son estructurales, acabados, exteriores e instalaciones eléctricas, hidráulicas y

4. RESULTADOS Y DISCUSIONES

Tabla 4.3: Reglas de Asociación para el sismo del 7 de septiembre de 2017

Antecedente		Consecuente
Tipo de construcción	Distancia Epicentral	
Muros de mampostería y losa de concreto	101-200 Km	Daño moderado, Zona C
		Daño estructural, acabados, instalaciones eléctricas
		Costo total de reparación \$32,643,208-\$44,884,410
	201-300 Km	Daño menor, Zona C
		Daño estructural, exteriores
		Costo total de reparación \$16,321,604-\$32,643,207
	301-400 Km	Daño moderado, Zona C, Daño estructural
		Costo total de reparación \$20,402,005-\$24,482,405
	401-800 Km	Daño menor, Zona C
		Daño acabados, estructural, mobiliario e instalaciones hidráulicas y sanitarias
		Costo total de reparación \$0-\$12,241,202
	Acero y losa de concreto	101-300 Km
Daño estructural, acabados, exteriores, equipo, instalaciones eléctricas, hidráulicas y sanitarias		
Costo total de reparación \$8,160,802-\$36,723,608		
301-800 Km		Daño menor, Zona C
		Daño estructural, acabados, exteriores, instalaciones hidráulicas y sanitarias
		Costo total de reparación \$0-\$12,241,202
Acero y techumbre ligera	101-150 Km	Daño menor, Zona C, Daño estructural
		Costo total de reparación \$12,241,203-\$16,321,603
	151-300 Km	Daño moderado, Zona C, Daño estructural
		Costo total de reparación \$20,402,005-\$36,723,608
	301-800 Km	Daño menor, Zona C
		Daño estructural, exteriores, mobiliario, acabados
	Costo total de reparación \$0-\$12,241,202	
Muros de mampostería y techumbre ligera	101-300 Km	Daño moderado, Zona C
		Daño estructural, acabados, exteriores
		Costo total de reparación \$8,160,802-\$28,562,806
	301-800 Km	Daño menor, Zona C
		Daño estructural, exteriores, acabados, mobiliario, instalaciones eléctricas, hidráulicas y sanitarias
		Costo total de reparación \$0-\$8,160,801

sanitarias. En cuanto a los costos aproximados de reparación para estas escuelas oscilan entre los \$8,160,304-\$36,723,608.

Por otro lado, para distancias epicentrales de 301-800 Km los daños fueron menores, presentando una predominancia en daños estructurales, acabados, equipo, mobiliario, exteriores, e instalaciones eléctricas hidráulicas y sanitarias, donde los costos aproximados de reparación están entre los \$0-\$12,241,202.

De las RA de la Tabla 4.3 también se puede observar que las escuelas que más variación de daño tuvo, fueron aquellas de muros de mampostería y losa de concreto ya que a distancias epicentrales de 101-200 Km y 300-400 Km los daños presentados fueron moderados, donde los daños predominantes son estructurales, acabados e instalaciones eléctricas, teniendo costos aproximados de reparación de \$20,402,005-\$44,884,410, y para distancias epicentrales de 201-300 Km y 401-800 Km el tipo de daños presentado fue de tipo menor, para daños estructurales, exteriores, acabados, mobiliario e instalaciones hidráulicas y sanitarias, donde los costos de reparación se encuentran entre \$0-\$32,643,207.

En la figura 4.1 se muestra una gráfica de las AR obtenidas, agrupadas por distancia epicentral y tipo de construcción, donde se observa la variación de los costos de reparación en función de la distancia epicentral y el tipo de construcción. En esta gráfica se puede observar que los mayores daños se produjeron en aquellas escuelas con muros de mampostería y losa de concreto. Del mismo modo se aprecia, en general, que los costos de reparación disminuyen en la medida que las escuelas se alejan del epicentro del sismo. De este modo, a menor distancia epicentral los costos de reparación son mayores a los costos de las escuelas que se ubican a una distancia epicentral mayor.

Este resultado que en apariencia es lógico, se debe al proceso de atenuación de las ondas sísmicas, que en general, a mayor distancia epicentral presentan menor intensidad, y por lo tanto, menores daños a las construcciones. De esta forma que se corrobora el correcto procesamiento de los datos por medio de las reglas de asociación utilizadas.

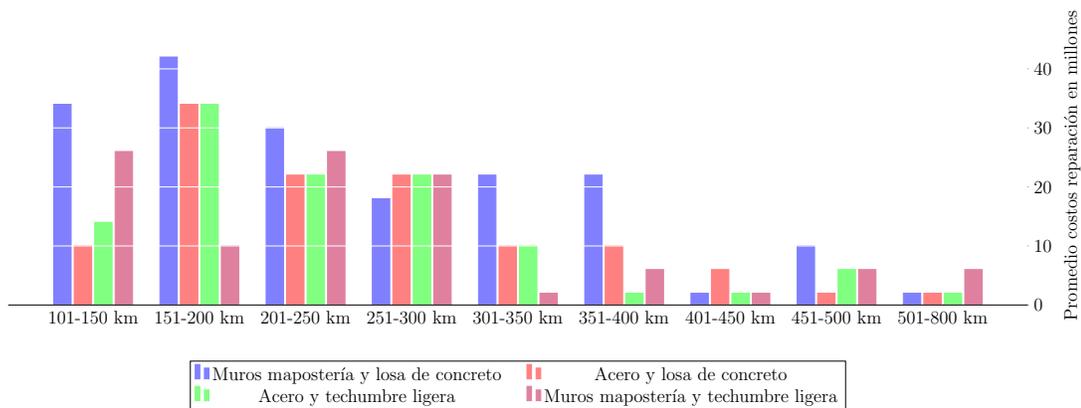


Figura 4.1: Variación de los costos individuales promedio de reparación para el sismo del 7 de septiembre de 2017 en función del tipo de construcción y la distancia epicentral.

4.1.2.2. Sismos del 19 de Septiembre de 2017

La Tabla 4.4 muestra las RA obtenidas para el sismo del 19 de septiembre de 2017. Estas RA estas agrupadas considerando la distancia epicentral y el tipo de daño que se fue presentando, las cuales fueron obtenidas con el algoritmo PSO-GES.

Con las RA mostradas en la Tabla 4.4 se observa que, todas las escuelas a distancias epicentrales menores a 100 Km, presentaron daños moderados, con daños

4. RESULTADOS Y DISCUSIONES

en acabados, estructurales y exteriores, teniendo costos de reparación que oscilan entre los \$28,562,807-\$40,804,009.

Un punto importante de observar, es que las escuelas se encuentra en las distancias epicentrales de 101-150 Km, con tipo de construcción de muros de mampostería y losa de concreto, acero y techumbre ligera y muros de mampostería y techumbre ligera, sufrieron daños graves durante este sismo, reportando daños predominantemente en acabados, estructurales, exteriores, instalaciones eléctricas, hidráulicas y sanitarias, donde los costos aproximados de reparación oscilan entre los \$40,804,010-\$53,045,212, por otro lado las estructuras de acero y losa de concreto, presentaron un daño moderado, dichos daños fueron en estructuras, acabados y exteriores, con costos aproximados de reparación de \$28,562,807-

Tabla 4.4: Reglas de Asociación para el sismo del 19 de septiembre de 2017

Antecedente		Consecuente
Tipo de construcción	Distancia Epicentral	
Muros de mampostería y losa de concreto	51-100 Km	Daño moderado, Zona C, Daño estructural
		Costo total de reparación \$32,643,208-\$36,723,608
	101-150 Km	Daño grave, Zona D
		Daño estructural, acabados, exteriores
	151-800 Km	Costo de reparación \$40,804,010-\$44,884,410
		Daño menor, Zonas B y C
Acero y losa de concreto	1-150 Km	Daño estructural, mobiliario, acabados, instalaciones eléctricas
		Costo total de reparación \$ 4,080,401-\$36,723,608
	151-800 Km	Daño moderado, Zonas B y C
		Daño estructural, acabados, exteriores
	151-800 Km	Costo total de reparación \$28,562,807-\$40,804,009
		Daño menor, Zonas B y C
Acero y techumbre ligera	51-100 Km	Daño estructural, acabados, equipo, instalaciones eléctricas
		Costo total de reparación \$0-28,562,806
	101-150 Km	Costo moderado, Zona C, Daño estructural
		Costo total de reparación \$32,643,208-\$36,723,608
	151-250 Km	Daño grave, Zona D
		Daño estructural, acabados, instalaciones eléctricas, hidráulicas y sanitarias
251-800 Km	Costo de reparación \$44,884,411- \$48,964,811	
	Daño moderado, Zonas B y C	
Muros de mampostería y techumbre ligera	1-100 Km	Daño estructural, exteriores, instalaciones eléctricas
		Costo total de reparación \$24,482,406-\$28,562,806
	101-150 Km	Daño menor, Zona C
		Daño estructural, exteriores, instalaciones hidráulicas y sanitarias
	151-200 Km	Costo total de reparación \$ 4,080,401-\$28,562,806
		Daño moderado, Zona C, Daño estructural, acabados
251-800 Km	Costo total de reparación \$32,643,208-\$36,723,608	
	Daño grave, Zona B, Daño estructural	
151-200 Km	Costo total de reparación \$48,964,812-\$53,045,212	
	Daño moderado, Zona B, Daño estructural	
251-800 Km	Costo total de reparación \$28,562,807-\$32,643,207	
	Daño menor, Zona C	
251-800 Km	Daño estructural, exteriores, acabados	
	Costo total de reparación \$8,160,802-\$32,643,207	

\$40,804,009.

Para distancias epicentrales mayores a 150 Km, se obtuvo que para los escuelas de muros de mampostería y losa de concreto y Acero y losa de concreto se tuvo un daño menor, en Zonas sísmicas B y C, donde los principales daños presentados son en instalaciones eléctricas, estructurales, acabados y exteriores, considerando costos de reparación de \$0-\$35,723,608.

Por otro lado, para las escuelas de acero y techumbre ligera y muros de mampostería y techumbre ligera el comportamiento del daño fue distinto ya que, para distancias epicentrales de 151-250 Km, el daño fue moderado, reportado daños estructurales, acabados y exteriores, considerando costos de reparación aproximados de \$24,482,406-\$32,643,207. De igual modo, para distancias epicentrales mayores a los 250 Km los daños en las escuelas son menores, teniendo un costos aproximado de reparación de \$4,080,401-\$32,643,207, para reparar daños estructurales, acabos, exteriores e instalaciones hidráulicas y sanitarias.

Por lo que para este sismo se puede decir que las escuelas que han sufrido menos variación de daños son aquellas que se encuentran hechas de acero y losa de concreto, teniendo daños moderados y menores.

En la figura 4.2 se muestra una gráfica con los resultados de todas las RA obtenidas, agrupadas por distancia epicentral y tipo de construcción, donde se relaciona la distancia del epicentro del sismo a las escuelas con el costo de reparación y el tipo de estructura. En este caso se aprecia que los mayores daños se presentaron en las escuelas que se encuentran a una distancia epicentral de 101 a 150 Km. En cuanto al tipo de construcción, al igual que para el sismo del 7 de septiembre, las escuelas que resultaron más afectadas fueron las construidas a base de muros de mampostería y techumbre ligera.

Nótese que en contraste con lo reportado para el sismo del 7 de septiembre, los mayores costos de reparación no se presentaron para las menores distancias epicentrales. En este caso, los mayores costos se presentaron para distancias epicentrales intermedias. Esto se explica en razón de que la Ciudad de México se localiza justo en esta distancia epicentral (100 a 150 km), por lo que al tener un gran número de escuelas en dicho sitio, es de esperarse que los costos de reparación sean mayores.

Desde el punto de vista sísmico, los mayores daños se explican debido a las características del suelo en ciertas zonas de la Ciudad de México, en las cuales las ondas sísmicas se amplifican. Desde el punto de vista práctico, este efecto

4. RESULTADOS Y DISCUSIONES

tiene como consecuencia que las aceleraciones producidas en dichas zonas sean comparables a las que se presentan en sitios más cercanos al epicentro del sismo.

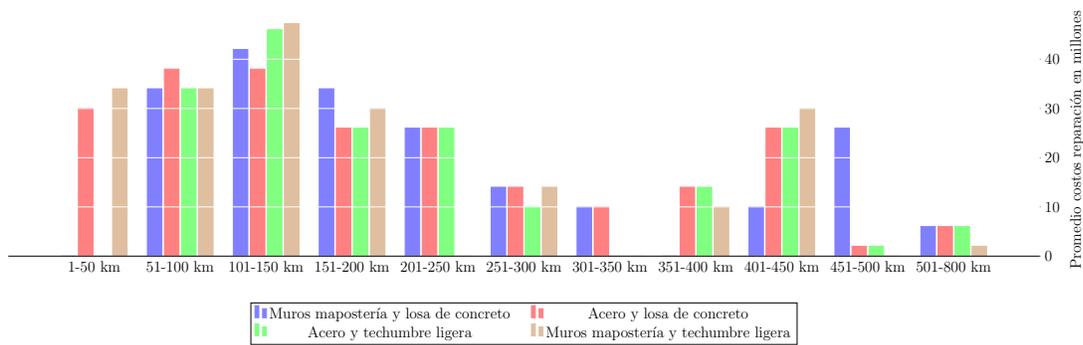


Figura 4.2: Variación de los costos individuales promedio de reparación para el sismo del 19 de septiembre de 2017 en función del tipo de construcción y la distancia epicentral.

Otros costos que son importantes de mencionar que se muestran en la gráfica 4.2, son para las distancias que van de los 401 a 450 Km, el costo alto de reparación para estas distancias se debe a que los mayores daños presentados en dichas escuelas son daños estructurales (reparación de muros, pisos y techos), además de presentar diferentes daños en acabados (pintura, vidrios, cerraduras, entre otros), por lo que los costos aproximados de reparación ascienden a los \$24,500,000, esto para el tipo de escuelas con: acero y losa de concreto, acero y techumbre ligera y muros de mampostería y techumbre ligera. Sin embargo también estos altos costos a distancias epicentrales grandes, se debieron al posible daño inicial que sufrieron las escuelas por el sismo del 7 de septiembre. El sismo del 19 de septiembre, quizá sólo evidenció buena parte del daño que ya tenían debido al sismo previo.

Derivado de los resultados obtenidos se puede decir que para el sismo del 7 de septiembre la construcción más afectada fue aquella de muros de mampostería y losa de concreto, mientras que para el sismo del 19 de septiembre la construcción con mayor afectación es aquella de muros de mampostería y techumbre ligera. En cuanto a la afectación por distancia epicentral, para el sismo del 7 de septiembre la construcción de muros de mampostería y losa de concreto es riesgosa para distancias que van de 101 -250 Km, 351-400 Km, y 451-500 Km, para una distancia epicentral de 401-500 Km la construcción más riesgosa es aquella de Acero y losa de concreto y para distancias mayores a 501 Km la construcción más riesgosa es aquella que está hecha con Muros de Mampostería y techumbre ligera.

En cuanto al sismo del 19 de septiembre para una distancia epicentral de 1-50 Km, 101-200 Km, 251-300 Km y 401-500 Km las escuelas más riesgosas son aquellas que están hechas de Muros de mampostería y techumbre ligera, para una distancia epicentral de 151-200 Km, 451-500 Km la escuela más riesgosa es aquella de Muros de mampostería y losa de concreto, mientras que para el resto de las distancias el riesgo de las construcciones sufrieron daños iguales por lo que todas tendrían un mismo riesgo.

4.2. Discusión

El análisis de riesgo sísmico es un tema de interés en México y el mundo, debido al impacto que los daños provocados tienen en la sociedad en general. Son diversas las estrategias que en la literatura se han sugerido para su estudio, simulación de Montecarlo, análisis probabilista de fallas, utilización de redes bayesianas y algoritmos de minería de datos, son solo algunos ejemplos.

En este capítulo se presentan los resultados obtenidos al aplicar técnicas descriptivas de minería de datos para el análisis de riesgo sísmico, en específico, el minado de reglas de asociación.

En la sección 2.5, se describen varios estudios que hacen uso de algoritmos para obtener RA en datos reales de sismos ocurridos al rededor del mundo. De estos estudios, el algoritmo más utilizado es el Apriori, situación por la que este algoritmo fue utilizado en esta investigación para obtener las RA.

Pocos algoritmos bioinspirados se reportan en el estado del arte para el análisis de riesgo sísmico, resaltando los algoritmos genéticos predominantemente. En esta investigación se hace uso del algoritmo enjambre de partículas PSO-GES, reportado en la literatura como uno de los más novedosos y potentes metaheurísticos aplicados a minado de reglas de asociación. Esto último fue posible constatarlo, ya que con los datos utilizados en esta Tesis, PSO-GES superó de forma considerable el rendimiento del algoritmo Apriori, cuya cantidad de RA obtenidas fue infima y prohibitiva para el estudio aquí presentado (ver anexo RA_Apriori).

La calidad de las RA obtenidas por un algoritmo es determinante para su aceptación o rechazo. En su mayoría el grado de confianza de la regla es el criterio de calidad más utilizado en la literatura (sección 2.5). Considerando este aspecto independientemente de los datos utilizados, las RA obtenidas y utilizadas para el análisis de riesgo sísmico en esta Tesis tienen una confianza superior al 75 %, a diferencia de algunas investigaciones del estado del arte que reportan

4. RESULTADOS Y DISCUSIONES

una confianza mínima de 50 % como: Mammani-Coaquira Y. et al. [39], Nivedhitha U. et al. [12], Aqdas Ikram, Usman Qamar [40], Martínez-Ballesteros [36], Martínez-Álvarez [11]

Otro aspecto importante de mencionar es en relación a los datos utilizados. Al utilizar datos de diferentes repositorios, es natural que las características del sismo utilizadas en el estudio sean diferentes, lo que dificulta que puedan ser comparados entre sí. Algunos estudios consideran características definitorias del sismo, tales como: latitud, longitud, magnitud, profundidad y fecha. Estos estudios distan de forma significativa en lo presentado en esta tesis, ya que ésta se enfoca a los daños estructurales producto del sismo, mucho más que al estudio del sismo en sí.

Estudios que sí consideran características estructurales como tipo de material, número de pisos, año de construcción y tipo de techo son los realizados por Lorenzo Diana, et al. [41], Ridiel Ismael et al. [42], Guettiche A. et al. [43] y Yaohui Liu et al. [44].

A diferencia de los estudios del estado del arte mostrados en la sección 2.5, el estudio realizado en esta Tesis considera características de ambos aspectos: características propias del sismo (fecha, magnitud del sismo, localización (magnitud y latitud), distancia epicentral y zona sísmica) y características estructurales de las edificaciones dañadas (tipo de construcción, detalles y tipo del daño de los edificios educativos, costos aproximando de reparación).

Al considerar ambos aspectos como parte del estudio, los resultados mostrados en este capítulo brindan una descripción más detallada de los daños ocasionados por los sismos del 7 y 19 de septiembre de 2019, así como los costos aproximados que estas escuelas han tenido para su reconstrucción.

Por último, si bien las características estructurales utilizadas en esta investigación son pocas para obtener un análisis sísmico completo, lo utilizado es de gran valor ya que en México el reportado en esta tesis es uno de los primeros registrados de los últimos sismos de gran relevancia en México.

Conclusiones y Trabajo futuro

Históricamente, la disposición de datos públicos relacionados a los daños producidos por los sismos en México, era casi nula hasta el pasado 2017. La dificultad de los levantamientos en campo, la veracidad, los tiempos de captura y procesamiento son algunas de las causas que en las décadas pasadas impidieron que se contara con esta base de conocimiento. Sin embargo, dados los avances de la tecnología y el amplio uso de las TIC, el registro de los daños producidos por los sismos del 7 y 19 de septiembre de 2017 constituye una fuente invaluable de información para el estudio del comportamiento, impacto y riesgo sísmico. No obstante, aún queda pendiente homologar estas bases de datos y replantear el tipo de datos que se recolectan, con el propósito de hacer estudios más formales referentes al comportamiento y riesgo sísmico de las construcciones.

Los resultados mostrados en esta Tesis fueron orientados al análisis de riesgo sísmico mediante el minado de Reglas de Asociación. Derivado de este minado y con apoyo de expertos en el área de Ingeniería estructural fue posible mostrar un panorama general del comportamiento de los costos de reparación de las escuelas afectadas por estos sismos, en función de la distancia epicentral y el tipo de construcción. Se pudo detectar que los mayores daños se produjeron durante el sismo del 19 de septiembre de 2017. En este caso, el costo individual promedio de reparación de las escuelas fue de \$43,000,000.00, mientras que para el sismo del 7 de septiembre fue de \$40,000,000.00. También se pudo detectar para el sismo del 7 de septiembre, con epicentro en las costas de Oaxaca, que a mayores distancias epicentrales los daños fueron menores en comparación a los costos registrados para las escuelas que se localizaron a las distancias epicentrales mas cercanas al sismo.

Por su parte, en el caso del sismo del 19 de septiembre, se pudo observar como los costos de reparación se incrementaron para distancias comprendidas entre 100 km y 150 km. Este comportamiento resultó congruente con los modelos sísmológicos existentes, los cuales reconocen una amplificación de las aceleraciones

5. CONCLUSIONES Y TRABAJO FUTURO

en ciertas zonas de la Ciudad de México, la cual se sitúa en este rango de distancias epicentrales. Esto explica en cierta medida este aumento de los costos de reparación.

En cuanto al tipo de construcción, se pudo identificar que las estructuras con muros de mampostería y techumbre ligera son las que presentaron mayores daños en ambos sismos. Por lo que se refiere al tipo de daño que más se presentó, correspondió al daño en la estructura y en los acabados de las escuelas. Es importante recalcar que la calidad de las RA utilizadas para el análisis realizado fue el eje rector del análisis a fin de garantizar el grado de confianza igual o superior al 75 %.

Derivado de los resultados de la investigación se pueden hacer algunas recomendaciones. Las escuelas hechas con muros de mampostería y losa de concreto son las que tienden a sufrir más daños en los sismos, por lo que durante un sismo se pueden dañar más fácilmente, lo cual implica gastos en reparación. Derivado de lo anterior, se recomienda poner especial cuidado en el diseño y construcción de este tipo de escuelas en zonas de alta o mediana sismicidad.

Adicional a los costos de reparación, el daño estructural también fue posible analizarlo. De este es posible concluir que para el sismo del 19 de septiembre las escuelas que tuvieron menores daños son aquellas que están hechas de muros de mampostería y techumbre ligera, por lo que para distancias epicentrales de 201 a 400 km y de 451 a 800 km es recomendable la construcción de escuelas con dichas características ya que fueron las que tuvieron menos costos de reparación después del sismo. Mientras que para distancias epicentrales menores a 200 km, las escuelas con menos afectación son aquellas que son de acero y techumbre ligera al igual que las de acero y losa de concreto, por lo que para estas distancias es recomendable tomar en cuenta la construcción de escuelas con estas características y evitar la construcción de escuelas de muros de mampostería y techumbre ligera ya que fueron las que sufrieron mayores daños.

Por su parte, para el sismo del 7 de septiembre se puede considerar que la mejor opción de construcción de escuelas son aquellas que están hechas de acero y losa de concreto y acero y techumbre ligera, ya que ante este sismo dichas construcciones fueron las que tuvieron menos costos de reparación en comparación con las otras escuelas, esto para la mayoría de las distancias epicentrales.

Por último se puede concluir que para ambos sismos las escuelas que menos variación de daños tuvieron, son las de acero y losa de concreto, ya que para los sismos del 7 y 19 de septiembre, los daños a partir de distancias epicentrales superiores a los 151 y 301 Km respectivamente, fueron menores, lo que implica que

si se consideran características similares a estos sismos ocurridos para la construcción de escuelas, este tipo de construcción podría ser una opción, ya que son las que menos presupuesto requieren para reparación, y menos daños han tenido durante los estos sismos ocurridos.

Las aportaciones de la investigación presentada en este Tesis pueden enumerarse como sigue:

1. Construcción de una base de conocimiento estandarizada para uso de algoritmos de aprendizaje automático y minería de datos. Esta base de conocimiento se pone a disposición de la comunidad científica para su explotación y estudio.
2. Para el análisis de riesgo sísmico, en la experimentación realizada se combina tanto información sísmica como información estructural. Este aspecto es de suma importancia, ya que la mayoría de los trabajos del estado del arte utilizan solamente datos sísmicos o datos estructurales en su análisis.
3. La calidad de las reglas obtenia es en su mayoría igual o superior a la reportada por otras investigaciones del estado del arte.

Derivado de lo antes mencionado, es posible afirmar que la hipótesis planteada en la sección 1.2, es aceptada debido a que mediante la generación de reglas de asociación con una confianza superior al 75 % se puede obtener una descripción del comportamiento de los daños en las construcciones del sector educativo, a causa de los sismos del 7 y 19 de septiembre del 2017.

Los resultados obtenidos en esta investigación han sido publicados en Mexican Conference on Pattern Recognition (12th Mexican Conference, MCPR 2020, Morelia, Mexico, June 24–27, 2020, Proceedings) [53].

Como trabajo futuro se pretende realizar diferentes análisis de riesgo sísmico, tomando distintos atributos de los datos, tanto para el antecedente como para el consecuente, esto con el fin de ver como se comportaron diferentes parámetros durante los sismos del 7 y 19 de Septiembre del 2020. Los parámetros que se pueden considerar pueden ser: entidad federativa, municipio, tipo y descripción del daño, zona sísmica, tipo de construcción, costos, entre otros.

Además el estudio aquí presentado podría replicarse para un análisis de otros sectores afectados durante los sismos antes mencionados como: sector salud, sector vivienda, sector patrimonio cultural e histórico.

5. CONCLUSIONES Y TRABAJO FUTURO

De igual modo, otra línea abierta de estudio apunta a la utilización de otros algoritmos para el minado de RA, así como utilizar la bae de conocimiento para fines predictivos con algoritmos de machine learning.

Bibliografía

- [1] P. Shenoy, J. R. Haritsa, S. Sudarshan, G. Bhalotia, M. Bawa, and D. Shah, “Turbo-charging vertical mining of large databases,” *Acm Sigmod Record*, vol. 29, no. 2, pp. 22–33, 2000.
- [2] R. H. León, *Descubrimiento de Conjuntos Frecuentes de Ítems en Datos Estáticos y Dinámicos*. Tesis de maestría, Instituto Nacional de Astrofísica, Óptica y Electrónica, 2008.
- [3] V. K. Pang-Ning Tan, Michael Steinbach, *Introduction to Data Mining*. Pearson Addison Wesley, 2006.
- [4] S. de Educación Pública Fuerza México FONDEN, “sep_reconstruccion.csv.” Obtenido de: https://www.transparenciapresupuestaria.gob.mx/work/models/PTP/Home/fuerzamexico/sep_reconstruccion.csv.
- [5] C. F. d. E. Instituto de Investigaciones Electricas, *Manual de diseño de Obras Civiles, Diseño por Sismo*. Comisión Federal de Electricidad, 2008.
- [6] S. Paniagua, *Desastres y emergencias: prevención, preparación y mitigación*. Editorial Tecnológica de CR, 2002.
- [7] F. Peña Mondragón, “Estrategias para el modelado y el análisis sísmico de estructuras históricas,” *Ingeniería sísmica*, vol. 1, no. 83, pp. 1–7, 2010.
- [8] J. G. Mayordomo, *Caracterización y Análisis de la Peligrosidad Sísmica en el Sureste de España*. Phd thesis, Universidad Complutense de Madrid, 2005.
- [9] R. C. J. Virseda Benito Fernando, *Minería de Datos y Aplicaciones*. UIAF:Colombia, 2014.
- [10] S. Gupta and R. Mamtora, “A survey on association rule mining in market basket analysis,” *International Journal of Information and Computation Technology*. ISSN:0974-2239, vol. 4, no. 4, pp. 409–414, 2014.

BIBLIOGRAFÍA

- [11] F. Martínez-Álvarez, A. Troncoso, A. Morales-Esteban, and J. C. Riquelme, “Computational intelligence techniques for predicting earthquakes,” in *International Conference on Hybrid Artificial Intelligence Systems*, pp. 287–294, Springer, 2011.
- [12] U. Nivedhitha and A. Krishna, “Development of a predictive system for anticipating earthquakes using data mining techniques,” *Indian J. Sci. Technol.*, vol. 9, pp. 1–7, 2016.
- [13] F. J. G. Montaña, *Metodología para el análisis de ocurrencias de terremotos de gran magnitud*. Tesis de maestría, Universidad de Sevilla, 2013.
- [14] G. Daniela, “17 años de desastres cuestan 470 mmdp, el universal.” Obtenido de: <http://www.eluniversal.com.mx/periodismo-de-datos/17-anos-de-desastres-naturales-cuestan-470-mmdp>. Consultada el 12 de Septiembre de 2020.
- [15] G. Ortiz, “Sismo de 8.2, el más intenso en casi un siglo en México, excelsior.” Obtenido de: <https://www.excelsior.com.mx/nacional/2017/09/08/1187054>. Consultada el 12 de Septiembre de 2020.
- [16] Milenio, “Los daños por el sismo: reporte del gobierno federal.” Obtenido de: <http://www.milenio.com/negocios/los-danos-por-el-sismo-reporte-del-gobierno-federal>. Consultada el 14 de Septiembre de 2020.
- [17] EFE, “Sep: reparar 12 mil 931 escuelas tras sismos costará 13 mil 650 mdp.” Obtenido de: <http://www.eluniversal.com.mx/nacion/sociedad/reparar-12-mil-931-escuelas-tras-sismos-en-mexico-costara-760-millones-de-dolares>. Consultada el 12 de Septiembre de 2020.
- [18] ElUniversal, “Los monumentos dañados por el sismo del 19 de septiembre de 2017.” Obtenido de: <http://www.eluniversal.com.mx/destinos/los-monumentos-danados-por-el-sismo-del-19-de-septiembre-de-2017>. Consultada el 14 de Septiembre de 2020.
- [19] BBCMundo, “Terremoto de magnitud 8,2, el mayor en un siglo, sacude el suroeste de México, deja al menos 61 muertos y miles de afectados.” Obtenido de: <https://www.bbc.com/mundo/noticias-america-latina-41197767>. Consultada el 19 de Julio de 2020.
- [20] S. K. y M. O. Víctor Cruz, “¿qué ocurrió el 19 de septiembre de 2017 en México?, unam.” Obtenido de: <http://ciencia.unam.mx/leer/652/-que-ocurrio-el-19-de-septiembre-de-2017-en-mexico->. Consultada el 14 de Septiembre de 2020.

- [21] Y. Vargas, L. Pujades, A. Barbat, and J. Hurtado, “Evaluación probabilista de la capacidad, fragilidad y daño sísmico de edificios de hormigón armado,” *Revista internacional de métodos numéricos para cálculo y diseño en ingeniería*, vol. 29, no. 2, pp. 63–78, 2013.
- [22] W. Claudio Palma, *Data Mining. El arte de anticipar. 10 casos reales*. RIL Editores, 2009.
- [23] R. Q. M. J. Hernández Orallo José, Ferri Ramírez Cesar, *Introducción a la Minería de Datos*. Pearson Prentice Hall, 2004.
- [24] J. M. Moine, *Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo*. Tesis de maestría, Universidad Nacional de La Plata, Facultad de Informática, 2013.
- [25] Y. Zhao and G. Karypis, “Evaluation of hierarchical clustering algorithms for document datasets,” in *Proceedings of the eleventh international conference on Information and knowledge management*, pp. 515–524, ACM, 2002.
- [26] J. Zyt, W. Klosgen, and J. Zytkow, *Handbook of data mining and knowledge discovery*. Oxford university press, 2002.
- [27] D. Sánchez, M. Miranda, and L. Cerda, “Reglas de asociación aplicadas a la detección de fraude con tarjetas de crédito,” in *Actas del XII Congreso Español sobre Tecnologías y Lógica Fuzzy*, pp. 15–17, 2004.
- [28] J. E. Rodríguez Rodríguez and M. E. De Moya Amaris, “La contribución de las reglas de asociación a la minería de datos,” *Tecnura*, vol. 7, no. 13, pp. 94–109, 2003.
- [29] D. Conti and F. J. M. de Pisón Ascacíbar, “Reglas de asociación en series temporales: panorama referencial y tendencias,” in *II Congreso Español de Informática*, pp. 213–221, 2010.
- [30] J. Kennedy and R. Everhart, “A new optimizer using particle swarm theory. in proceedings of the sixth international symposium on micro machine and human science. nagoya japon,” *IEEE service center Piscataway, NJ*, vol. 43, no. 1, pp. 39–43, 1995.
- [31] B. B. Gretel, *Diseño de una Metaheurística para el minado de reglas de asociación en bases de datos transaccionales*. Tesis de maestría, Universidad Autónoma del Estado de México, Facultad de Ingeniería, 2018.

BIBLIOGRAFÍA

- [32] E. G. García, R. J. López, J. J. M. Moreno, A. S. Abad, B. C. Blasco, and A. P. Pol, “La metodología del data mining. una aplicación al consumo de alcohol en adolescentes,” *Adicciones*, vol. 21, no. 1, pp. 65–80, 2009.
- [33] O. Mario, “Algunos conceptos del análisis de riesgo,” *Prevención*, vol. 1, no. 14, pp. 6–11, 1996.
- [34] I. T. G. de España, C. de Liñán, J. sobre Reducción de Riesgos Geológicos en España, and F. y N. Real Academia de Ciencias Exactas, *Reducción de riesgos geológicos en España*. Instituto Tecnológico Geominero de España, España, 1995.
- [35] E. Florido-Navarro, *Metodología basada en minería de datos para el descubrimiento de patrones precursores de terremotos de magnitud media y elevada*. Phd thesis, Universidad Pablo de Olavide, 2013.
- [36] M. Martínez-Ballesteros, F. Martínez-Álvarez, A. Troncoso, and J. C. Riquelme, “Quantitative association rules applied to climatological time series forecasting,” *International Conference on Intelligent Data Engineering and Automated Learning*, vol. 1, no. 1, pp. 284–291, 2009.
- [37] J. D. V. Mendoza, J. F. G. Mejía, and J. A. Garcia, *Caracterización de la Actividad Sísmica de la República Mexicana: Una Primera Aproximación*. Tesis de maestría, Universidad Autónoma del Estado de México, 2011.
- [38] P. M. Álvaro, *Una metaheurística para la extracción de reglas de asociación. Aplicación a Terremotos*. Tesis de maestría, Escuela Técnica Superior de Ingeniería Informática, 2012.
- [39] Y. Mamani-Coaquira, J. S. Aguilar-Ruiz, M. J. Ibarra, and J. L. Soncco-Álvarez, “Searching for association rules to forecast earthquakes,” in *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–6, IEEE, 2020.
- [40] A. Ikram and U. Qamar, “Developing an expert system based on association rules and predicate logic for earthquake prediction,” *Knowledge-Based Systems*, vol. 75, pp. 87–103, 2015.
- [41] L. Diana, J. Thiriot, Y. Reuland, and P. Lestuzzi, “Application of association rules to determine building typological classes for seismic damage predictions at regional scale: the case study of basel,” *Frontiers in Built Environment*, vol. 5, p. 51, 2019.

- [42] I. Riedel, P. Guéguen, M. Dalla Mura, E. Pathier, T. Leduc, and J. Chanussot, “Seismic vulnerability assessment of urban environments in moderate-to-low seismic hazard regions using association rule learning and support vector machine methods,” *Natural hazards*, vol. 76, no. 2, pp. 1111–1141, 2015.
- [43] A. Guettiche, P. Guéguen, and M. Mimoune, “Seismic vulnerability assessment using association rule learning: application to the city of constantine, algeria,” *Natural hazards*, vol. 86, no. 3, pp. 1223–1245, 2017.
- [44] Y. Liu, Z. Li, B. Wei, X. Li, and B. Fu, “Seismic vulnerability assessment at urban scale using data mining and giscience technology: application to urumqi (china),” *Geomatics, Natural Hazards and Risk*, vol. 10, no. 1, pp. 958–985, 2019.
- [45] A. Morales-Esteban, F. Martínez-Álvarez, A. Troncoso, J. Justo, and C. Rubio-Escudero, “Pattern recognition to forecast seismic time series,” *Expert Systems with Applications*, vol. 37, no. 12, pp. 8333–8342, 2010.
- [46] G. Asencio-Cortés, F. Martínez-Álvarez, A. Troncoso, and A. Morales-Esteban, “Medium–large earthquake magnitude prediction in tokyo with artificial neural networks,” *Neural Computing and Applications*, vol. 28, no. 5, pp. 1043–1055, 2017.
- [47] H. Camargo and M. Silva, “Dos caminos en la búsqueda de patrones por medio de minería de datos: Semma y crisp,” *Tecnol Journal of Technology*, vol. 9, no. 1, pp. 18–25, 2010.
- [48] C. Cobos, J. Zuñiga, J. Guarín, and E. León, “Cmin-herramienta case basada en crisp-dm para el soporte de proyectos de minería de datos,” *Ingeniería e investigación*, vol. 30, no. 3, pp. 45–56, 2010.
- [49] J. M. Moine, A. S. Haedo, and S. E. Gordillo, “Estudio comparativo de metodologías para minería de datos,” *XIII Workshop de Investigadores en Ciencias de la Computación*, vol. 1, no. 1, pp. 278–281, 2011.
- [50] UNAM, “Reporte especial 07 de septiembre de 2017 tehuantepec(m 8.2) y 19 de septiembre de 2017 puebla-morelos (m 7.1). servicio sismológico nacional - unam.” Obtenido de: <http://www.ssn.unam.mx/sismicidad/reportes-especiales/>. Consultada el 14 de Septiembre de 2020.
- [51] C. C. Robusto, “The cosine-haversine formula,” *The American Mathematical Monthly*, vol. 64, no. 1, pp. 38–40, 1957.

BIBLIOGRAFÍA

- [52] J. Han, J. Pei, and Y. Yin, “Mining frequent patterns without candidate generation,” *ACM sigmod record*, vol. 29, no. 2, pp. 1–12, 2000.
- [53] G. García-Rueda, R. M. Valdovinos, J. Valdés-González, R. Alejo, J. L. González-Ruiz, and J. R. Marcial-Romero, “Analysis of repair costs of scholar buildings affected by earthquakes using data mining. case study: Earthquakes of 2017 in mexico,” in *Mexican Conference on Pattern Recognition*, pp. 45–56, Springer, 2020.

Anexos

Anexo A

Formato de dictamen A

Dictamen en el cual se colocan en el cuadro rojo la ubicación de los detalles estructurales como por ejemplo: tipo de escuela RC y Atipica.



SECRETARÍA DE EDUCACIÓN PÚBLICA
INSTITUTO NACIONAL DE LA INFRAESTRUCTURA FÍSICA EDUCATIVA
GOBIERNO DEL ESTADO DE CHIAPAS
INSTITUTO DE LA INFRAESTRUCTURA FÍSICA EDUCATIVA DEL ESTADO DE CHIAPAS



INSTITUTO DE LA INFRAESTRUCTURA FÍSICA EDUCATIVA DEL ESTADO DE CHIAPAS

CEDULA DE INFORMACION DE VALUACION DE DAÑOS POR DESASTRE NATURAL DENOMINADO:
"Daños ocasionados por el sismo de magnitud 8.2 ocurrido el 7 de septiembre en 82 municipios en el estado de Chiapas, se corrobora con fecha 8 de septiembre de 2017"

1.- DATOS GENERALES

NOMBRE: D.N. Felipe Carrillo Rendo NIVEL EDUCATIVO: Basico

LOCALIDAD: Toluca MUNICIPIO: Tapuachula

NÚMERO DE VALIDACIÓN DE CÉDULA: E-145 CON DAÑO SIN DAÑO

COORDENADAS: _____ ÁREA TOTAL DEL PLANTEL: _____

FECHA: 15- Sep - 2017

CCT (s): 07E2N04890

ALUMNOS: 60

DOCENTES: 2

ADMINISTRATIVO Y DE APOYO: 1

2.- INFRAESTRUCTURA MOBILIARIO, EQUIPO Y CONTENIDO

EDIFICIO	ESTRUCTURA	No. E. E.	ESPERIFICACION MATERIAL Y/O ACABADOS			INSTALACIONES				DESCRIPCIÓN GENERAL DEL DAÑO
			MUROS	CUBIERTA	PISOS	ELEC.	HID.-SAN.	AIRE-A.	ESP.	
A	R	1	Tabique	Concreto	Concreto	Si	NO	NO	NO	Sin daños
B	Atipica	1	Madera	Lamina	Concreto	Si	NO	NO	NO	Sin daños
C	Atipica	1	Concreto	Lamina	Concreto	Si	NO	NO	NO	Sin daños

3.- ÁREAS E INSTALACIONES EXTERIORES A LOS EDIFICIOS, ACCESOS Y ELEMENTOS DELIMITANTES DEL PLANTEL:

DESCRIPCIÓN DEL DAÑO: Con daños en barda perimetral

ELEMENTO	EXISTE	CIDANO	ELEMENTO	EXISTE	CIDANO	CROQUIS DE LOCALIZACIÓN	SELLO DE AUTENTICIDAD ESCOLAR
PLAZA CIVICA	Si	NO	LETRINA	Si	NO		
ANDADORES	Si	NO	ASTA BANDERA	Si	NO		
MURO DE ACCESO	Si	NO	BARDA PERIMETRAL	Si	Si		
MURO DE ACOMETIDA	Si	NO	PÓRTICO O PTA DE ACCESO	NO	NO		
REJA	Si	NO	TECHUMBRE CANCHA	NO	NO		
CANCHAS DEPORTIVAS	NO	NO	TECHUMBRE PLAZA CIVICA	Si	NO		
CERCA DE MALLA	Si	NO	RED ELÉCTRICA	Si	NO		
CISTERNA	NO	NO	RED SANITARIA	NO	NO		
FOSA SÉPTICA	Si	NO	RED HIDRAÚLICA	Si	NO		
POZO DE ABSORCIÓN	NO	NO	BEBEDEROS	NO	NO		
CUBO DE TINACOS	NO	NO	TAMPAS	NO	NO		
TANQUE ELEVADO	NO	NO	OTROS	NO	NO		
ILUMINACIÓN EXTERIOR	NO	NO					

ELABORÓ: Monte Carlo Fernando Madero

Vo.Bo. Ing. Rafael Rosas Sánchez

FORMATO FONDEM-1, 2, 3 1075

Formato de dictamen B

Dictamen en el cual la información estructural se encontraba de forma libre en la última sección de este.

30

Formato de captura de datos para evaluación estructural

Nombre del evaluador: _____

INFORMACIÓN GENERAL

Ingeniero o arquitecto Estudiante Ing/Arq Otro

Fecha: 2/09/17

Coordenadas: (_____ N _____ O _____ ms nm)

Nombre del inmueble: E. P. I. Leona Vicario

Calle y número: Montes de la Buena Vista 12, D.F. P.R. 211531

Código postal: 211531

Ciudad o estado: Mérida, Yucatán

Delegación/Municipio: H. Ayuntamiento de Mérida

Estado: Yucatán

Referencias: Frente al Escudo

Persona contactada/propietario: Br. ...

entre calles "A" y "B" un sitio notable: ...

Teléfono: 441 231 1234

Uso: Vivienda Hospital Oficina Iglesia Comercio Escuela Otro Desocupado

Reunión (cine/estadio/salón) Industrial (fábrica/bodega) No. niveles: 4

No. sótanos: 0

No. ocupantes: 477 m²

Dimensiones: Frente X = _____ m Fondo Y = _____ m

SISTEMA ESTRUCTURAL

La dirección X es paralela a la fachada, indicar X y Y en el croquis

Dirección X: Marcos de concreto Muros de concreto Columnas y losa plana Muros de carga de mampostería Muros de adobe o bahareque Muros de madera, lámina, otros

Dirección Y: Marcos de acero Marcos de concreto Columnas y losa plana Muros de carga de mampostería Muros de adobe o bahareque Muros de madera, lámina, otros

Muros de mampostería: Confinada Refuerzo interior Simple

Sistema de piso: Losa maciza Losa reticular Viguetas y bovedilla No se sabe

Sistema de techo: Igual al de piso Lámina Teja Otro: ...

Cimentación: Zapatas aisladas Zapatas corridas Cimiento de piedra Losa de cimentación

VULNERABILIDAD

Irregular en planta: Asimetría por muros, cubos, cargas Grandes aberturas, entrantes/salientes Geometría irregular en planta

Irregular en elevación: Planta baja de doble altura Muros no llegan a cimentación Planta baja flexible Columna corta

Posición en manzana: Esquina Medio Aislado

Grandes masas en pisos superiores Reducción brusca de pisos superiores Separación edif. vecino: _____ cm

EVALUACIÓN DE DAÑOS

Geotécnicos: Grietas en el terreno Hundimientos Inclinación del edificio: _____ %

Losas: Colapso Grietas máx. _____ mm Flecha máx. _____ cm

Conexiones: Falla

Otros daños: Vidrios Acabados Plafones Fachadas Bardas y pretiles Cubos (escalera/elevador) Instalaciones

Entrepiso crítico (más débil y/o más dañado): No. de columnas (o muros) daño severo = 5 (colapso, aplastamiento, pandeo, grietas > 3 mm)

Total de columnas (muros) en el entrepiso = _____

NIVEL DE DAÑO DE LA ESTRUCTURA

Colapso total Daño severo Daño medio Daño ligero

CRÓQUIS DEL INMUEBLE

Existen planos: Arquitectónico Estructural Ninguno

El inmueble data de año 1947 y 2 años de altura en (Marcar el Norte) N

con los muros que separan los muros de carga de la manzana de las aulas y la escuela, esta aula se considera de alta flexión por lo que no debe ocuparse. En cualquier momento las aulas hasta que se estén ocupando.

Anexo C

Formato de dictamen C

Dictamen en el cual la información estructural por secciones y cada una se encontraba marcada por cada característica correspondiente a la escuela.



22/septiembre/2017

Nombre de la Escuela:	" TELESECUNDARIA 290 "		
Clave:	13DTV0047U	Nivel educativo:	TELESECUNDARIA
Dirección:	TLAXCO		
Municipio:	METZITLAN	Estado:	Hidalgo

CÉDULA PARA EL DIAGNÓSTICO O EVALUACIÓN DE DAÑOS OCASIONADOS A LAS ESCUELAS PÚBLICAS DE EDUCACIÓN BÁSICA POR LOS SISMOS DE SEPTIEMBRE DE 2017

Señale las anomalías existentes en la escuela:

Accesos	Muros	Techos	Pisos
Puertas atascadas	Grietas		
Ventanas atascadas	Inclinados		
Rejas atascadas	Colapsados		
	Con desprendimientos		
Cables tirados	Grietas		
Postes tirados o inclinados	Inclinados		
Lámparas colgadas	Colapsados		
Focos colgados	Con desprendimientos		
Olor a gas	Grietas		
Mangueras rotas	Hundimientos		
Mangueras desconectadas	Encharcamientos severos		
Estructuras metálicas sueltas			
Deslaves aledaños			
Desbordamiento de ríos cercanos			
Colinda con edificios en mal estado			

Descripción de los daños:

Valoración de los daños conforme a la calificación establecida (seleccione la opción):

- Reconstrucción total
- Reconstrucción parcial
- Daño menor

Observaciones del evaluador: El daño fue ocasionado el 19 de Septiembre por Sismo

El que suscribe, con Cédula Profesional No. 4705776, realizó la evaluación de daños en el Plantel cuyos datos se anotan al inicio de este instrumento.	Responsable de la evaluación:
	T.C. CARLOS RIVERA REYES Nombre y firma
	Instituto Hidalguense de la Infraestructura Física Educativa