



UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA
FACULTAD DE INGENIERÍA
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO
MAESTRÍA EN INGENIERÍA EN COMPUTACIÓN

“ANÁLISIS MULTIDIMENSIONAL DE DATOS SISMOLÓGICOS”

TESIS

PARA OBTENER EL GRADO DE:

MAESTRÍA EN INGENIERÍA EN COMPUTACIÓN

PRESENTA

ELIZABETH BURROLA MÁRQUEZ



UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA
FACULTAD DE INGENIERÍA
MAESTRÍA EN INGENIERÍA EN COMPUTACIÓN

ANÁLISIS MULTIDIMENSIONAL DE DATOS SISMOLÓGICOS

APROBADO:

M.I. Jesús Roberto López Santillán, director

M.A. José Saúl de Lira Miramontes, sinodal

Dra. Graciela María de Jesús Ramírez Alonso,
sinodal

CHIHUAHUA, CHIH.

AGOSTO 2016

Derechos reservados
© Elizabeth Burrola Márquez
Facultad de Ingeniería, Circuito Universitario Campus II
2016

Copyright ©

por

Elizabeth Burrola Márquez

2016

Agradecimientos

Agradezco a Dios por haberme acompañado y guiado, por ser mi fortaleza en los momentos de debilidad y por brindarme una vida llena de aprendizajes y experiencias.

Le doy gracias a mis padres Guillermo y Lucina por apoyarme en todo momento, por los valores que me han inculcado, por su paciencia y palabras sabias, pero sobre todo, por ser un ejemplo de vida.

A mis hermanos, por ser parte importante de mi vida, por haberme apoyado en las buenas y en las malas.

Gracias a mi director de tesis M.I. Roberto López, a mis asesores la Dra. Graciela Ramírez y M.A. Saúl de Lira, gracias por su entrega y dedicación.

Sin el apoyo del Dr. Ignacio Reyes, M.I. Rodrigo de la Garza, M.I. Isabel Vázquez esto no hubiera sido posible.

Por último, pero no menos importante, quiero agradecer a las instituciones que hicieron posible esta tesis, a CONACyT por darme la oportunidad de seguir con mis estudios de una manera más fácil otorgándome una beca la cual fue necesaria para el autodesarrollo y sustento dentro del tiempo requerido de mis estudios. A la Universidad Autónoma de Chihuahua, por su excelente trabajo y nivel académico lo cual abrió una Maestría que atrajo mi atención. A la Facultad de Ingeniería por haberme recibido con las puertas abiertas, por sus instalaciones y su equipo de docente especializados en cada área para aprender de ellos. Al Servicio Sismológico Nacional, por las dudas resueltas y el acceso a los datos.

Resumen

En la actualidad los daños ocasionados por los sismos han dejado grandes pérdidas tanto materiales como humanas, esto obliga a buscar herramientas y/o diseñar nuevos sistemas eficaces para el análisis de datos sismológicos.

Aunque existen procesos que permiten realizar estas tareas, no siempre se obtienen los resultados adecuados y en algunas ocasiones éstos se basan en un análisis de datos no confiables, esto es, la fuente de los datos proviene de un muestreo del cual no se puede tener una perspectiva amplia y definitiva, debido a que la cantidad de datos que se obtienen es demasiado voluminosa (millones de registros por día), por ello se realiza un muestreo. Sin embargo, en ocasiones es necesario hacerlo debido a la poca disponibilidad de equipo de cómputo que soporte las grandes cantidades de datos que se deben analizar, aunque hacer esto conlleva una pérdida de datos, ya que procesar la totalidad de datos requiere grandes capacidades de equipo, además de una inversión fuerte de tiempo para el procesamiento de los mismos.

La destreza que se debe manejar para tener un análisis oportuno de los sismos, está basada en las herramientas con que se cuenta. Uno de los aspectos que debe considerarse es la cantidad de datos que se pueden consultar en un determinado tiempo, otro aspecto es la capacidad para visualizar la información desde varios puntos de vista. Involucrando éstas alternativas se puede tener un sistema integral para representar los datos y analizarlos dependiendo del contexto que se quiera manejar, y con ello, determinar las características geológicas, como puede ser la rigidez del suelo por donde se propagan las ondas, entre otras propiedades; así mismo existen otros aspectos a tomar en cuenta de otras disciplinas, es decir, teniendo datos sismológicos, se puede obtener como resultado qué tipos de técnicas es conveniente utilizar en determinadas áreas de estudio, o los daños en las estructuras están siendo causadas por los sismos. En esta propuesta de tesis se plantea el desarrollo de una herramienta que permita el análisis de grandes cantidades de datos con el fin de obtener conocimiento sobre los movimientos telúricos y la localización de los

mismos. Para ello, es vital conformar una base de datos que almacene esta información de una manera estructurada, teniendo sólo aquellos datos que son útiles para la determinación de los aspectos antes mencionados.

Una herramienta muy útil en el caso de tener cantidades de datos muy grandes es el conocido *cubo OLAP*, ya que optimiza la capacidad de análisis a través de la visualización gráfica de la relación de los datos. Además de este tipo de implementaciones, se puede obtener información más detallada mediante la implementación de algoritmos de clasificación y técnicas de reconocimiento de patrones, como *Random Forest*, *Decision Trees*, *K Nearest Neighbors*, *Support Vector Machine*, *Artificial Neural Network*, aunque existen muchos más, estos son algunos de los más utilizados.

Índice de contenido

Agradecimientos	3
Resumen	5
Índice de contenido.....	7
Índice de Figuras	9
Índice de Tablas.....	10
Capítulo.....	11
1.1. Antecedentes.....	12
1.2. Problema de investigación.....	16
1.3. Pregunta de investigación.....	16
1.4. Objetivo general	17
1.5. Objetivos específicos.....	17
1.7. Justificación.....	17
1.8. Aportación práctica.....	18
1.9. Objeto de estudio	18
1.10. Ideas a defender	18
Capítulo II.....	19
2.1. Sismos.....	19
2.2. Data Warehouse.....	29
2.3. <i>Business Intelligence</i>	32
2.4. OLAP.....	35
2.5. Bag of Words.....	37
Capítulo III.....	45
3.1. Actividades para el pre procesamiento de datos.....	45
3.2. Actividades para la creación del almacén de datos	46

3.3. Actividades para la conformación del <i>cubo OLAP</i>	47
3.4. Actividades para la búsqueda de patrones	48
Capítulo IV.....	49
4.1. Resultados de Actividades del pre procesamiento de los datos.....	49
4.2. Resultados de las Actividades de la creación del almacén de datos.....	57
4.3. Resultados de Actividades de la construcción del cubo OLAP.....	62
4.4. Resultados de Actividades de la aplicación de técnicas de reconocimiento de patrones.....	66
Capítulo V.....	74
Referencias.....	76

Índice de Figuras

Figura 2. 1: Distribución de placas tectónicas.....	19
Figura 2.2: Zona de convergencia con Subducción.....	20
Figura 2.3: Frontera del tipo divergente, o zona de “acreción”.....	20
Figura 2.4: Ejemplo de frontera transcurrente.....	21
Figura 2.5: Placas tectónicas de México	22
Figura 2.6: Ejemplo de Epicentro.....	24
Figura 2. 7: Relación del epicentro con la estación sísmica.....	25
Figura 2.8: Obtención de la magnitud de un sismo a través del método de Richter por medio de un sismograma.....	27
Figura 2.9: Comparación entre estilos Inmon y Kimball.....	32
Figura 2.10: Segmentación de la señal	38
Figura 2.11: Formación de los CodeWords.....	38
Figura 2.12: Representación del CodeBook.....	39
Figura 2.13: Aproximación del algoritmo KNN	41
Figura 2. 14: Ejemplo de árbol de decisión.....	42
Figura 2.15: Ejemplos de separación mediante SVM	43
Figura 2.16: Diagrama de la red neuronal artificial.....	44
Figura 3.1: Actividades para el pre procesamiento de datos.....	45
Figura 3.2: Actividades para la creación del almacén de datos.....	46
Figura 3.3: Actividades para la conformación del cubo OLAP	47
Figura 3.4: Actividades para la búsqueda de patrones.....	48
Figura 4.1: Aceleraciones de enero, febrero y marzo de 2013, estación HPIG.....	55
Figura 4.2: Aceleraciones de enero, febrero y marzo de 2013, estación HSIg.....	55
Figura 4.3: Aceleraciones de enero, febrero y marzo de 2013, estación CGIG.....	56
Figura 4.4: Aceleraciones enero-marzo, estaciones HPIG, CGIG, HSIg.....	56
Figura 4.5: Aceleraciones enero estación CGIG.....	57
Figura 4.6: Carga de los archivos diarios al almacén de datos.....	58
Figura 4.7: Creación de la dimensión de Aceleraciones	60
Figura 4.8: Diagrama del almacén de datos	61
Figura 4.9: Datos de la tabla de hechos.....	62
Figura 4.10: Esquema del cubo	63
Figura 4.11: Esquema de relación Resultante-Magnitud-Profundidad	63
Figura 4.12: Esquema de relación Magnitud-Latitud-Longitud-Profundidad.....	64
Figura 4.13: Resultante vectorial.....	68
Figura 4.14: Ejemplo de histograma.....	68
Figura 4.15: Segmentos de señales.....	69
Figura 4.16: Matrices de confusión para los distintos algoritmos.....	71
Figura 4.17: Estructura de la red neuronal	72
Figura 4.18: Diagrama de relación entre datos y procesos.....	73

Índice de Tablas

Tabla 2.1: Grados y características habituales en la escala de Mercalli	27
Tabla 4.1: Ejemplo de datos de aceleraciones	49
Tabla 4.2: Datos menores a la moda, 1 de enero de 2013, estación HPIG	51
Tabla 4.3: Datos interpolados, día 1 de enero de 2013, estación HPIG	54
Tabla 4.4: Parámetros de configuración de los algoritmos	69
Tabla 4.5: Resultados de exactitud para los distintos algoritmos	70

Índice de Gráficas

Gráfica 4.1: Relación Magnitud-Profundidad en el cubo de datos	64
Gráfica 4.2: Relación Magnitud-Profundidad publicadas por el SSN	65
Gráfica 4.3: Comparación de eventos sísmicos en los años 2013 y 2014	65
Gráfica 4.4: Comparación de las resultantes vectoriales	66

Capítulo I

Introducción

En la naturaleza existen fenómenos que están ocurriendo a cada instante y que pueden estar provocando, directa o indirectamente, algún perjuicio en contra de casas o edificios. Algunos de estos fenómenos pueden ser, si se tratara de un área rural, el caminar del ganado, el paso del tren, la explosión controlada de una mina, entre otros sucesos. Todas estas actividades que pueden estar ocurriendo constantemente, sin duda están provocando movimientos telúricos, debido a la fricción, o al golpe contra la estructura terrestre. Si se tuviera en este lugar un sismómetro, éste estaría midiendo todos estos movimientos, los cuales tienen longitudes de onda muy pequeñas en comparación con los movimientos telúricos ocasionados por perturbaciones en la corteza terrestre (sismos). Las perturbaciones antes mencionadas, pueden tener varios orígenes, desde el choque de placas tectónicas, hasta la liberación de energía acumulada en alguna de las capas interiores de la tierra. El sistema terrestre está formada por cuatro capas: hidrósfera, atmósfera, biósfera y geósfera. La hidrósfera está formada por el agua en sus tres estados e incluye el hielo, el vapor de agua y el agua líquida en la atmósfera, el océano, los lagos y los ríos, así como el agua subterránea. La atmósfera es la capa de gas que rodea la Tierra. La biósfera incluye todos los seres vivos; se encuentran en muchas partes de la hidrósfera y la atmósfera así como en la zona superficial de la geósfera. El ser humano forma parte de la biósfera. La geósfera incluye un núcleo metálico, magma, rocas, sedimentos y suelo. La geósfera, a su vez, se puede visualizar en capas; litósfera, astenósfera, mesósfera y endosfera. La litósfera comprende los cien primeros kilómetros, que involucra toda la corteza además de la parte más externa y sólida del manto. Astenósfera, se encuentra por debajo de la litósfera hasta unos 700 km. de profundidad desde la superficie. Mesósfera, bajo la astenósfera se encuentra la mesósfera hasta unos 2,700 km. de profundidad. Endósfera, Corresponde a unos 200 km del manto (su parte más profunda) y a la totalidad del núcleo. Los sismos ocurren en la geósfera (Emilio Pedrinaci, 2013).

Los eventos sísmicos pueden suceder en cualquiera de las capas de la geósfera antes mencionadas, aunque cabe señalar que a mayor profundidad, mayor será el impacto que tendrá (Zúñiga, 2011).

Independientemente de las causas, la cantidad de información que se genera cuando ocurre un evento sísmico es comparable en cantidad, a la que emiten las estaciones meteorológicas, algunos procesos físicos, sistemas biológicos, entre otros; que puede llegar a cientos o miles de datos por segundo (Fernández, 2015). Con esta gran cantidad de información, es complicado realizar una toma de decisiones rápida y eficaz que abarque todos los posibles riesgos que ella implica.

1.1. Antecedentes

En los últimos años, grandes terremotos han afectado a las ciudades en países de todo el mundo: China en 2009, Japón en 2011, y en Italia en 2009 y 2012, también en el caso del terremoto de magnitud 7 en Haití en el año 2010, donde el gobierno haitiano reportó un estimado de 230 000 muertos, 300 000 heridos, y aproximadamente 1,3 millones de personas desplazadas; el daño económico fue estimado entre \$8 000 y \$14 000 millones de dólares (Brett, 2013). En México, el Servicio Sismológico Nacional (SSN) reporta en su página de internet <http://www.ssn.unam.mx/> los sismos que ocurren diariamente en el país, así como también ofrece los sismos históricos. Según Quintana (2008) en el estado de Chihuahua han ocurrido sismos de gran importancia, como el sucedido en la ciudad de Hidalgo del Parral en 1928 (Quintana, 2008).

Detectar sismos implica hacerlo en cuestión de segundos, debido a que los sensores ubicados en ciertos lugares específicos no permiten la detección inmediata, sino hasta que el sensor ha recibido la señal electromagnética, y aunque esta señal dura unos cuantos segundos, es necesario realizar el análisis del tipo de onda, frecuencia, y posible origen de la misma, el determinar si ésta fue generada por un sismo o alguna otra causa como pueden ser rayos, explosiones mineras, entre otros, lo cual puede tardar más de unos segundos (Tsutsui, 2014). Estrada (2012) llama microsismos a todos aquellos movimientos que

aparecen débilmente y que son causados por el movimiento de los árboles debido al viento, algunos otros por el paso de los vehículos o trenes, en cambio, los terremotos son causados por la liberación de energía acumulada. Estos últimos son capaces de causar grandes daños y pérdidas humanas (Estrada, 2012).

Tradicionalmente los sismos se detectan utilizando herramientas topográficas analizadas por expertos. Como lo menciona Wang (2011) las ondas generadas en las capas terrestres se examinan utilizando el algoritmo FDTD (*finite-difference time-domain*) donde las ondas se filtran y normalizan para obtener una mejor visualización del tiempo de atenuación (Wang Y. , 2011). Estos métodos tradicionales son empleados para la detección de los sismos lo más tempranamente posible. Con ello se pretende realizar acciones oportunas que prevean la seguridad de las personas al momento en que está ocurriendo un sismo, y en la medida de lo posible, la detección de éste unos instantes anteriores. El SSN detecta los sismos de acuerdo a la distancia del epicentro respecto a la base sismográfica, la cual es calculada a partir de las ondas registradas por un sismógrafo. Con la distancia se genera un círculo de radio igual a esta. La dirección del epicentro es obtenida a partir de tres estaciones, para las cuales también se generó este círculo. Teniendo la intersección de estos tres círculos, se obtiene el área posible de ocurrencia del sismo. Este método es impreciso, y en la práctica resulta tardío. Para aumentar la precisión, es necesario aumentar la cantidad de estaciones que se analizan, con lo que se reduce el área de ocurrencia del sismo (SSN).

Existen otras investigaciones encaminadas al análisis posterior a los eventos sismológicos. Debido a que estos eventos provocan daños no solo al área terrestre más próxima donde ocurren, sino que afectan a una gran cantidad de bienes tanto muebles como inmuebles, es necesario determinar el daño que está siendo provocado a causa de estos fenómenos. Ya que los sismos comúnmente son de diferentes magnitudes, por lo tanto la frecuencia y amplitud de las ondas varía constantemente (Hovem, 2012). Este es el caso de las investigaciones realizadas por Monteiro (2015) en donde presenta dos estudios

mostrando que influyen de manera directa la medida de la intensidad del sismo, en especial los movimientos que estaban teniendo las capas terrestres, así como las características que presentan los edificios (Monteiro, 2015).

Por otro lado, se han realizado análisis a las señales percibidas por acelerómetros, como es el caso de la investigación realizada por González y otros (2016), donde analiza las ondas captadas por los acelerómetros instalados en celulares, para caracterizar las señales de acuerdo a las perturbaciones percibidas cuando los automóviles pasan por irregularidades del camino, como son topes, baches o boyas (González, Moreno, Escalante, Martínez, & Carlos, 2016).

Pocas investigaciones emplean sistemas informáticos para la detección de sismos, una de ellas la propone Konstantaras (2007), donde utiliza dos modelos neuro-fuzzy o neuro-difuso (redes neuronales con habilidades de lógica difusa). El primer modelo se utiliza para recuperar las señales del campo electromagnético en periodos de severas perturbaciones y un segundo modelo empleado para detectar posibles señales precursoras de un sismo, comparando ambos modelos y obteniendo una señal resultante de campo paralela a la distorsión detectada, mostrando como resultado el origen de la onda, que puede ser producida por un sismo, explosión minera, o algún otro fenómeno (A. Konstantaras, 2007). Otro modelo presentado es el que propuso Roberto (1992). Es un sistema de reconocimiento de patrones utilizando redes neuronales en el cual implementó una estación de trabajo para la detección de sismos, analizando la señal recibida por un sensor sismográfico. Los resultados de este proyecto fueron afectados por la incompatibilidad de señales detectadas, análisis incorrecto de los datos, provocando con esto falsas alarmas y en algunos casos no se detectó el sismo que en ese momento estaba ocurriendo (Roberto, 1992).

Estos modelos, solo analizan las señales actuales de los sismógrafos, y no proporcionan una relación con datos obtenidos en fechas anteriores, otra desventaja que presentan es la cantidad de datos que analizan, (aproximadamente 10% del total de datos)

respecto al total de datos obtenidos de un sismógrafo (millones de datos), aunado a esto, los resultados de estos estudios demuestran la falta de precisión y el deficiente análisis de los datos, así como también la ausencia de una aplicación práctica de estos estudios. Por lo tanto, tomar una decisión teniendo en cuenta los millones de datos que se tienen realmente es una tarea arriesgada para los encargados de realizarla. Debido a esto, en esta investigación se propone la implementación de herramientas informáticas aplicando técnicas de análisis de la información a través de inteligencia de negocios teniendo en cuenta una gran cantidad de datos almacenada.

Uno de los objetivos del almacenamiento de datos es proveer disponibilidad de la información que dé apoyo a las gestiones de toma de decisiones debido al impacto que estas tienen una vez que se ha resuelto llevarlas a cabo (Sen, 2006). Toda organización debe disponer, en general, de la habilidad de reconocer nuevos valores de información (externos o internos), asimilarlos y aplicarlos efectivamente. Sin embargo, se pueden tener datos empíricos, datos de eventos reales o una combinación de ellos y en base a estos realizar el análisis (Jazdani, 2013). También se puede calcular la probabilidad de que los eventos descritos por estos datos, vuelvan a ocurrir.

Una vez que se tienen los datos ordenados en un almacén, es posible aplicar una serie de herramientas informáticas (*business intelligence*) que faciliten la visualización y relación que existe entre ellos, pudiendo ser usadas para extraer de manera significativa la información más relevante y precisa aumentando la aceptación de la calidad del soporte de toma de decisiones (Nofal, 2013). Tal es el caso del estudio realizado por Hallo Carrasco (2006), donde muestra la implementación de un sistema informático que permite visualizar los eventos sísmicos mediante la especificación de algunos parámetros como la zona geográfica, o el tiempo en que ocurrieron los eventos (Hallo Carrasco, 2006).

Con el fin de reducir costos, tiempo de entrega y de incrementar la eficiencia y calidad, las organizaciones están forzadas a presentar un sistema integral de información

que les permita realizar estas actividades de la manera más inteligente posible (Villard, 2013).

En el estudio realizado por Zhiliang(2008) señala que integrar un sistema de soporte para la solución de proyectos basados en estandarización de documentos utilizando inteligencia de negocios, almacenamiento de grandes volúmenes de datos y otras herramientas es de gran utilidad para la toma de decisiones fiable y efectiva por parte del personal (Zhiliang, 2008).

Por todo lo antes expuesto, en esta propuesta de tesis se propone la aplicación de *business intelligence* y técnicas de reconocimiento de patrones a un conjunto de datos considerable (millones de registros) almacenados en una base de datos, para realizar un análisis y en base a ello, los expertos geólogos o civiles puedan tomar decisiones.

1.2. Problema de investigación

Sin duda los constantes movimientos sísmicos están afectando la vida diaria de las personas, causado afecciones que van desde el miedo, hasta el daño provocado a los edificios y todo bien material. Sin embargo, esta situación, aunque es recurrente en algunas zonas, no tiene la atención debida en cuanto al análisis que de ella pudiera surgir, desaprovechando la oportunidad de obtener indicios que lleven a una mejor calidad de vida en esas regiones.

Por ello, si se analiza cada situación en donde se está presentando un evento sísmico, se puede llegar a tener estimaciones que faciliten la comprensión de la situación actual e histórica de los sismos entre 3 y 6 en la escala de Richter ocurridos en el estado de Chihuahua. Cabe señalar que estas evaluaciones serán realizadas por los expertos geólogos y civiles, pudiendo tomar en cuenta los resultados obtenidos en los análisis de las técnicas aquí presentadas.

1.3. Pregunta de investigación

¿Cómo analizar los datos sísmicos para obtener información precisa de actividades sísmicas?

1.4. Objetivo general

Desarrollar un sistema que muestre distintas visualizaciones basadas en el análisis de datos obtenidos de eventos sísmológicos.

1.5. Objetivos específicos

- Pre procesar los datos para tomar en cuenta solo la información necesaria.
- Establecer un almacén de datos que concentre una gran cantidad de información sobre sismos. En base a datos de sensores instalados en lugares predeterminados se concentrará la información requerida.
- Generar una estructura multidimensional que permita visualizar desde distintas perspectivas (fecha, magnitud, etc.) la información del almacén de datos.
- Buscar patrones intentando encontrar tendencias en los datos del almacén.

1.6. Hipótesis

Con un almacén de datos de aceleraciones de sismos, con magnitudes entre 3 y 6 en la escala de Richter, es posible visualizar la información relevante.

1.7. Justificación

Los sismos son fenómenos que no se pueden prevenir y que causan grandes pérdidas, tanto materiales como humanas, es por ello que el análisis de los datos debe hacerse de forma oportuna y eficaz. Los estudios realizados en este aspecto muestran un abanico muy angosto de opciones en la utilización de sistemas informáticos para el análisis de datos. Basándose en estos argumentos se considera primordial el desarrollo de un sistema que presente información basándola en un análisis sobre una cantidad considerable de datos que provea un compendio de resultados que muestren la cantidad de sismos, así como también los movimientos más relevantes que presentan en determinado tiempo.

En esta investigación se trabaja en conjunto con expertos (geólogos y de otras áreas) de la Facultad de Ingeniería de la UACH para llevar a cabo las investigaciones convenientes respecto a este tema.

1.8. Aportación práctica

La habilidad de detectar sismos con gran exactitud es elemental al realizar planificaciones y toma de decisiones en aspectos interdisciplinarios, como de ingeniería civil, geológica y áreas de gobierno, entre otros. Analizar los eventos sísmicos en un área geográfica, nos permite llevar a cabo estas tareas de una manera eficaz, lo que apoyará en gran medida a empresas privadas y de gobierno en la realización y mantenimiento de obras civiles.

1.9. Objeto de estudio

Aceleraciones sísmicas de estaciones del estado de Chihuahua.

1.10. Ideas a defender

Teniendo en cuenta las aceleraciones sísmicas es posible visualizar el comportamiento de los eventos sísmicos para determinar la cantidad de perturbaciones que se está teniendo en alguna zona particular.

Con esto, se puede tener un sistema integral para las aceleraciones con la capacidad de análisis en distintas dimensiones.

Capítulo II

Marco Teórico

2.1. Sismos

Los sismómetros son instrumentos utilizados para medir las perturbaciones que se generan en cualquier capa de la geósfera, siendo más común observar movimientos en la litósfera. Estas mediciones se llevan a cabo registrando los valores de las aceleraciones que se producen cuando hay movimientos de las placas tectónicas, o cuando se libera energía acumulada en las capas terrestres. El planeta Tierra está formado por diversas placas tectónicas, que con el paso de los años han estado moviéndose, lo que provoca que algunas placas choquen entre sí (Quintana, 2008).

Se tienen principalmente dos tipos de placas tectónicas: continentales y oceánicas, como su nombre lo indica, las placas continentales son aquellas que se encuentran dentro de la parte terrestre de los continentes; las placas oceánicas pertenecen a la parte que corresponde a los océanos (Zúñiga, 2011). La figura 2.1 muestra la distribución de las principales placas.

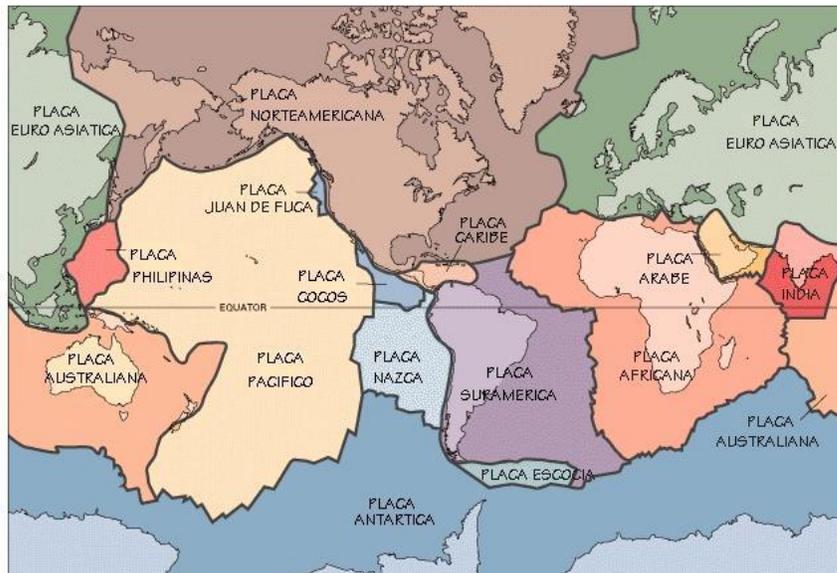


Figura 2. 1: Distribución de placas tectónicas.

Se pueden tener tres casos principales de fronteras entre placas:

- a) *Convergentes*: Las placas han tenido una “colisión”, y por lo general, la placa más densa penetra por debajo de la otra. La figura 2.2 muestra la colisión entre las placas.

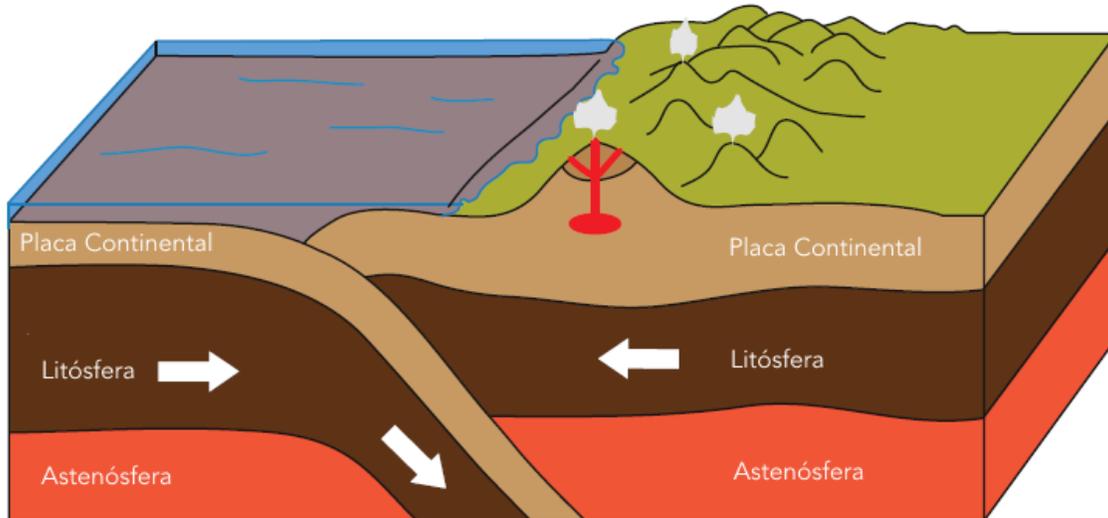


Figura 2.2: Zona de convergencia con Subducción.

- b) *Divergentes*: Las placas se separan de la frontera en dirección opuesta, partiendo la frontera. La figura 2.3 presenta la manera en que se separan las placas.

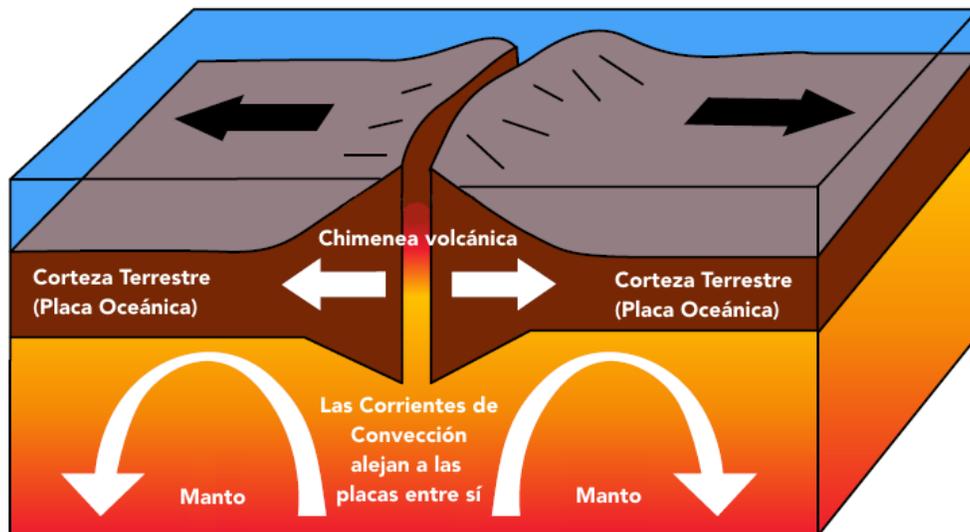


Figura 2.3: Frontera del tipo divergente, o zona de “acreción”.

c) *Transcurrentes*: En este tipo de fronteras, las placas se mueven con desplazamientos laterales, es decir, pasan una junto a la otra. En la figura 2.4 se ilustra este tipo de eventos.

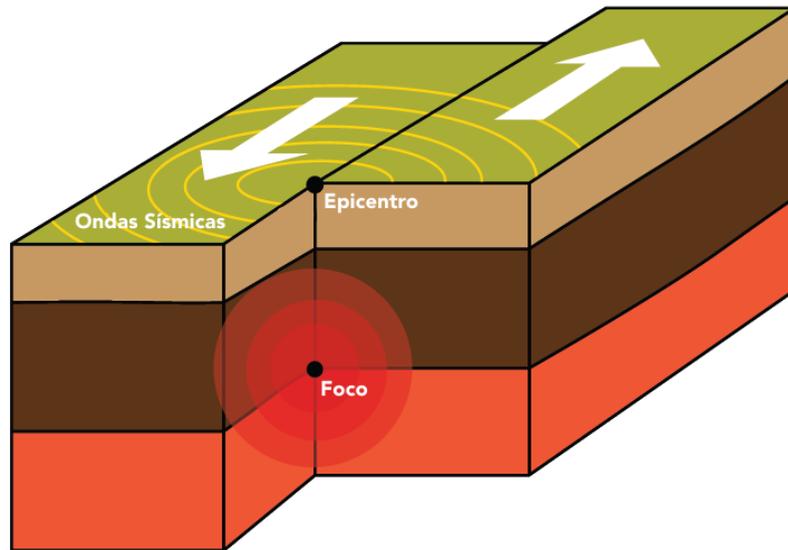


Figura 2.4: Ejemplo de frontera transcurrente.

Los sismos son rupturas en las rocas que provocan movimientos de las capas terrestres, como menciona Zúñiga (2011) existen varios tipos de sismos, como son telúricos (movimiento de placas tectónicas), algunos otros causados por la actividad volcánica o el colapso de subsuelo debido a la extracción de fluidos o materiales (Zúñiga, 2011). También menciona que los sismos pueden ser clasificados, debido a que cada sismo puede tener distintos tipos de fallas o medios de propagación por el cual viajen las ondas generadas por el movimiento de la ruptura. Se clasifican, dependiendo de la zona de generación y la profundidad, en los siguientes:

- a) Sismos de subducción someros.
- b) Sismos de subducción profundos.
- c) Sismos intraplaca de profundidad intermedia.
- d) Sismos de zonas de acreción.

- e) Sismos de fallas de transcurrencia.
- f) Sismos corticales intracontinentales.

Debido a que las placas tectónicas están en continuo movimiento, se pueden presentar temblores o sismos tectónicos en cualquier momento. Dado esto, se analiza, según Quintana (2008) la posibilidad de ocurrencia de algún sismo, en este caso, en la ciudad de Chihuahua, donde se han presentado históricamente sismos en esta zona a causa del movimiento de las placas de Norteamérica y Pacífico, que generan fallas que continúan hasta Estados Unidos, donde se conoce como Falla de San Andrés (Quintana, 2008). Estas placas se presentan en la figura 2.5, mostrada a continuación.

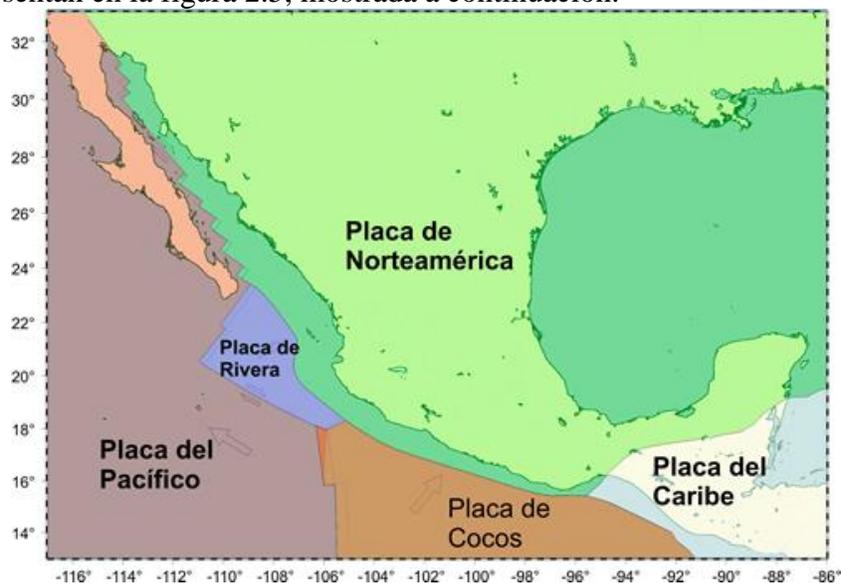


Figura 2.5: Placas tectónicas de México

El riesgo de que ocurra un sismo de magnitud grande (mayor a 4, según los expertos) que afecte considerablemente la entidad de Chihuahua es relativamente pequeño, considerando su ubicación respecto a las fronteras de las placas Norteamérica y Pacífico. Sin embargo, como se mencionó anteriormente, las perturbaciones que se producen diariamente pueden ocasionar rupturas o daños a las estructuras de casas y edificios, lo que genera un riesgo para los habitantes de dicha localidad. Dado que los sismos pueden ocurrir no solo en los límites de las placas tectónicas, sino en cualquier lugar donde se esté acumulando energía en la litósfera.

Teniendo en cuenta esto, se analizan los posibles orígenes de los sismos, para con ello, tomar las debidas precauciones.

- *Ondas sísmicas*

Los aparatos para medir los sismos pueden variar. Se pueden utilizar, tanto sismómetros de alcance local (datos a nivel nacional), sismómetros de alcance mundial (datos de estaciones lejanas), así como acelerómetros. Estos últimos tienen un alcance muy corto, pero de igual forma pueden ser utilizados dentro de un área reducida, como puede ser una ciudad (Tsutsui, 2014).

Dependiendo del tipo de instrumento que se utilice, es la capacidad de respuesta que se tendrá para generar una alerta sísmica, en el caso en que se lleve a cabo un análisis en tiempo real de las ondas. A mayor distancia las estaciones del lugar de origen del sismo, mayor será el tiempo que se deba tomar para determinar el lugar donde ocurrió dicho sismo, su magnitud e intensidad.

En México, el SSN proporciona las aceleraciones obtenidas por los sismómetros de alcance local, respecto a las estaciones sísmicas con que cuenta. En este trabajo de investigación se tomarán en consideración tres estaciones sísmicas:

a) CGIG.

Ubicación: Latitud: 30.365044, Longitud: -107.977352

Ciudad: Casas Grandes, Chihuahua.

b) HPIG.

Ubicación: Latitud: 26.93679, Longitud: -105.66399

Ciudad: Parral, Chihuahua.

c) HSIg.

Ubicación: Latitud: 29.019754, Longitud: -110.94928

Ciudad: Hermosillo, Sonora.

Cada una de estas estaciones registra constantemente las aceleraciones que se producen. Los sismómetros instalados registran cien conteos por segundo, es decir, cada

segundo registran cien valores para la aceleración. Esto lo hacen en tres ejes; el eje Norte (N), eje Este (E) y eje perpendicular a la superficie terrestre (Z). Con esto se forma un espacio tridimensional, del cual, el sismómetro registra las aceleraciones de cada eje, y en base a ellas, se pueden calcular los datos antes mencionados (Estrada, 2012). Es necesario tener por lo menos 3 estaciones sísmicas para determinar la localización del origen del movimiento sísmico.

Cuando se presenta un movimiento sísmico, éste se produce en la parte interior de las capas terrestres, el punto exacto donde se origina se conoce como *hipocentro* o *foco*, la proyección de éste sobre la superficie terrestre recibe el nombre de *epicentro*. Figura 2.6.

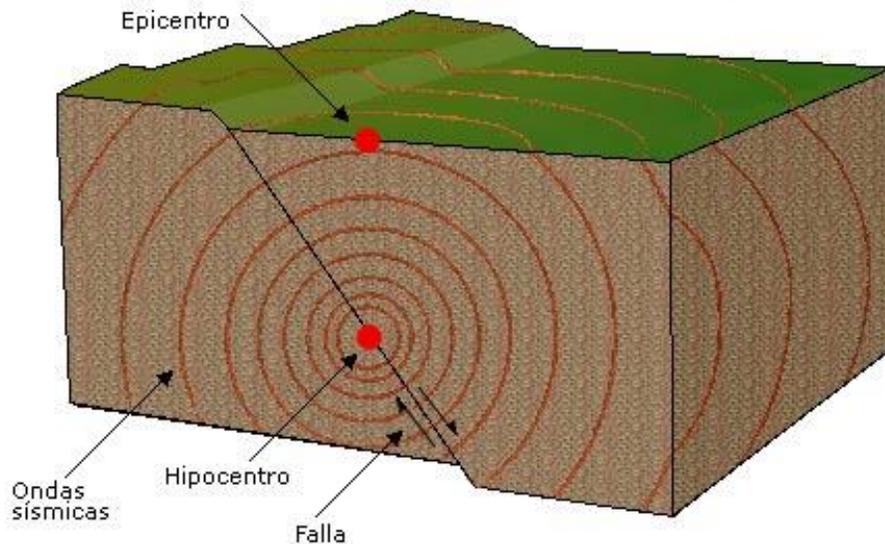


Figura 2.6: Ejemplo de Epicentro.

A partir de que las ondas sísmicas son generadas, viajan hacia la estación donde son registradas. La figura 2.7 muestra la relación de las ondas sísmicas con el epicentro. Existen dos tipos de ondas, las ondas *primarias* (P) y las ondas *secundarias* (S), las ondas P viajan más rápidamente y son registradas primero que las ondas S.

- *Magnitud*

Respecto a la medición de los sismos, existen diferentes técnicas para medirlos. La magnitud consiste en una medida cuantitativa de la energía intrínseca que cada evento tiene (Zúñiga, 2011). Según los expertos, la magnitud debe ser un valor único en cada evento, y

este número no debe verse afectado por las consecuencias que tenga el sismo, ya que éstas dependen de la zona en la que ocurra el sismo.

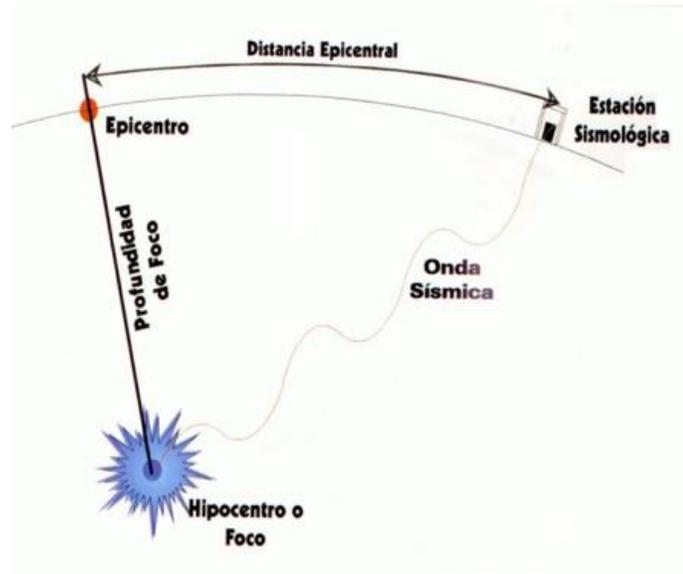


Figura 2. 7: Relación del epicentro con la estación sísmica.

La magnitud, aunque es un valor único, puede cambiar dependiendo de la técnica utilizada y del lugar donde se calcule, es decir, si la estación que detecta el sismo se encuentra demasiado lejos del lugar donde ocurrió, la medición se verá afectada. Esto complica la obtención de una medida certera, por ello, los expertos utilizan solo una de las distintas técnicas y respecto a ella deben informar a la población del suceso ocurrido. La elección de la técnica puede variar, comúnmente se utiliza la escala de Richter.

- *Escala de Richter*

Una de las contribuciones más importantes que hizo Charles Richter fue el descubrir que las ondas sísmicas propagadas por todos los terremotos pueden proporcionar buenas estimaciones de sus magnitudes. La obtención de la magnitud mediante esta escala consiste en asociar la magnitud con la amplitud de la onda sísmica (Edwards & Douglas, 2014). Encontrar la longitud, profundidad y desplazamiento de una falla puede llevar varios días, semanas o incluso meses después de un gran sismo. El mapeo de las fallas sísmicas que hacen los geólogos o el dibujo de la distribución espacial de las réplicas como hacen los

sismólogos, puede proporcionar estos parámetros después de un considerable esfuerzo. Pero, dado que algunos sismos, no presentan ni fallas superficiales ni tienen suficientes réplicas para poder estimar su magnitud de una forma convencional, resulta de utilidad poder calcular rápidamente la magnitud de un evento, así que ahí es donde se verifica la relevancia de obtener datos precisos en el menor tiempo posible.

La figura 2.8 muestra cómo se utiliza el método original de Richter para calcular la magnitud por medio de un sismograma, teniendo en cuenta que el procedimiento se realiza una vez que ha ocurrido el sismo, es decir, no se realiza el cálculo a la hora de llegada la onda sísmica.

Este procedimiento se lleva a cabo de la siguiente manera: se obtiene la máxima amplitud de la onda en el sismograma (parte superior de la figura), se debe calcular su logaritmo, y escalarlo según la distancia que exista entre el sismómetro y el evento sísmico, después se calcula la diferencia de tiempo de las ondas S-P. Las escalas en el diagrama de abajo forman un nomograma que permite realizar el cálculo matemático rápidamente mediante un vistazo. La ecuación (1) representa este nomograma:

$$M = \log_{10} A (mm) + 3 \log_{10} (8 \Delta t (s)) - 2.92 \quad (1)$$

Donde A es la "amplitud" en milímetros, medida directamente del registro del sismograma, el tiempo S - P en segundos, da como resultado Δt . En la figura la diferencia de llegada de estas ondas es de 20 segundos y la amplitud de 23 milímetros.

El nomograma representa gráficamente los posibles valores que pueden tomar las variables, para este caso, la máxima amplitud, distancia y magnitud. Teniendo la distancia del epicentro a la estación y la amplitud máxima de la onda, se traza una línea recta tomando como puntos estos valores y se determina la magnitud. Este método también puede ser calculado con la fórmula antes mencionada.

Se debe obtener una estimación de la magnitud independientemente para cada estación sísmica donde se registre el evento, para luego obtener un promedio. Cabe señalar que entre más lejos estén las estaciones de los sismos, mayor será la atenuación que

presenten las ondas registradas por el sismómetro. La manera ideal de calcular la magnitud, sería teniendo varias estaciones cerca de donde ocurre el evento, para poder determinar con mayor precisión el punto exacto del epicentro, por supuesto, esto es una tarea difícil de realizar, debido al costo que presenta la instalación de los sensores.

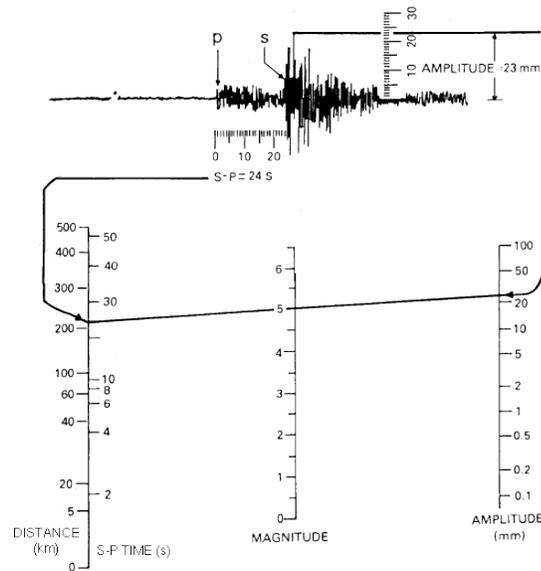


Figura 2.8: Obtención de la magnitud de un sismo a través del método de Richter por medio de un sismograma.

- *Intensidad*

La intensidad de un sismo está asociada a la zona geográfica donde ocurre el evento, esto es, la intensidad mide la cantidad de daño causado a consecuencia de la energía liberada. Esta manera de medir la cantidad de daño está dada en la escala de Mercalli modificada, la cual consta de los números romanos del I al XII (CENAPRED, 2013). La tabla 2.1 muestra esta clasificación.

Tabla 2.1: Grados y características habituales en la escala de Mercalli

Grado	Características
I	Apreciable únicamente por sismógrafos.
II	Percibidos por personas especialmente en pisos altos.

III	Dentro de edificios, pocas personas se dan cuenta. Se siente como el trepidar de un carro.
IV	Poco perceptible fuera de construcciones. Dentro de éstas, se observa ligero movimiento de lámparas, vajillas y muebles. Algunas personas dormidas se despiertan.
V	Fuera de construcciones es percibido con claridad; el follaje y hierbas se mecen por un viento inexistente. Oscilan en forma peligrosa, y a veces caen, lámparas, vajillas y muebles; se baten puertas y ventanas. Se rompen vidrios. Casi todas las personas dormidas se despiertan.
VI	Todo el mundo se da cuenta del sismo. Gran parte de las vajillas se rompe; los cuadros se desprenden de las paredes; caen muchos libros y objetos colocados en estantes. Se producen grietas.
VII	Suenan solas las campanas de las iglesias; el agua de los ríos y lagos se encrespan y enturbian; caen todos los objetos colgados o adosados a las paredes de las construcciones; se producen daños ligeros en éstas, particularmente en torres y chimeneas, y se desprenden algunos adornos arquitectónicos.
VIII	Los troncos de los árboles oscilan y algunos se revientan. Los muebles recorren grandes distancias o se desploman; las torres y chimeneas se rajan y destrozan. Los edificios sufren daños considerables. En pendientes el terreno se presentan grietas y en suelos húmedos brota agua con lodo y arena.
IX	Las construcciones antisísmicas también son dañadas. Las construcciones en general sufren daños severos; las de armazón de madera se desquician y desprenden de sus basamentos de mampostería.
X	La mayor parte de los edificios no antisísmicos son destruidos desde sus cimientos. Los rieles de ferrocarril se curvan ligeramente. El pavimento de las calles se agrieta y ondula. En ríos, lagos y similares, el agua golpea fuertemente

	las orillas, desmoronándolas.
XI	Todos los edificios de mampostería y casi todos los de madera son destruidos; los rieles de ferrocarril se curvan. Hay desplazamientos; se producen anchas grietas y hendiduras; ocurren aludes de peñascos; en los terrenos blandos o húmedos surgen manantiales de agua y lodo.
XI	Todas las obras hechas por el hombre quedan destruidas. Aparecen fallas de gran destaque en áreas pétreas; se desmoronan orillas de lagos y ríos; se forman cataratas, se desvían ríos; se trasladan lagos, etc.

Cabe mencionar que la diferencia entre la escala de Mercalli y la escala de Richter es que la segunda mide la cantidad de energía liberada y la primera mide los efectos materiales producidos en cierto evento, esta diferencia se hace notoria si los eventos suceden en un campo árido o en una ciudad, ya que en la ciudad la intensidad será más alta, aunque el sismo tenga la misma magnitud.

2.2. Data Warehouse

Un *data warehouse* o *almacén de datos* es una base de datos relacional diseñada para consultar y analizar datos así como para procesar transacciones (Vora, 2015). Usualmente contiene datos históricos de transacciones, pero también puede contener datos de otras fuentes. En esta investigación se analizarán datos provenientes de estaciones sísmicas, así como datos publicados por el SSN, los archivos donde se guardan los datos de las estaciones son almacenados en archivos de texto, mientras que los segundos, son almacenados en archivos de Excel ®.

Las bases de datos que comúnmente se conocen, procesan diariamente transacciones, es decir, las transacciones representan las operaciones que se realizan cotidianamente respecto al objetivo de la base de datos. Si la base de datos representa las ventas realizadas en alguna tienda departamental, las transacciones serían todas aquellas ventas que se efectúan diariamente, por el contrario, el almacén de datos contiene los datos históricos que se han realizado en dicha tienda departamental (Hoffer, 2013).

Los *data warehouses* son diseñados para ayudar a analizar datos. Por ejemplo, conocer más sobre los datos de venta de una compañía a través de un *data warehouse* concentrando en ventas. Usando este *data warehouse* se pueden resolver preguntas como “¿Quién fue el mejor cliente del año pasado?”, “¿En qué periodo se vendió mayormente un producto específico?”, entre otras preguntas (Database, 2005).

Una de las principales características de un almacén de datos es su no volatilidad, es decir, todos aquellos registros que se guardan en él no pueden ser borrados ni actualizados como normalmente se haría en una base de datos transaccional. Esto es lo que permite tener datos históricos para analizarlos. Cabe señalar que cada almacén de datos tiene un tiempo específico de duración de la información, pasado este tiempo, se borran los registros para dar cabida a la nueva información generada por las distintas fuentes de datos transaccionales (Gosain & Heena, 2015).

El *data warehouse* se enfoca en la relación de los cambios de los datos respecto al tiempo, esto requiere de grandes (o enormes) cantidades de datos, debido a que los fenómenos que son medidos respecto al tiempo, cambian su estado en cada momento dependiendo del lapso en que se necesite actualizar la base de datos, lo que provoca un aumento considerable de datos respecto a los fenómenos que son medidos de acuerdo a su ocurrencia. Este cambio de estado en la base de datos, puede variar, dependiendo el negocio que represente la misma, en el caso de las ventas de una tienda, la base de datos puede cambiar incluso cada segundo, conforme ocurran las ventas. Esto se define como la *granularidad* del almacén, es decir, que tan periódicamente se va almacenar la información (Golfarelli & Rizzi, 2009). La granularidad de los datos a analizar será de un dato por segundo, para cada una de las tres estaciones contempladas.

Los almacenes de datos están constituidos de tablas *dimensión* y tablas *de hechos*. Las tablas dimensionales son las que poseen información descriptiva, como puede ser el nombre de los empleados, el puesto, la compañía, entre otros datos. Estos campos suelen

ser llamados *atributos* y tienen relación entre ellos. *Clientes* sería la dimensión para este ejemplo.

Los hechos representan el nivel de detalle más fino, como una transacción simple o la ocurrencia de un evento. Consiste de una o varias medidas, tales como indicadores de rendimiento, junto con sus propiedades descriptivas (Mansmann, Rehman, Weiler, & Scholl, 2013).

Los datos de estas dimensiones pueden ser extraídas desde varias fuentes de datos, conformando, luego, todos los datos en cada dimensión, según corresponda. Algunas fuentes de datos son: bases de datos, archivos de texto, fuentes externas, etcétera.

Existen varios estilos de arquitecturas de *data warehouse*, una de estas clasificaciones es:

- *Bill Inmon Style*

En esta arquitectura el *data warehouse* debe estar almacenado en tercera forma normal, es decir, debe cumplir ciertas reglas básicas para evitar redundancia de datos, disminuir problemas de actualización, proteger la integridad de datos, entre otras.

Su principal característica es que provee un enfoque general del negocio teniendo el *data warehouse* centralizado, dependiendo de él todos los *data marts*.

La figura 2.9 muestra las arquitecturas de los estilos de *data warehouse*.

- *Ralph Kimball Style*

Define que la información almacenada es “una copia de los datos transaccionales estructurados específicamente para consultas y análisis”. Define un arreglo multidimensional en el que almacena los datos extraídos y transformados. Este estilo se define comúnmente como un “esquema estrella”, debido a la disposición típica de los elementos en torno a una entidad central. En este enfoque, los *data marts* son creados antes que el *data warehouse*, teniendo *data mart* para cada área del negocio (por ejemplo, *data mart* de inventarios, ventas, mercadotecnia, entre otros), proporcionando una imagen parcial del campo completo.

Ambas arquitecturas presentan los datos y permiten realizar análisis, la diferencia radica en la manera en que organizan la información (Chowdhury, 2004).

La arquitectura a utilizar en este trabajo es la antes descrita *Kimball Style*. Se pretende conformar el almacén de datos a partir de las dimensiones de Magnitud y Aceleraciones, que son obtenidas de distintas fuentes de datos, por un lado, las magnitudes que son proporcionadas por el SSN en formato Excel, por otro lado, las aceleraciones en formato de archivo de texto plano. Se extraerán estos datos y se conformará el almacén de datos a partir de ellos.

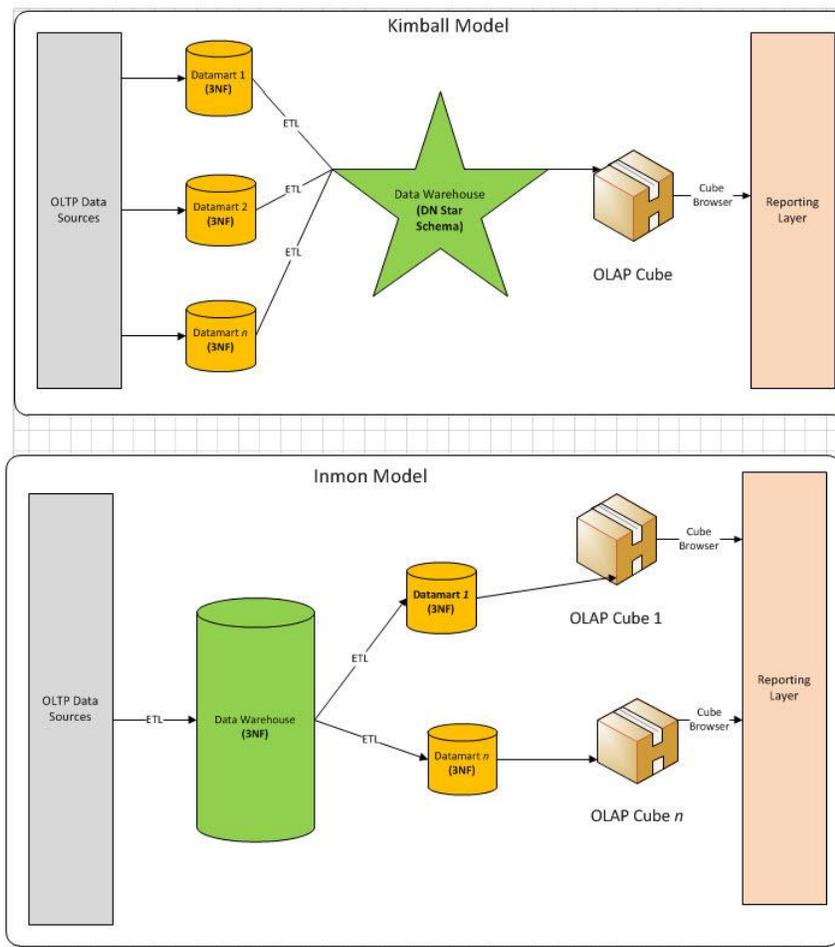


Figura 2.9: Comparación entre estilos Inmon y Kimball.

2.3. Business Intelligence

La inteligencia de negocios es el conjunto de técnicas y herramientas para la transformación de datos en información significativa y útil para fines de análisis de

negocios. Se trata del acceso, análisis y descubrimiento de nuevas oportunidades para lograr obtener conocimiento que pueda ser aplicado mediante estrategias (Ahumada Tello & Perusquia Velasco, 2014).

Los sistemas de inteligencia de negocios se centran en mejorar el acceso y la entrega de información tanto a los proveedores de información como a los consumidores de la misma. Esto es logrado debido al procesamiento avanzado utilizando representación gráfica, herramientas de minería de datos, entre otras. Algunas herramientas de inteligencia de negocios son utilizadas para analizar datos como la minería de datos, sistemas de soporte de decisiones o predicción, manejo de conocimiento, mapeo, visualización de la información o tableros de mando, sistemas de manejo de información, sistemas de información geográfica, análisis de tendencias, entre otros muchos (Ranjan, 2009).

Estas aplicaciones comúnmente necesitan procesar y analizar grandes volúmenes de información utilizando diferentes herramientas (Mehta, 2000). Un sistema de inteligencia de negocios debe, por lo tanto, proporcionar escalabilidad y ser capaz de apoyar e integrar productos de múltiples proveedores. Estos sistemas también pueden simplificar el acceso a la información a través del uso de catálogos de información que pueden ser utilizados por los documentos de ayuda en la toma de decisiones y que también pueden ser accedidos por los consumidores de información que pueden responder preguntas que surgen en las operaciones diarias en las empresas.

La importancia de estos sistemas radica en el impacto que tiene el desarrollo de la competitividad entre los sistemas empresariales del negocio. Tanto más útil y novedoso sea el conocimiento obtenido a través de los sistemas de inteligencia de negocio, mayor será el provecho que se pueda obtener de dichos sistemas. La innovación juega un papel importante en la generación del conocimiento, ya que está definida como las acciones u operaciones que promueven la mejora de cualquiera de los productos o servicios con que cuenta el negocio (Ahumada Tello & Perusquia Velasco, 2014).

La declaración de inteligencia de negocios “Obtener la información correcta para las personas correctas en el momento adecuado” se centra en el hecho de que la inteligencia de negocios utiliza información y no datos debido a las capacidades incluidas para el procesamiento de datos brutos en información inteligente, que es válida y aceptada por toda el área de negocio y que puede ser utilizada constantemente en el modelado de procesos (Lungu, 2013).

Una manera de implementar la inteligencia de negocios es llevar a cabo, en una etapa previa, el planteamiento de los posibles factores incluidos en la implementación, un ejemplo de esta etapa sería la realización de las siguientes preguntas agrupadas en subcategorías (Ranjan, 2009):

- Alineación de Objetivos. ¿Qué objetivo estratégico de negocio aborda el programa? ¿La hipótesis detalla el mejoramiento o rendimiento?
- Consultas de referencia. ¿Se tiene la capacidad de vigilar las fuentes de información importantes? ¿Qué datos se recopilan y cómo se almacenan? ¿Cuáles son los parámetros estadísticos de estos datos?
- Costo y riesgo. ¿Es necesario evaluar el costo de las operaciones actuales y el aumento de los costos dada la implementación de inteligencia de negocios? ¿Cuál es el riesgo de que exista un error? Esta evaluación de riesgos debe ser convertida en una métrica financiera e incluirla en la planificación.
- Clientes y accionistas. ¿Quién tiene una participación en el procedimiento actual? ¿Quién se beneficia directa e indirectamente con el proyecto? ¿Cuáles son los beneficios cualitativos y cuantitativos?
- Cuestionamientos relacionados con las métricas. ¿Cuáles son los mejores indicadores? ¿Cómo puede determinarse esto? ¿Cuántas métricas deben ser controladas?
- Metodología de medición. ¿Qué métodos se utilizarán y qué periodicidad tendrá la recopilación de datos? ¿Existen estándares para esto?

- Resultados. ¿Cómo se puede mostrar que las herramientas de inteligencia de negocios (y no otros factores) contribuyeron a un cambio en los resultados?

Una vez aclaradas estas interrogantes, y dado que se haya determinado que las herramientas de inteligencia de negocios pueden proveer la mejora de la obtención de conocimiento respecto al problema de negocio, es viable realizar la implementación de estas tecnologías para intentar obtener los requisitos expresados como objetivos.

A través de las herramientas de *Business Intelligence* se pretende mejorar el acceso a la información a través del almacén de datos teniendo conglomerada la información a analizar. El fácil acceso a los datos permitirá llevar a cabo distintos análisis, generando resultados visuales, como gráficas, que permitan a los expertos geólogos y civiles tomar decisiones de acuerdo a los resultados obtenidos. Estas herramientas posibilitan un análisis más rápido de los datos, por lo que tienen ventajas sobre las tradicionales.

2.4. OLAP

On-Line Analytical Processing (Procesamiento Analítico en Línea) es el conjunto de herramientas para el análisis de información, en su mayoría, histórica. Estas incluyen planificación y medición del desempeño, entre otras, las cuales proveen interfaz amigable con el usuario, que le permite configurar el desarrollo de análisis y muestra de resultados, haciendo que estas tareas puedan ser analizadas a distintos niveles entre las distintas dimensiones y hechos (Dehne, Kong, Rau-Chaplin, Zaboli, & Zhou, 2014). Esto facilita visualizar la información desde distintas perspectivas, dicho de otra manera, se puede utilizar un *Cubo OLAP multidimensional*, que contiene una serie de dimensiones jerárquicas clasificadas. Las dimensiones en un cubo representan características ortogonales de las medidas. Cada dimensión es un eje en un espacio multidimensional con los valores como coordenadas. Cada celda contiene un valor de la medida definida por las coordenadas respectivas (Mansmann, Rehman, Weiler, & Scholl, 2013).

Un *cubo OLAP* y su instancia se definen como:

- i. Un esquema de dimensión D_i ($1 \leq i \leq n$) y sus medidas M son conjuntos disjuntos de atributos.
- ii. Un esquema de cubo OLAP $C = D_1 \cup D_2 \cup \dots \cup D_n \cup M$, donde $D_1 \dots D_n$ son esquema de dimensión, M es el conjunto de medidas de atributos con sus unidades y escalas estáticas, $m_1, m_{1u}, m_{1s}, \dots, m_p, m_{pu}, m_{ps}$.
- iii. En cada dimensión, los valores de atributos, deben formar una jerarquía completa y discontinua. Esto significa que existe una relación de muchos a uno desde el atributo de más bajo nivel para el atributo de nivel más alto. Los niveles de dimensión forman una ordenación lineal.
- iv. Una instancia de *cubo OLAP* “c” es una relación sobre el esquema de *cubo OLAP* C y una relación “d” sobre D se denomina instancia de la dimensión.
- v. Se supone que la medida escrita M no está vacía y el cubo tiene al menos una dimensión.

La definición sigue el modelo de base de datos relacional suponiendo que no existen tuplas o registros duplicados (Niemi, Niinimäki, Thasnisch, & Nummenmaac, 2014).

Cada registro en la tabla de hechos corresponde a un evento de medición. Cada dato en cada registro es un nivel específico de detalle, referido anteriormente como grano, como puede ser cada producto vendido en una transacción de venta. Usualmente las medidas son numéricas y aditivas, como la cantidad de dólares de las ventas (Kimball & Ross, 2013).

Un sistema de almacenamiento de información en múltiples capas optimiza el rendimiento y minimiza el esfuerzo en la carga de datos desde las distintas fuentes de datos. Una arquitectura de este tipo de sistemas es el propuesto por Mansman (2013), el cual consta de cinco capas (Mansmann, Rehman, Weiler, & Scholl, 2013):

- Fuentes de datos, como pueden ser bases de datos internas y externas.
- ETL, puede ser un sistema de Extracción, Transformación y Carga de los datos (*Extract, Transform and Load*).
- *Data Warehouse*. En cualquier implementación.

- Análisis, en esta capa puede realizarse cualquier operación de inteligencia de negocios, como puede ser cubos OLAP, minería de datos, y algunos otros métodos.
- Presentación, o muestra de los datos mediante tableros de mando, gráficas o reportes.

Los *cubos OLAP* pueden ser utilizados para realizar análisis, reportes, modelar y planear procesos de optimización, se requieren técnicas y herramientas en conjunto con sofisticados almacenes de datos. La implementación del *cubo* se llevará a cabo mediante dos herramientas similares en cuanto a la función principal, SQL Server Analysis Service ® y Pentaho ®, las cuáles aunque, implementan el mismo cubo en cuanto a definición, reportan distintos tipos de resultados, la complementación dará una mayor calidad de información.

2.5. Bag of Words

Bag of Words (BoW) es un método propuesto por Wang, Liu, She, & Nahavandi (2012), el cuál fue elaborado para clasificar series de tiempo biomédicas (Wang, Liu, She, & Nahavandi, 2012). Esto se lleva a cabo segmentando la señal. La segmentación se realiza tomando un valor predefinido de “*ventana*”, el cual determina la longitud del desplazamiento que se hará a través de la señal. Cada segmento contiene una parte de la señal, como se muestra en la figura 2.10.

El *codebook* (vocabulario) es formado de la siguiente manera: el método de BoW cuenta el número de cada *codeword* (palabra) que existe en un documento y provee una representación en forma de histograma de *codewords*, es decir, se cuenta la cantidad de señales que son similares en un documento a través de los centroides de un clustering (Wang, Liu, She, & Nahavandi, 2012). La figura 2.11 presenta gráficamente la manera en que se obtiene la palabra representativa de los segmentos de señal similares.

Clustering es la agrupación de elementos que tienen características en común, cada cluster es diferente a los demás. Existe una distancia que se minimiza entre cada elemento del cluster y una distancia que se maximiza entre cada cluster (Ding & Fu, 2014). Un algoritmo comúnmente utilizado es el *k-means*, a través de cual se obtienen los centroides para formar las palabras.

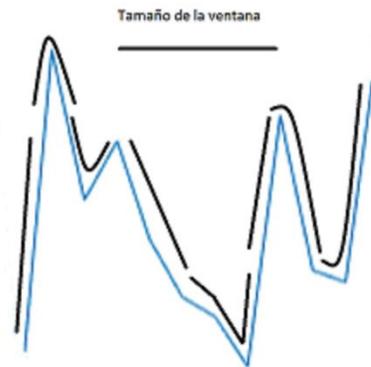


Figura 2.10: Segmentación de la señal

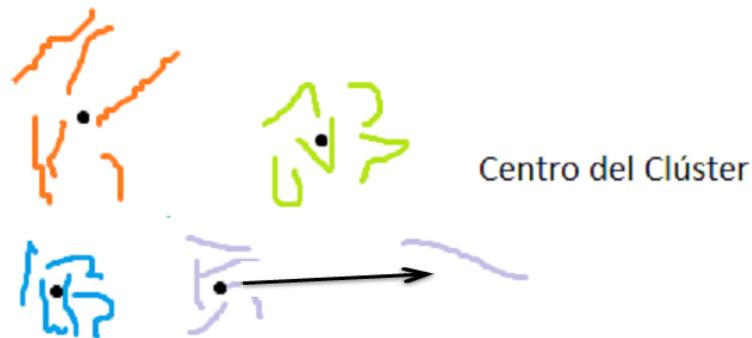


Figura 2.11: Formación de los CodeWords

El *codebook* se forma del conjunto de todas las palabras, es decir, contiene todos los *codewords* formados de las señales de entrenamiento de cada una de las clases. Para ello, se especifica la cantidad de palabras que se desean formar (González, Moreno, Escalante, Martínez, & Carlos, 2016). Este proceso puede ser ejecutado varias veces y con ello

obtener un promedio, debido a que cada ejecución genera resultados diferentes, y de esta manera se tiene un resultado más confiable.

La figura 2.12 muestra hipotéticamente la representación gráfica del *codebook*. Cada punto representa un *codeword* y están agrupados según la clase que corresponden.

Algunos algoritmos útiles para clasificar los histogramas son los descritos a continuación:

- *Random Forest (RF)*.

Es un meta clasificador que se basa en la idea de que si un árbol de decisión es bueno, varios (un bosque) serán mejor. Se ajusta al número (es un parámetro que puede variar) de clasificadores de árboles de decisión en varias sub-muestras del conjunto de datos y el uso de un promedio de los resultados de cada árbol para mejorar la exactitud de predicción. El tamaño de la sub-muestra es siempre el mismo que el tamaño original de la muestra de entrada, pero las muestras se extraen con reemplazo del conjunto de entrenamiento en caso de que el parámetro `bootstrap = True`, lo cual es predeterminado. Dicho de otra manera, cada muestra se toma de manera aleatoria y es devuelta a la población, lo que le permite tener las mismas posibilidades de volver a ser elegida en el resto de las muestras (James, Witten, Hastie, & Tibshirani, 2013).

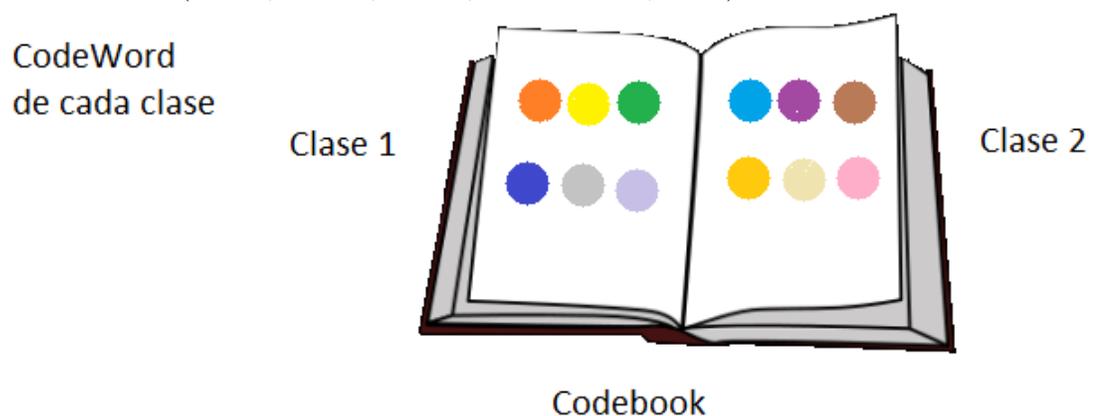


Figura 2.12: Representación del CodeBook

Cuando se divide un nodo, la división elegida para la construcción del árbol es la mejor entre un subconjunto al azar entre las características, además, esta división ya no puede ser elegida como la mejor, lo que aumenta la aleatoriedad (Pedregosa, y otros, 2011).

En este algoritmo se crean los árboles de decisión de la siguiente manera:

Sea N la cantidad total de casos de prueba, T el número de variables del clasificador, t el total de entradas o características que serán utilizadas para determinar la decisión de un nodo dado, se tiene que cumplir que t sea mucho menor que T . Cumpliendo estos requisitos se procede a elegir un conjunto de entrenamiento para el árbol, los demás casos se utilizarán para las pruebas, donde se estimará el error. Para cada nodo del árbol se eligen aleatoriamente t características para basar la decisión en ellas y se calcula la mejor partición de las variables del conjunto de entrenamiento (James, Witten, Hastie, & Tibshirani, 2013).

En cada nodo del árbol de decisión, al seleccionar las t características, se calcula el coeficiente de Gini, el cuál determina si los datos de una hoja pertenecen todos a una misma clase, esto se conoce como pureza. Este coeficiente se calcula de acuerdo a la fórmula (2), para una característica en particular h , donde g es el número de clases (Marsland, 2015).

$$G_h = 1 - \sum_{i=1}^g N(i)^2 \quad (2)$$

- *K Nearest Neighbors(KNN)*

Dado un número entero positivo K y un punto de observación x_0 , el clasificador *KNN* identifica K puntos en el conjunto de datos de entrenamiento que están más cercanos a x_0 utilizando una distancia euclidiana métrica, eso se realiza hasta tener los K vecinos (James, Witten, Hastie, & Tibshirani, 2013).

Este algoritmo utiliza un conjunto de datos de prueba para determinar la precisión del modelo, debido a que cada conjunto de datos pertenece a una clase identificada con alguna etiqueta.

Teniendo como registros (individuos) $E = \{e_1, e_2, \dots, e_n\}$ y un conjunto de clases $G = \{G_1, G_2, \dots, G_m\}$, el problema de la clasificación es encontrar una función $f: E \rightarrow G$ tal que cada e_i es asignada una clase G_j . De esta manera se establece a qué grupo pertenece cada individuo (Marsland, 2015).

La figura 2.13 muestra una aproximación del algoritmo. Los puntos dentro del círculo representan los K vecinos a un punto dado x_0 . Los elementos a tomarse en cuenta para la formación del grupo son aquellos que se encuentran menos distantes.

La elección del valor de K tiene un efecto drástico en los resultados obtenidos por el clasificador. Si K es muy pequeño, el modelo será muy sensitivo a los puntos que son atípicos o ruido, por otro lado, si K es muy grande, el modelo tiende a asignar los puntos a la clase más grande.

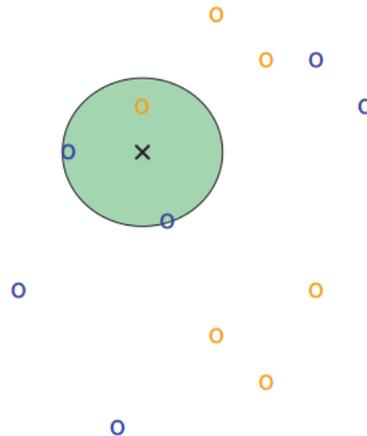


Figura 2.13: Aproximación del algoritmo KNN

- *Decision Trees*

El objetivo de este algoritmo es determinar qué característica predice el valor de una variable objetivo, para con ello, determinar un modelo, lo anterior, aprendiendo las reglas de decisión simples deducidas a partir de las características de los datos.

La entropía se define como la medida de incertidumbre, y es usada para decidir el atributo que presenta mayor cantidad de información al momento de la construcción del árbol de decisión. Para calcularla se utiliza la fórmula (3)

$$\text{Entropía}(S) = \sum_{i=1}^n -p_i \log_2 p_i \quad (3)$$

donde p forma parte del conjunto de probabilidades de estados que puede adquirir el nodo, S es la colección de elementos, n es el número de elementos que pertenecen a la colección (Marsland, 2015). Las probabilidades se obtienen de la proporción de un cierto elemento respecto al total de ellos en la colección.

Los árboles comienzan a construirse a partir de la raíz, quien en un principio es el nodo con mayor entropía, a partir de él, comienzan a descender los niveles construyendo los nodos hijos y hojas. La figura 2.14 muestra un ejemplo de un árbol. Cada elemento del conjunto de datos se evalúa en el árbol y con este se determina a que clase pertenece. Las hojas en el árbol representan la clase, mientras que los nodos son las condiciones que deben cumplir los elementos, de esta manera se dirigen los elementos a una parte u otra del árbol, según las características que cumplan.



Figura 2. 14: Ejemplo de árbol de decisión

- *Support Vector Machine(SVM)*

Es un algoritmo supervisado, es decir, cada conjunto de ejemplos de entrenamiento se puede etiquetar la clase a la que pertenece. El objetivo de este algoritmo es construir un hiperplano, o conjunto de ellos en un espacio dimensional grande. La correcta separación entre las clases, permitirá la adecuada clasificación.

El algoritmo toma en cuenta un conjunto de puntos dado, que pertenece a un subconjunto del espacio dimensional, en el que cada uno de ellos pertenece a una clase, de

dos posibles categorías, mientras que un algoritmo basado en *SVM* construye un modelo capaz de predecir si un nuevo punto dado pertenece a una clase u otra.

El hiperplano que se busca es de un grado menor que el espacio que lo contiene, es decir, hablando de un plano x - y el hiperplano es una línea que divide a este plano. En la figura 2.15 se muestran tres ejemplos de la separación de dos clases en un espacio como el antes descrito.

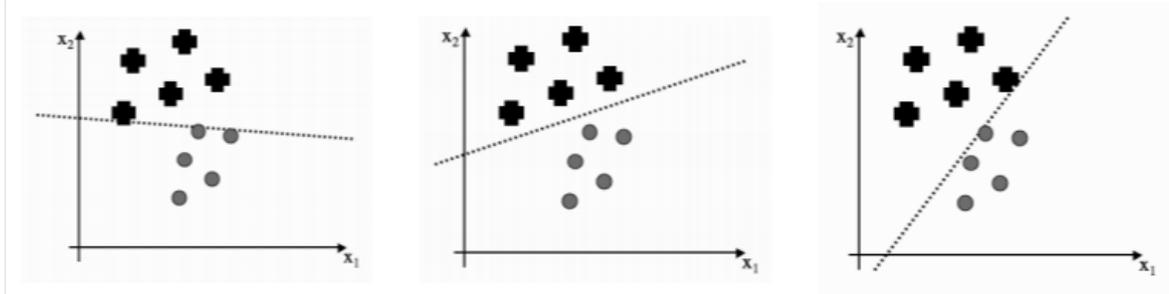


Figura 2.15: Ejemplos de separación mediante SVM

Dado que no siempre es posible tener una línea recta que separe los datos en las distintas clases, existen funciones kernel que auxilian en la creación de divisores más precisos. Estas funciones pueden ser: polinomial, sigmoideal y radial. No existe un método exacto en la elección del kernel, algunos experimentos han sido ejecutados cambiando entre kernels, así como los parámetros de estos, hasta determinar la configuración más apropiada (Marsland, 2015).

- *Artificial Neural Network(ANN)*

Una red neuronal tiene mucha semejanza con las neuronas existentes en el cerebro humano. Su función es responder a los cambios del entorno (Marsland, 2015). Las redes neuronales artificiales están modeladas de acuerdo a un modelo matemático propuesto por McCulloch y Pitts, que se muestra en la figura 2.16. Este modelo consta de:

- a) Entradas. Un conjunto de valores de entrada x_i que corresponden a la sinapsis, o unión entre las neuronas.

- b) Un sumador que suma los productos de las entradas con los pesos (equivalente a la membrana de la célula que se acumula carga eléctrica).
- c) Una función de activación (inicialmente una función de umbral). Que decide si la neurona responde para las entradas.

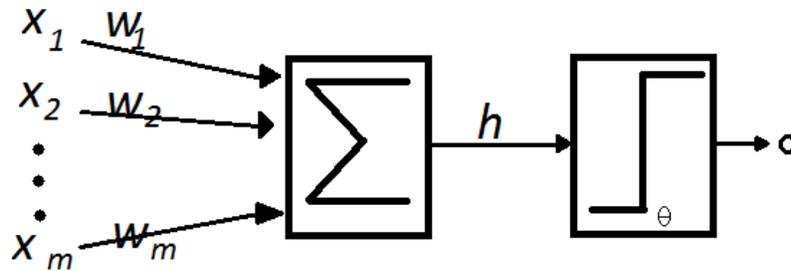


Figura 2.16: Diagrama de la red neuronal artificial

El conjunto de nodos de entrada (x_1, x_2, \dots, x_m) se encuentra en la parte izquierda del diagrama, estos valores pueden ser los valores de salida de otras neuronas. Un valor de 0 indica que la neurona no reacciona al cambio en el entorno, mientras que el valor de 1 indica que si lo hace.

Cuando una neurona descarga fluido a lo largo de una sinapsis para llegar a otra neurona, este tiene una fuerza, llamada peso (w_1, w_2, \dots, w_m). La fuerza de la sinapsis afecta la fuerza de la señal, por lo que multiplica la entrada por el peso de la sinapsis (por lo que se tiene $x_1 * w_1$ y $x_2 * w_2 \dots x_m * w_m$). Cuando todas estas señales llegan a la neurona destino, se suman para comprobar si la neurona reaccionará. Esto se puede ver como la fórmula (4).

$$h = \sum_{i=1}^m w_i * x_i \quad (4)$$

donde la cantidad de entradas van desde i hasta m .

Capítulo III

Metodología

3.1. Actividades para el pre procesamiento de datos

Para lograr el pre procesamiento de datos se pretende realizar un muestreo de los datos de cada estación uniformizando los datos para tener un valor de onda por segundo. Esta estandarización se llevará a cabo mediante el uso de Matlab ®. La figura 3.1 muestra los pasos a seguir en la obtención de este objetivo.

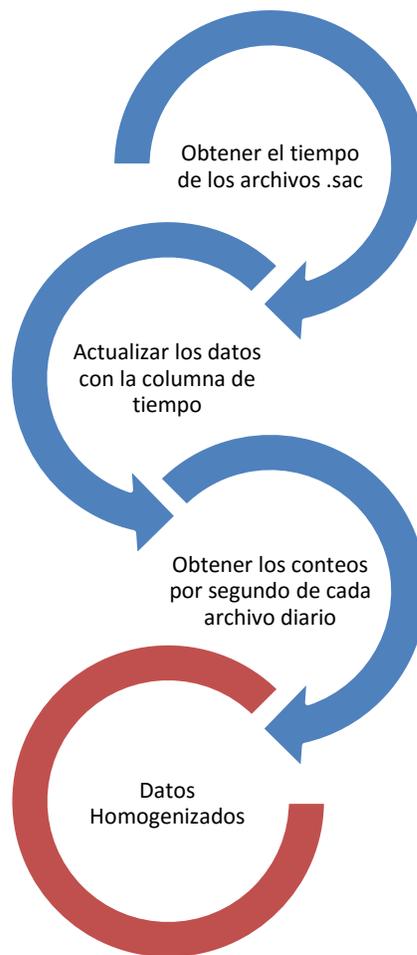


Figura 3.1: Actividades para el pre procesamiento de datos.

3.2. Actividades para la creación del almacén de datos

El almacén de datos se creará teniendo como dimensiones Fecha, Tiempo, Aceleraciones y Magnitudes. La carga de los datos al almacén de datos se realizará en SQL Server Integration Services (SSIS), teniendo la base de datos en SQL Server ®. Las actividades se presentan en el esquema de la figura 3.2

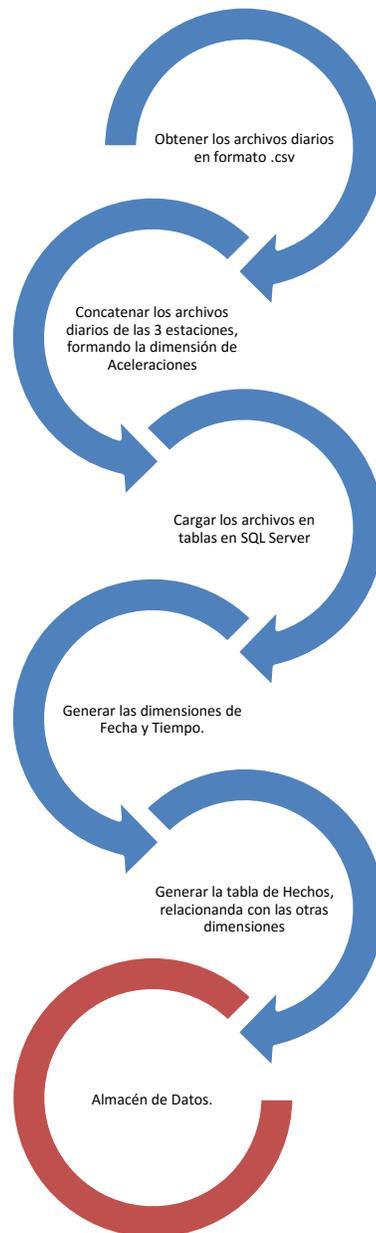


Figura 3.2: Actividades para la creación del almacén de datos

3.3. Actividades para la conformación del *cubo OLAP*

Con el almacén de datos, se implementará una estructura multidimensional para llevar a cabo visualizaciones, teniendo en cuenta las actividades mostradas en la figura 3.3



Figura 3.3: Actividades para la conformación del cubo OLAP

3.4. Actividades para la búsqueda de patrones

Se intenta buscar una tendencia entre los datos de las estaciones sismológicas, siguiendo las actividades de la figura 3.4.

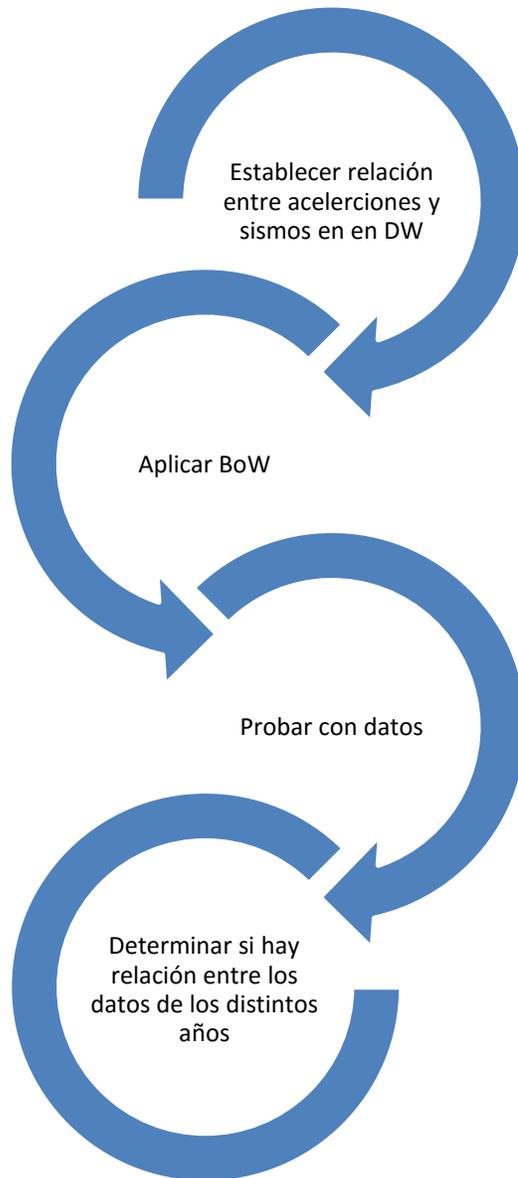


Figura 3.4: Actividades para la búsqueda de patrones.

Capítulo IV

Resultados

4.1. Resultados de Actividades del pre procesamiento de los datos

En el pre procesamiento de los datos se pretende eliminar los datos que no son relevantes para su análisis. Para lograr esto, se tiene que cada estación registra los valores de los tres ejes, junto con el valor del tiempo, medido en segundos. La tabla 4.1 muestra un ejemplo de los valores obtenidos por la estación HPIG del día 1 de enero de 2013, el total de registros es de 8 640 000.

Tabla 4.1: Ejemplo de datos de aceleraciones

Z	N	E	Tiempo
-857	-844	-866	0
-804	-954	-841	0.01
-954	-951	-725	0.02
-950	-775	-863	0.03
-802	-891	-976	0.04
-821	-997	-719	0.05
-850	-934	-651	0.06
-851	-872	-775	0.07
-995	-845	-806	0.08
-930	-656	-947	0.09
-965	-677	-959	0.10
-797	-770	-776	0.11
-672	-997	-723	0.12
-918	-985	-687	0.13
-1023	-749	-777	0.14
-890	-707	-745	0.15

Para poder analizar estas aceleraciones respecto a cuáles de ellas representan la generación de un sismo, es necesario tomar en cuenta sólo aquellas aceleraciones que podrían ser precursoras del sismo. Como se mencionó anteriormente, el sismómetro registra todas las aceleraciones que se producen, inclusive aquellas que no están generando sismos, debido a ello, se necesita realizar una selección de los valores significativos, es decir, aquellos valores que realmente provocan el movimiento de las placas tectónicas, o que son provocados por un fenómeno intraplacas.

Para realizar este procedimiento de selección de valores significativos, se toma en cuenta el valor del vector resultante de las tres aceleraciones, **N**, **E**, **Z**. Este cálculo se realiza para cada aceleración, debido a que se cuenta con 8 640 000 registros por cada estación, este proceso se realiza en Matlab®, ya que provee herramientas robustas que permiten el manejo de grandes volúmenes de datos.

La base de este procedimiento es la estadística, la metodología que se siguió es la siguiente:

1. Se calcula la resultante vectorial para cada registro mediante la fórmula
$$R = \sqrt{N^2 + E^2 + Z^2}.$$
2. Se eliminan los valores menores a la moda (valor más común) de esta resultante, así como también los valores de los ejes **N**, **E**, **Z**. La moda se calcula con la siguiente fórmula: $M = L_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)A_i$. Donde L_i es el límite inferior, Δ_1 es la diferencia entre la frecuencia absoluta modal y la frecuencia absoluta premodal, Δ_2 es la diferencia entre la frecuencia absoluta modal y la frecuencia absoluta postmodal y A_i es la amplitud del intervalo modal.
3. Se aplica interpolación al vector resultante. Con la configuración “*nearest*”, que toma en cuenta los valores más cercanos a un punto dado para obtener el nuevo valor, se reduce cada conjunto de datos diarios a dos veces su desviación estándar.

Con esto se eliminan los datos más comunes y los menores a ellos para cada eje coordenado, lo que proporciona datos relevantes, es decir, se tienen los datos más altos que se han registrado diariamente.

Un ejemplo del día 1 de enero de 2013 de la estación HPIG seleccionando los datos menores a la moda es el que se muestra en la tabla 4.2.

Tabla 4.2: Datos menores a la moda, 1 de enero de 2013, estación HPIG

Z	N	E	Tiempo	Resultante
-857	-844	-866	0	1482.14
-804	-954	-841	0.01	1504.597
-954	-951	-725	0.02	1529.752
-950	-775	-863	0.03	1499.297
-802	-891	-976	0.04	1545.852
-821	-997	-719	0.05	1478.178
-850	-934	-651	0.06	1420.794
-851	-872	-775	0.07	1444.025
-995	-845	-806	0.08	1534.172
-930	-656	-947	0.09	1480.555
-965	-677	-959	0.10	1519.616
-797	-770	-776	0.11	1352.88
-672	-997	-723	0.12	1402.968
-918	-985	-687	0.13	1511.594
-1023	-749	-777	0.14	1487.03
-890	-707	-745	0.15	1359.034
-824	-753	-740	0.16	1339.247
-984	-749	-841	0.17	1495.505
-863	-786	-938	0.18	1497.467

El resultado de la interpolación de los datos para este día en particular se muestra en la tabla 4.3.

Del total de 8 640 000 registros que se tenían originalmente, mediante la reducción de datos, para este día en concreto, se lograron reducir a 629 datos relevantes. A partir de estos, se puede realizar un análisis detallado de las aceleraciones, obteniendo con ello, una noción aproximada de la orientación (Norte, Sur, Este u Oeste) respecto a los movimientos registrados por el sismómetro.

Un análisis más completo requiere la concatenación de las aceleraciones de varios días, incluso meses para tener una mejor estimación de la orientación. Por ello, se toman en cuenta tres meses consecutivos de cada estación y se comparan los resultados de todas las estaciones respecto a un periodo de tiempo.

De esta manera, en la figura 4.1, se tiene la gráfica de los puntos graficados individualmente correspondiente a los meses de enero, febrero y marzo de 2013 de la estación HPIG. Esta imagen muestra los eventos individuales que ha registrado la estación HPIG, donde el origen de los movimientos proviene de la orientación Noroeste. El análisis se hace debido a que si en algún momento ha habido ruptura de roca, es probable que con eventos sísmicos posteriores, la energía pueda ser liberada a través de la ruptura que se ha presentado, por ello, se hacen análisis con tiempos consecutivos, para determinar si se está liberando energía por alguna ruptura de roca.

Estas aceleraciones pueden compararse con aquellas obtenidas por las otras dos estaciones, confrontando el comportamiento de los movimientos detectados.

La interpolación se realiza al vector resultante, mientras que los datos a analizar para determinar si hubo rupturas de rocas se hace mediante el estudio de los puntos individualmente graficados de la resultante vectorial de las aceleraciones.

La figura 4.2 presenta las aceleraciones registradas en los mismos tres meses de la estación HSI, así como la figura 4.3 muestra las correspondientes de la estación CGIG.

Como se puede observar, cada estación, aunque presenta un comportamiento similar respecto a la orientación, presenta diferentes cantidades de datos relevantes. Lo que lleva a determinar que existen movimientos fuertes más cerca de la estación HSIG, ubicada en Hermosillo, Sonora. Esta comparación puede hacerse visualmente de manera más fácil al presentar los tres conjuntos de registros en una misma gráfica, mostrada en la figura 4.4.

El procedimiento antes realizado genera una escala de tiempo en el cual los valores de tiempo no coinciden entre cada una de las estaciones, por lo que realizar el análisis comparando los eventos sísmicos entre las estaciones se ve afectado, para ello, se reducen los datos de acuerdo a otros métodos como el descrito a continuación, comprobando que la pérdida de información no sea significativa.

Una vez obtenidas estas aceleraciones, es posible reducir los datos, intentando que el ruido captado por las señales no afecte el análisis. Para esto se lleva a cabo la reducción de los datos a través de la obtención del promedio de las aceleraciones por cada segundo. Es decir, en cada segundo se tienen 100 registros de aceleraciones, las cuales se promedian para obtener una aceleración por segundo.

Además de esto, es necesario eliminar los datos de aquellas fechas en donde no se tenga registro en las tres estaciones.

La figura 4.5 muestra los datos del mes de enero de la estación CGIG, la cual muestra que los datos tienen el mismo sentido después de aplicar la reducción de los datos mediante el promedio.

Tabla 4.3: Datos interpolados, día 1 de enero de 2013, estación HPIG

Z	N	E	Tiempo	Resultante
-587	-701	-813	26.22	1223.494
-932	-1109	-904	78.01	1707.548
-989	-1019	-833	130.2	1646.32
-1250	-1241	-961	182.11	2006.514
-802	-813	-637	232.88	1307.647
-1046	-783	-914	282.91	1594.553
-745	-894	-915	333.24	1480.366
-821	-1011	-916	384.08	1592.236
-992	-1009	-977	434.25	1719.498
-778	-1180	-1193	487.68	1849.576
-798	-924	-1074	536.72	1626.055
-842	-522	-860	588.96	1311.887
-869	-756	-1112	639.37	1601.012
-770	-912	-831	690.27	1454.374
-1009	-1055	-976	740.94	1756.041
-1069	-1017	-935	791.83	1746.789
-1163	-1187	-1211	841.68	2056.224
-1077	-730	-297	891.92	1334.555
-974	-833	-795	943.61	1508.174
-1131	-919	-1144	994.69	1852.689

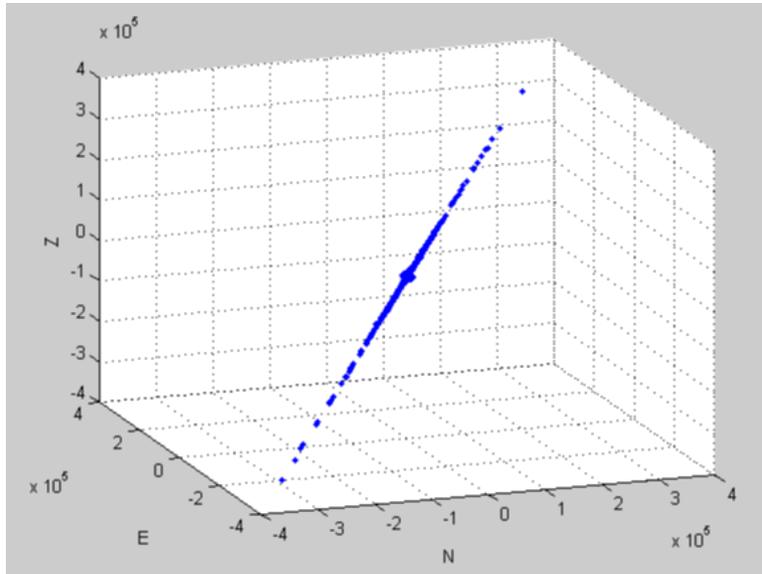


Figura 4.1: Aceleraciones de enero, febrero y marzo de 2013, estación HPIG

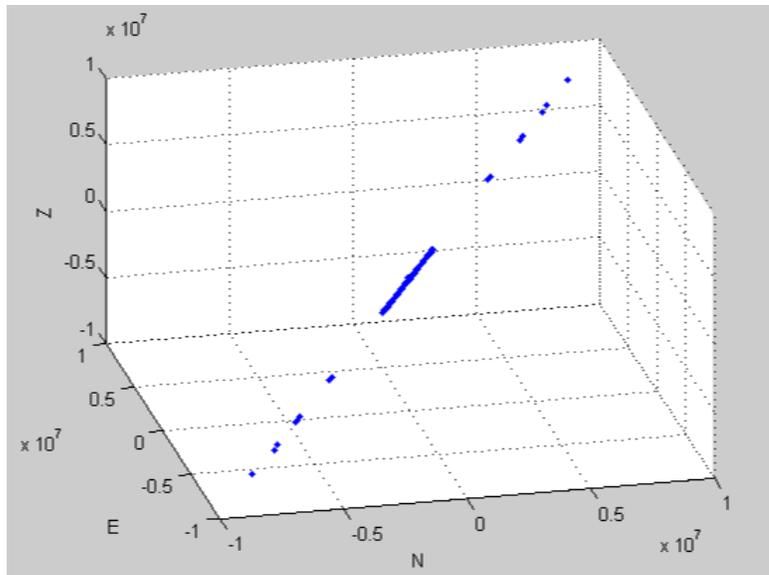


Figura 4.2: Aceleraciones de enero, febrero y marzo de 2013, estación HSIG.

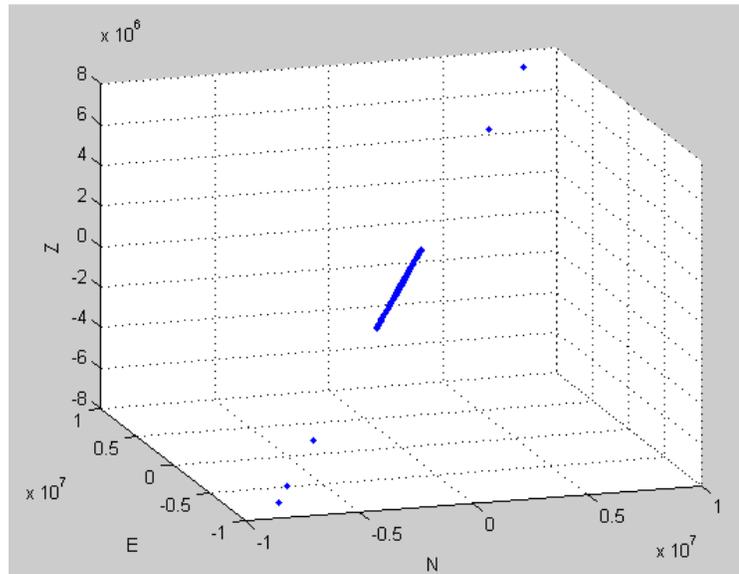


Figura 4.3: Aceleraciones de enero, febrero y marzo de 2013, estación CGIG.

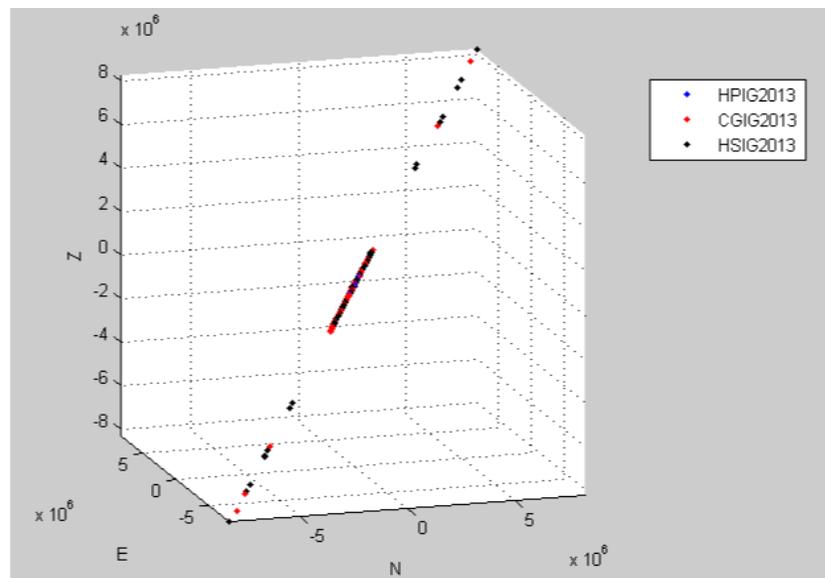


Figura 4.4: Aceleraciones enero-marzo, estaciones HPIG, CGIG, HSI2013.

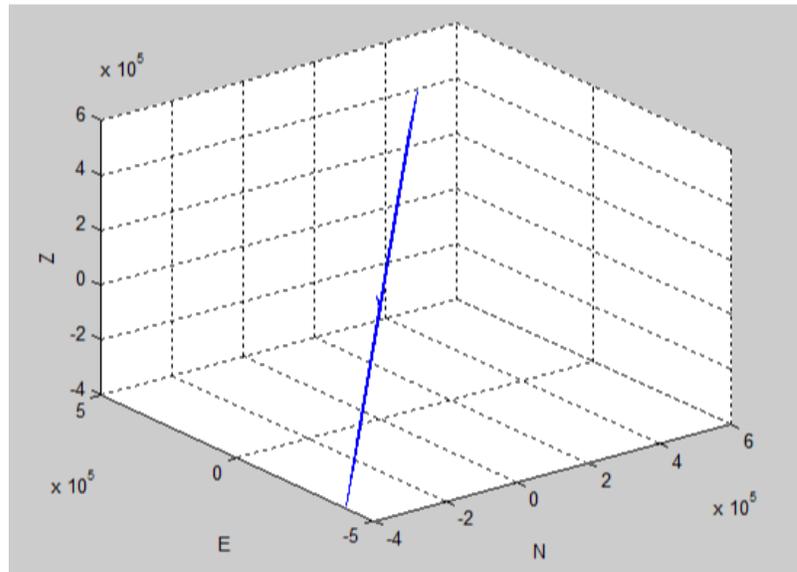


Figura 4.5: Aceleraciones enero estación CGIG.

4.2. Resultados de las Actividades de la creación del almacén de datos.

Para establecer un almacén de datos, que concentre una gran cantidad de información sobre sismos en base a datos de sensores instalados en lugares predeterminados, es necesario establecer las dimensiones que se tendrán en el almacén de datos. Dado que los análisis en bases de datos de este tipo se lleva a cabo teniendo en cuenta los datos históricos, es necesario tener una dimensión de fecha.

Los datos sismológicos almacenados se diferencian en tiempo de un segundo, por lo que se tienen 86 400 registros por día, teniendo aproximadamente 30 días por mes, durante dos años. Por lo que se cuenta con más de 53 millones de registros. Cabe señalar que la cantidad de registros está dada por las condiciones anteriores, además de la eliminación de registros que se llevó a cabo en los casos que una o más estaciones no hubieran registrado aceleraciones en determinado tiempo.

Teniendo en cuenta que los registros están tomados en intervalos de tiempo de segundos, se tiene una dimensión de tiempo. La diferencia entre las dimensiones de tiempo y fecha, reside en que la primera se especifica como los campos correspondientes al tiempo, como son: hora, minuto, segundo, y algunos otros que pueden ser distintas maneras de especificar el tiempo, por ejemplo, am o pm, tiempo en 24 horas, entre otros.

Por otro lado, la dimensión de fecha contiene los campos referentes a la misma, como son: día, año, mes, nombre del día, día de la semana, entre otros.

Existen otras dos dimensiones: Aceleraciones y Sismos. La primera contiene las aceleraciones tomadas de los sensores sismológicos, referentes a los datos de cada eje, **N**, **E**, **Z** de cada estación CGIG, HPIG, HSIG. La construcción de la dimensión de Aceleraciones se realizó de la relación entre las aceleraciones con cada uno de los sismos, donde las primeras se etiquetaron con números positivos de acuerdo al sismo que representan en la dimensión de Magnitudes, aunque pertenecen a la misma clase: sismos.

La figura 4.6 muestra la manera en que los datos de cada archivo con extensión .csv fueron cargados a la base de datos. Los archivos contienen los datos de cada día para cada estación. Los datos se convierten de texto a datos numérico.

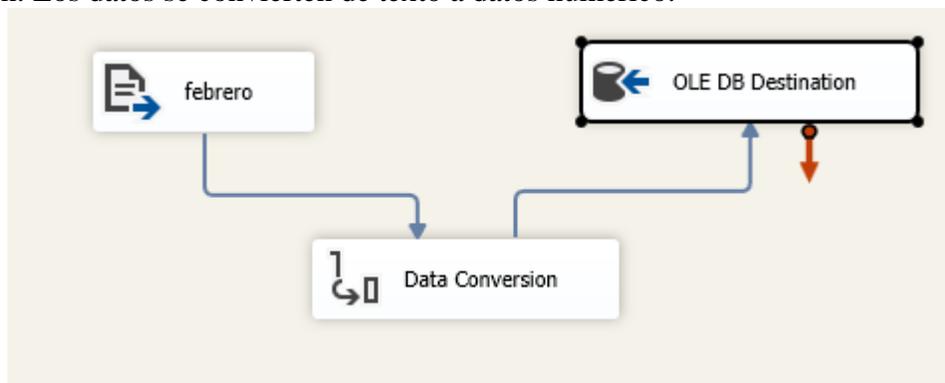


Figura 4.6: Carga de los archivos diarios al almacén de datos.

La carga del almacén de datos se realizó a través de SQL Server Integration Services®, la figura 4.7 muestra la estructura de los elementos utilizados para la carga de las aceleraciones, la creación de la dimensión de magnitudes se realizó de manera similar. Donde primero se realizó la carga de los valores de los datos de las aceleraciones, luego se conformó un atributo de fecha, la cual consta del año, mes y día, dado que los días y meses de un solo dígito debían representarse con un cero antes de este valor, se calculó la longitud de estos campos para determinar la forma correcta de expresarlos, siendo así que el día 1 de enero de 2013 se expresaría como “20130101”. Después de esto, se calculó la resultante de las sumas vectoriales de la resultante de cada estación y con esto se conformó la dimensión

de Aceleración. La dimensión de Magnitudes se cargó directamente al almacén de datos, teniendo en cuenta la conversión de la hora de manera similar a la conversión de la fecha antes mencionada.

Cada valor de aceleración representa el promedio de los cien valores que se presentaron en cada segundo, como se mencionó anteriormente. Por otro lado, las aceleraciones que no representan un sismo se etiquetaron con el valor de cero. Dado que cada sismo tiene distinta duración, se tomaron valores arbitrarios para determinar el rango de señales que serían tomadas como parte de un sismo. Los valores arbitrarios son: 15 segundos antes del sismo reportado por el SSN y 45 segundos después de que ocurrió el evento. Debido a que cada sismo tiene duración distinta, se toman estos valores tratando de uniformizar los tiempos tomados para su análisis.

La dimensión de sismos, representa los datos obtenidos del SSN, los cuales constan de los atributos: magnitud, profundidad, zona geográfica, latitud y longitud.

La tabla de hechos corresponde a la parte central del almacén de datos, quien tiene la relación de cada una de las dimensiones. La figura 4.8 muestra el diagrama del almacén de datos y cada uno de los atributos de las dimensiones.

La figura 4.9 muestra un ejemplo de los datos pertenecientes a la tabla de hechos del almacén de datos.

La columna de "*IdAcel*", en la tabla de hechos, corresponde a la relación con la dimensión de aceleraciones, la cual tiene 61 registros por cada sismo. La columna "*DateKey*" es el enlace con la dimensión de fecha, este campo está compuesto por el año, mes y día del evento sísmico. De la misma manera los atributos "*TimeKey*" y "*IdSismo*" se referencian a las dimensiones tiempo y magnitudes, respectivamente. Los atributos de medición, "*Magnitud*" y "*Resultante*" son aquellos de los cuales se desea observar el comportamiento.

El campo “*Resultante*” corresponde a la suma de las resultantes de cada estación. La resultante vectorial de cada estación se calcula de la raíz cuadrada de la suma de los cuadrados de cada eje, como se mencionó anteriormente.

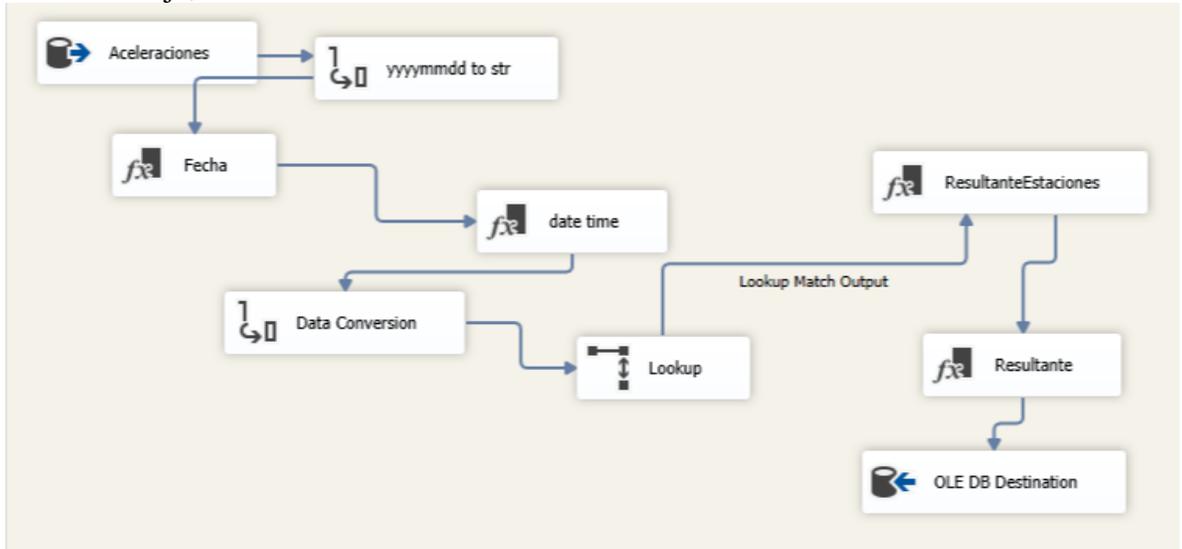


Figura 4.7: Creación de la dimensión de Aceleraciones



Figura 4.8: Diagrama del almacén de datos

IdAcel	DateKey	TimeKey	IdSismo	Resultante	Magnitud
1	20130827	01:57:56	1	5282.3876	3.9
2	20130827	01:57:57	1	5621.5723	3.9
3	20130827	01:57:58	1	5336.3226	3.9
4	20130827	01:57:59	1	5465.9541	3.9
5	20130827	01:58:00	1	5227.7987	3.9
6	20130827	01:58:01	1	5400.5009	3.9
7	20130827	01:58:02	1	5645.3410	3.9
8	20130827	01:58:03	1	5413.0265	3.9
9	20130827	01:58:04	1	5392.0833	3.9
10	20130827	01:58:05	1	5506.2313	3.9
11	20130827	01:58:06	1	5563.1742	3.9
12	20130827	01:58:07	1	5411.0572	3.9
13	20130827	01:58:08	1	5323.0480	3.9
14	20130827	01:58:09	1	5604.2444	3.9
15	20130827	01:58:10	1	5298.1238	3.9
16	20130827	01:58:11	1	5367.3233	3.9

Figura 4.9: Datos de la tabla de hechos.

Este almacén de datos será la base para los próximos análisis. La tabla de hechos contiene la información relevante respecto a la relación que existe entre las aceleraciones y los reportes de sismos.

4.3. Resultados de Actividades de la construcción del cubo OLAP

Para establecer un *cubo OLAP* y realizar análisis multidimensionales a los datos que represente las relaciones entre las dimensiones, observando el comportamiento de los campos de interés, se tendrá en cuenta que las medidas a observar son la magnitud y la suma de los vectores resultantes de cada estación. La figura 4.10 muestra la estructura del cubo realizado en *SQL Server Analysis Service (SSAS)* ®.

Una vez especificado el cubo, es posible realizar análisis a los datos. Se aplica minería de datos para determinar cuál es la relación de los campos Magnitud, Resultante y Profundidad, para esto se selecciona el clasificador *Árboles de Decisión*, tomando el 30% de los datos para realizar las pruebas de clasificación. Los análisis realizados corresponden a la clasificación de las magnitudes teniendo en cuenta la profundidad que registraron. La

figura 4.11 muestra el diagrama de árbol resultante de este análisis, donde se muestra que existe una relación directa entre la profundidad y la magnitud, así como el vector resultante con la magnitud.

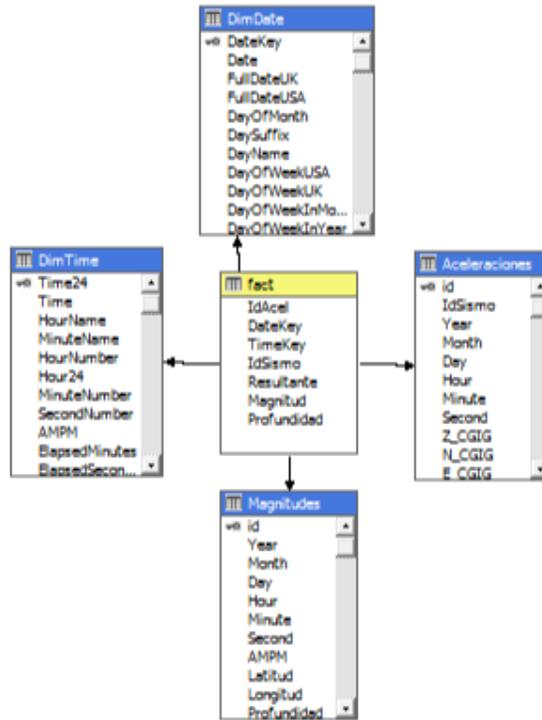


Figura 4.10: Esquema del cubo

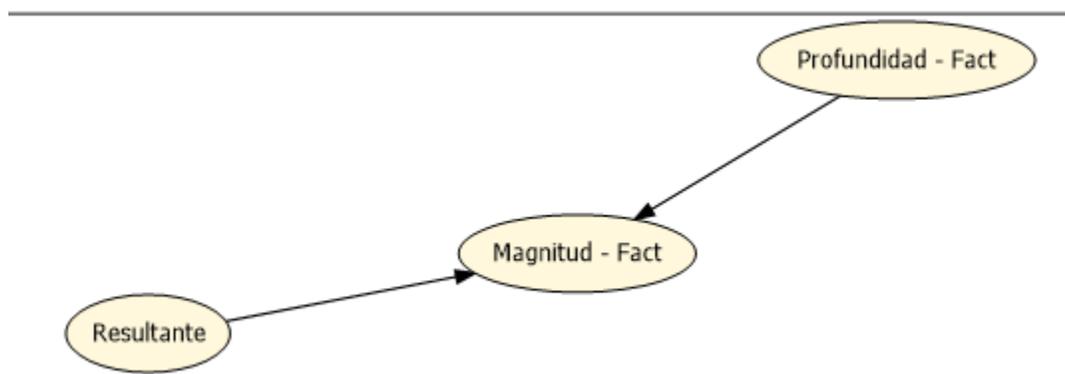


Figura 4.11: Esquema de relación Resultante-Magnitud-Profundidad

Se realizó el mismo análisis a través de árboles de decisión, tomando en cuenta los valores de latitud y longitud proporcionados por el SSN y la magnitud de cada evento,

teniendo como resultado el esquema presentado en la figura 4.12, donde se muestra que solo existe relación entre la profundidad y la magnitud.

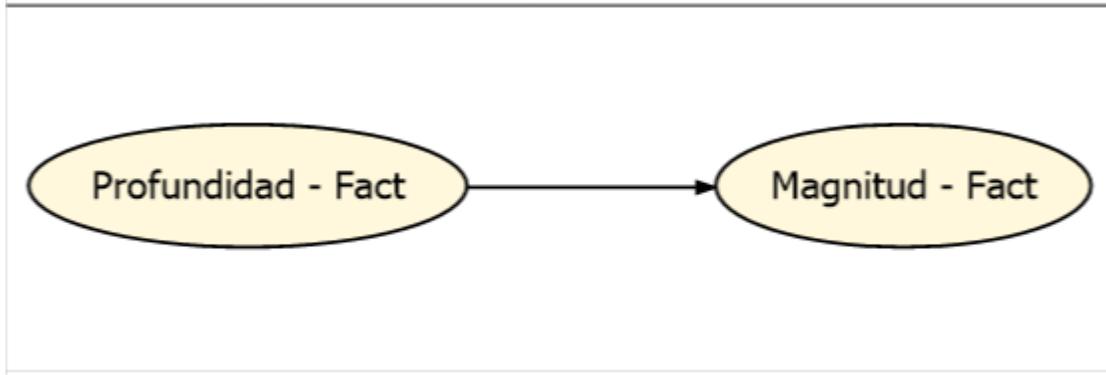
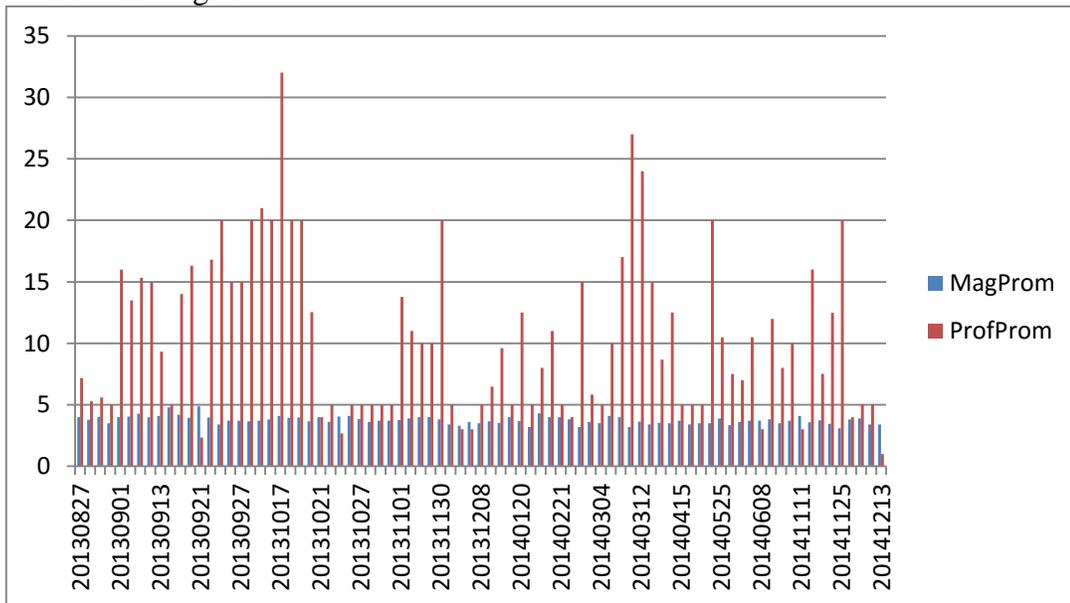
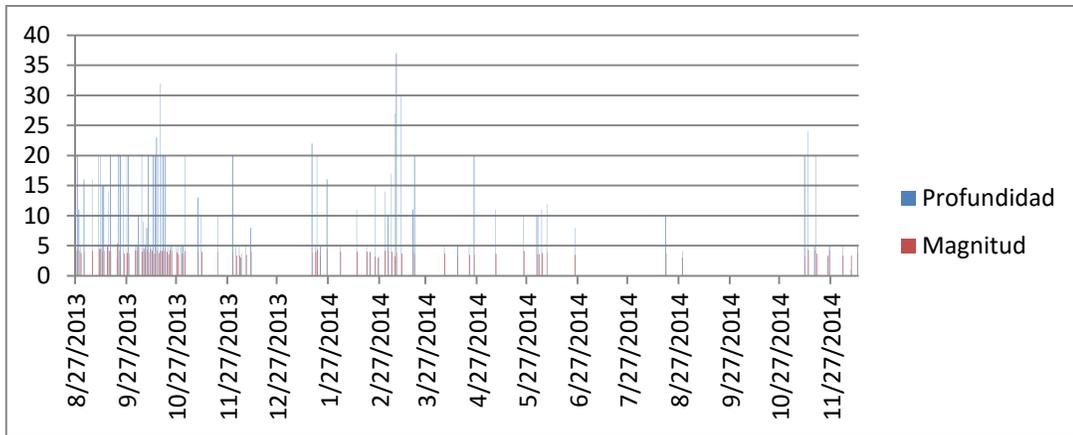


Figura 4.12: Esquema de relación Magnitud-Latitud-Longitud-Profundidad

Otro análisis fue realizado en Excel ®, donde se visualizó mediante una gráfica la relación del promedio de magnitudes por día y la profundidad del sismo. Esta gráfica se muestra en la gráfica 4.1. Comparando los datos de los promedios de magnitudes de cada día con los eventos publicados por el SSN, mostrados en la gráfica 4.2, se tiene que existe similitud entre ambas gráficas.

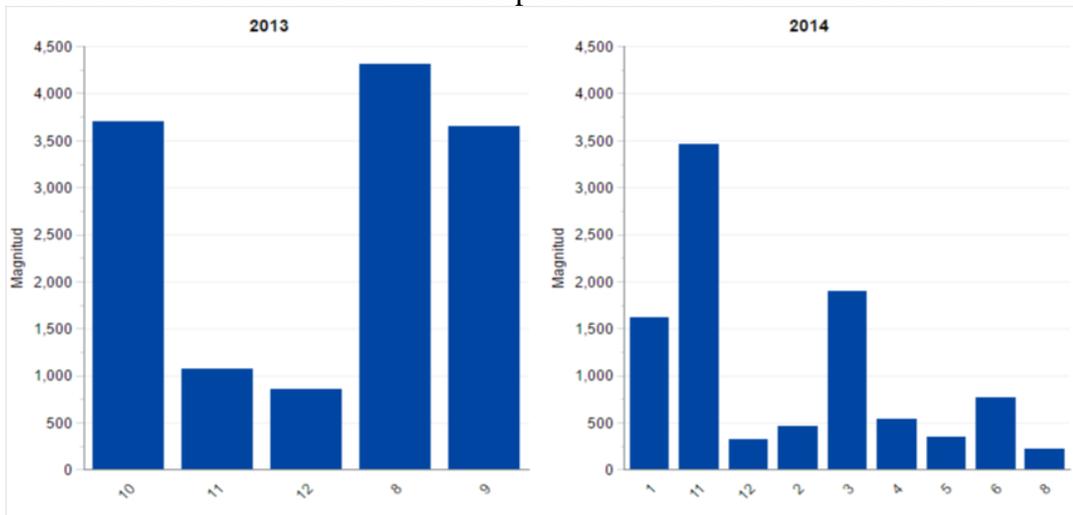


Gráfica 4.1: Relación Magnitud-Profundidad en el cubo de datos



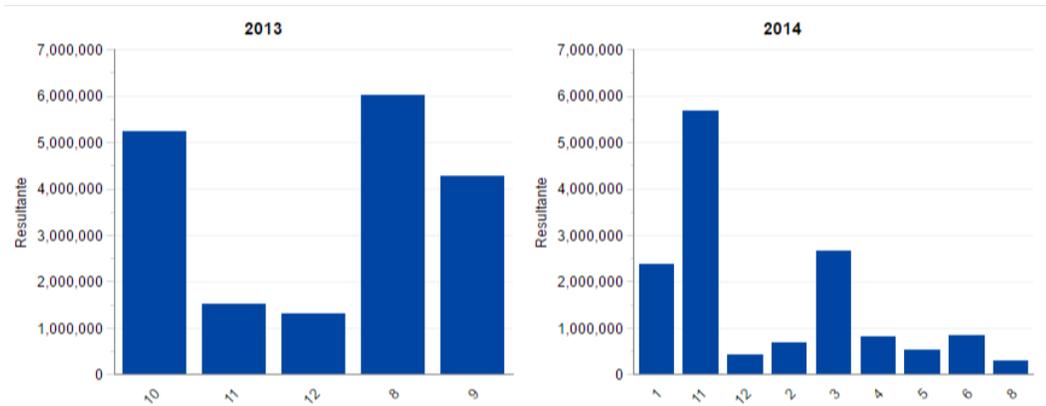
Gráfica 4.2: Relación Magnitud-Profundidad publicadas por el SSN

Por otro lado, se realizó un análisis a los datos a través de un *cubo OLAP* en Pentaho®, en el cual se estableció un reporte para realizar una comparación de los eventos registrados en los dos años, teniendo la gráfica 4.3 con dicha comparación, donde el eje horizontal representa el número de mes respecto a cada año. Se puede observar que existe mayor actividad en el año 2014, la escala de las gráficas muestra la suma total de las magnitudes de todas las aceleraciones tomadas para conformar el almacén de datos.



Gráfica 4.3: Comparación de eventos sísmicos en los años 2013 y 2014

La gráfica 4.4 muestra la comparación de los valores de las resultantes vectoriales. A través de esta gráfica se puede observar la relación que existe entre la magnitud y la profundidad, como se había mencionado anteriormente a través de la clasificación del árbol de decisión realizado en SSAS.



Gráfica 4.4: Comparación de las resultantes vectoriales

A partir de los resultados antes presentados, los expertos geólogos, civiles y de otras áreas, pueden analizar la actividad sísmica que se presenta en el estado de Chihuahua.

4.4. Resultados de Actividades de la aplicación de técnicas de reconocimiento de patrones

Para realizar búsquedas de tendencias de los datos a través de la aplicación de las técnicas de *Bag of Words*, se clasificaron las señales de las aceleraciones, éstas se tomaron del almacén de datos, de la dimensión de Aceleraciones, que contiene la relación de estas con los eventos sísmicos, así como también aceleraciones que no generaron ningún sismo.

Dado que las aceleraciones que pertenecen a algún sismo están tomadas en intervalos de 61 segundos, para este objetivo se tomaron periodos de tiempo similares para las aceleraciones que no pertenecen a algún sismo.

Cabe señalar que aunque existe una relación entre las dimensiones de Aceleraciones y Magnitudes, éstas no están relacionadas directamente en la base de datos. La relación proviene de la etiqueta para los eventos sísmicos, es decir, existe un número que indica a cada aceleración a que señal de sismo pertenece, como se mencionó anteriormente.

Debido a que es necesario agrupar los datos de prueba y entrenamiento, se toman los datos del año 2013 como datos de entrenamiento, tanto de sismos y no sismos, y los datos del año 2014 como datos de prueba.

Las señales de sismo y no sismo del año 2013 son utilizadas como señales de entrenamiento, se segmentaron deslizando continuamente una ventana con un tamaño de 5

segundos, para formar segmentos locales de señal. Una vez obtenidos estos segmentos, se generan los *codewords*, con un total de 10 *codewords* por cada archivo de valores de aceleraciones sísmicas. Tanto el valor de la ventana, como la cantidad de palabras formadas son valores arbitrarios.

Una vez que se tienen los segmentos, las 10 palabras o *codewords* son formadas por los centroides generados a través de la clasificación a través del algoritmo *k-means*, el cual obtiene una palabra representativa con la menor distancia al centro de las representaciones de todas las palabras pertenecientes a un archivo de datos.

Los archivos utilizados para formar estos segmentos contienen 61 registros formados por la suma vectorial de las resultantes vectoriales de las aceleraciones de cada estación sísmica, es decir, cada estación cuenta con tres ejes, **N**, **E**, **Z**. Cuando sucede un sismo, provoca un movimiento en cada eje, con lo cual se puede obtener un vector resultante de cada estación, la suma de los tres vectores, da el vector representativo del movimiento en las tres estaciones. En la figura 4.13 se puede observar la representación de un fragmento de 61 segundos de la resultante antes mencionada, el eje horizontal representa el tiempo medido en segundos y el eje vertical el valor de la resultante vectorial de aceleraciones.

El *codebook* o alfabeto es formado al juntar los centroides de todos los segmentos utilizados para entrenamiento. Los histogramas son formados de la segmentación de los archivos de prueba junto con el *codebook*, al contar la cantidad de segmentos similares. La figura 4.14 muestra un histograma. Al momento de que los *codewords* son generados se colocan aleatoriamente garantizando la aleatoriedad. En la figura 4.15 se muestran ejemplos de segmentos de la resultante.

Tanto el entrenamiento como las pruebas se realizaron 20 veces con los mismos conjuntos de datos, obteniendo en cada uno de ellos diferentes resultados, teniendo un histograma por cada una de estas veces que se aplicó la técnica a los conjuntos de datos. Una vez promediados estos resultados se calcula la exactitud con que fueron clasificados

las señales tanto de sismos como de no sismos, respecto a las señales de entrenamiento. Este procedimiento se llevó a cabo con cinco algoritmos: *K-Nearest Neighbors (KNN)*, *Decision Tree (DT)*, *Support Vector Machines (SVM)* y *RandomForest (RF)*. Estos métodos se utilizaron basados en la implementación de *scikit-learn* basado en el paquete de Python (Pedregosa, y otros, 2011).

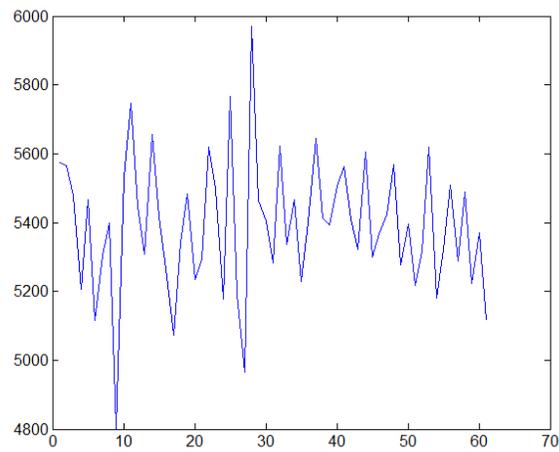


Figura 4.13: Resultante vectorial

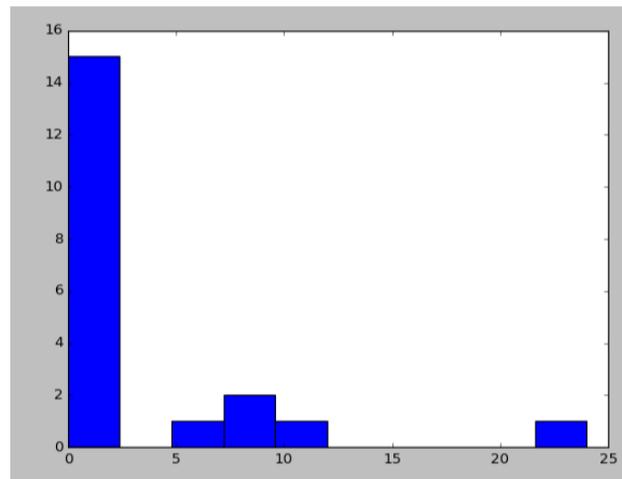


Figura 4.14: Ejemplo de histograma

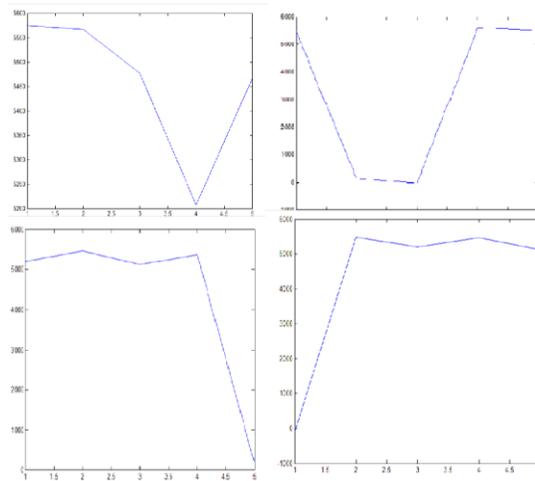


Figura 4.15: Segmentos de señales

La tabla 4.4 muestra las configuraciones utilizadas para cada algoritmo.

Tabla 4.4: Parámetros de configuración de los algoritmos

KNN	n_neighbors=2, weights='distance', algorithm='auto', leaf_size=30, p=2, metric='minkowski', metric_params=None, n_jobs=1
DT	criterion='gini', splitter='best', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=None, random_state=None, max_leaf_nodes=None, class_weight=None, presort=False
SVM	C=1.0, cache_size=200, class_weight=None, coef0=0.0, decision_function_shape=None, degree=3, gamma='auto', kernel='poly', max_iter=-1, probability=False, random_state=None, shrinking=True, tol=0.001, verbose=False
RF	n_estimators=10, criterion='gini', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto', max_leaf_nodes=None, bootstrap=True, oob_score=False, n_jobs=1, random_state=None, verbose=0, warm_start=False, class_weight=None

La figura 4.16 muestra las matrices de confusión de cada algoritmo. La exactitud representa aquellas instancias clasificadas correctamente respecto al total de las mismas analizadas. A continuación, en la tabla 4.5, se presentan los promedios de las exactitudes para cada uno de los algoritmos analizados. La red neuronal artificial (ANN, por sus siglas en inglés) se implementó de acuerdo a Masrland (2015).

Los algoritmos utilizados dieron resultados similares. La cantidad de eventos correctamente clasificados se obtiene sumando los eventos de sismo clasificados correctamente, que se muestran en el cuadrante superior izquierdo, más los eventos de no sismo correctamente clasificados que se encuentran en el cuadrante inferior derecho, este cálculo es conocido como exactitud.

Tabla 4.5: Resultados de exactitud para los distintos algoritmos

Algoritmo	Exactitud (%)
<i>DecisionTreeClassifier</i>	67.7
<i>KNeighborsClassifier</i>	74.9
<i>RandomForestClassifier</i>	73.1
<i>SupportVectorClassification</i>	72.7
<i>ArtificialNeuralNetwork</i>	73.78

La totalidad de estos valores representan la cantidad de instancias que fueron utilizadas para la realización de pruebas.

Por otro lado, se realizó otro análisis en Matlab ®, donde se analizaron los histogramas obtenidos en la técnica de *Bag of Words*, los cuales se examinaron a través de 2 algoritmos de redes neuronales.

El primer caso fue el algoritmo *Levenberg-Marquardt backpropagation*. Por medio de la instrucción

```
net.trainFcn = 'trainlm';
```

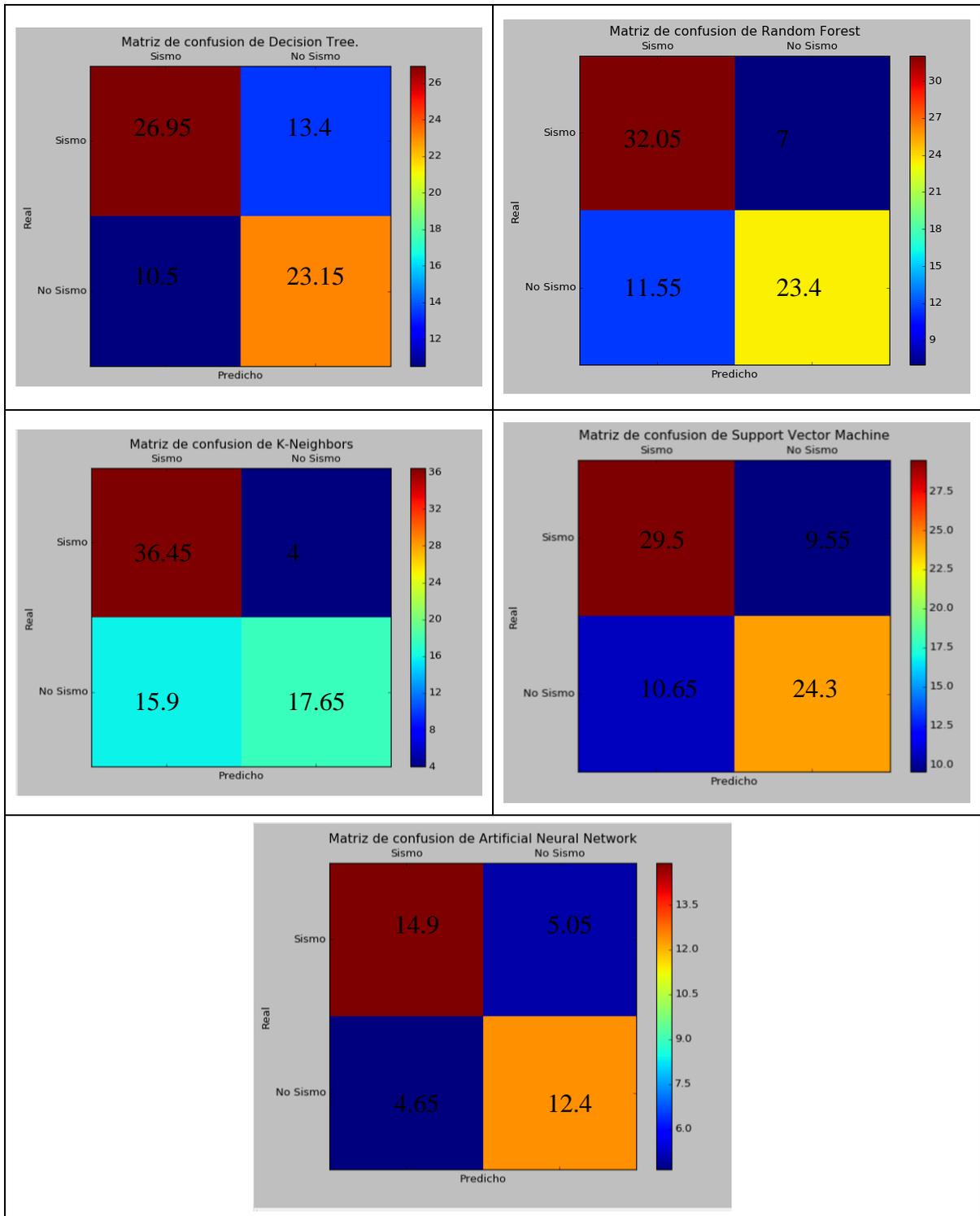


Figura 4.16: Matrices de confusión para los distintos algoritmos

donde previamente se había creado una red neuronal, la cual contiene como entradas los histogramas y como *targets* los valores que especifican si el histograma representa un sismo o no. Se tiene una red neuronal con 20 características, que se especificaron al momento de segmentar las señales. Otra capa, la capa oculta, contiene 40 neuronas y la última capa contiene 4 neuronas, las cuales determinan si los histogramas han sido clasificados como representantes de un sismo o no. La estructura de la red neuronal se muestra en la figura 4.17.

Este algoritmo clasificó correctamente el 78.5% de los datos en promedio de los 20 histogramas.

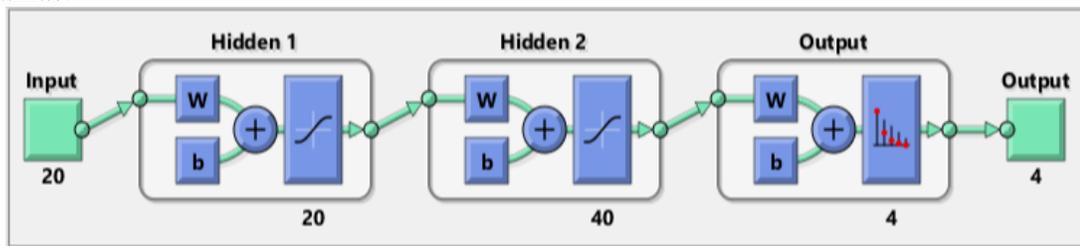


Figura 4.17: Estructura de la red neuronal

Además de este algoritmo, se realizaron pruebas con el algoritmo *LVQ2*, mediante la instrucción

```
net.inputWeights{1}.learnFcn = 'learnlv2';
```

dando como resultado de clasificación 74.9%, la estructura de la red neuronal es igual a la mostrada anteriormente.

Para ambos algoritmos se utilizó el 60% de los datos para entrenamiento, el 20% para pruebas y el 20% para validación, con un máximo de 500 iteraciones.

La figura 4.18 muestra la manera en que se relacionan los datos utilizados para los análisis antes mencionados y las herramientas utilizadas en esta investigación.

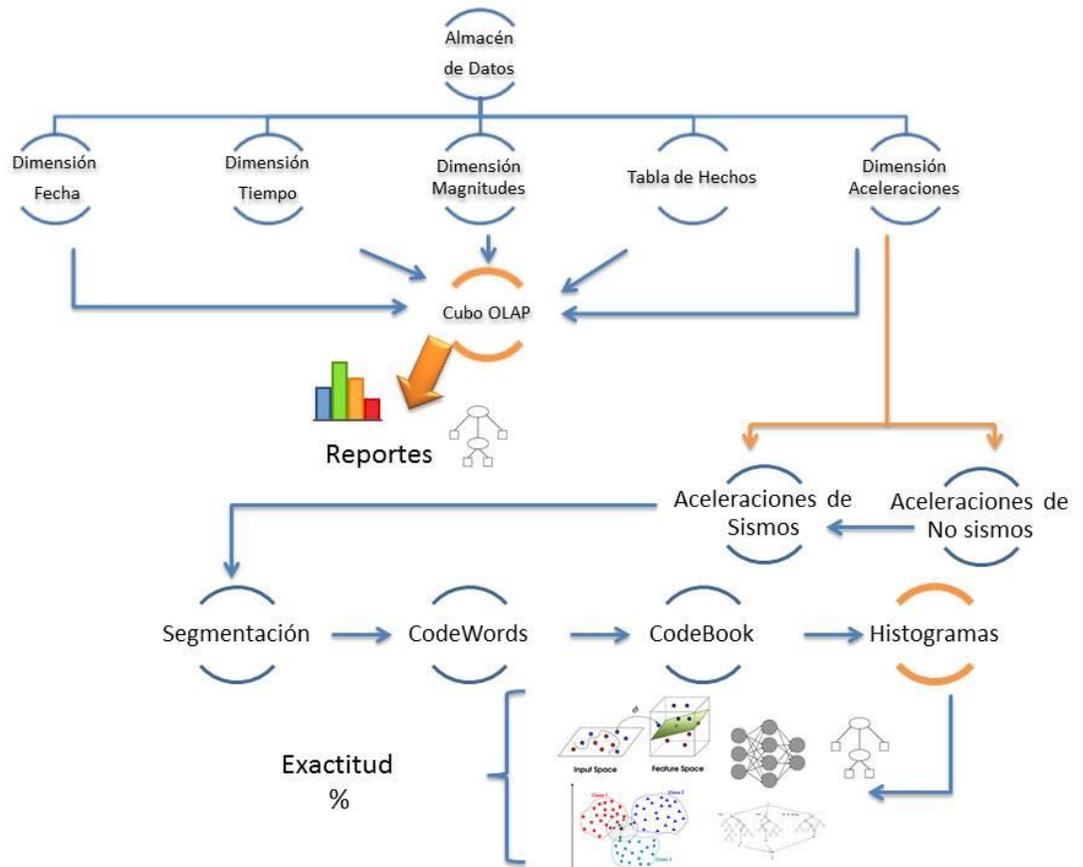


Figura 4.18: Diagrama de relación entre datos y procesos

Capítulo V

Conclusiones

Se logró conformar un almacén de datos representativo de las señales obtenidas de tres sismógrafos, para con él realizar distintos análisis. Este almacén contiene la relación entre las aceleraciones y los sismos reportados por el SSN. Para llevar a cabo la conformación de este almacén se realizó la reducción de los datos, comprobando que ambos conjuntos de datos representan señales similares, por lo tanto, es posible utilizar la señal reducida para su análisis.

Por otro lado, la creación del almacén de datos permite realizar consultas con menor tiempo en comparación con una base de datos relacional, lo que facilitó la creación del cubo de datos. Con este último, a través de la integración de distintos análisis se logró comprobar que no existe relación entre el vector resultante y la magnitud, sino que esta última está más bien determinada por la profundidad del evento sísmico.

A partir de este almacén se aplicaron técnicas de reconocimiento de patrones, lo que permitió establecer un punto de partida para su posterior análisis a profundidad. Dado que no existía la aplicación de técnicas similares, no existe punto de comparación, por lo tanto, se establece que el porcentaje de clasificación obtenido es considerado como porcentaje alto.

Los resultados aquí expuestos son de utilidad para expertos geólogos y civiles en cuanto a la cantidad de movimiento telúrico de magnitudes pequeñas en la escala de Richter, sin embargo, cabe destacar que los análisis antes mencionados se aplicaron a un modelo matemático en base a los datos de aceleraciones obtenidas de tres estaciones sísmicas. Sin embargo, para mejorar este análisis y su aplicación, en futuros trabajos sería importante incluir aspectos geológicos como tipo de suelo, distancia al epicentro, magnitud momento, entre otras, para con ello obtener resultados más certeros.

La calidad de los resultados obtenidos en esta investigación no fueron relevantes comparados con otros trabajos en la misma área, dado que se utiliza como datos a analizar

únicamente las aceleraciones. Otro factor a considerar en estos resultados es la cantidad de tiempo que se tomó como señal precursora de un sismo, sin duda, tomar la señal completa del sismo sería un aporte esencial para mejorar estos resultados.

Otro aspecto es la brecha que deja abierta esta investigación para realizar distintos análisis en la Facultad de Ingeniería a través de las distintas áreas de investigación que conjuntamente pueden realizar trabajos de campo con las estaciones sísmicas pudiendo mejorar los resultados, así mismo, puede llevarse a cabo la clasificación de las señales de acuerdo a la magnitud en escala de Richter que representan.

Referencias

- A. Konstantaras, M. R. (2007). Detection of Weak Seismo-Electric Signals Upon the Recordings of the Electrotelluric Field by Means of Neuro-Fuzzy Technology. *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*.
- Ahumada Tello, E., & Perusquia Velasco, J. (2014). Inteligencia de negocios: estrategia para el desarrollo de competitividad en empresas de base tecnológica. *ScienceDirect*.
- Almeida, M. S. (1999). Getting Started with Data Warehouse and Business Intelligence. *International Technical Support Organization*.
- BISHOP, C. M. (1995). *Neural Networks for Pattern Recognition*.
- Brett, P. T. (2013). *Earthquake Damage Detection in Urban Areas Using Curvilinear Features*.
- CENAPRED. (2013). Guía de prevención de desastres. *Centro Nacional de Prevención de Desastres*.
- Chowdhury, J. L. (2004). Best Practices in Data Warehousing to Support Business Initiatives and Needs. *Proceedings of the 37th Hawaii International Conference on System Sciences*.
- Database, O. (2005). Data Warehousing Guide. *Data Warehousing Guide*.
- de Souza, F. T. (2010). A Data Mining approach to predict mass movements induced by seismic events in Sichuan, China. *2010 Sixth International Conference on Natural Computation*.
- Dehne, F., Kong, Q., Rau-Chaplin, A., Zaboli, H., & Zhou, R. (2014). Scalable real-time OLAP on cloud architectures. *Elsevier*.
- Ding, Y., & Fu, X. (2014). Neurocomputing. *Science direct*.
- Edwards, B., & Douglas, J. (2014). Magnitude scaling of induced earthquakes. *Elsevier*.
- Emilio Pedrinaci, S. A. (2013). Alfabetización en Ciencias de la Tierra. *Enseñanza de las ciencias de la tierra*.
- Estrada, L. (2012). APUNTES DE SISMOLOGÍA. *FACET – UNT – Sismología para Geólogos*.
- Fernández, A. (2015). Redes de Observación Atmosférica y Ambiental. *Centro de Ciencias de la Atmósfera, UNAM*.
- Golfarelli, M., & Rizzi, S. (2009). *Data Warehouse Design: Modern Principles and Methodologies*. McGraw Hill.
- González, L., Moreno, R., Escalante, H., Martínez, F., & Carlos, R. (2016). Learning Roadway Surface Disruption pattern using the Bag of Words representation. *IEEE TRANSACTIONS ON INTELLIGENT SYSTEMS*.
- Gosain, A., & Heena. (2015). Literature Review of Data model Quality metrics of Data Warehouse. *ScienceDirect*.
- Hallo Carrasco, M. (2006). ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART PARA EL ÁREA DE SISMOLOGÍA DEL DEPARTAMENTO DE GEOFÍSICA DE LA ESCUELA POLITÉCNICA NACIONAL. *ESCUELA POLITÉCNICA NACIONAL*.
- Han, J. (2012). *Data Mining Concepts and Techniques*.
- Hoffer, J. A. (2013). *Modern Database Management*.
- Hovem, J. (2012). *Modeling Propagation of Seismic Airgun Sounds and the Effects on Fish Behavior*.

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*.
- Jazdani, A. (2013). Bayesian estimation of seismic hazards in Iran. *Scientia Iranica*.
- Kamal Boulil, F. L. (2014). Multidimensional modeling and analysis of large and complex. *Ecological Informatics*, 90-106.
- Kevin Chihcheng Hsu, M.-Z. L. (2011). Techniques for finding similarity knowledge in OLAP reports. *Expert Systems with Applications*.
- Kimball, R., & Ross, M. (2013). *The data Warehouse toolkit*.
- Kou-Yuan Huang. (1999). Neural Networks for Seismic Principal Components Analysis. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*.
- Lungu, I. (2013). Business Intelligence, ERP, Data Mining and Data Warehouse. *Database Systems Journal*.
- Madrid, R. Z. (2011). Notas introductorias Sismología. *POSGRADO EN CIENCIAS DE LA TIERRA CENTRO DE GEOCIENCIAS*.
- Mansmann, S., Rehman, N., Weiler, A., & Scholl, M. (2013). Discovering OLAP dimensions in semi-structured data. *Elsevier*.
- Marsland, S. (2015). *Machine Learning, an algorithmic perspective*.
- Mehta, A. (2000). Automated Workload Management for Enterprise Data Warehouses. *Hewlett Packard Labs*.
- Monteiro, R. (2015). Probabilistic Seismic Assessment of RC Bridges: Part I — Uncertainty Models. *Elsevier B.V.*
- Niemi, T., Niinimäki, M., Thasnisch, P., & Nummenmaa, J. (2014). Data and Knowledge Engineering. *Elsevier*.
- Nofal, M. I. (2013). Integration of Business Intelligence and Enterprise Resource Planning within Organizations. *The 4th International Conference on Electrical Engineering and Informatics*.
- Oleg Aulov, M. H. (2012). Human Sensor Networks for Improved Modeling of Natural Disasters. *Proceedings of the IEEE*.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michael, V., Thirion, B., Giel, O., y otros. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*.
- PONNIAH, P. (2001). DATA WAREHOUSING FUNDAMENTALS. *John Wiley & Sons, Inc.*
- Quintana, A. (2008). sismos en Chihuahua. *Universidad Autónoma de Chihuahua*.
- Ranjan, J. (2009). BUSINESS INTELLIGENCE: CONCEPTS, COMPONENTS, TECHNIQUES AND BENEFITS. *Journal of Theoretical and Applied Information Technology*.
- Roberto, V. (1992). Seismic Signal Understanding: A Knowledge-Based Recognition System. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*.
- Sen, A. (2006). data warehousing process maturity: an exploratory study of factors influencing user perceptions. *ieee transactions on engineering management*.
- SSN. (s.f.). *SSN UNAM*. Recuperado el 2016, de <http://www2.ssn.unam.mx/>
- Tsutsui, M. (2014). Behaviors of Electromagnetic Waves Directly Excited by Earthquakes. *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*.
- Villard, A. U. (2013). The Effects of Management Information and ERP Systems on Strategic Knowledge Management and Decision-Making. *Social and Behavioral Sciences* .

- Vora, M. (2015). Modelling Architecture for Multimedia Data. *International Journal of Innovative Research in Science, Engineering and Technology*.
- Wang, J., Liu, P., She, M., & Nahavandi, S. (2012). Bag-of-words Representation for Biomedical Time Series Classification. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*.
- Wang, Y. (2011). Analysis of seismic electromagnetic phenomena using the FDTD method. *IEEE transactions on antennas and propagation*.
- Zhiliang, M. (2008). A Decision Support System for Construction Projects Based on Standardized Exchanged Documents. *TSINGHUA SCIENCE AND TECHNOLOGY*.
- Zúñiga, F. R. (2011). NOTAS INTRODUCTORIAS SISMOLOGÍA. *POSGRADO EN CIENCIAS DE LA TIERRA CENTRO DE GEOCIENCIAS*.
- ZÚÑIGA, F. R. (2011). NOTAS INTRODUCTORIAS SISMOLOGÍA. *POSGRADO EN CIENCIAS DE LA TIERRA CENTRO DE GEOCIENCIAS*.