

Descubrimiento de Reglas de Asociación de eventos eruptivos del volcán Galeras*¹

[Discovery of Association Rules from eruptive events of Galeras volcano]

**RICARDO TIMARÁN PEREIRA², LISBETH ROSERO LEGARDA³,
YEHIMY CABRERA CABRERA⁴**

RECIBO: 20.08.2014 – APROBACIÓN: 20.04.2015

Resumen

En este artículo se presentan los primeros resultados del proyecto de investigación que tuvo como objetivo detectar patrones de eventos eruptivos del volcán Galeras con técnicas de minería de datos, a partir de los datos almacenados en el Observatorio Vulcanológico y Sismológico de Pasto - OVSP (Colombia), aplicando la metodología CRISP-DM. Se construyó, limpió y transformó un repositorio de datos con la información de los eventos eruptivos del volcán Galeras registrados desde 1989 hasta 2013. A partir de este repositorio, se detectaron patrones asociados a estos eventos, utilizando la tarea de minería de datos asociación. El conocimiento generado se integrará al existente con el fin de ayudar al OVSP y a los organismos gubernamentales de prevención de desastres a tomar decisiones eficaces en lo relacionado

* Modelo para la citación de este artículo:

TIMARÁN PEREIRA, Ricardo; ROSERO LEGARDA, Lisbeth & CABRERA CABRERA, Yehimy (2015). Descubrimiento de Reglas de Asociación de eventos eruptivos del volcán Galeras. En: Ventana Informática No. 32 (ene-jun). Manizales (Colombia): Facultad de Ciencias e Ingeniería, Universidad de Manizales. p. 183-197. ISSN: 0123-9678

1 Artículo de investigación científica y tecnológica proveniente del proyecto *Construcción de un Repositorio Limpio de Datos para la Detección de Patrones de Eventos Eruptivos del Volcán Galeras con Técnicas de Minería de Datos*, ejecutándose en el periodo 06.2013-06.2014, financiado por el *Sistema de Investigaciones de la Universidad de Nariño* y con la asesoría del Observatorio Vulcanológico y Sismológico de Pasto, OVSP.

2 PhD en Ingeniería, MSc. en Ingeniería, Especialista en Multimedia, Ingeniero de Sistemas y Computación. Director grupo de investigación GRIAS, Profesor Asociado, Departamento de Sistemas, Facultad de Ingeniería, Universidad de Nariño (Pasto, Nariño, Colombia). Correo electrónico: ritimar@udenar.edu.co

3 Ingeniera de Sistemas, Universidad de Nariño (Pasto, Nariño, Colombia). Correo electrónico: lrosero17@udenar.edu.co

4 Ingeniera de Sistemas, Universidad de Nariño (Pasto, Nariño, Colombia). Correo electrónico: ycabrera2@udenar.edu.co

a la implementación de planes de prevención ante una posible erupción del volcán Galeras.

Palabras Clave: *Patrones de Eventos Eruptivos, Volcán Galeras, Minería de Datos*

Abstract

In this paper, the first results of a research project that aimed to detect patterns of eruptive events of Galeras volcano with data mining techniques from the data stored in the Volcanological and Seismological Observatory of Pasto - VSOP (Colombia), applying CRISP-DM methodology, are presented. A data repository with the information of the eruptive events of Galeras volcano recorded from 1989 to 2013 was built, cleaned and transformed. Using the data mining task association were detected patterns associated with these events. The knowledge generated will be integrated to the existing order to help VSOP and government agencies of disaster prevention to take effective decisions related to the implementation of prevention plans for a possible eruption of the Galeras volcano.

Keywords: *Patterns of eruptive events, Galeras volcano, Data Mining.*

Introducción

Las erupciones volcánicas son explosiones o emanaciones de lava, ceniza y gases tóxicos desde el interior de la Tierra a través de los volcanes. Son uno de los fenómenos más impredecibles de la naturaleza. Sin embargo, casi todas ellas tienen algo en común, no importa el tipo de volcán o su localización, en la mayoría de los casos, la erupción es precedida por temblores que se producen minutos, días o semanas antes de que el volcán entre en actividad. También pueden estar precedidas por actividad fumarólica y por áreas nuevas o más grandes de suelo caliente, según señalan Tilling & Lipman (1993, 277). La mayoría de los cambios precursores son demasiado sutiles para notarlos, y es esencial utilizar los medios más efectivos en la vigilancia del volcán, que incluyen una amplia gama de técnicas geofísicas, geodésicas y geoquímicas. Afirman Simkin & Siebert (1994, 2), que se utilizan sismógrafos para detectar y localizar temblores relacionados con la ascensión del magma. Otras técnicas incluyen la medición de los cambios en el flujo de calor en el volcán, así para Tilling & Lipman (1993, 280), los cambios en la

composición o abundancia relativa de gases fumarólicos también pueden preceder erupciones volcánicas y se pueden detectar a través de frecuentes análisis de gases.

Estas técnicas pueden ser útiles a la hora de detectar signos de advertencia de una erupción inminente, pero el éxito global de un sistema de vigilancia depende de la detección e interpretación de los signos precursores con suficiente antelación como para advertir y evacuar a la gente de las áreas amenazadas e iniciar otras medidas para mitigar los efectos de la erupción. Aunque estos sistemas de vigilancia pueden ser útiles para indicar un aumento en la probabilidad de actividad volcánica y su localización, explican Tilling & Lipman (1993, 280), no indican el tipo o escala de una erupción inminente, ni el Índice de explosividad volcánica⁵ que alcanzará.

Por su parte, Timarán (2009, 44), señala que con la minería de datos es posible construir modelos a partir de grandes volúmenes de datos que permitan predecir, asociar o relacionar, en el caso de los volcanes, posibles eventos eruptivos con la debida antelación.

En este artículo se presentan los primeros resultados del proyecto de investigación que tuvo como objetivo detectar patrones de eventos eruptivos del volcán Galeras con técnicas de minería de datos a partir de los datos almacenados en el Observatorio Vulcanológico y Sismológico de Pasto, OVSP, aplicando la metodología CRISP-DM. Se construyó, limpió y transformó un repositorio de datos con la información de los eventos eruptivos del volcán Galeras registrados desde 1989 hasta 2013. Se detectaron patrones asociados a estos eventos, utilizando la técnica de minería de datos asociación con el algoritmo Apriori, presente en la herramienta Weka (Hall et al., 2011). El conocimiento generado se integrará al existente con el fin de ayudar al OVSP y a los organismos gubernamentales de prevención de desastres a tomar decisiones eficaces en lo relacionado a la implementación de planes de prevención ante una posible erupción del volcán Galeras.

El resto del artículo se organiza de la siguiente manera: En la sección 1, se presentan los fundamentos teóricos sobre el volcán Galeras. En la sección 2, se describe la metodología CRISP-DM. En la sección 3, se presentan los resultados obtenidos y finalmente, las conclusiones y trabajos futuros.

5 Según Spalletti & D'Elia (s.f.), el Índice de Explosividad Volcánica o IEV es una escala de ocho grados con la que los vulcanólogos miden la magnitud de una erupción volcánica.

1. Fundamento Teórico

1.1 Volcán Galeras

«El volcán Galeras se localiza en el Departamento de Nariño (Colombia), aproximadamente a 9 km al occidente de la Ciudad de San Juan de Pasto, capital de este departamento, en las coordenadas 1°13'43,8» de latitud norte y 77°21'33,0» de longitud oeste y con una altura de 4276 msnm. De acuerdo con el Catálogo de Volcanes Activos del Mundo (CAVW) de la Asociación Internacional de Vulcanología y Química del Interior de la Tierra (IAVCEI), su código es el 1501-08. (...) El volcán Galeras que conocemos hoy en día es el centro eruptivo más reciente y actualmente activo del denominado Complejo Volcánico Galeras (CVG), el cual posee una forma cónica con su edificio destruido en la parte occidental» (SGC, 2014).

De acuerdo con OVSP (s.f.), el actual volcán tiene una edad estimada de 4500 años, que corresponde al centro del complejo volcánico Galeras, CVG, con una evolución de más de un millón de años, donde se han identificado siete estados eruptivos, reconocidos del más antiguo al más reciente como: Cariaco, Pamba, Cobanegra, La Guaca, Genoy, Urcunina y el actual Galeras, como se puede observar en la figura 1.

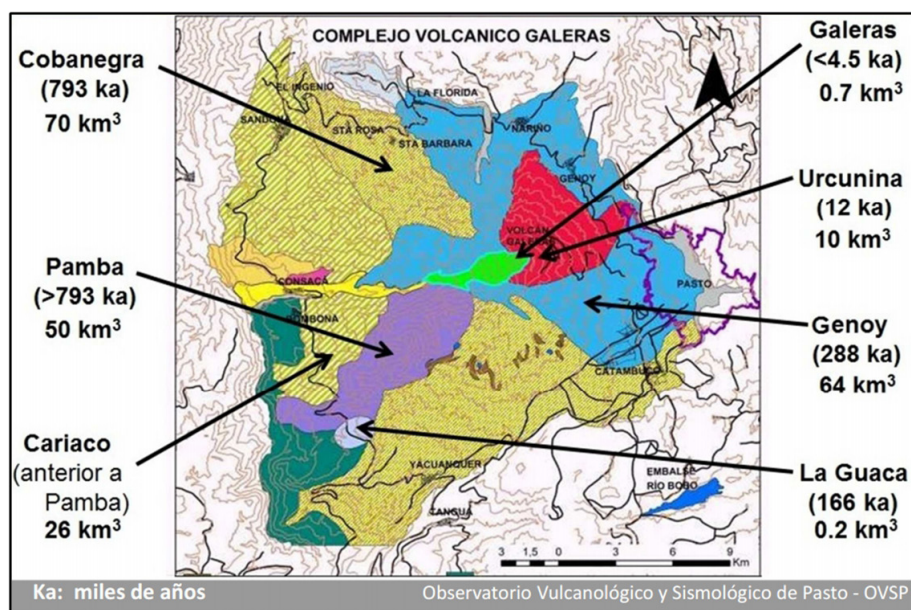


Figura 1. Evolución geológica del Complejo Volcánico Galeras (OVSP, s.f.).

1.2 Vigilancia del Volcán Galeras

El monitoreo o vigilancia volcánica de Galeras se realiza a través de la utilización de métodos geofísicos, geodésicos, geoquímicos y geológicos, apoyados por tecnologías de la Electrónica, las Comunicaciones y la Informática. Según Ingeominas (2013), para medir los cambios horizontales y verticales que ocurren por las deformaciones en la superficie del edificio volcánico debido a su actividad interna, conocida como área de deformación volcánica (Geodesia), se emplean métodos de inclinometría seca, inclinometría electrónica, líneas cortas de nivelación, medidas electrónicas de distancias y sistemas de posicionamiento global (GPS).

Para el caso de los sismos generados por la actividad volcánica, se utilizan redes con equipos portátiles y telemétricos localizadas sobre el volcán y en sus cercanías. Los instrumentos utilizados para captar y registrar las señales sísmicas son conocidos como sismógrafos (SGC, s.f.). Las señales sismológicas de origen volcánico se deben al movimiento de fluidos en el sistema volcánico (desde la cámara magmática, hasta el cráter). Normalmente se clasifican como:

- Sismos volcanotectónicos (VT), de acuerdo a Stephens et al. (1994,156), su origen es asociado a fracturas que ocurren como respuesta a cambios de esfuerzos en las áreas activas por movimiento de fluidos, su frecuencia generalmente es >5 Hz.
- Sismos de largo período (LPS), que para Chouet (1992,137) se atribuyen a la resonancia en grietas, cavidades y conductos, debido a cambios de presión en los fluidos que existen en los volcanes predominan las bajas frecuencias.
- Tremor volcánico (tremor), se caracteriza por el registro de formas de onda de manera persistente o sostenida en el tiempo. El tremor refleja una vibración continua del suelo o pequeños sismos muy frecuentes cuyas ondas se traslapan. Sanders & Ryall (1984), señalan que si la señal es monotónica (de frecuencia constante) se denomina tremor armónico; de otra parte si la señal varía significativamente en frecuencia o amplitud, se llama tremor espasmódico.
- Sismos tipo híbridos (HYB), son una combinación entre un LPS y un volcanotectónico.
- Sismos tornillo (TOR), para Kumagai & Chouet (2000, 25501), son un tipo particular de eventos sísmicos de periodo largo registrado en volcanes andesíticos.

«Para el estudio geoquímico en vigilancia de volcanes se recurre a muestreos sistemáticos de las diferentes

emanaciones gaseosas y líquidas características de la actividad, tales como aguas termales y gases volcánicos, incluidos radón y dióxido de carbono (CO₂) contenidos en suelos (...). Tales muestras, señala son sometidas a análisis para conocer las condiciones en su origen y en su camino a la superficie.

El COSPEC es un espectrómetro que mide la emisión de SO₂ de los volcanes». SGC (s.f.).

La aplicación de la geología en el conocimiento e investigación de los volcanes se hace a través de estudios de las rocas y depósitos originados en eventos eruptivos pasados y actuales, lo cual permite conocer la historia eruptiva del volcán, las fechas de ocurrencia de las erupciones, cómo fueron dichas erupciones y el posible comportamiento en eventos futuros. Así, dice Pyle (1999, 265), se hace posible la elaboración de los mapas de amenaza volcánica, herramienta indispensable en la planificación y en prevención de los desastres ocasionados por las erupciones volcánicas que afectan las poblaciones y el ambiente. Un complemento necesario de la geología son las investigaciones históricas sobre erupciones observadas y registradas por cronistas o insertas en leyendas.

Debido a que la masa y la tasa de erupción varían en diferentes órdenes de magnitud, una escala logarítmica es necesaria para categorizar los tamaños de las erupciones volcánicas. En esta área se mide la llamada Índice de Explosividad Volcánica (VEI), que usa una escala integradora desde un valor de 0 a 8 para describir el volumen y la altura de la pluma de una erupción dada. Para el caso del Volcán Galeras si el VEI está entre 0 y 1 se considera una emisión y si el VEI está entre 2 y 8 se considera una erupción, expresa SGC (2014).

2. Metodología

Con el fin de descubrir patrones asociados a eventos eruptivos del volcán Galeras se utilizó CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Esta es la metodología de referencia más utilizada en el desarrollo de proyectos de minería de datos en los ambientes académico e industrial, afirma Gallardo (2009, 5). Comprende seis fases: Análisis del problema, análisis de los datos, preparación de los datos, modelado, evaluación y explotación.

2.1 Análisis del Problema

En esta fase se requiere comprender con exactitud el problema al cual se le va a dar solución utilizando la minería de datos. Esto permitirá

recolectar la información necesaria para interpretar con asertividad los resultados encontrados, según Gallardo (2009, 17).

El OVSP es un instituto de investigación cuya principal actividad es la vigilancia del comportamiento del volcán Galeras. Para realizar este proceso se cuenta con equipos que registran los cambios de deformación, que se pueden presentar en el interior del volcán y del entorno. Estos registros son transmitidos al OVSP, donde son almacenados en una base de datos. A través del sistema experto *Sistema para la Integración, Análisis e Interpretación de la Información Generada por el Proceso de Vigilancia y Monitoreo del Volcán Galeras*, los datos son procesados con el fin de generar modelos de comportamiento. Con el análisis de la información obtenida con estos modelos, se establecen escenarios de evolución de la actividad al interior y en el entorno del Volcán Galeras. Descubrir factores asociados a los eventos eruptivos, se convirtió en un problema a tratar con la minería de datos.

2.2 Análisis de los Datos

En esta fase, siguiendo a Gallardo (2009, 18), se realiza la recolección inicial de datos, con el objetivo de establecer un primer contacto con el problema, familiarizándose con ellos, identificar su calidad y establecer las relaciones más evidentes que permitan definir las primeras hipótesis.

La única fuente de datos para la construcción inicial del repositorio de datos fue la base de datos del OVSP, de donde se escogieron 32 tablas que contienen la información registrada de la actividad volcánica del Galeras desde el año 1989 hasta el 2013. Para entender mejor las relaciones entre estas tablas y sus atributos, y con la asesoría del personal del OVSP, se construyó un diccionario de datos, con el fin de entender cada una de las variables a considerarse en este estudio. Con estas tablas se construyó una base de datos denominada *Repo_Galeras*, utilizando el *SGBD PostgreSQL*.

Por medio de consultas SQL ad-hoc y a través de histogramas, se analizó minuciosamente la calidad de los datos contenidos en cada uno de los atributos de las tablas de la base de datos *Repo_Galeras*. Como resultado de esta fase, se identificaron datos para dos tipos de eventos: las emisiones de gases o ceniza y las erupciones.

2.3 Preparación de los Datos

Una vez efectuada la recolección inicial de datos, de acuerdo con Gallardo (2009, 19), se procede a su preparación para adaptarlos a las técnicas de minería de datos que se aplican. La preparación de datos incluye las tareas generales de selección de datos a los que se va a

aplicar una determinada técnica de modelado, limpieza de datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato.

De la base de datos *Repo_Galeras* se seleccionaron inicialmente 59 atributos considerados por los investigadores como relevantes para el proyecto. Una vez realizado el análisis de la calidad de los datos, se eliminaron aquellos atributos, con un alto porcentaje de valores nulos, por la imposibilidad de encontrar sus valores, que correspondían generalmente a mediciones no registrados por los aparatos de vigilancia del Volcán Galeras. Por otra parte, para facilitar la extracción de patrones, se discretizaron los valores de los atributos numéricos que no fueron eliminados y se crearon nuevos atributos a partir de los existentes.

Como resultado de esta fase, se crearon dos conjuntos de datos limpios: el repositorio TEMI649A19, compuesto por 649 registros y 19 atributos que almacena todas las emisiones del volcán Galeras y el repositorio TERU17A18 que registra las 17 erupciones del volcán en el periodo 1989 -2013 caracterizadas con 18 variables. Los atributos que forman parte de los repositorios se describen en la tabla 1.

2.4 Modelado

En esta fase se seleccionan las técnicas de modelado más apropiadas para el proyecto de Minería de Datos, de acuerdo con Gallardo (2009, 21). Con el fin de detectar patrones asociados a los eventos eruptivos del volcán Galeras, se seleccionó la tarea de minería de datos asociación a partir de los datos de los repositorios TEMI649A19 y TERU17A18.

Para la tarea de asociación se utilizó el algoritmo *Apriori* (Agrawal & Srikant, 1994, 487), implementado en la herramienta de minería de datos *Weka*, presentada por García (s.f., 22) y Witten & Frank (2000, 294), en el paquete *WEKA associations.Apriori*. Para evaluar las regla de asociación resultantes se utilizaron los parámetros soporte y confianza, dos métricas que permiten conocer la calidad de la regla. El soporte o cobertura de una regla se define como el número de instancias en las que la regla se puede aplicar. La confianza o precisión mide el porcentaje de veces que la regla se cumple cuando se puede aplicar, como lo señalan Adamo (2001, 7), Hernández, Ramírez & Ferri (2005, 238) y Han & Kamber (2001, 228).

Tabla 1. Atributos de los repositorios TEMI649A19 y TERU17A18

Atributo	Tipo	TEMI649A19	TERU17A18
Fecha	Date	Fecha de registro del evento	Fecha de registro del evento
comportamiento_n_v	Character varying	Subió, estable o bajó la medida de so2 según el anterior	Subió, estable o bajó la medida de so2 según el anterior
Tipo	Character varying	Tipo de evento localizado.	Tipo de evento localizado.
Qm	Character varying	Calidad de la localización [A,B,C,D] A: epicenter excelent, focal depth good B: epicenter good, focal depth fair C: epicenter fair, focal depth poor D: epicenter poor, focal depth poor	Calidad de la localización [A,B,C,D] A: epicenter excelent, focal depth good B: epicenter good, focal depth fair C: epicenter fair, focal depth poor D: epicenter poor, focal depth poor
comportamiento_v	Character varying	Subió, estable o bajó según la Velocidad del sismo anterior	Subió, estable o bajó según la Velocidad del sismo anterior
Comportamiento_r	Character varying	Subió, estable o bajó según el comportamiento del radón anterior	Subió, estable o bajó según el comportamiento del radón anterior
Incandescencia	Character varying		Si se observa incandescencia [S,N]
Altura	Double precision	Altura máxima de la emisión (metros sobre la cima).	Altura máxima de la columna
caída_piroclastos	Character varying		Si se observa caída de piroclastos [S, N].
onda_choque	Character varying		Si se observa onda de choque [S, N].
proyectil_balístico	Character varying		Si se registran proyectiles balísticos.
Dirección	Character varying	Dirección a la cual se desplaza la columna.	
Color	Character varying	Se observó color de la emisión [S, N].	
material_gas	Character varying	Si se observa salida de gas [S, N]	
Cenizas	Character varying	Si se observó salida de ceniza[S,N]	
fenomenos_olor	Character varying	Si se asocia algún tipo de olor a la emisión [S, N].	
Clase_emision	Character	Si hay emisión [S, N].	
LPS	character varying	Se observó eventos LPS clasificables para la fecha dada. [S,N]	Se observó eventos LPS clasificables para la fecha dada. [S,N]
Vt	character varying	Se observó eventos VT clasificables para la fecha dada. [S,N]	Se observó eventos VT clasificables para la fecha dada. [S,N]
Tre	character varying	Se observó eventos tipo TREMOR para la fecha dada. [S,N]	Se observó eventos tipo TREMOR para la fecha dada. [S,N]
Tor	character varying	Se observó eventos tipo TOR para la fecha dada. [S,N]	Se observó eventos tipo TOR para la fecha dada. [S,N]
Hyb	character varying	Se observó eventos HYB para la fecha dada [S,N]	Se observó eventos HYB para la fecha dada [S,N]
Clase_sismos	character varying	Se relacionó algún sismo [S,N]	Se relacionó algún sismo [S,N]
Clase_erupcion	character varying		Si hay erupción [S,N]

Con el fin de generar reglas de asociación fuertes (*strong rules*), i.e. reglas que superen el soporte y la confianza mínima, se estableció el soporte mínimo en 0.1 y la confianza en 80%. Se generaron 100 reglas, de las cuales se escogieron las reglas con una confianza mínima del 80%. Las reglas más representativas de acuerdo al soporte, para emisiones se muestran en la figura 2 y para erupciones, en la figura 3.

2.5 Evaluación e Interpretación

En esta fase, de acuerdo con Gallardo (2009, 22), se interpretan los patrones descubiertos con el fin de consolidar el conocimiento descubierto e incorporarlo en otro sistema para posteriores acciones o para confrontarlo con conocimiento previamente descubierto.

Se utilizó el conjunto de datos TEMIT648A19 para obtener reglas de asociación que asocien 19 factores, antes de una emisión del volcán Galeras, con 648 emisiones registradas en el OVSP en el periodo 1989 al 2013. Entre las reglas de asociación más representativas están:

- Regla 1. Probablemente hay una emisión si se registra sismos tipo LPS de 5 a 10 segundos, hay registro de una energía sísmica tipo tremor y hay un sismo tipo híbrido. El 38% de las emisiones asocian estos factores.
- Regla 2. Probablemente hay una emisión si el comportamiento de la intensidad del sismo con relación al anterior subió, hay un registro de una energía sísmica tipo tremor de larga duración y no hay un sismo tipo tornillo en el mismo día. El 31% de las emisiones asocian estos factores.

```

Best rules found:
1. tre=S 452 ==> clase_emision=S 452 conf:(1)
2. tre=S tor=N 427 ==> clase_emision=S 427 conf:(1)
3. lps=S 416 ==> clase_emision=S 416 conf:(1)
4. lps=S tor=N 387 ==> clase_emision=S 387 conf:(1)
5. lps=S tre=S 371 ==> clase_emision=S 371 conf:(1)
6. qm=Ci 366 ==> clase_emision=S 366 conf:(1)
7. qm=Ci tor=N 349 ==> clase_emision=S 349 conf:(1)
8. lps=S tre=S tor=N 346 ==> clase_emision=S 346 conf:(1)
9. hyb=S 345 ==> clase_emision=S 345 conf:(1)
10. vt=S 329 ==> clase_emision=S 329 conf:(1)
11. tor=N hyb=S 317 ==> clase_emision=S 317 conf:(1)
12. vt=S tor=N 313 ==> clase_emision=S 313 conf:(1)
13. qm=Ci tre=S 298 ==> clase_emision=S 298 conf:(1)
14. lps=S hyb=S 285 ==> clase_emision=S 285 conf:(1)
15. tre=S hyb=S 285 ==> clase_emision=S 285 conf:(1)
16. qm=Ci tre=S tor=N 283 ==> clase_emision=S 283 conf:(1)
17. comportamientorad=subió 277 ==> clase_emision=S 277 conf:(1)
18. comportamientovel=bajó 274 ==> clase_emision=S 274 conf:(1)
19. qm=Ci lps=S 272 ==> clase_emision=S 272 conf:(1)
20. comportamientovel=subió 265 ==> clase_emision=S 265 conf:(1)
21. comportamientorad=subió tor=N 264 ==> clase_emision=S 264 conf:(1)
22. comportamientoton_v=sin_medida 261 ==> clase_emision=S 261 conf:(1)
23. vt=S tre=S 261 ==> clase_emision=S 261 conf:(1)
24. tre=S tor=N hyb=S 261 ==> clase_emision=S 261 conf:(1)
25. comportamientovel=bajó tor=N 257 ==> clase_emision=S 257 conf:(1)
26. lps=S tor=N hyb=S 257 ==> clase_emision=S 257 conf:(1)
27. qm=Ci lps=S tor=N 255 ==> clase_emision=S 255 conf:(1)
28. comportamientorad=bajó 254 ==> clase_emision=S 254 conf:(1)
29. comportamientovel=subió tor=N 253 ==> clase_emision=S 253 conf:(1)
30. comportamientoton_v=sin_medida tor=N 248 ==> clase_emision=S 248 conf:(1)
31. vt=S tre=S tor=N 247 ==> clase_emision=S 247 conf:(1)
32. lps=S tre=S hyb=S 246 ==> clase_emision=S 246 conf:(1)
33. lps=S vt=S 245 ==> clase_emision=S 245 conf:(1)
34. qm=Ci hyb=S 243 ==> clase_emision=S 243 conf:(1)
35. comportamientorad=bajó tor=N 240 ==> clase_emision=S 240 conf:(1)
36. qm=Ci lps=S tre=S 240 ==> clase_emision=S 240 conf:(1)
37. comportamientovel=bajó tre=S 239 ==> clase_emision=S 239 conf:(1)
38. comportamientorad=subió tre=S 229 ==> clase_emision=S 229 conf:(1)
39. lps=S vt=S tor=N 229 ==> clase_emision=S 229 conf:(1)

```

Figura 2. Mejores Reglas de Asociación para emisiones.

```

Best rules found:

1. vt=S 14 ==> clase_erupcion=S 14    conf:(1)
2. lps=S vt=S 14 ==> clase_erupcion=S 14    conf:(1)
3. vt=S tre=S 14 ==> clase_erupcion=S 14    conf:(1)
4. lps=S vt=S tre=S 14 ==> clase_erupcion=S 14    conf:(1)
5. hyb=S 13 ==> clase_erupcion=S 13    conf:(1)
6. lps=S hyb=S 13 ==> clase_erupcion=S 13    conf:(1)
7. hyb=S tre=S 13 ==> clase_erupcion=S 13    conf:(1)
8. lps=S hyb=S tre=S 13 ==> clase_erupcion=S 13    conf:(1)
9. vt=S hyb=S 12 ==> clase_erupcion=S 12    conf:(1)
10. lps=S vt=S hyb=S 12 ==> clase_erupcion=S 12    conf:(1)
11. vt=S hyb=S tre=S 12 ==> clase_erupcion=S 12    conf:(1)
12. lps=S vt=S hyb=S tre=S 12 ==> clase_erupcion=S 12    conf:(1)
13. comportamientorad=subió 11 ==> clase_erupcion=S 11    conf:(1)
14. tor=N 11 ==> clase_erupcion=S 11    conf:(1)
15. comportamientorad=subió lps=S 11 ==> clase_erupcion=S 11    conf:(1)
16. comportamientorad=subió tre=S 11 ==> clase_erupcion=S 11    conf:(1)
17. lps=S tor=N 11 ==> clase_erupcion=S 11    conf:(1)
18. tre=S tor=N 11 ==> clase_erupcion=S 11    conf:(1)
19. comportamientorad=subió lps=S tre=S 11 ==> clase_erupcion=S 11    conf:(1)
20. lps=S tre=S tor=N 11 ==> clase_erupcion=S 11    conf:(1)
21. comportamientorad=subió vt=S 10 ==> clase_erupcion=S 10    conf:(1)
22. comportamientorad=subió hyb=S 10 ==> clase_erupcion=S 10    conf:(1)
23. vt=S tor=N 10 ==> clase_erupcion=S 10    conf:(1)
24. hyb=S tor=N 10 ==> clase_erupcion=S 10    conf:(1)
25. comportamientorad=subió lps=S vt=S 10 ==> clase_erupcion=S 10    conf:(1)
26. comportamientorad=subió lps=S hyb=S 10 ==> clase_erupcion=S 10    conf:(1)
27. comportamientorad=subió vt=S tre=S 10 ==> clase_erupcion=S 10    conf:(1)
28. comportamientorad=subió hyb=S tre=S 10 ==> clase_erupcion=S 10    conf:(1)
29. lps=S vt=S tor=N 10 ==> clase_erupcion=S 10    conf:(1)
30. lps=S hyb=S tor=N 10 ==> clase_erupcion=S 10    conf:(1)
31. vt=S tre=S tor=N 10 ==> clase_erupcion=S 10    conf:(1)
32. hyb=S tre=S tor=N 10 ==> clase_erupcion=S 10    conf:(1)

```

Figura 3. Mejores Reglas de Asociación para erupciones.

- Regla 3. Probablemente hay una emisión si hay un sismo tipo LPS de 5 a 10 segundos, hay un registro de una energía sísmica tipo tremor de larga duración, no hay un sismo tipo tornillo y hay un sismo tipo híbrido. El 34% de las emisiones cumplen con este patrón.

- Regla 4. Probablemente hay una emisión si hay sismo LPS de 5 a 10 segundos de duración, hay una energía sísmica tipo volcanotectónica y hay un registro de una energía sísmica tipo tremor de larga duración. El 33% de las emisiones cumplen con este patrón.

- Regla 5. Probablemente hay una emisión si el epicentro y la profundidad focal del sismo es pobre, hay un registro de una energía sísmica tipo tremor de larga duración y hay un sismo tipo híbrido. El 31% de las emisiones asocian estos factores.

- Regla 6. Probablemente hay una emisión si el comportamiento del radón subió, hay un sismo tipo LPS de duración de 5 a 10 segundos y hay un registro de una energía sísmica tipo tremor de larga duración. El 29% de las emisiones cumplen con este patrón.

Se utilizó el conjunto de datos TERU17A18 para obtener reglas de asociación que asocien 18 factores, antes de una erupción del volcán Galeras, con 17 eventos registrados en el OVSP en el periodo 1989 al 2013. Entre las reglas de asociación más representativas están:

- Regla 1. Probablemente hay una erupción del volcán Galeras si el comportamiento de la intensidad del sismo subió con relación

al anterior, se registra sismos tipo LPS de 5 a 10 segundos de duración, existen sismos provocados por movimientos de fluidos y volcanotectónicos a la vez llamados híbridos, existe registro de una energía sísmica tipo tremor de larga duración y no se registran eventos sísmicos tipo tornillo en el mismo día. El 44% de las erupciones asocian estos parámetros.

- Regla 2. Probablemente hay una erupción del volcán Galeras si el comportamiento del radón subió, existe un registro de sismo tipo LPS, el cual puede tener una duración de 5 a 10 segundos, existen sismos provocados por movimientos de fluidos y volcanotectónicos a la vez llamados híbridos y no se registran eventos sísmicos tipo tornillo en el mismo día. El 44% de las erupciones cumplen este patrón.

- Regla 3. En el 100% de las erupciones del volcán Galeras el comportamiento del radón subió, existe registro de una energía sísmica tipo tremor de larga duración, existen sismos producidos por movimiento de fluidos y volcanotectónicos a la vez llamados híbridos y no se registran eventos sísmicos tipo tornillo en el mismo día. El 44% de las erupciones cumplen este patrón.

2.6 Explotación o Implementación

En esta fase se trata de explotar la potencialidad de los modelos, integrarlos en los procesos de toma de decisión de la organización y difundir informes sobre el conocimiento extraído, siguiendo a Gallardo (2009, 23).

El conocimiento generado se integrará al existente con el fin de ayudar al OVSP y a los organismos gubernamentales de prevención de desastres a tomar decisiones eficaces en lo relacionado a la implementación de planes de prevención ante una posible erupción del volcán Galeras.

3. Resultados

De acuerdo con los resultados obtenidos en las diferentes pruebas realizadas en la etapa de minería de datos con la tarea de asociación, donde se analizaron 648 casos de emisiones del volcán Galeras, el patrón general es que hay probabilidad de que exista una emisión si no se presentan sismos tipo tornillo en el mismo día, se presentan sismos tipo tremor de larga duración y adicional a esto se presentan sismos tipo LPS de 5 a 10 segundos.

En el caso de erupciones, donde se analizaron 17 casos, el patrón general es que en una erupción se presentan sismos tipo LPS y sismos tipo tremor. En el caso de sismos tipo híbridos y volcanotectónicos se

evidenció que están presentes en un 88,23%, así como también el comportamiento del radón subió en un 70% de los casos.

4. Conclusiones y trabajos futuros

Los patrones asociados a eventos eruptivos del volcán Galeras, descubiertos a través de la tarea de minería de datos asociación, indican que esta es capaz de generar un modelo consistente con la realidad observada y el respaldo teórico, basándose únicamente en los datos que se encontraron almacenados en las bases de datos del Observatorio Vulcanológico y Sismológico de Pasto en el periodo comprendido entre 1989 y 2013.

Una de las grandes dificultades que se presentó en esta investigación fue la mala calidad de los datos de la bases de datos del OVSP la cual llevó a que, después del proceso de limpieza de datos, se descartaran ciertas variables por la imposibilidad de obtener sus valores y que de alguna manera influyen en los resultados sobre patrones asociados a eventos eruptivos del volcán Galeras.

Como trabajos futuros está la aplicación de otras técnicas de minería de datos como *clustering*, con el fin de encontrar grupos similares entre los diferentes registros de emisiones o erupciones del volcán Galeras y analizar los resultados. Por otra parte es necesario que el OVSP y los organismos de prevención de desastres del municipio de Pasto, analicen los patrones obtenidos y los integren al conocimiento del volcán Galeras para la implementación de planes de evacuación temprana, ante una posible erupción.

5. Agradecimientos

Al Observatorio Vulcanológico y Sismológico de Pasto por facilitar las bases de datos de los eventos eruptivos del volcán Galeras y la asesoría por parte de su director ingeniero Diego Gómez y el profesional de Sistemas, ingeniero John Meneses.

Referencias Bibliográficas

- ADAMO, J.M. (2001). *Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms*. New York (USA): Springer-Verlag. 253 p. ISBN: 0-387-95048-6.
- AGRAWAL, R. & SRIKANT, R. (1994). Fast Algorithms for Mining Association Rules. In: 20th International Conference on Very Large Data Bases, VLDB 1994 (12-15/09/1994). Santiago de Chile (Chile): VLDB. BOCCA, J.B.; JARKE, M. & ZANIOLO, C. (ed.) (1994). *Proceedings of the VLDB'94*. Santiago (Chile): Morgan Kaufmann. p. 487-499. ISBN: 1-55860-153-8.
- CHOUET, B. (1992). A Seismic Model for the Source of Long-Period Events and Harmonic Tremor. In: GASPARINI, P., SCARPA, R. & AKI, K. (Eds). *Volcanic Seismology, IAVCEI Proceedings in Volcanology, Vol 3*. Berlin (Germany): Springer Berlin Heidelberg. p. 133-156. ISBN: 978-3-642-77008-1.
- GALLARDO ARANCIBIA, J.A. (2009). *Metodología para la Definición de Requisitos en Proyectos de Data Mining (ER-DM)* [en línea]. Tesis Doctoral (Doctor en Informática). Madrid (España): Universidad Politécnica de Madrid, Facultad de Informática. 266 p. <http://oa.upm.es/1946/1/JOSE_ALBERTO_GALLARDO_ARANCIBIA.pdf> [consulta: 18/02/2014].
- GARCÍA, D. (s.f.). *Manual de Weka* [en línea]. Valladolid (España): MetaEmotion S.L. 42 p. <<http://www.metaemotion.com/diego.garcia.morate/download/weka.pdf>> [consulta: 15/06/2014].
- HALL, M.; FRANK, E. & WITTEN, I. (2011). *Practical Data Mining: Tutorials* [en línea]. Hamilton (Nueva Zelanda): University of Waikato. <<http://www.micai.org/2012/tutorials/Weka%20tutorials%20Spanish.pdf>> [consulta: 20/02/2014].
- HAN, J. & KAMBER, M. (2001). *Data Mining: Concepts and Techniques*. San Diego (USA): Morgan Kaufmann Publishers. 745 p. ISBN: 978-1-55860-901-3.
- HERNÁNDEZ, O.J.; RAMÍREZ, Q.M. & FERRI, R.C. (2005). *Introducción a la Minería de Datos*. Madrid (España): Pearson Prentice Hall. 656 p. ISBN: 84-205-4091-9.
- INGEOMINAS (2013). *Actividad Histórica Volcán Galeras* [en línea]. Pasto (Colombia): Instituto Colombiano de Geología y Minería <www.sgc.gov.co/getattachment/Pasto/Volcanes/Volcan-Galeras/Actividad-historica/Actividad_historica_galeras.pdf.aspx> [consulta: 10/02/2014].
- KUMAGAI, H. & CHOUET, B. (2000). Acoustic properties of a crack containing magmatic or hydrothermal fluids [online]. In: *Journal of Geophysical Research: Solid Earth*, Vol. 105, No. B11 (Nov). Washington D.C. (USA): AGU Publications. p. 25493-25512. e-ISSN: 2169-9356 <DOI: 10.1029/2000JB900273> <<http://onlinelibrary.wiley.com/doi/10.1029/2000JB900273/pdf>> [consult: 11/10/2014]
- OBSERVATORIO VULCANOLÓGICO Y SISMOLÓGICO DE PASTO, OVSP (s.f.). *El Complejo Volcánico Galeras (CVG): Generalidades sobre el CVG a partir de las investigaciones geológicas de maestría y doctorado* [en línea]. Pasto (Colombia): Servicio Geológico Colombiano. <http://www.sgc.gov.co/Pasto/Volcanes/Volcan-Galeras/Generalidades/Complejo_volcanico_Galeras-Modo-de-compatibilidad.aspx> [consulta: 11/10/2014].
- PYLE, D.M. (1999). Sizes of Volcanic Eruption. In: SIGURDSSON, H.; HOUGHTON, B.; RYMER, H.; STIX, J. & McNUTT, S. (eds.) (1999). *Encyclopedia of Volcanoes, Part II*. San Diego (CA, USA): Academy Press. p. 263-269. ISBN: 978-012-64-3140-7
- SANDERS, C.O. & RYALL, F. (1984). Geometry of magma bodies beneath Long Valley, California determined from anomalous earthquake signals [online]. In: *Journal of Geophysical Research*, Vol. 89, No. B10 (Sep), Washington D.C. (USA): Wiley Online Library. p. 8287-8302. <DOI: 10.1029/JB089B10p08287> [consult: 18/10/2014].
- SERVICIO GEOLÓGICO COLOMBIANO, SGC (2014). *Volcán Galeras-Generalidades* [en línea]. Pasto (Colombia): Servicio Geológico Colombiano <<http://www.sgc.gov.co/Pasto/Volcanes/Volcan-Galeras/Generalidades.aspx>> [consulta: 10/02/2014].
- SERVICIO GEOLÓGICO COLOMBIANO, SGC (s.f.). *Observatorios Vulcanológicos: Metodología para la vigilancia volcánica* [en línea]. Bogotá (Colombia). <<http://www.sgc.gov.co/Observatorios-Vulcanologicos/Metodologia-para-la-vigilancia-volcanica.aspx>> [consulta: 11/10/2014].
- SIMKIN, T. & SIEBERT, L. (1994). *Volcanoes of the World: A Regional Directory, Gazetteer, and Chronology of Volcanism during the last 10,000 years*. 2 ed. Tucson (Arizona, USA): Geoscience Press. 349 p. ISBN: 978-094-50-0512-4.
- SPALLETI, L.A. & D'ELIA, L. (s.f.). *Rocas piroclásticas y sedimentación en ambientes volcánicos* [en línea]. La Plata (Buenos Aires, Argentina); Universidad Nacional de La Plata, Facultad

- de Ciencias Naturales y Museo, Cátedra de Sedimentología. <<http://www.fcnym.unlp.edu.ar/catedras/sedimentologia/pdf/piroclasticas.pdf>> [consulta: 04/04/2015]
- STEPHENS, C.D., CHOUET, B.A., PAGE, R.A., LAHR, J.C. & POWER, J.A. (1994). Seismological aspects of the 1989-1990 eruptions at Redoubt Volcano, Alaska: the SSAM perspective. In: *Journal of Volcanology and Geothermal Research*, Vol. 62, No. 1-4, (Aug). Philadelphia (PA, USA): Elsevier. p. 153-182. ISSN: 0377-0273 <DOI: 10.1016/0377-0273(94)90032-9> [consult: 18/10/2014]
- TILLING, R.I. & LIPMAN, P.W. (1993). Lessons in reducing volcanic risk [online]. In: *Nature International weekly journal of science*, Vol. 364, No. 6435, (Jul). London (UK): Nature Publishing Group, NPG. p. 277-280. e-ISSN: 1476-4687 <DOI: 10.1038/364277a0> <<http://www.nature.com/nature/journal/v364/n6435/pdf/364277a0.pdf>> [consult: 11/10/2014]
- TIMARÁN, R. (2009). Una mirada al descubrimiento de conocimiento en bases de datos. En: *Ventana Informática*. No 20 (ene-jun). Manizales (Colombia): Universidad de Manizales, Facultad de Ingeniería. p 39-58. ISSN: 0123-9678.
- WITTEN, I. & FRANK, E. (2000). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco, CA (USA): Morgan Kaufmann Publishers. 365 p. ISBN: 1-55860-552-5.