

Caracterización de la Actividad Sísmica de la República Mexicana: Una Primera Aproximación

Jesus David Velasco Mendoza, Juan Fernando Garcia Mejia, Jose Antonio Garcia Mejia

Centro Universitario UAEM Atlacomulco
Carretera Toluca-Atlacomulco Km.60 Atlacomulco Estado de Mexico Tel. 01 712 12 2 04 46, jesus_david_velasco@hotmail.com

Resumen. En el periodo que comprende los años 2006 - 2010 la incidencia de terremotos con magnitud mayor a 6.0 grados en escala Richter han sido frecuentes por lo tanto es de interés la predicción de la intensidad, ubicación temporal y espacial de estos fenómenos naturales. Con esto en mente en el presente trabajo se muestra una primera caracterización de la actividad sísmica en la República Mexicana; mediante el estudio de una base de datos en la cual se encuentran almacenados registros sísmicos de 2006 a 2010, para este fin se empleó Weka -el cual es un entorno para análisis del conocimiento-, Cabe destacar que la caracterización que se pretende realizar está fundamentada en ideas casuísticas.

Abstract. From 2006 to 2010 the incidence of earthquakes with magnitude greater than 6.0 degrees on Richter scale has been frequent that is why the interest to predict the intensity, time and space of these natural phenomenon. Having this in mind, it is presented an initial characterization of the seismic activity in the Mexican Republic, by studying a database in which are stored seismic records from 2006 to 2010, Weka which is an Environment for Knowledge Analysis was used for this purpose. It is important to mention the characterization is based on casuistry ideas.

Keywords: Clustering, Minería de datos, Weka, Actividad Sísmica

1 Introducción

La sismología es la disciplina que estudia las causas y el mecanismo de los terremotos, la definición sobre bases teóricas y experimentales de los parámetros que caracterizan la aparición y transmisión de los choques sísmicos, y también la predicción de dichos fenómenos. [1]

En 1947 Housner consideró al terremoto o sismo como un proceso de naturaleza aleatoria [2] El cual es un fenómeno natural ocasionado por el rompimiento de la corteza terrestre, causado por la tensión entre las placas tectónicas.[3]

En el estado del arte sobre predicción sísmica se cuentan con métodos estadísticos y probabilísticos sin embargo estos no han logrado dar un rango de tiempo corto en el

cual se puede presentar un evento sísmico y no se tiene precisión en la ubicación y zona exacta.

En el artículo “The Third Asymptotic Distribution of Largest Magnitudes in the Balkan Earthquake Provinces” se crea una predicción para el retorno de los grandes sismos en base de una distribución asintótica, donde se trabajó bajo la premisa de que “no existe nada estocástico sobre el origen de los terremotos”. En el trabajo se consideran tres supuestos que debieran como lo son:

- Las condiciones del pasado deben ser iguales a las del futuro
- Los terremotos con mayor magnitud son independientes uno de otro
- El comportamiento de los terremotos más fuertes será similar a la del pasado

Además se hace mención sobre el tratamiento gráfico o numérico de la información sobre los periodos de retorno de los terremotos para la posibilidad de estimar su magnitud.

Otra propuesta teórica se documenta en el artículo “Anomalous land surface temperature and outgoing long-wave radiation observations prior to earthquakes in India and Romania” donde se hace mención de algunos precursores en la aparición de un terremoto como lo son las mediciones del infrarrojo térmico (TIR) donde los parámetros observados son la temperatura de la superficie de la tierra (land surface temperature) y la salida de radiación de onda larga (outgoing long-wave radiation) por sus siglas en inglés LST y OLR respectivamente, las cuales pueden ser captadas con satélites especialmente equipados, para su explicación se vale de dos terremotos moderados como lo fueron los presentados el 22 de julio de 2007, en Yamnotri, India y el 27 de octubre de 2004 en Vrancea, Rumania cuyas anomalías se presentaron entre los 7 y 8 días antes del evento principal cabe destacar que estas anomalías desaparecieron junto con el terremoto.

En el artículo “Parameter estimation of the stochastic AMR model and its application to the study of several strong earthquakes” fue elaborado en base a los terremotos ocurridos en China y Nueva Zelanda con el objetivo de analizar si existe o no una posible aceleración sísmica antes de un al igual se mencionan algunas características de los terremotos como la distribución de dona o la quietud sísmica antes de los grandes terremotos. Como resultado del trabajo se pudo estimar el tiempo de ocurrencia de un terremoto fuerte con un error aproximado de ± 1.28 años.

En la actualidad no se utilizan técnicas de minería de datos para la caracterización de terremotos puesto que estas son relativamente nuevas y los sismólogos las desconocen parcial o totalmente y lo mismo sucede con los especialistas en dichas técnicas que desconocen o no le toman importancia a la caracterización sísmica.

Por tal motivo en este trabajo se propone implementar minería de datos para caracterizar la actividad sísmica en México la cual se entiende como: una alternativa que ayuda a identificar, analizar y extraer la información relevante que no siempre es clara en grandes volúmenes de datos [4] a partir de la información almacenada en esta, se identifican de patrones, tendencias, desviaciones, etcétera, con la finalidad de crear un mejor criterio para la toma de decisiones en súper mercados, escuelas, empresas, industrias, y otras áreas del saber.

Actualmente la minería de datos es aplicada en almacenamiento de texto he imágenes, bioinformática, imágenes médicas, en el cuidado de la salud, en las finanzas e investigaciones, fabricación y producción, negocios y marketing, telecomunicaciones y en la world wide web (WWW) [5].

Actualmente existen aplicaciones de datamining como Weka el cual es un entorno para análisis del conocimiento, desarrollado en la universidad de Waikato la cual cuenta con herramientas de pre-procesamiento, clasificación, regresión, clustering, reglas de asociación y de visualización, es software de código abierto bajo los términos de la GNU GPL y es un banco de herramientas para aprendizaje automático y minería de datos. Este software cuenta con 4 opciones posibles de trabajo.

- Simple CLI: en la cual se trabaja por medio de líneas de comando.
- Explorer: permite llevar a cabo la ejecución de los algoritmos.
- Experimenter: permite definir experimentos complejos, con objeto de ejecutar algoritmos sobre conjuntos de datos de entrada, y comparar estadísticamente los resultados.
- KnowledgeFlow: permite llevar a cabo las mismas acciones del "Explorer", con una configuración totalmente gráfica.

También alberga seis tareas fundamentales de la minería de datos como lo son: Preprocess, Classify, Cluster, Associate, Select attributes y Visualize

2 Técnicas de clasificación

Existen dos formas de clasificación de algoritmos en el entorno de Weka, los supervisados y las no supervisadas.

Algoritmos de clasificación supervisada: Se encuentran en el panel de trabajo “Classify” además de algunas tareas de regresión, en él se encuentran gran número de algoritmos que se dividen en siete familias como lo son:

- *Bayes.*
- *Functions.*
- *Lazy.*
- *Meta.*
- *Miscellaneous.*
- *Rules.*
- *Trees.*

De los cuales se originan los algoritmos *Logistic Model Trees* y *Naive Bayes* que a continuación se explican:

El algoritmo ***Logistic Model Trees (LMT)***, es una combina modelos de regresión logística con árboles de inducción. Es un modelo de árbol con funciones de regresión en las hojas, Como en un árbol de decisión ordinario, una prueba en uno de los atributos está asociada con cada nodo interno. Para un atributo nominal con “k” valores, el nodo tiene “k” nodos secundarios y los casos son clasificados hacia abajo en una de las “k” ramas dependiendo del valor del atributo. Para atributos numéricos, el nodo tiene dos nodos secundarios y la prueba consiste en comparar el valor del atributo con un umbral: un caso es ordenado hacia abajo en la rama izquierda si su valor para el atributo es más pequeño que el umbral, y en la rama derecha en el caso contrario. [6]

El algoritmo ***Naive Bayes (NB)***, es considerado como el modelo más simple, eficiente y eficaz para el aprendizaje automático y minería de datos es ampliamente usado en procesos de clasificación, basado en una red Bayesiana (utiliza la

distribución de Gauss para la estimación de la densidad), trabaja sobre el principio que todos los atributos son independientes entre sí y parte de nodo raíz, correspondiente a la clase. Pero su capacidad de representación es muy limitada ya que sólo puede representar ciertas funciones. [7]

Algoritmos de clasificación no supervisada: Estos se encuentran en el panel de trabajo “Cluster” Aquí se encuentran algoritmos tales como:

- Cobweb.
- EM.
- Farthest First.
- Filtered Clusterer.
- Hierarchical Clusterer.
- Make Density Based Clusterer.
- Simple KMeans.

Los algoritmos de clustering tiene la función de agrupar los datos de entrada originando como salida un patrón de comportamiento. [8] De los cuales destacan: Cobweb, EM, K-mens, entre otros.

Cobweb: Es un algoritmo de clustering de sistema jerárquico incremental el cual organiza las observaciones en un árbol de clasificación, Cada nodo de un árbol de clasificación representa una clase (concepto) y la etiqueta es un concepto probabilístico que resume la distribución de atributo-valor de los objetos clasificados en el nodo. Este árbol de clasificación se puede utilizar para predecir atributos que faltan o no a la clase de un nuevo objeto.

EM: Es una aproximación repetitiva para el cálculo de la máxima veracidad, utilizado para encontrar una estimación del conjunto de parámetros del modelo y así maximizar la probabilidad de generación de los datos de entrenamiento, $O = \{O_0, \dots, O_k\}$, partiendo del modelo λ , $P(O | \lambda)$ de tal modo que la probabilidad del nuevo modelo (λ^*) sea mayor o igual a la del modelo anterior.

$$P(O | \lambda^*) \geq P(O | \lambda) . \quad (1)$$

Es decir trata de maximizar la probabilidad del modelo partiendo de datos incompletos. [9]

K-mens: Es un algoritmo de clustering el cual consiste en la presentación de clusters y sus centros, automatiza una mezcla de atributos categóricos y numéricos. Además, el algoritmo normaliza automáticamente los atributos numéricos al hacer los cálculos de distancia de Euclides para calcular distancias entre casos y grupos.

Es un método de análisis de conglomerados que tiene como objetivo la partición de “N” observaciones en “K” grupos en los que cada observación pertenece al grupo más cercano con la media. Los pasos a seguir cuando se utiliza el cluster K-means son:

- Aplicar el cluster en el conjunto de entrenamiento en cada clase usando “N” prototipos.
- Asignar una etiqueta de clase de cada uno de los $N * K$ prototipos.
- Clasificar a una nueva función a la clase del prototipo más cercano.[10]

Desarrollo

En la historia sísmica de la República Mexicana se han presentado alrededor de 38 terremotos devastadores (de los que se tiene algún tipo de registro), los cuales fueron relacionándolos con las fases lunares en cuya relación se pudo observar que en luna llena se presentaron 16 sismos, en luna nueva 7, en cuarto menguante 8 y en cuarto creciente 7. (Tabla 1) Con la finalidad de verificar esa incidencia de sismos en luna llena se capturaron 9493 mayores a 2.1 grados en escala de Richter de enero del año 2006 al 31 de Diciembre de 2010, de la página del Servicio Sismológico Nacional. Al igual que se almaceno su relación con la fase lunar y el campo magnético interplanetario del sol.

Tabla 1. Sismos con magnitud mayor a los 6.4 grados Richter en la república mexicana.

Fecha	Lugar	Fase Lunar	Magnitud
Septiembre 23, 1902	Chiapas	Cuarto menguante	8.2
Octubre 16, 1902	Baja California	Luna llena	7.8
Enero 14, 1903	Oaxaca	Luna llena	8.2
Abril 14, 1907	Costa de Guerrero	Luna nueva	7.9
Julio 31, 1909	Costa de Guerrero	Luna llena	7.5
Junio 11, 1911	No precisado	Luna llena	7.8
Noviembre 19, 1917	Acambay	Cuarto creciente	7.0
Enero 2, 1920	Puebla	Cuarto menguante	6.4
Abril 30, 1921	Ciudad Guzmán	Luna llena	7.8
Marzo 22, 1928	Oaxaca	Luna nueva	7.7
Junio 17, 1928	Pinotepa Nacional	Luna nueva	8.0
Agosto 4, 1928	Oaxaca	Luna llena	7.4
Octubre 9, 1928	Puerto Escondido	Cuarto menguante	7.8
Junio 3, 1932	Jalisco	Luna nueva	8.4
Julio 26, 1937	Orizaba Veracruz	Luna llena	7.7
Abril 15, 1941	Colima	Luna llena	7.9
Febrero 22, 1943	Petatlan Guerrero	Luna llena	7.5
Julio 28, 1957	Guerrero	Luna llena	7.7
Agosto 26, 1959	Veracruz	Cuarto menguante	6.9
Mayo 11, 1962	Guerrero	Cuarto creciente	7.0
Julio 6, 1964	Michoacán	Cuarto creciente	6.9
Agosto 23, 1965	Oaxaca	Cuarto menguante	7.3
Agosto 2, 1968	Oaxaca	Cuarto creciente	7.1
Agosto 28, 1973	Veracruz	Luna nueva	7.3
Noviembre 29, 1978	Oaxaca	Luna nueva	7.6
Marzo 14, 1979	Guerrero	Luna llena	7.6
Junio 22, 1979	Oaxaca	Cuarto menguante	7.1
Octubre 24, 1980	Huajaan de León	Luna llena	7.0
Octubre 24, 1981	Colima	Cuarto creciente	7.3
Septiembre 19, 1985	Michoacán	Cuarto menguante	8.1

Octubre 9, 1995	Colima	Luna llena	8.0
Febrero 25, 1996	Oaxaca	Cuarto creciente	7.1
Junio 15, 1999	Puebla	Luna nueva	7.0
Septiembre 30, 1999	Oaxaca	Cuarto menguante	7.5
Agosto 9, 2000	Guerrero	Cuarto creciente	7.0
Enero 21, 2003	Colima	Luna llena	7.6
Febrero 12, 2008	Oaxaca	Cuarto creciente	7.0
Abril 4, 2010	Baja California	Luna llena	7.2

Posteriormente como herramienta para el estudio de los datos se utilizó el entorno de desarrollo “Weka” el cual no reconoce archivos de algunos gestores de bases de datos, dado lo anterior se migro del gestor de bases de datos Oracle 10g a un archivo reconocido por el software de desarrollo el cual reconoce de archivos del tipo .arff (Attribute-Relation File Format). En el cual los campos son: día, mes, hora, año, hora, latitud, longitud, profundidad, magnitud, fase lunar y El campo magnético interplanetario.

En el momento de cargar en el modo Experimenter se analizaron diferentes graficas de las cuales se podría inferir fácilmente que en el mes cuatro se tienen mayor número de posibilidades que ocurra un sismo fuerte ya que en este mes se ha tenido más número de terremotos durante 4 años. (Fig. 1.)

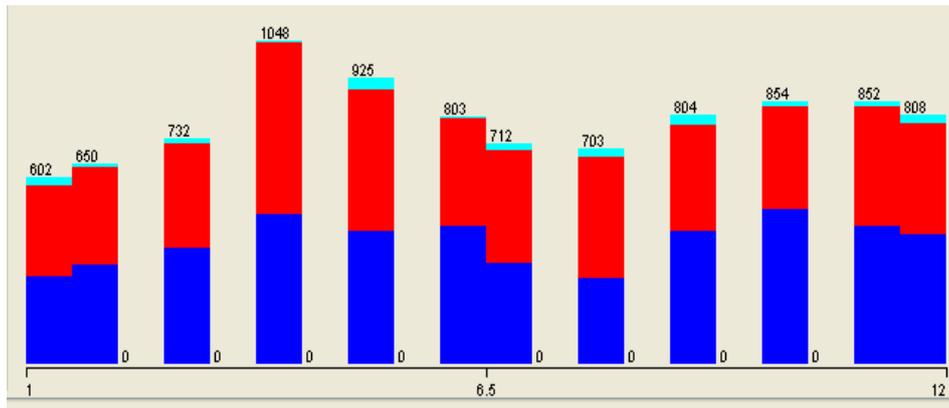


Fig. 1. Gráfica del (total de terremotos) presentados de 2006 a 2010 agrupados por el número de (mes) donde se muestra el mes 4 con una mayor incidencia de sismos.

Al contraponer las fases lunares contra el número de sismos es difícil determinar en qué fase podría presentarse un terremoto, Como se muestra en la (Fig. 2.) No hay un patrón completamente definido entre fase y recurrencia de la sismicidad

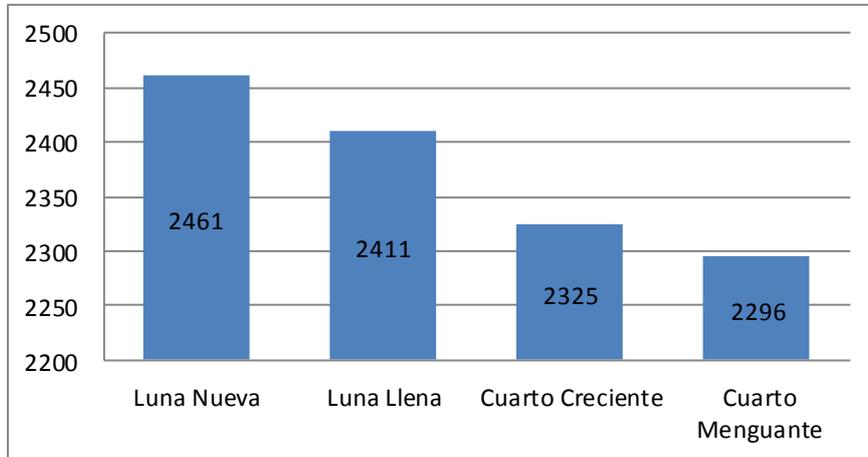


Fig. 2. Presentación de las (*fases de la luna*) y el (número de *sismos*) ocurridos en cada una de ellas siendo en la fase lunar “Luna nueva” donde un mayor número de sismos se registraron.

Se considera de importancia mencionar el incremento de la sismicidad presentada en los últimos 4 años. (Fig. 3.) En el año 2006 en la Republica Mexica se presentaron 971 terremotos, en 2007 se contabilizaron 1258, para 2008 el total fue de 1815, para 2009 hubo 2047 terremotos y para 2010 dio un salto abrupto hasta llegar a los 3402 terremotos.

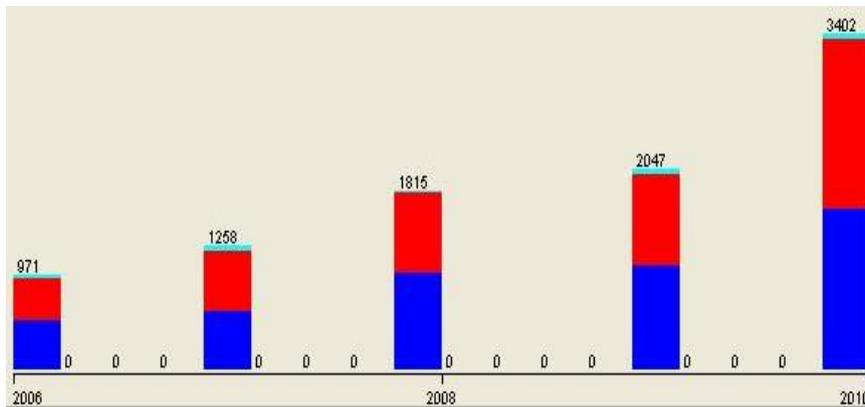


Fig. 3. (*Número de terremotos*) registrados en el periodo de tiempo comprendido de (2006 a 2010) demostrando el incremento del número de sismos que se presentan por año.

El campo magnético interplanetario se representa mediante B_t y B_z . Donde B_t : es la actividad total del campo magnético interplanetario y B_z es la actividad que afecta directamente a la tierra, Superponiendo las gráficas para comparar la magnitud de los terremotos con B_z se puede observar una mayor concentración de sismos cuando B_z es menor de 7.55. (Fig. 4.)

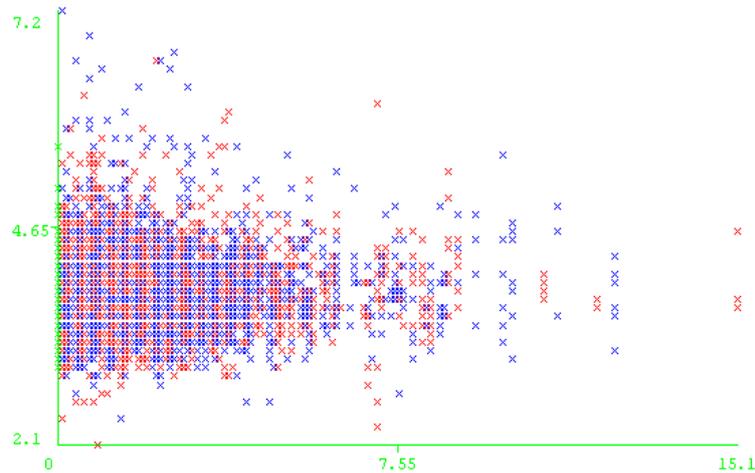


Fig. 4. Grafica donde se observa en el eje “y” la (*magnitud*) de los sismos de 2006 a 2010 y en el eje “x” se tiene la actividad solar que afecta directamente a la tierra representada por (*Bz*)

En el entorno grafico de Weka se aplicaron algoritmos de clasificación supervisada a el archivo .arff el cual fue dividido en dos partes para trabajar de una manera más completa con la información. [11] Siendo la primer parte de entrenamiento del algoritmo registros y el segundo como comprobación.

El algoritmo de clasificación supervisada implementado en la aplicación “KnowledgeFlow” fue el LMT de la cual se muestra la secuencia de procesamiento de los datos. (Fig. 5)

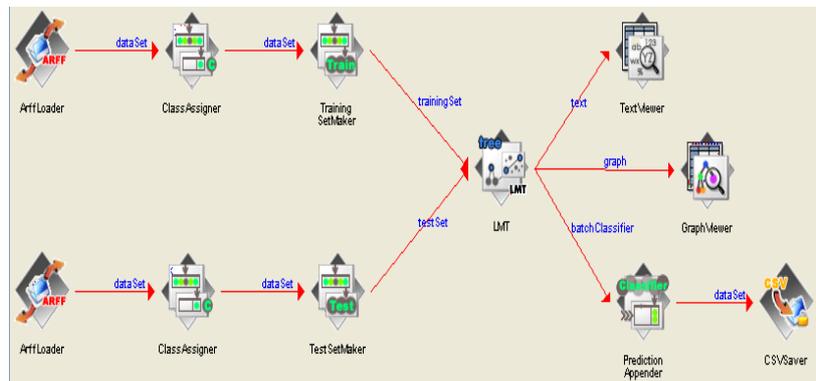


Fig. 5. Implementación del Algoritmo (*Logic Model Tree*) en el entorno “KnowledgeFlow” de Weka el cual permite la visualización de texto y la del árbol.

El cual permite ver en modo texto una estructura de árbol de decisión estándar con funciones de regresión logística en las hojas de la cual solo se muestra un fragmento

ya que el diagrama de árbol original tiene un tamaño de 257 y un número de hojas aproximado en 129.(Fig. 6)

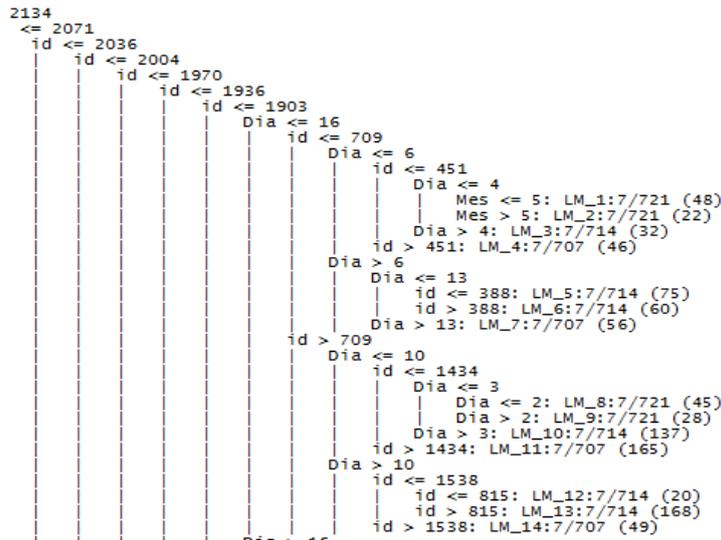


Fig. 6. (Fragmento de la visualización de texto del algoritmo LMT.)

El cluster LMT permite observar dicho árbol en un modo más gráfico. Aun que al igual que el anterior se debió recortar la imagen para poder dar una representación de este.(Fig. 7)

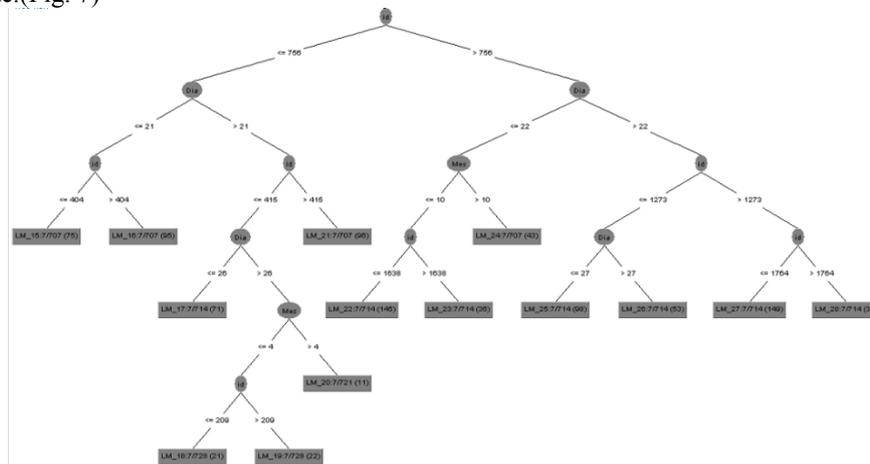


Fig. 7. (Árbol modo gráfico del algoritmo Logic Model Trees.)

Se realizó otro experimento mediante el algoritmo clasificación supervisada Naive Bayes con la misma secuencia de procesamiento de los datos.(Fig. 8.) Sin embargo este solo permite la visualización de un archivo de texto donde se muestra la

desviación estándar y la media de cada uno de los atributos respecto a las fases lunares. (Fig. 9.)

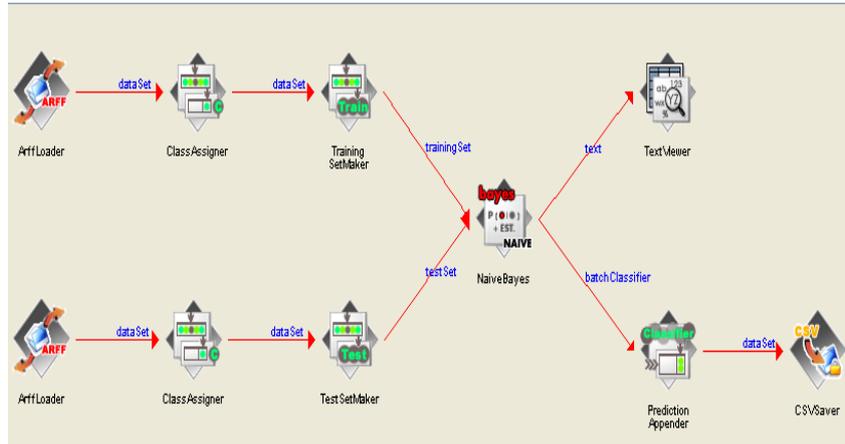


Fig. 8. Aplicación del Algoritmo de (Naive Bayes) el cual solo permite una visualización de texto.

Naive Bayes Classifier				
Attribute	Class			
	luna_nueva (0.26)	cuarto_menguante (0.24)	cuarto_creciente (0.25)	luna_llena (0.25)
Latitud				
mean	18.7378	18.2608	18.0914	18.4407
std. dev.	5.5343	5.0404	4.8856	5.2688
weight sum	2392	2201	2253	2296
precision	0.0199	0.0199	0.0199	0.0199
Longitud				
mean	-100.1828	-99.5761	-99.186	-99.6937
std. dev.	6.9193	6.4055	6.3629	6.6864
weight sum	2392	2201	2253	2296
precision	0.0162	0.0162	0.0162	0.0162
Profundidad				
mean	34.5472	36.9578	38.6386	37.0445
std. dev.	41.2731	42.3343	43.956	42.9688
weight sum	2392	2201	2253	2296
precision	1.2026	1.2026	1.2026	1.2026
Magnitud				
mean	3.8745	3.8495	3.8631	3.8647
std. dev.	0.3945	0.4091	0.4178	0.3993
weight sum	2392	2201	2253	2296
precision	0.1565	0.1565	0.1565	0.1565
Bt				
mean	4.4705	4.2345	4.4041	4.2748
std. dev.	2.1313	1.9706	2.3413	2.2656
weight sum	2392	2201	2253	2296
precision	0.2008	0.2008	0.2008	0.2008
Bz				
mean	1.8358	1.8131	1.8549	1.7798
std. dev.	1.6827	1.6694	1.7834	1.6752
weight sum	2392	2201	2253	2296
precision	0.1641	0.1641	0.1641	0.1641

Fig. 9. (Visualización del texto del algoritmo Naive Bayes) donde se muestra la media, desviación estándar y otros indicadores.

El algoritmo de clasificación no supervisado que se utilizó fue el cluster Simple K-means. (Fig. 10.) con la misma secuencia de procesamiento que los ejercicios

anteriores sin embargo este nos dio como resultado un archivo de texto con los centroides de cada atributos.

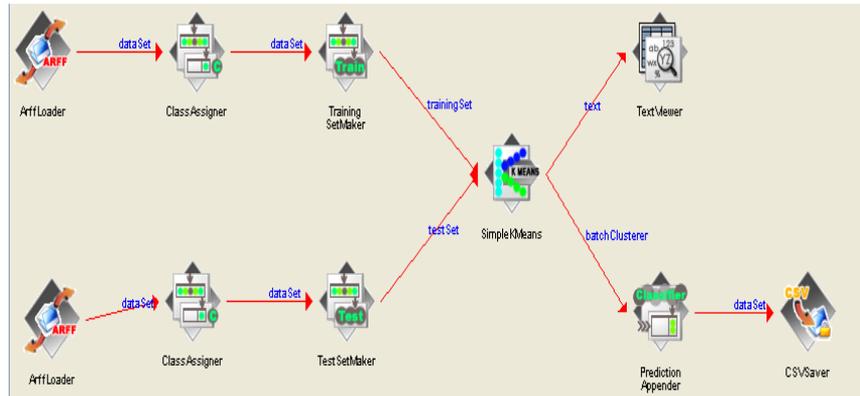


Fig. 10. Aplicación del algoritmo clustering (*Simple K-Means*) del cual se extraerán los centroides de cada atributo de los archivos arff.

El resultado del Cluster Simple K-means fue un archivo de texto donde se muestran los centroides de cada atributo (Fig. 11.) en el que se dividen en 5 grupos donde se analizaron los 9142 registros. De donde se puede extraer la información tal como el centroide de la latitud y longitud el cual es 18.3894 y -99.6682 respectivamente o el centroide de la magnitud de los sismos del país el cual es de 3.861.

Attribute	Full Data (9142)	0 (2253)	1 (1019)	2 (1277)	3 (2201)	4 (2392)
id	4571.5	4720.2006	2065.2365	6596.112	4335.4053	4635.4946
Dia	15.7627	16.9987	17.0049	14.7666	13.8019	16.4055
Mes	6.4602	6.5273	6.5594	6.2835	6.4134	6.4921
Año	2008.5413	2008.6023	2007.2345	2009.6053	2008.4334	2008.5719
Latitud	18.3894	18.0917	16.5198	19.9741	18.2609	18.7384
Longitud	-99.6682	-99.1861	-97.471	-101.467	-99.5761	-100.1827
Profundidad	36.8366	38.6973	41.2885	33.7917	37.0373	34.6283
Magnitud	3.861	3.8593	3.9665	3.7777	3.8491	3.8733
Fase_lunar	luna_nueva cuarto_creciente	luna_llena	luna_llena cuarto_menguante	luna_nueva		
Bt	4.3821	4.4384	4.2982	4.3151	4.2713	4.5023
Bz	1.8213	1.8581	1.7896	1.7699	1.8139	1.8343
Posicion	1.446	1.447	1.5005	1.4174	1.4993	1.388

Fig. 11. Resultados del cluster de clasificación no supervisado K-means donde se muestran los centroides de los atributos.

Conclusiones

Mediante la aplicación del cluster Neive-Bayes se encuentran algunos hechos importantes:

- Existe una incidencia de sismos en luna nueva
- El parámetro Bt presenta su valor promedio más alto en luna nueva
- Los sismos de mayor magnitud en la República Mexicana se encuentran dentro de la clase luna nueva
- La latitud y longitud de los sismos presentados en luna nueva se localiza en el estado de Guerrero cerca de El Monteal, Chapuluapan y El Veladero

Con base a los resultados, es posible teorizar que existe una incidencia entre el parámetro Bt y la luna nueva, este hecho se confirma con la actividad solar de los días (25 al 28 de Septiembre de 2011) donde se presentaron sismos de magnitud considerable en la república mexicana con influencia en la luna nueva y el parámetro Bt, con un valor alto histórico en últimas fechas.

Bibliografía

- [1] Barbat, H.B.: *Calculo Sísmico de las Estructuras*. Ed, Reverte, Barcelona (1982)
- [2] Yuxian Hu., Shih Chi Liu., Weimin Dong.: *Earthquake Engineering*. Ed, Routledge Taylor and Francis, USA (1996)
- [3] Francesco, M., Robert, J.G.: *Earthquake Science and Seismic Risk Reduction*. Ed Springer, USA (2003)
- [4] Nikraves, M., Aminzadeh, F., Zadeh, L.A.: *Soft Computing and Intelligent Data Analysis in Oil Exploration*, Volumen 228, Volumen 230, Ed, Elsevier (2003)
- [5] Sushmita, M., Tinku, Ac.: *Data mining: multimedia, soft computing, and bioinformatics*. Ed. John Wiley and Sons (2003)
- [6] Rui, X., Donald C.W.: *Clustering*. Ed John Wiley and Sons (2009)
- [7] Huajie, Z., Ling, X.C., Zhiduo, Z.: *The Learnability of Naive Bayes*. Ed. Springer (2000)
- [8] Barragán, A. J.: *Síntesis de sistemas de control borroso estables por diseño*. Universidad de Huelva (2009)
- [9] Sierra, A.B.: *Aprendizaje Automático: conceptos básicos y avanzados*. Ed, Prentice Hall. Madrid (2006)
- [10] Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning*. Ed, Springer. USA (2008)
- [11] González, J.C., Castellón, M., Castejón, M.J.: *Técnicas de Clasificación en el Entorno de Weka para la Determinación de Cultivos de Regadío (Cítricos) En Librilla, Murcia (Se España)*. XII Congreso de la Asociación Española de Teledetección, (2009)