

Identificación de lesiones no fatales en la cartografía del municipio de pasto con la técnica de grupamiento

Identification of non-fatal injuries in the cartography of the municipality of pasto with the clustering technique

Ricardo Timarán-Pereira¹
Gonzalo Hernández-Garzón²
Nelson Quemá-Taimbud³

Recibido: febrero 6 de 2017
Aceptado: junio 29 de 2017

Resumen

En este artículo se presenta uno de los resultados del proyecto de investigación que tuvo como objetivo detectar patrones de eventos violentos y georreferenciados, a partir de la información almacenada en el Observatorio del Delito del Municipio de Pasto (Colombia) con técnicas de minería de datos. Para georreferenciar cada evento y procesarlo geográficamente, se construyó un geocodificador de direcciones urbanas del municipio de Pasto, bajo software libre. Este geocodificador se integró a un visor cartográfico, que al aplicar la técnica de agrupamiento con el algoritmo k-means, permitió visualizar e identificar patrones en zonas del municipio donde suceden los diferentes eventos violentos. Esta información facilitará a los organismos gubernamentales y de seguridad la toma de decisiones eficaces relacionadas con la seguridad ciudadana y la prevención de este tipo de eventos.

Palabras clave: descubrimiento de patrones, eventos violentos, georreferenciación, agrupamiento, minería de datos.

Abstract

This paper presents one of the results of the research project that aims to detect patterns of violent and georeferenced events based on information stored in the Crime Observatory of the Municipality of Pasto (Colombia) with data mining techniques. In order to georeference each event and process it geographically, a geocoder of urban addresses of the municipality of Pasto was built, using a free software. This geocoder was integrated to a cartographic viewer, and when applying the clustering technique with the algorithm k-means, allowed to visualize and to identify patterns in areas of the municipality where the different violent events happen. This information will make it easier for government and security agencies to make effective decisions related to citizen security and the prevention of such events.

Keywords: pattern discovery, violent events, georeferencing, clustering technique, data mining.

¹ Ingeniero de Sistemas y Computación, Doctor en Ingeniería énfasis Ciencias de la Computación, Universidad de Nariño, Pasto, Colombia. E-mail: ritimar@udenar.edu.co

² Ingeniero de Sistemas, Magíster en Ingeniería, Universidad de Nariño, Pasto, Colombia. E-mail: gonzalohernandez@udenar.edu.co

³ Ingeniero de Sistemas, Universidad de Nariño, Pasto, Colombia. E-mail: nelson8533@gmail.com

1. Introducción

La Organización Mundial de la Salud, OMS, define la violencia como «el uso deliberado de la fuerza física o el poder, ya sea en grado de amenaza o efectivo, contra uno mismo, otra persona o un grupo o comunidad, que cause o tenga muchas probabilidades de causar lesiones, muerte, daños psicológicos, trastornos del desarrollo o privaciones» (OMS, 2002). La Organización Panamericana de la Salud, OPS, desde el año 1993 y la OMS en 1996, aceptaron que la violencia es un problema de salud pública, situación que se corrobora en el Informe de Violencia y Salud. En el mismo informe, América Latina presentó una tasa de homicidios que se encontraba alrededor de 18 por cada 100.000 personas, considerada como una de las regiones más violentas del mundo (OPS, 2008a).

En América Latina se reconoce que la ausencia de información confiable y oportuna es una limitante para avanzar en la identificación de la magnitud y características de las diferentes formas en que se expresa la violencia, así como el monitoreo y evaluación de los programas y proyectos para su prevención y control (OPS, 2008b). La OMS busca incrementar la capacidad de los sistemas de información existentes para avanzar en la vigilancia en temas de interés de salud pública de los países (WHO, 2012). Para ello es importante difundir información y gestionar conocimiento a través de observatorios que resulten de una necesidad compartida (agendas), concentrados para ser complementados unos con otros, y mediante una verdadera colaboración y cooperación en materia de información, permitiendo finalmente, una visión de la salud pública de un país, de sus grupos de interés y/o vulnerables y de cada individuo desde diversas dimensiones (MinSalud, 2013).

Una de las estrategias implementadas en vigilancia en salud pública, corresponde a los Observatorios de Muertes de Causa Externa (OPS, 2008a), los cuales se han instaurado para el seguimiento y análisis en el nivel local –municipal-, en casos de mortalidad por causa externa como: homicidios, suicidios, eventos de tránsito y muertes no intencionales. De esta forma a 2007 se reporta que en diferentes países de América Latina, ya se han creado observatorios y sistemas de información que tienen como propósito fundamental

el de recolectar información relacionada con actividades delictivas (OEA, 2013).

En Colombia se cuenta con diversas experiencias a nivel municipal y departamental. En el año 2012 el Ministerio de Salud y Protección Social propuso la organización y fortalecimiento de los Observatorios en salud en el país, para garantizar su sostenibilidad en el tiempo y mejorar la disponibilidad y uso de información necesaria para orientar desde una adecuada gestión de conocimiento las decisiones del Sistema de Salud (Rivillas et al., 2013).

En el municipio de Pasto (Colombia), el observatorio de muertes por causa externa, denominado Observatorio del Delito, nace en el segundo semestre del año 2002, como resultado de un proyecto conjunto con el Programa Colombia de la Universidad de Georgetown (Betancourth-Salazar, 2005). Actualmente el Observatorio del delito del municipio de Pasto cuenta con un sistema de vigilancia de eventos violentos, a partir de la implementación de un sistema de información georreferenciado denominado SIGEODEP, soportado en un mercado de datos. En este sistema se encuentra almacenada información histórica, limpia y transformada sobre lesiones de causa externa, fatales y no fatales, registradas en el Observatorio en un periodo de 11 años; esto permite disponer de información confiable, oportuna, de buena calidad y representativa de las lesiones de causa externa que ocurren en el municipio de Pasto (Timarán, Barón, Hernández, Hidalgo, & Betancourth, 2012), (Timarán, Calderón, Hidalgo, Barón, & Hernández, 2013).

El sistema SIGEODEP georreferencia las lesiones de causa externa fatales y no fatales que ocurrieron en el municipio de Pasto desde el año 2002 hasta la actualidad, a nivel de barrios y comunas como el gránulo más pequeño de georreferenciación y no a nivel de direcciones urbanas. Entre las lesiones fatales asociadas a las líneas de vigilancia del Observatorio del delito están los homicidios, suicidios, muertes por accidentes de tránsito y muertes no intencionales; mientras que entre las lesiones no fatales están: la Violencia Interpersonal en comunidad, la violencia autoinflingida, las lesiones por accidentes de tránsito, las lesiones accidentales y la violencia intrafamiliar. Los datos anteriores se basan en información procesada median-

te un análisis estadístico básico, donde se consideran fundamentalmente variables y relaciones primarias, sin tener en cuenta las verdaderas interrelaciones, que por lo general están ocultas y que únicamente se pueden descubrir utilizando un tratamiento de los datos más complejo, que solo es posible con la minería de datos (Timarán-Pereira, & Yepes-Chamorro, 2016).

En este contexto, la minería de datos emerge como el siguiente paso evolutivo en el proceso de análisis de datos (Perversi et al., 2007) (Valenga et al, 2008), como la tecnología que ayuda a comprender los datos de una base de datos. En general, los datos son la materia prima bruta. En el momento que el usuario les atribuye algún significado especial pasan a convertirse en información. Cuando los especialistas elaboran o encuentran un modelo, haciendo que la interpretación conjunta entre la información y ese modelo represente un valor agregado, entonces se convierte en conocimiento (Molina, 2002).

En este artículo se presentan los resultados de detectar patrones de eventos violentos no fatales y georreferenciados, utilizando la técnica de minería de datos clustering y su posterior visualización. Con este fin, se

construyó un geocodificador bajo código abierto, de direcciones urbanas del municipio de Pasto, que permitió georreferenciar cada evento violento. Este geocodificador se integró a un visor cartográfico, que al aplicar clustering con el algoritmo k-means, permitió visualizar e identificar patrones en zonas del municipio donde suceden los diferentes eventos violentos. Esta información facilitará a los organismos gubernamentales y de seguridad la toma de decisiones eficaces relacionadas con la seguridad ciudadana y la prevención de este tipo de eventos.

2. Materiales y métodos

La investigación fue de tipo descriptivo bajo el enfoque cuantitativo, aplicando un diseño no experimental. Para la identificación georreferenciada de patrones de lesiones no fatales se siguieron las cinco fases que se muestran en la figura 1, a saber: construcción del repositorio de información urbana, construcción del geocodificador de direcciones urbanas, construcción del visor cartográfico, geocodificación de eventos delictivos y descubrimiento de patrones de lesiones no fatales.

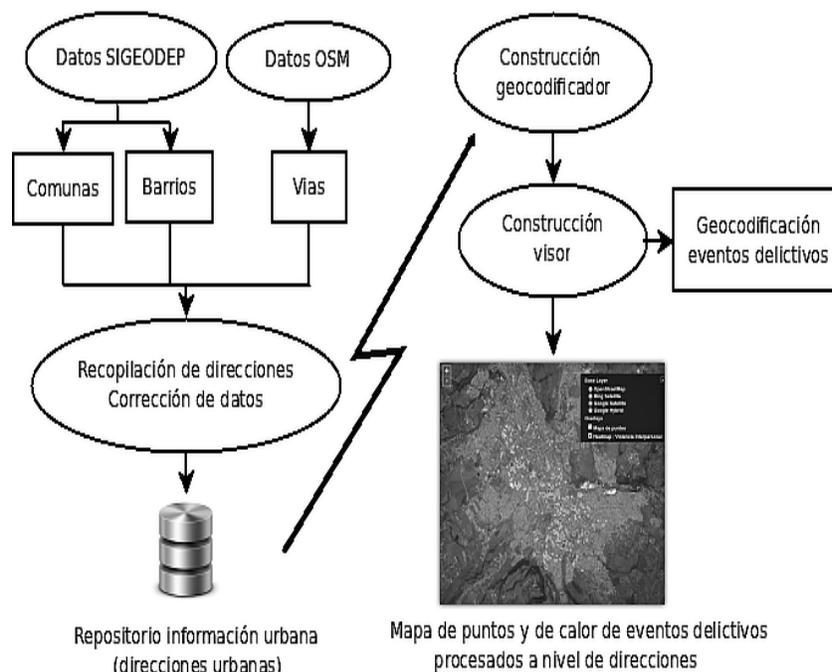


Figura 1. Metodología aplicada en la investigación.

2.1 Construcción del repositorio de información urbana

Inicialmente se obtuvo la capa de datos espaciales vectoriales que corresponden a la malla vial del área urbana del municipio de Pasto, disponibles en el servicio de edición y compartición libre de mapas cartográficos en la web Open Street Maps. Para la descarga de estos datos fue utilizada la herramienta de edición de datos cartográficos Java Open Street Maps Editor. Se obtuvo un archivo con extensión OSM que almacena información georreferenciada del área urbana del municipio de Pasto, bajo una estructura de datos XML.

2.2 Construcción del geocodificador de direcciones urbanas

El proceso de geocodificación de direcciones urbanas dependió de si la dirección tiene la nomenclatura de la malla vial o la de barrio, manzana, predio, cuyos procesos se muestran en las figuras 2 y 3, respectivamente.

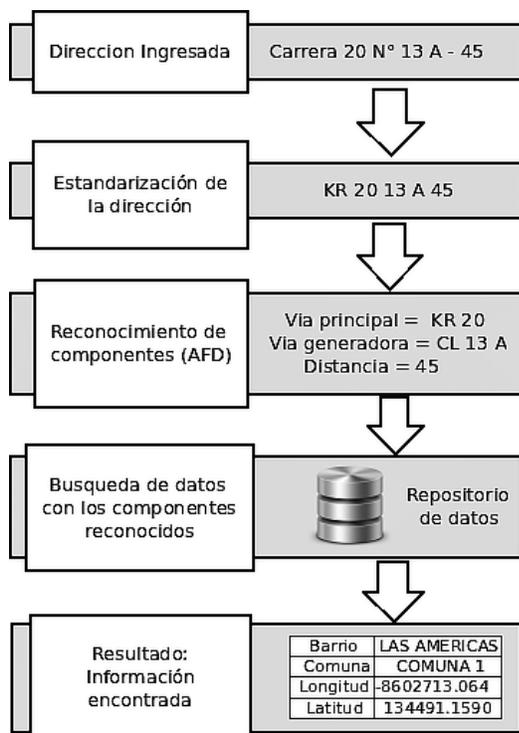


Figura 2. Proceso de geocodificación de direcciones urbanas con nomenclatura de malla vial.

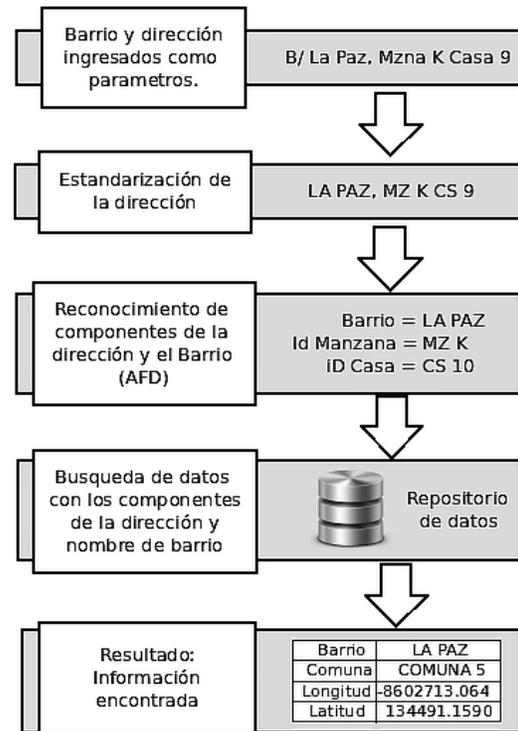


Figura 3. Proceso de geocodificación de direcciones urbanas con nomenclatura de barrio-manzana-predio.

2.3 Construcción del visor cartográfico SIGEODEP SIG

En esta fase el geocodificador fue acoplado con SIGEODEP, el sistema georreferenciado del Observatorio del Delito del municipio de Pasto, mediante la construcción del visor cartográfico SIGEODEP-SIG, permitiendo visualizar los eventos violentos fatales y no fatales registrados por este organismo gubernamental, a nivel de direcciones urbanas, mediante mapas de puntos y de calor, proceso que anteriormente no era posible realizar. La construcción del visor cartográfico fue realizada en su totalidad bajo el sistema operativo Linux Ubuntu en su versión 14.04. El lenguaje de programación utilizado fue Java, con el entorno de desarrollo integrado, IDE, Netbeans versión 8.1. El servidor de aplicaciones utilizado fue Glassfish en su versión 3.2.

Los procesos de construcción del repositorio de información urbana, del geocodificador y del visor cartográfico SIGEODEP SIG, se describen con mayor detalle en (Timarán, Hernández, & Quemá, 2016). Los procesos de geocodificación de eventos violentos y de descubrimiento de patrones de lesiones no fatales y su visualización se explican en las siguientes secciones.

3. Resultados y discusión

3.1 Geocodificación de eventos violentos

Al contar con un geocodificador de direcciones urbanas del municipio de Pasto, como resultado de las fases anteriores, en esta fase se realizaron las actividades que permitieron traducir las direcciones de los eventos violentos que se encuentran

almacenados en el mercado de datos del sistema SIGEODEP del Observatorio del Delito, a coordenadas geográficas, asignándoles a cada uno la longitud y latitud donde sucedió el evento violento fatal o no fatal.

Como resultado de esto y con la construcción del visor cartográfico SIGEODEP SIG, se puede visualizar cada evento violento en la cartografía del municipio de Pasto. En el caso de mapas de puntos (figura 4) el usuario puede navegar sobre los mapas generados y visualizar datos relacionados con cada evento violento georreferenciado. En el caso de los mapas de calor generados (figura 5) el usuario puede seleccionar áreas de interés donde se visualicen zonas o patrones visuales de interés y generar gráficos estadísticos sobre el conteo de eventos violentos presentes en esa zona.

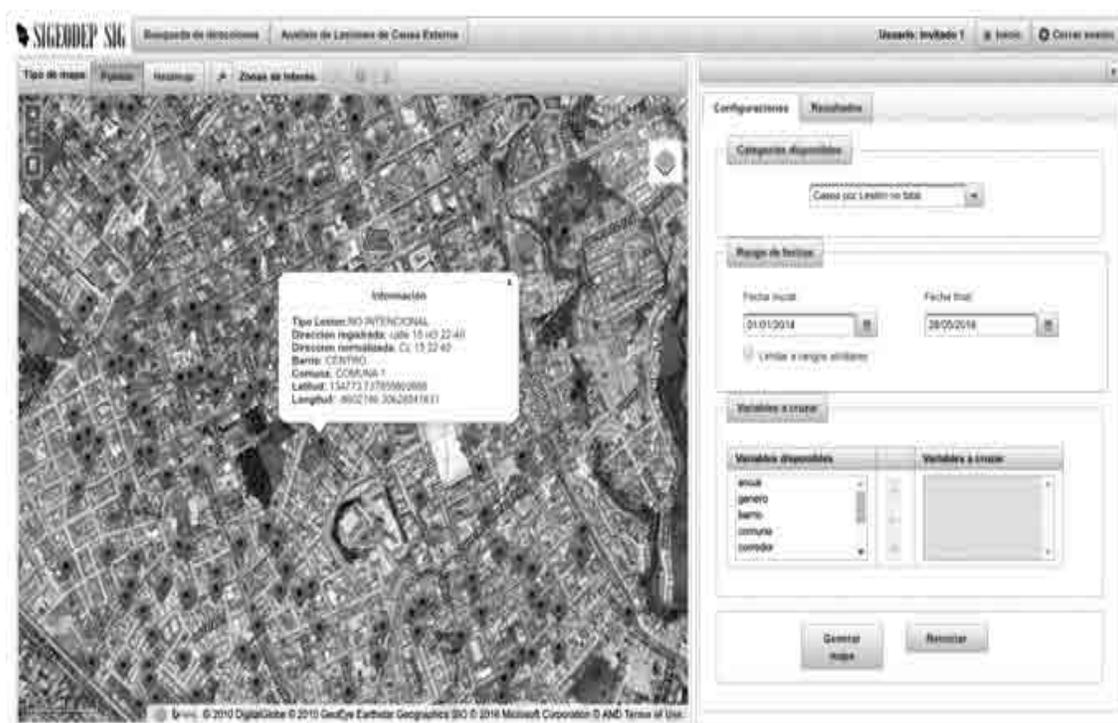


Figura 4. Visualización de eventos delictivos con mapa de puntos.

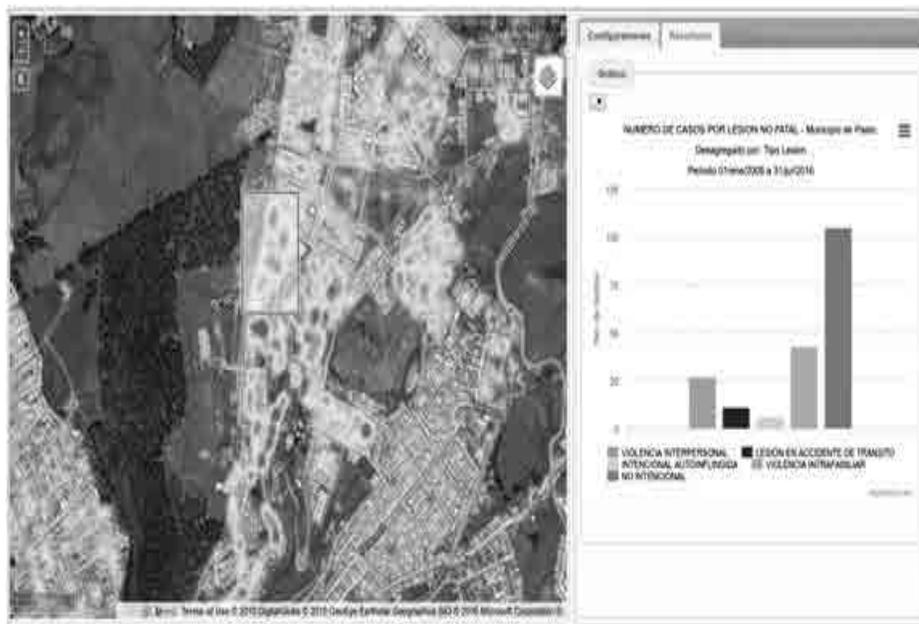


Figura 5. Visualización de eventos delictivos con mapas de calor.

3.2 Descubrimiento de patrones de lesiones no fatales

Con el fin de detectar patrones de eventos violentos no fatales y georreferenciados, se siguió como metodología las etapas del proceso de descubrimiento de conocimiento en base de datos, que contempla: la selección de datos, el preprocesamiento, la minería de datos y la visualización (Hernández, Ramírez, & Ferri, 2005; Han, & Kamber, 2001).

En la etapa de selección se escogió del almacén de datos de SIGODEP, la información de las lesiones no fatales que ocurrieron en el municipio de Pasto, entre los años 2011 y 2014, con direcciones asignadas según la malla vial del municipio de Pasto y con coordenadas geográficas asignadas por el geocodificador. Los atributos seleccionados se muestran en la tabla 1.

ATRIBUTO	DESCRIPCIÓN
Tipo delito	Tipo de lesión de causa externa no fatal
Día	Día de ocurrencia del delito, calculado a partir de su fecha de ocurrencia
Trimestre	Trimestre del año en el que ocurrió el delito, calculado a partir de su fecha de ocurrencia
Barrio	Barrio donde ocurrió el delito, dato asignado por el geocodificador de direcciones.
Comuna	Comuna donde ocurrió el delito, dato asignado por el geocodificador de direcciones.
Longitud	Valor numérico decimal, dato asignado por el geocodificador de direcciones.
Latitud	Valor numérico decimal, dato asignado por el geocodificador de direcciones.

Tabla 1. Atributos seleccionados del repositorio de eventos violentos no fatales.

En la etapa de preprocesamiento se eliminaron aquellos registros de las lesiones no fatales que carecían de dirección y que por lo tanto el geocodificador no le pudo asignar coordenadas geográficas. Como resultado de esta etapa se obtuvo el repositorio T8911A7 compuesto por 8911 registros y 7 atributos, de los cuales 5127 casos pertenecen a lesiones no intencionales, 2618 casos pertenecen a violencia intrafamiliar, 670 a violencia interpersonal, 333 a lesiones en accidentes de tránsito y 163 a lesiones autoinflingidas.

En la etapa de minería de datos se seleccionó la tarea de aprendizaje no supervisado agrupamiento o clustering, como la técnica más adecuada para solucionar el problema objeto de la investigación. Se utilizó la herramienta Weka (Hall, Frank, & Witten, 2011) para la identificación de los clusters.

En la tarea de clustering se trata de encontrar grupos similares entre un conjunto de datos basado en el concepto de distancia (Hernández, Ramírez, & Ferri, 2005), (Han, & Kamber, 2001). Los clusters tienen una alta homogeneidad interna (dentro del cluster) y una alta heterogeneidad externa (entre cluster) (Chen, Han, & Yu, 1996). El algoritmo más popular de los métodos de clustering particionales es el K-means (Han, & Kamber, 2001). La idea de este algoritmo es tomar como entrada el parámetro k y particionar el conjunto de datos en k grupos o clusters, donde los casos o instancias pertenecientes al mismo grupo tengan características similares. Todo caso nuevo es comparado con los grupos formados y asociado a aquél que sea el más próximo, en los términos de una distancia que normalmente es la Euclidiana (Sharma, Bajpai, & Litoriya, 2012).

En la herramienta Weka, para ejecutar K-means, se escoge la opción del algoritmo SimpleKmeans y

se configura número de clusters (NumClusters) o grupos a formar y la semilla (seed), que se utiliza en la generación de un número aleatorio, el cual es usado para hacer la asignación inicial de instancias a los clusters. Para evaluar los resultados del agrupamiento, está la opción Cluster Mode. Lo más sencillo es utilizar el propio conjunto de entrenamiento, Use training set, que indica que utilizará todo el conjunto de datos disponibles para entrenamiento y luego aplicará otra vez sobre los mismos. El resto de opciones realizan un entrenamiento con un conjunto, sobre el que construyen los clusters y a continuación aplican estos clusters para clasificar un conjunto independiente que puede proporcionarse aparte (Supplied test), o ser un porcentaje del conjunto de entrada (split). Existe también la opción de comparar los clusters con un atributo de clasificación (Classes to clusters evaluation) que no se considera en la construcción de los clusters.

Específicamente, en esta investigación se utilizó el propio conjunto de entrenamiento (Use training set) para evaluar los resultados del agrupamiento. Se configuró K-means para encontrar $K=2$, $K=4$, $K=6$ clusters, con una semilla por defecto de 10. Con $k=2$ se obtienen en los dos clusters (0 y 1) únicamente lesiones no intencionales. Con $K=4$, los tres primeros clusters (0, 1 y 2), caracterizan a las lesiones no intencionales y el último (3) a violencia intrafamiliar, como se muestra en la tabla 2. Con $K=6$, los clusters (0, 1, 4 y 5,) agrupan a lesiones no intencionales, y los clusters 2 y 3 a violencia intrafamiliar. No se produce un nuevo conocimiento acerca de las lesiones no fatales. Por estas razones, se escoge $K=4$ como el parámetro que encuentra los cuatro clusters con mayor cohesión. En la tabla 3 se muestran las características de los centroides de cada cluster.

CLUSTER	CLUSTER0	CLUSTER1	CLUSTER2	CLUSTER3
Cantidad	2.425	2.870	1.808	1.808
Porcentaje	27	33	20	20

Tabla 2. Aplicación del algoritmo K-means con k=4.

Atributo	Cluster0	Cluster1	Cluster2	Cluster3
tipo_delito	No intencional	No intencional	No intencional	Violencia intrafamiliar
día	Martes	Lunes	Miércoles	Domingo
trimestre	II	III	III	IV
barrio	Aire libre	Centro	El Pilar	El Progreso
comuna	Comuna 2	Comuna 1	Comuna 5	Comuna 5
longitud	-8602800.89	-8602981.94	-9602257.84	-8602398.7591
latitud	134700.2291	135385.6017	133695.6267	134188.3962

Tabla 3. Características de cada cluster.

Analizando los resultados obtenidos, de acuerdo con las tablas 2 y 3, en el cluster 0 se identifican las características predominantes del 27% de las lesiones no fatales seleccionadas, las cuales se asocian a lesiones de tipo no intencional que ocurren los días martes del segundo trimestre del año, en la comuna 2, específicamente en el área del barrio Aire Libre. En el cluster 1 se identifican las características predominantes del 33% de las lesiones no fatales procesadas, las cuales se asocian a lesiones de tipo no intencional que ocurren los días lunes del tercer trimestre del año, en la comuna 1, específicamente en la zona centro de la ciudad. En el cluster 2 se identifican las características predominantes del 20% de las lesiones no fatales seleccionadas, las cuales se asocian a lesiones de tipo no intencional que ocurren los días miércoles del tercer trimestre del año, en la comuna 5, específicamente en el área del barrio El Pilar. Finalmente, en el cluster 3 se identifican las características predominantes del 20% de las lesiones no fatales seleccionadas, las cuales se asocian a lesiones de tipo violencia intrafamiliar, que ocurren los días domin-

go del cuarto trimestre del año, en la comuna 5, específicamente en el área del barrio El Progreso.

3.3 Visualización de patrones

El objetivo de esta etapa fue ubicar y visualizar en la cartografía del municipio de Pasto, cada una de las lesiones no fatales agrupadas en los diferentes clusters encontrados en la etapa de minería de datos, a partir de las coordenadas geográficas latitud y longitud.

Para hacer esto posible, se utilizó la funcionalidad de visualización de asignación de clusters implementada en la herramienta Weka [9], que permite asignar a cada una de las instancias o registros, a que cluster pertenece. Algunos de estos datos pueden observarse en la figura 6.

Instance_n	delito	dia	trimestre	cuadra	barrio	comuna	longitud	latitud	Cluster
0	'NO INTENCIONAL'	Sabado	IV	'KR 12 CL 15'	'LAS VIOLETAS'	'COMUNA 2'	-8602203.75	133795.59	cluster0
1	'NO INTENCIONAL'	Martes	III	'KR 22 A CL 9'	'LOS ALAMOS'	'COMUNA 2'	-8603199.89	134465.78	cluster0
2	'NO INTENCIONAL'	Domingo	I	'KR 37 CL 18'	PALERMO	'COMUNA 9'	-8603252.09	136458.76	cluster1
3	'VIOLENCIA INTRAFAMILIAR'	Domingo	I	'CL 18 KR 11'	FATIMA	'COMUNA 2'	-8601972.12	133894.59	cluster3
4	'NO INTENCIONAL'	Jueves	II	'KR 33 CL 16 A'	MARIDIAZ	'COMUNA 9'	-8603187.75	135869.95	cluster0
5	'NO INTENCIONAL'	Domingo	III	'CL 18 KR 56'	TOROBAJO	'COMUNA 9'	-8604877.66	137417.67	cluster1
6	'NO INTENCIONAL'	Sabado	IV	'KR 18 CL 20'	'EL PRADO'	'COMUNA 2'	-8602170.74	134509.66	cluster0
7	'VIOLENCIA INTERPERSONAL'	Domingo	I	'KR 13 A CL 20'	'EL RECUERDO'	'COMUNA 2'	-8601841.09	134070.93	cluster3
8	INTENCIONAL AUTOINFLINGIDA	Sabado	II	'CL 5 KR 22 D'	SAN JOSE OBRERO'	'COMUNA 1'	-8603486.53	134499.19	cluster0

Figura 6. Asignación de cluster a cada evento violento no fatal.

Para la visualización geográfica de clusters, se utilizó la funcionalidad de procesamiento y agrupación de datos georreferenciados almacenados en archivos CSV, implementada en el visor carto-

gráfico SIGODEP SIG. En la figura 7 se muestra la ubicación geográfica de las lesiones no fatales, en la cartografía del municipio de Pasto, que fueron agrupadas en el cluster 0.

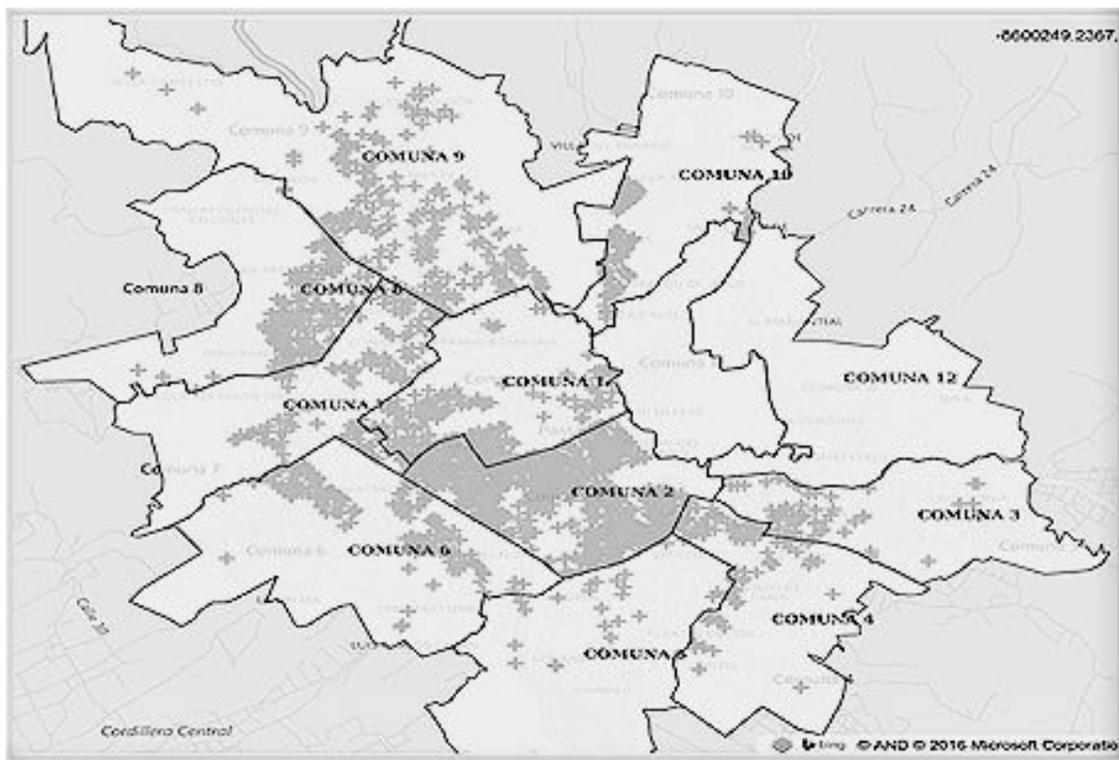


Figura 7. Visualización de las lesiones no fatales, georreferenciadas en el cluster 0.

Igualmente, en la figura 7 se puede identificar la tendencia descrita por el patrón encontrado en el cluster 0, ubicando una mayor cantidad de lesiones de tipo no intencional en la comuna 2 del municipio de Pasto. De igual forma se identifican las

comunidades 8 y 9 con una presencia de este tipo de lesiones. Adicionalmente, en la figura 8 se muestra la ubicación geográfica de las lesiones no fatales, que fueron agrupadas en el cluster 1.

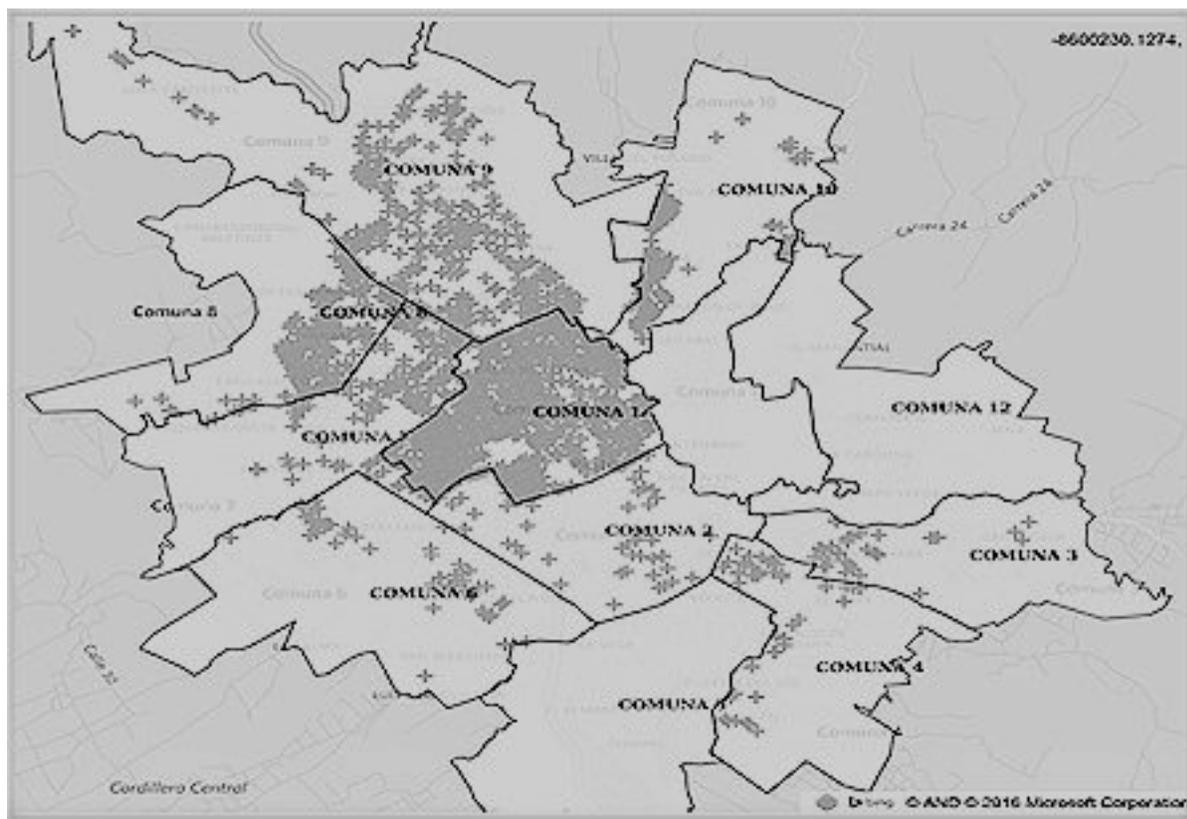


Figura 8. Visualización de las lesiones no fatales, georeferenciadas en el cluster 1.

En la figura 8 se ratifica la tendencia descrita por el patrón encontrado en el cluster 1, en la etapa de minería de datos, como es el de ubicar una mayor cantidad de lesiones de tipo no intencional en la comuna 1 del municipio de Pasto, aunque también hay presencia de esta lesión en las comunas 7, 8 y 9.

En la figura 9 se muestra la ubicación geográfica de las lesiones no fatales, en la cartografía del municipio de Pasto, que fueron agrupadas en el cluster 2.

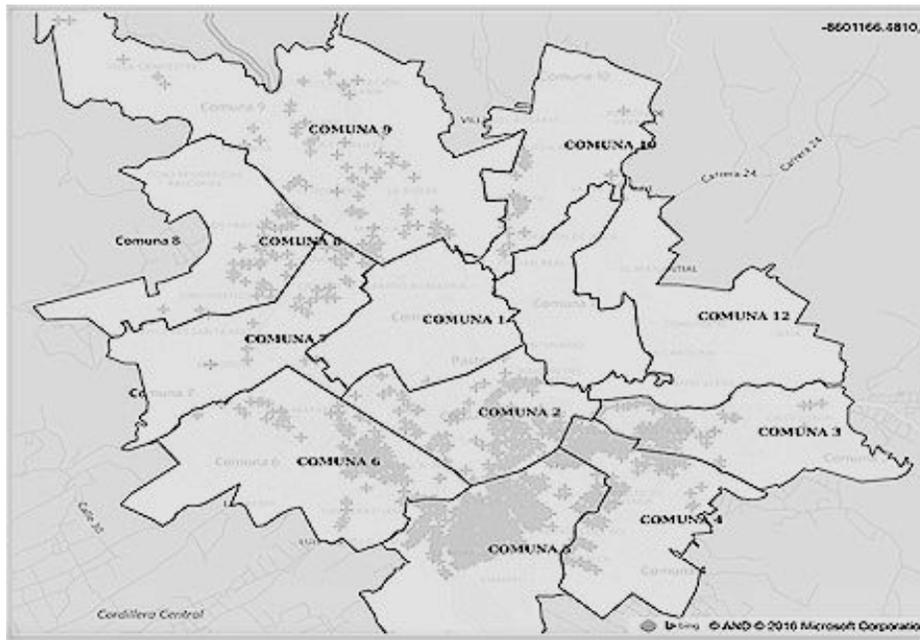


Figura 9. Visualización de las lesiones no fatales georreferenciadas en el cluster 2.

En la figura 10 se muestra la ubicación geográfica de las lesiones no fatales, en la cartografía del

municipio de Pasto, que fueron agrupadas en el cluster 3.

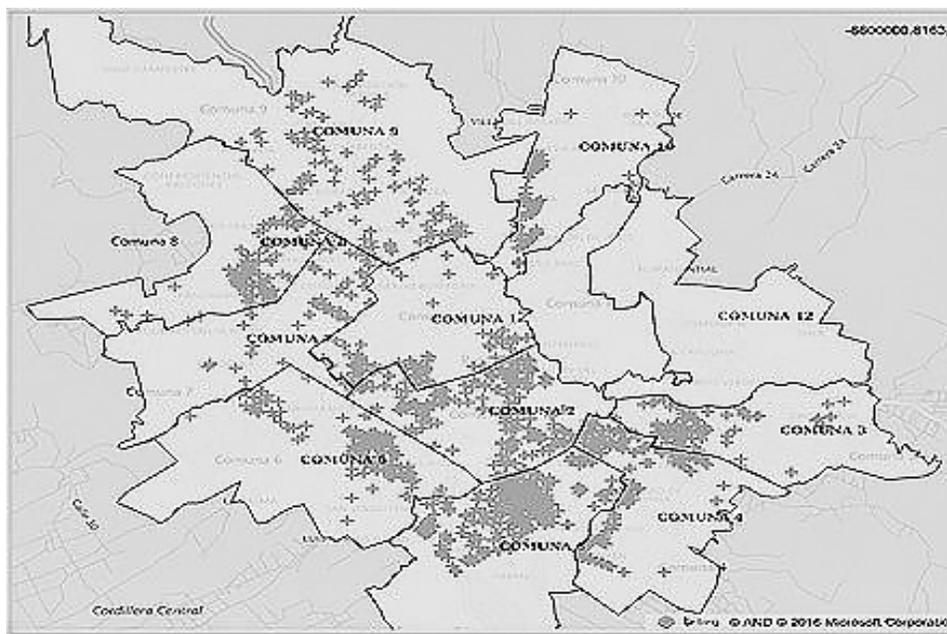


Figura 10. Visualización de las lesiones no fatales georreferenciadas en el cluster 3.

Como se puede observar en la figura 10, en las comunas 5 y 2 del municipio de Pasto suceden la mayoría de lesiones por violencia intrafamiliar.

4. Conclusiones

Teniendo en cuenta los resultados obtenidos con la técnica de aprendizaje no supervisado clustering, la mayoría de las lesiones no fatales son relacionadas por similitud como el tipo no intencional o accidental, en un 79.7 % (clusteres 0, 1 y 2) y el resto de los 8911 casos registrados se asimilan como del tipo violencia intrafamiliar, con un 20.3% (cluster 3).

Al georreferenciar cada lesión no fatal y al aplicarles el algoritmo k-means, se pudo obtener mapas de riesgos de lesiones no fatales que presentan un panorama sobre el comportamiento del riesgo de este tipo de evento, en todas las comunas del municipio de Pasto, en el periodo analizado.

Se plantea como trabajo futuro complementar este estudio utilizando otras técnicas de minería de datos que permitan relacionar qué lesiones fatales y no fatales se presentan juntas, así como predecir con técnicas supervisadas las zonas de riesgo del municipio de Pasto para las diferentes lesiones.

Agradecimientos

Al Sistema de Investigaciones de la Universidad de Nariño por financiar esta investigación.

Referencias

Betancourt-Salazar, C. (2005). *Vigilancia de lesiones de causa externa para la toma de decisiones en el nivel local: experiencia de Pasto, Colombia, año 2005*. San Juan de Pasto, Colombia: Instituto Nacional de Salud – SEA, Instituto Cisalva. Recuperado de: http://www.javeriana.edu.co/fcea/convocatorias/memorias_1congreso_sp/vigilancia_salud_publica/103w.pdf

Chen, M., Han, J., & Yu, P. (1996). Data Mining: An Overview from Database Perspective. *Journal IEEE Transactions on Knowledge Data Engineering*, 8, (6), 866-883.

Hall, M., Frank, E., & Witten, I. (2011). *Practical Data Mining: Tutorials*. University of Waikato. Recuperado de: <http://www.micai.org/2012/tutorials/Weka%20tutorials%20Spanish.pdf>.

Han, J., & Kamber, M. (2001). *Data Mining: Concepts and Techniques*. San Francisco, USA: Morgan Kaufmann Publishers.

Hernández, J., Ramírez, M., & Ferri, C. (2005). *Introducción a la Minería de Datos*. Madrid, España: Editorial Pearson Educación S.A.

Ministerio de Salud y Protección Social, MinSalud (2013). *Guía Metodológica para Registros, Observatorios, Sistemas de Seguimiento y Salas Situacionales Nacionales en Salud*. Bogotá, Colombia: Dirección de Epidemiología y Demografía.

Molina, J., (2002). *Data Mining: torturando a los datos hasta que confiesen*. Recuperado de: <http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>.

Organización de Estados Americanos, OEA. (2013). *Manual para la creación de Observatorios Nacionales sobre criminalidad y violencia*. Recuperado de: <http://www.oas.org/dsp/alertamerica/documents/manualobservatoriosnacionales.pdf>

Organización Mundial de la Salud, OMS. (2002). *Informe mundial sobre la violencia y la salud: SINOPSIS*. Recuperado de: http://apps.who.int/iris/bitstream/10665/67411/1/a77102_spa.pdf?ua=1

Organización Panamericana de la Salud, OPS. (2008a). *Sistematización de Experiencias sobre Sistemas de Vigilancia, Observatorios o Sistemas de Información de Violencia en América Latina*. Recuperado de: <http://apps.who>

int/iris/bitstream/10665/173208/1/OPS%-2520CISALVA%2520Manual%2520Observatorios%2520%282008%29.pdf

Organización Panamericana de la Salud, OPS. (2008b). *Guía Metodológica para la Replicación de Observatorios Municipales de Violencia*. Cali, Colombia: Centro Editorial CATORSE SCS.

Rivillas, J. C., Montañó-Caicedo, J. I., Cuéllar-Segura, C. M., & Ospina, M. L. (2013). Registros, observatorios y sistemas de seguimiento en salud en Colombia: Orientación de políticas basadas en la evidencia y gestión del conocimiento. Superintendencia Nacional de Salud. *Monitor Estratégico*, 4, 56-62.

Sharma, N., Bajpai, A., & Litoriya, M. R. (2012). Comparison the various clustering algorithms of weka tools. *International Journal of Emerging Technology and Advanced Engineering*, 2, (5) 73-80. Recuperado de: <https://pdfs.semanticscholar.org/eca2/5eb78be04ffe09b029dd1d36f5ba66749f29.pdf>

Timarán, R., Barón, A., Hernández, G., Hidalgo, A., & Betancourth, C. (2012). SIGODEP: Un primer paso para la Detección de Patrones Delictivos con Técnicas de Minería de Datos. *IX Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento (JIISIC 2012)*. Pontificia Universidad Católica del Perú. Lima, Perú.

Timarán, R., Calderón, A., Hidalgo, A., Barón, A., & Hernández, G. (2013). Construcción de un Mercado de Datos para soportar la Detección de Patrones. *Primer Congreso Andino de Computación,*

Informática y Educación CACIED 2013. Universidad de Nariño. San Juan de Pasto, Colombia.

Timarán, R., Hernández, G., & Quemá, N. (2016). Geocodificador de eventos delictivos georreferenciados a nivel de direcciones urbanas en el municipio de Pasto. *5º Congreso Internacional de Gestión Tecnológica y de la Innovación COGESTEC 2016*. Universidad Industrial de Santander. Bucaramanga, Colombia.

Timarán-Pereira, R., & Yépez-Chamorro, M. (2016). Caracterización de la supervivencia de mujeres con cáncer invasivo de cuello uterino usando minería de datos. *Revista de Investigación, Desarrollo e Innovación*, 7 (1), 127-139. doi: <http://dx.doi.org/10.19053/20278306.v7.n1.2016.4315>.

Perversi, I., Valenga, F., Fernández, F., Britos, P., & Garcia-Martinez, R. (2007). Identificación y Detección de Patrones Delictivos basada en Minería de Datos. *En memorias de IX Workshop de Investigadores en Ciencias de la Computación, Universidad Nacional de la Patagonia San Juan Bosco, Trelew, Argentina*.

Valenga, F., Fernández, E., Merlino, H., Procopio, C., Britos, P., & Garcia-Martinez, R., (2008). Minería de Datos Aplicada a la Detección de Patrones Delictivos en Argentina. *En memorias de VI Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento (JIISIC 2008)*. Guayaquil, Ecuador.

Global Health Observatory, WHO. (2012). *Geneva: WHO*. Recuperado de: <http://www.who.int/research-observatory/en/>.