

# Identificación Georreferenciada de Patrones de Lesiones no Fatales con Técnicas de Aprendizaje no Supervisado

Ricardo Timarán Pereira, Ph.D.<sup>1</sup>, Gonzalo Hernandez, Mg.<sup>1</sup>, Nelson Quemá Taimbud, Ing.<sup>2</sup>

<sup>1</sup>Universidad de Nariño, Colombia, {ritimar,gonzalohernandez}@udenar.edu.co

<sup>2</sup>Universidad de Nariño, Colombia, nelson8533@gmail.com

**Abstract—** This article presents one of the results of the research project that aimed to identify in a georeferenced way at the level of urban addresses, patterns of fatal and non-fatal criminal events from the information stored in the Crime Observatory of the Municipality of Pasto (Colombia) with data mining techniques. In order to analyze a criminal event and process it geographically, a geocoder was built under open source code, of urban addresses of the municipality of Pasto that allowed georeferencing each criminal event. This geocodificador was integrated to a cartographic viewer, that when applying the technique of non-supervised learning named clustering, allowed to visualize and to identify patterns in zones of the municipality where the different criminal events happen. This information will make it easier for government and security agencies to make effective decisions related to citizen security and crime prevention.

**Keywords—** Patterns Discovery, Georeferencing, Non-Fatal Injuries, Non-Supervised Learning, DataMining.

**Resumen—** En este artículo se presenta uno de los resultados del proyecto de investigación que tuvo como objetivo identificar de manera georreferenciada a nivel de direcciones urbanas, patrones de eventos delictivos fatales y no fatales a partir de la información almacenada en el Observatorio del Delito del Municipio de Pasto (Colombia) con técnicas de minería de datos. Para analizar un evento delictivo y procesarlo geográficamente, se construyó un geocodificador bajo código abierto, de direcciones urbanas del municipio de Pasto, que permitió georreferenciar cada evento delictivo. Este geocodificador se integró a un visor cartográfico, que al aplicar la técnica de aprendizaje no supervisado llamada clustering, permitió visualizar e identificar patrones en zonas del municipio donde suceden los diferentes eventos delictivos. Esta información facilitará a los organismos gubernamentales y de seguridad la toma de decisiones eficaces relacionadas con la seguridad ciudadana y la prevención de delitos.

**Palabras Clave—** Descubrimiento de Patrones, Georreferenciación, Lesiones no Fatales, Aprendizaje no Supervisado, Minería de datos.

## I. INTRODUCCIÓN

La 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»[1]. La Organización Panamericana

de la Salud, OPS desde el año 1993 y la Organización Mundial de la Salud, 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 [2].

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 [3].

Una de las estrategias implementadas en vigilancia en salud pública, corresponde a los Observatorios de Muertes de Causa Externa [2], 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. En Colombia se cuenta con diversas experiencias a nivel municipal y departamental.

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.

Actualmente el Observatorio cuenta con un sistema de vigilancia de eventos violentos a partir de la implementación de un sistema de información georreferenciado denominado SIGODEP soportado en un mercado de datos, donde 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, que le permite disponer de información confiable, oportuna, de buena calidad y representativa de las lesiones de causa externa que ocurren en el municipio de Pasto [4], [5].

Este sistema georreferencia las lesiones de causa externa fatales y no fatales, (ver tabla 1), que ocurrieron en el municipio de Pasto desde el año 2002 hasta el presente año a nivel de barrios y comunas como el gránulo más pequeño de georreferenciación y no a nivel de direcciones urbanas.

En este artículo se presenta los resultados de identificar de manera georreferenciada patrones de lesiones no fatales utilizando la técnica de minería de datos clustering y su

Digital Object Identifier (DOI): <http://dx.doi.org/10.18687/LACCEI2017.1.1.93>

ISBN: 978-0-9993443-0-9

ISSN: 2414-6390

**13<sup>th</sup> LACCEI Annual International Conference:** “Engineering Education Facing the Grand Challenges, What Are We Doing?”  
July 29-31, 2015, Santo Domingo, Dominican Republic

posterior visualización. Con este fin, se construyó un geocodificador bajo código abierto, de direcciones urbanas del municipio de Pasto, que permitió georreferenciar cada evento delictivo. Este geocodificador se integró a un visor cartográfico, que al aplicar la técnica de aprendizaje no supervisado clustering, permitió visualizar e identificar patrones en zonas del municipio donde suceden los diferentes eventos delictivos. Esta información facilitará a los organismos gubernamentales y de seguridad la toma de decisiones eficaces relacionadas con la seguridad ciudadana y la prevención de delitos.

TABLA I  
TIPOS DE EVENTOS DELICTIVOS ASOCIADOS A LAS LÍNEAS DE  
VIGILANCIA DEL OBSERVATORIO DEL DELITO

| LESIONES FATALES                  | LESIONES NO FATALES                  |
|-----------------------------------|--------------------------------------|
| Homicidios                        | Violencia Interpersonal en comunidad |
| Suicidios                         | Violencia Autoinflingida             |
| Muertes por accidente de tránsito | Lesiones por accidentes de tránsito  |
| Lesiones no intencionales         | Lesiones no Intencional o Accidental |
|                                   | Violencia Interpersonal en Familia   |

El resto del artículo se organiza de la siguiente manera: en la sección II se describe la metodología utilizada en esta investigación. En la sección III se explica los procesos de geocodificación de eventos delictivos y de descubrimiento de patrones de lesiones no fatales. En la sección IV se visualizan los patrones encontrados en la cartografía del municipio de Pasto y finalmente en la sección V se presentan las conclusiones y futuros trabajos.

## II. METODOLOGÍA

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: 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.

### A. 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.

### B. Construcción del geocodificador de direcciones urbanas

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

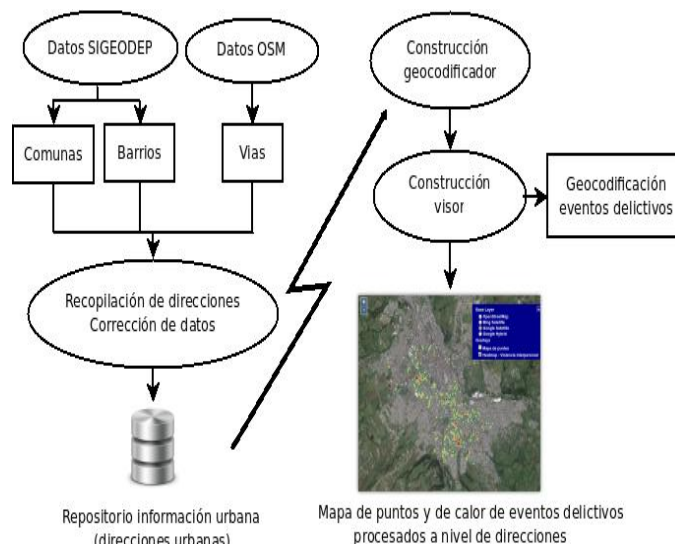


Fig. 1. Metodología aplicada en la investigación

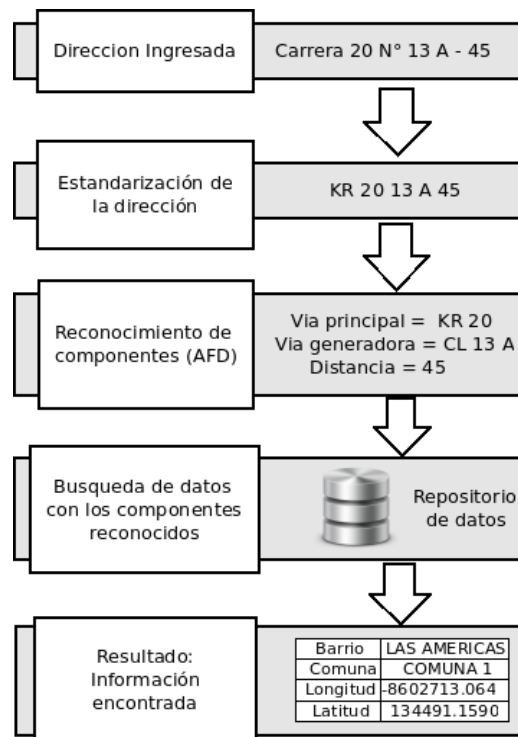


Fig. 2. Proceso de geocodificación de direcciones urbanas con nomenclatura de malla vial

### C. 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

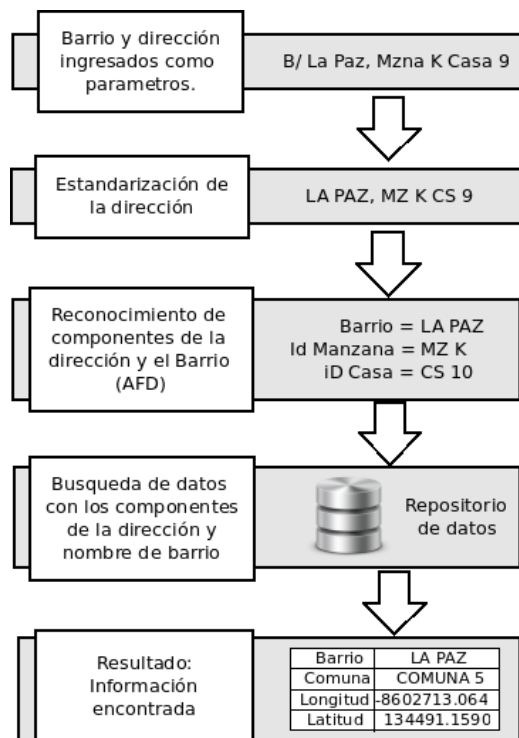


Fig. 3. Proceso de geocodificación de direcciones urbanas con nomenclatura de barrio-manzana

eventos delictivos fatales y no fatales registrados por este ente 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 [6]. Los procesos de geocodificación de eventos delictivos y de descubrimiento de patrones de lesiones no fatales y su visualización se explican en las siguientes secciones.

### III. RESULTADOS

#### A. Geocodificación de eventos delictivos

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 delictivos 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 delictivo fatal o no fatal.

Como resultado de ello y con la construcción del visor cartográfico SIGEODEP SIG, se puede visualizar cada evento delictivo 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 delictivo 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 delitos presentes en esa zona.

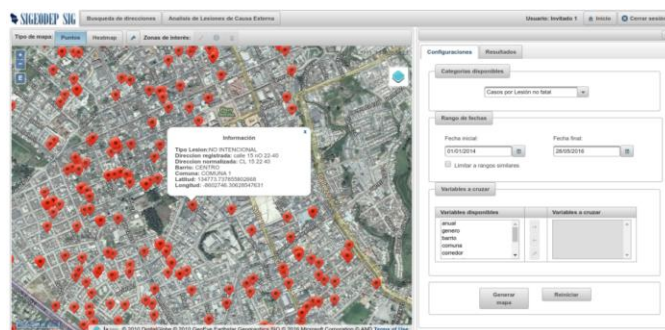


Fig. 4. Visualización de eventos delictivos con mapa de puntos

#### B. Descubrimiento de patrones

Con el fin de identificar patrones georreferenciados de lesiones no fatales, se siguió como metodología las etapas del proceso de descubrimiento de conocimiento en base de datos que por lo general contempla la selección de datos, el preprocesamiento, la minería de datos y la visualización [7][8].

En la etapa de selección, se escogieron del almacén de datos de SIGEODEP, 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 II.

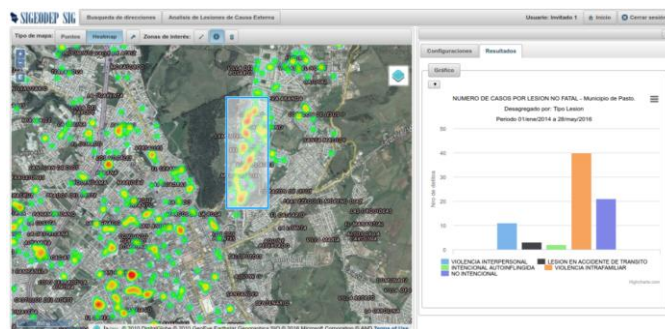


Fig. 5. Visualización de eventos delictivos con mapas de calor

TABLA II  
ATRIBUTOS SELECCIONADOS DEL REPOSITORIO DE LESIONES NO FATALES

| 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.                 |

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, listo para aplicarle técnicas de minería de datos.

En la etapa de minería de datos se seleccionó la tarea de aprendizaje no supervisado *clustering*, como la técnica más adecuada para solucionar el problema objeto de la investigación. Se utilizó la herramienta *Weka* [9] para el descubrimiento 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 [7][8]. Los *clusters* tienen una alta homogeneidad interna (dentro del *cluster*) y una alta heterogeneidad externa (entre *cluster*) [10]. Para esta tarea, se utilizó la técnica particional con el algoritmo *K-means* [8], en el cual se configura el número de grupos (*NumClusters*) 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 grupos. Para evaluar los resultados del agrupamiento, se utilizó el propio conjunto de entrenamiento (*Use training set*), que indica que porcentaje de instancias se van a cada grupo. Se configuró *K-means* para encontrar  $K=2$ ,  $K=4$  y  $K=6$  *clusters* con una semilla por defecto de 10. Para evaluar los resultados del agrupamiento se utilizó un porcentaje del 66% del propio conjunto de entrenamiento. Con el parámetro  $K=4$  se encontraron los cuatro clusters más homogéneos como se muestra en la tabla III. En la tabla IV están las características de los centroides de cada cluster. De esta manera, se pudo encontrar cuales son las similitudes particulares de cada uno de estos grupos, en un proceso no supervisado.

TABLA III  
CLUSTERS ENCONTRADOS CON K-MEANS CON  $K=4$

| CLUSTER    | CLUSTER0 | CLUSTER1 | CLUSTER2 | CLUSTER3 |
|------------|----------|----------|----------|----------|
| CANTIDAD   | 2.425    | 2.870    | 1.808    | 1.808    |
| PORCENTAJE | 27       | 33       | 20       | 20       |

TABLA IV  
CARACTERÍSTICAS DE CADA CLUSTER

| 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             |

Analizando los resultados obtenidos, de acuerdo con las tablas III y IV, 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 domingo del cuarto trimestre del año, en la comuna 5, específicamente en el área del barrio El Progreso.

#### IV. 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 tabla V.

Para la visualización geográfica de *clusters*, fue utilizada la funcionalidad de procesamiento y agrupación de datos georreferenciados almacenados en archivos CSV, implementada en el visor cartográfico SIGEODEP SIG. A continuación se muestra y describe la visualización de cada cluster encontrado.



TABLA V  
ASIGNACIÓN DE CLUSTERS A CADA LESION

| Instance n | delito                    | dia     | trimestre | cuadra          | barrio          | comuna     | longitud    | latitud   | Cluster  |
|------------|---------------------------|---------|-----------|-----------------|-----------------|------------|-------------|-----------|----------|
| 0          | 'NO INTENCIONAL'          | Sabado  | IV        | 'KR 12 CL 15'   | 'LAS VIOLETAS'  | 'COMUNA 2' | -8602103.75 | 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'   | TORONAJA        | '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 AUTONFLINGIDA | Sabado  | II        | 'CL 5 KR 22 D'  | SAN JOSE OBRERO | 'COMUNA 1' | -8603486.53 | 134499.19 | cluster0 |

En la figura 6 se muestra la ubicación geográfica de los patrones de lesiones no fatales, en la cartografía del municipio de Pasto, que fueron agrupadas en el cluster 0.

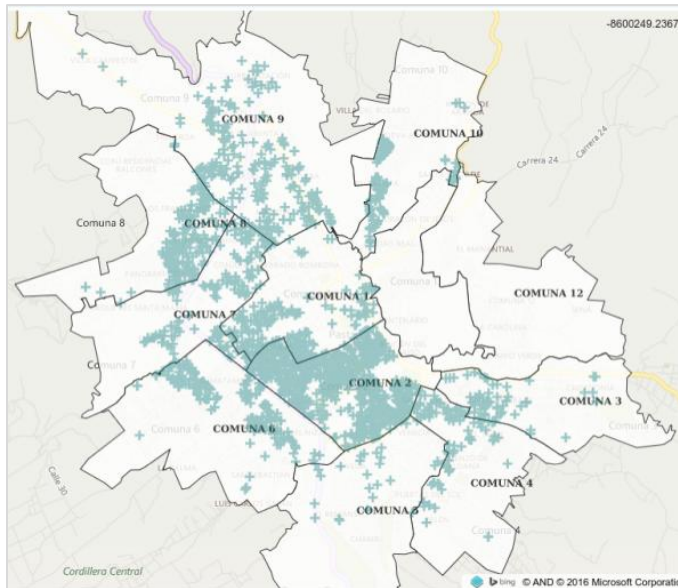


Fig. 6. Visualización de patrones cluster 0

Como se puede observar en la figura 6, 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 comunas 8 y 9 con una presencia de este tipo de lesiones.

En la figura 7 se muestra la ubicación geográfica de los patrones de lesiones no fatales, en la cartografía del municipio de Pasto, que fueron agrupadas en el cluster 1.

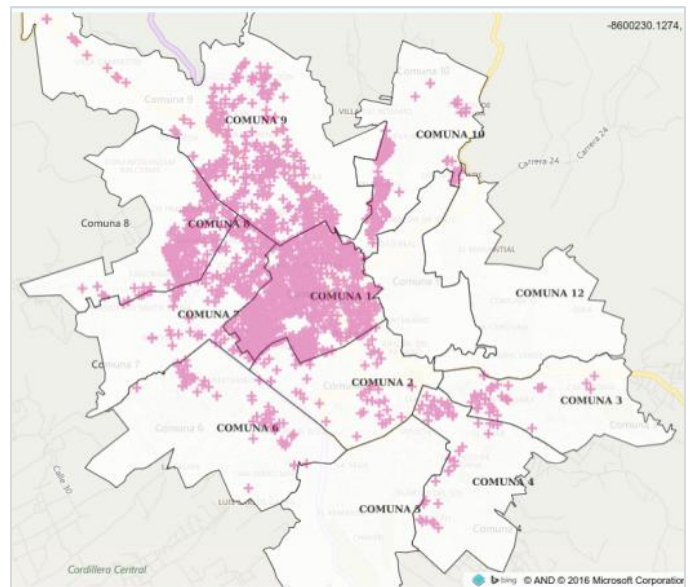


Fig. 7. Visualización de patrones cluster 1

Como se observa en la figura 7, 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 8 se muestra la ubicación geográfica de los patrones de lesiones no fatales, en la cartografía del municipio de Pasto, que fueron agrupadas en el cluster 2.

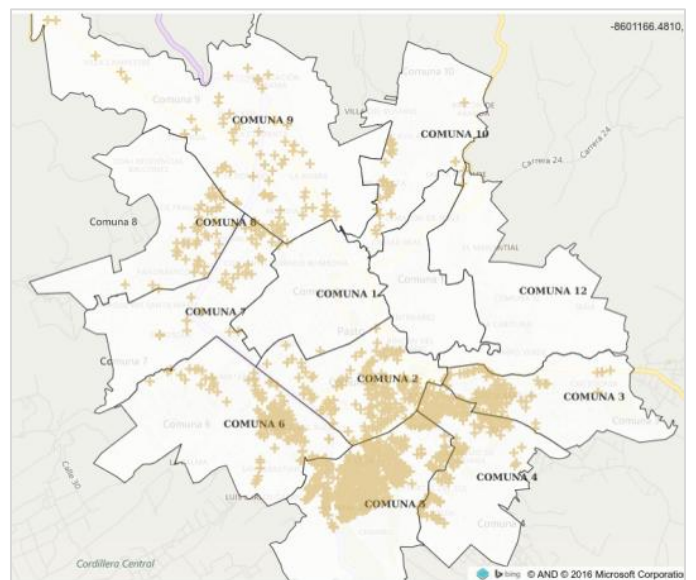


Fig. 8. Visualización de patrones cluster 2

Como se observa en la figura 8, se ratifica la tendencia descrita por el patrón encontrado en el cluster 2, en la etapa de

minería de datos, como es el de ubicar una mayor cantidad de lesiones de tipo no intencional en la comuna 5 del municipio de Pasto.

En la figura 9 se muestra la ubicación geográfica de los patrones de lesiones no fatales, en la cartografía del municipio de Pasto, que fueron agrupadas en el cluster 3.

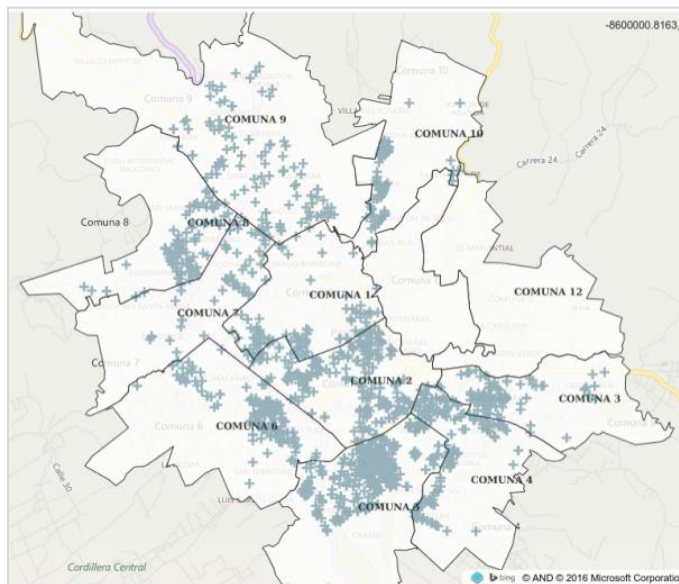


Fig. 9. Visualización de patrones cluster 3

Como se puede observar en la figura 9, en las comunas 5 y 2 del municipio de Pasto, suceden la mayoría de lesiones por violencia intrafamiliar.

## V. CONCLUSIONES Y TRABAJOS FUTUROS

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 % (clusters 0, 1 y 2) y el resto de los 8911 casos registrados se asimilaran 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 trabajos futuros complementar este estudio utilizando otras técnicas de minería de datos que permitan relacionar que lesiones fatales y no fatales se presentan juntas, así como predecir con técnicas supervisadas las zonas de riesgo del municipio de Pasto de las diferentes lesiones.

## AGRADECIMIENTOS

Este proyecto de investigación se financió con recursos del sistema de investigaciones de la Universidad de Nariño (Colombia).

## REFERENCIAS

- [1] OMS, 2002. Informe mundial sobre la violencia y la salud., SINOPSIS, Organización Mundial de la Salud, Ginebra, Suiza, 2002.
- [2] OPS, 2008, Sistematización de Experiencias sobre Sistemas de Vigilancia, Observatorios o Sistemas de Información de Violencia en América Latina. Observatorios de Violencia: mejores prácticas. Organización Panamericana de la Salud. ISBN 978-958-9481-99-8. Cali, Colombia, 2008.
- [3] OPS, 2008, Guía Metodológica para la Replicación de Observatorios Municipales de Violencia. Organización Panamericana de la Salud. Centro Editorial CATORSE SCS ISBN 978-958-8404-00-4. Cali, Colombia, 2008.
- [4] R. Timarán, A. Barón, G. Hernández, A. Hidalgo y C. Betancourth, 2012, SIGODEP: Un primer paso para la Detección de Patrones Delictivos con Técnicas de Minería de Datos, en: Memorias de IX Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento (JIISIC 2012), Pontificia Universidad Católica del Perú, ISBN No 978-612-4057-85-4, Lima, Perú, 2012.
- [5] R. Timarán, A. Calderón, A. Hidalgo, A. Barón, y G. Hernández, 2013, Construcción de un Mercado de Datos para soportar la Detección de Patrones, en: Primer Congreso Andino de Computación, Informática y Educación, Universidad de Nariño, ISBN No. 978-958-8609-67-6, San Juan de Pasto, Colombia, 2013.
- [6] R. Timarán, G. Hernández y N. Quemá, 2016, Geocodificador de eventos delictivos georreferenciados a nivel de direcciones urbanas en el municipio de Pasto, en: Memorias de 5º Congreso Internacional de Gestión Tecnológica y de la Innovación (COGESTEC 2016), Universidad Industrial de Santander, Bucaramanga, Colombia, 2016.
- [7] J. Hernández, M. Ramírez, and C. Ferri, 2005. Introducción a la Minería de Datos. Editorial Pearson Educación SA, Madrid. Recuperado a partir de <http://dspace.ucbscz.edu.bo/dspace/handle/123456789/526>
- [8] J. Han, and M. Kamber, 2001. *Data Mining: Concepts and Techniques, Third Edition* (3 edition.). Burlington, MA: Morgan Kaufmann.
- [9] M. Hall, E. Frank, and I. Witten, 2011. Practical Data Mining: Tutorials. University of Waikato. Recuperado a partir de <http://www.micai.org/2012/tutorials/Weka%20tutorials%20Spanish.pdf>
- [10] M. Chen, J. Han and P. Yu, 1996, Data Mining: An Overview from Database Perspective, Journal IEEE Transactions on Knowledge Data Engineering, Vol. 8, No. 6, pp. 866-883, ISSN 1041-4347. New Jersey, USA.