



Prototipo de un modelo clasificador para la toma de decisiones a partir de la percepción actual de la seguridad en Bogotá

Prototype of a Classification Model for Decision Making Based on the Current Perception of Security in Bogotá

ARTÍCULO DE INVESTIGACIÓN

Fecha de recepción:
26-11-2015

Fecha de aceptación:
22-05-2018

ISSN: 2344-8288

Vol. 6 No. 2

Julio - Diciembre 2018

Bogotá-Colombia

Raúl Andrés Velandia Rodríguez¹, Andrés Guataquira², Jehison Prada³

Para citar este artículo: Velandia, R., Guataquira, A. y Prada, J. (2018). Prototipo de un modelo clasificador para la toma de decisiones a partir de la percepción actual de la seguridad en Bogotá. *TIA*, 6(2), pp. 18-24.

Resumen

Este artículo describe la metodología y los resultados obtenidos al analizar un conjunto de datos recolectados con el objetivo de determinar las posibles causas de mayor relevancia en los actos delictivos de Bogotá; se busca clasificar la información pertinente para obtener información adicional, necesaria en algunos planes sociales, dando ventaja a la hora de incorporar medidas preventivas contra robos o cualquier situación de delincuencia en cada una de las localidades de la ciudad. Como herramienta principal se incorpora la minería de datos, que permite obtener patrones por medio de la clasificación de información relacionada con la criminalidad; se toma como base el modelo de *clustering* y el algoritmo k-means para realizar minería de datos, posteriormente se evalúan factores o patrones que a simple vista no son perceptibles para identificar las causas de este tipo de actos delictivos. Finalmente, se analizan datos concluyentes del proceso de *clustering* y se presentan algunas soluciones para el problema de la delincuencia en Bogotá.

Palabras clave: delincuencia, *clustering*, minería de datos, algoritmo k-means, patrones delictivos.

¹ Ingeniero de Sistemas; Especialista en Ingeniería de Software, Universidad Distrital Francisco José de Caldas. Correo electrónico: ravelandiar@correo.udistrital.edu.co

² Ingeniero de Sistemas, Fundación Universidad Panamericana; Especialista en Ingeniería de Software de la Universidad Distrital Francisco José de Caldas. Desarrollador de software, ASAI [Automated Systems America, Inc.]. Correo electrónico: jaguataquira@correo.udistrital.edu.co

³ Ingeniero de Sistemas; Especialista en Ingeniería de Software, Universidad Distrital Francisco José de Caldas. Ingeniero de software, Seven4N Colombia. Correo electrónico: jepradau@correo.udistrital.edu.co

Abstract

This paper describes the methodology and results obtained by analyzing a set of data collected in order to determine possible causes that are most relevant in Bogota's criminal acts. The relevant information is classified to obtain additional information, necessary in some social programs, giving an advantage when incorporating preventive measures against theft or any crime situation in each of the city's locations. As a main tool, datamining is incorporated, which allows obtaining patterns through the classification of information related to crime; the clustering model and the k-means algorithm are used as a basis to carry out datamining; later, factors or patterns that at first sight are not perceptible to identify the causes of such criminal acts are evaluated. Finally, clustering process conclusive data is analyzed and some solutions for the problem of delinquency in Bogota are presented.

Keywords: delinquency, clustering, data mining, k-means algorithm, crime patterns.

INTRODUCCIÓN

Bogotá, así como las grandes metrópolis del mundo, incorporan una serie de problemáticas que van desde la pobreza hasta problemas de delincuencia, que se unen a la falta de recursos y de atención oportuna por parte de los gobiernos [1], [2], [3]. En la actualidad, la delincuencia uno de los principales problemas y más difíciles de erradicar en la capital colombiana, desde robos con el uso de arma blanca o de fuego, hasta el robo de carros, hogares y celulares es el diario vivir de los ciudadanos, la inseguridad ha pasado de ser un problema de coyuntura nacional y de búsqueda de soluciones por parte de los entes nacionales [4], [2].

Con base en lo citado anteriormente, el presente documento busca ofrecer una herramienta de análisis y toma de decisiones frente a una problemática social que ha venido en crecimiento (criminalidad en Bogotá), a la cual se debe ofrecer

soluciones a partir de mecanismos modernos sin dejar de un lado la experiencia y metodologías tradicionales.

La minería de datos (*datamining*) es un conjunto de técnicas y herramientas que permiten, a partir de un gran conjunto de datos, la obtención de nuevo conocimiento o patrones de comportamiento que no son visibles a simple vista, se caracterizan por trabajar con datos históricos incorporando análisis de bodegas de datos, buscando asociaciones, detección de ciclos temporales, y una de las más importantes, predecir comportamientos [5].

Para la práctica se buscó obtener dos tipos de soluciones con respecto a la problemática de inseguridad en la ciudad, ya mencionada con anterioridad, que permitieran tener una ventaja a la hora de proponer mecanismos que pudieran ofrecer soluciones más radicales a los diferentes tipos de delitos mencionados a continuación. En primer lugar se busca hacer una clasificación que permita detectar aquellas zonas de mayor índice de delitos a partir del algoritmo de k-means (utilizado en las técnicas de minería de datos para la segmentación de poblaciones a partir de la información recolectada), una vez se obtuvo dicha clasificación se optó por buscar aquellas causas que son percibidas por los ciudadanos como las que incentivan este tipo de delitos, desde la pobreza hasta la falta de compromiso por parte del gobierno nacional.

Como mecanismo de obtención de datos se utilizaron una serie de encuestas que fueron realizadas a 49 personas en diferentes localidades de la ciudad, con el fin de encontrar la percepción actual que se tiene por el ciudadano de a pie acerca de la seguridad, más exactamente en cada localidad dependiendo de su clase económica-social. Una vez se obtuvieron los datos, se optó por trabajar con Weka (plataforma para el aprendizaje automático y la minería de datos) como herramienta o entorno para el análisis del conocimiento, el cual es un *software* para el aprendizaje automático y la minería de datos escrito en Java que nos permite implementar el

algoritmo de k-means, el cual facilita el proceso de *clustering* a partir de los datos obtenidos de las diferentes encuestas.

LA MINERÍA DE DATOS EN EL CAMPO DE LA SEGURIDAD

Unos de los campos de la minería de datos es la capacidad de clasificar perfiles, ya sea de usuarios o tendencias, es por eso que los diferentes gobiernos incorporan dichas tecnologías con el objetivo de analizar las bases de datos comerciales para detectar terroristas.

Utilización de minería de datos por el FBI

A principios del mes de julio de 2002, el director del FBI en los Estados Unidos, John Ashcroft, anunció que el departamento de justicia comenzaría a introducirse en la enorme cantidad de datos comerciales en los que se revelan hábitos y costumbres de la población, con el fin de poder identificar a potencialidades terroristas con antelación a que puedan cometer algún atentado. Por su parte, algunos expertos (como Virseda y Román) se encargaban de asegurar que, con esta información, el FBI unirá todas las bases de datos mediante el número de seguridad de la seguridad social y permitirá saber si una persona fuma, qué talla y qué tipo de ropa usa, si ha sido arrestado, cuántas veces, el barrio donde vive, si en este se cometen más o menos delitos, en pocas palabras todos los datos que de cierta forma fueron proporcionados a entidades públicas o gubernamentales [6].

Ante este tipo de aplicaciones surge nuevamente el tema ético de la privacidad y libertad de los individuos. En el fondo la aplicación es un gran hermano que observa cualquier movimiento que se haga, pudiendo con estar ser utilizado para otros fines diferentes a los que tenía al ser inicialmente diseñado.

Minería de datos aplicada en la detección de intrusos en Colombia

Actualmente la seguridad de la información es un tema que compete a cualquier entidad u organización, por tal motivo es de vital importancia tener un sistema que permita detectar comportamientos anómalos en los usuarios que acceden a la información de dichas compañías para así poder ser detectados a tiempo en caso de que puedan ser partícipes de hurtos o daños en la información. Actualmente en Colombia se han realizado proyectos enfocados no a la seguridad social de la ciudadanía, sino a la seguridad de los datos almacenados en la red por diferentes entidades públicas (empresas de servicios públicos, entidades de salud, seguridad social, entre otros), utilizando como herramienta principal la minería de datos [7].

Minería de datos aplicada a información criminal en Argentina

A partir de la crisis de finales del 2001, Argentina se vio afectada por una creciente ola de inseguridad, caracterizada por un aumento en los índices delictivos y los niveles de violencia. Esta situación, por su parte, se presentaba con mayor profundidad en los centros urbanos y llevó a tomar decisiones coordinadas a nivel nacional tendientes a prevenir el delito.

Una de estas medidas fue el impulso del sistema de alerta temprana (SAT) por parte del Ministerio de Justicia y Derechos Humanos. En el plano internacional, los ataques terroristas del 11 septiembre han aumentado significativamente la preocupación por la seguridad de los EE.UU. En general, el tamaño de las bases de datos está basado en aspectos como la capacidad y eficiencia del almacenamiento y no en su posterior uso o análisis [8].

AGRUPACIÓN DE DATOS (CLUSTERING)

La agrupación o *clustering* consiste en agrupar un conjunto de datos basándose en la similitud de los valores de sus atributos. El *clustering* identifica regiones densamente pobladas, denominadas *clusters* de acuerdo con alguna medida de distancia establecida. La técnica de *clustering* ha sido estudiada en las áreas de la estadística, *machine learning*, bases de datos. Entre los algoritmos más destacados que permiten realizar *clustering* se encuentran *Self Organizing Maps* (SOM) y k-means [9].

Algoritmo k-means

Es un método iterativo que busca formar k-clusters, predeterminando la variable *k* antes del inicio del proceso del algoritmo. k-means comienza particionando los datos en *k* subconjuntos no vacíos, calcula el centroide de cada partición como el punto medio del *cluster* y asigna cada dato al *cluster* cuyo centroide sea el más próximo. Luego vuelve a particionar los datos iterativamente, hasta que no haya más datos que cambien de *cluster* de una iteración a la otra.

SOM

También denominado redes de Kohonen, fue creado por Teuvo Kohonen en 1982, se trata de

un modelo de red neuronal con capacidad para formar mapas con características de manera similar a como ocurre en el cerebro. SOM está basado en el aprendizaje no supervisado y competitivo, lo cual quiere decir que no necesita intervención humana durante este [10].

DESARROLLO

A continuación se hará una descripción de cada uno de los procesos y técnicas utilizadas para el desarrollo y obtención de resultados, partiendo desde una fase inicial de obtención de datos, limpieza de datos, preparación de datos, aplicación de una técnica de minería de datos y obtención de conocimiento o patrones de comportamiento.

Obtención de datos

Para el proceso de obtención de datos se procedió a realizar una serie de encuestas que permitían obtener la percepción del ciudadano común, a partir de esto se logró conseguir una serie de datos que posteriormente fueron analizados.

Como características cabe mencionar que se utilizó la plataforma Google Drive (Formularios) donde se plantearon una serie de preguntas como mecanismo de obtención de datos (Figura 1).

La encuesta fue aceptada por 49 personas de diferentes localidades de la ciudad de Bogotá y de diferentes edades.

Considera usted que la seguridad proporcionada por el gobierno actual es :

- Excelente
- Buena
- Regular
- Mala

¿Cuál es su localidad de residencia?

Considera usted que en los últimos años la seguridad de su localidad a:

- Mejorado
- Sigue igual
- Ha venido en decadencia
- Es pésima
- Otro:

Considera que el tipo de delincuencia más común en su localidad es:

- Asalto a transeúntes
- Carterismo
- Violación
- Robo de bienes y artículos menores
- Robo a casas habitación
- Robo de vehículos
- Vandalismo
- Graffiti y pintura de muros y monumentos
- Otro:

De acuerdo a la respuesta anterior, califique de 1 a 5 (1 Frecuencia Baja, 5 Frecuencia Alta) la frecuencia con que se comete el delito :

1 2 3 4 5

Considera usted que las franjas en las que más se comete este tipo de delitos es:

- 6:00 PM - 12:00 PM
- 12:00 PM - 6:00 AM
- 6:00 AM - 11:59 AM
- 11:59 AM - 6:00 PM

Los sitios donde más se ven este tipo de delitos es:

- Vías Públicas
- Casas de Habitación
- Local Comercial
- Frente Residencias

Una posible causa para este tipo de delitos es:

- El desempleo
- La falta de oportunidades educativas
- La pobreza
- La falta de castigo a quienes cometen delitos
- La drogadicción
- El mal funcionamiento de la justicia
- Falta de programas para los jóvenes
- El narcotráfico
- Las diferencias entre ricos y pobres
- La escasez de policías

Si usted es víctima de un delito, ¿Usted lo denunciaría?

- Si
- No

Figura 1. Encuesta realizada a los ciudadanos de Bogotá, sobre la delincuencia.

Fuente: elaboración propia.

Resultados obtenidos por la encuesta



Figura 2. RTA sobre seguridad para la encuesta

Fuente: elaboración propia.

Respecto a los usuarios encuestados, se obtuvo que un 46.9% considera que la seguridad

proporcionada por los entes gubernamentales es regular (Figura 2).

Cabe resaltar que para esta encuesta existió un mayor índice de participación de las localidades de Bosa con un 14,9%, de donde se puede obtener mayor garantía de los datos, seguido de Ciudad de Kennedy con un 10,6% (Figura 3).

La percepción frente a la seguridad para los encuestados se pudo detectar que es 42,6% “sigue igual” y 42,6% “Ha venido en decadencia” con los porcentajes más altos (Figura 4).

Para la mayoría de localidades se puede observar que el mayor índice de incidencia frente a la delincuencia está con 56,3% asalto a transeúntes (Figura 5).

¿Cuál es su localidad de residencia?

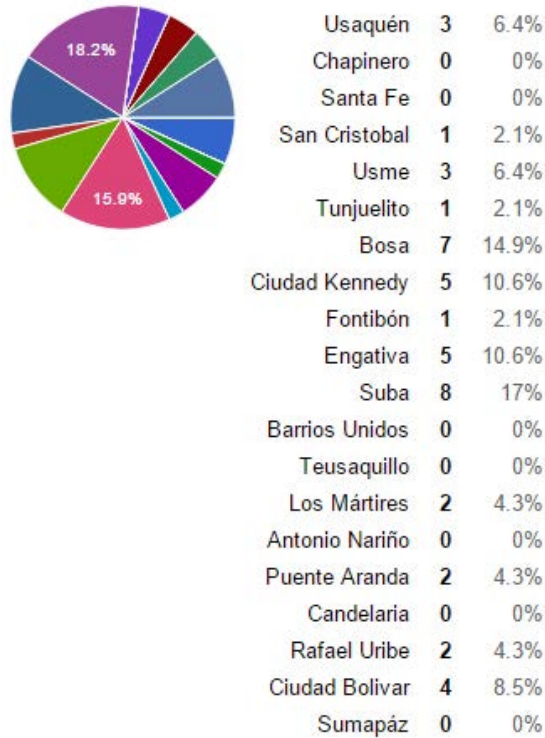


Figura 3. RTA sobre localidades

Fuente: elaboración propia.

Otro dato a destacar es que la frecuencia con la que se percibe la intensidad de los delitos es bastante alta, para una calificación de cuatro se puede observar un porcentaje del 41,7% (Figura 6).

De acuerdo a la respuesta anterior, califique de 1 a 5 (1 Frecuencia Baja, 5 Frecuencia Alta) la frecuencia con que se comete el delito :

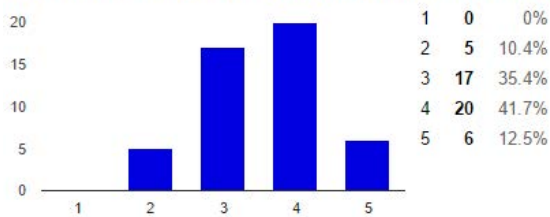


Figura 6. RTA sobre calidad

Fuente: elaboración propia.

Considera usted que en los últimos años la seguridad de su localidad a:



Figura 4. RTA sobre seguridad

Fuente: elaboración propia.

Considera que el tipo de delincuencia más común en su localidad es:

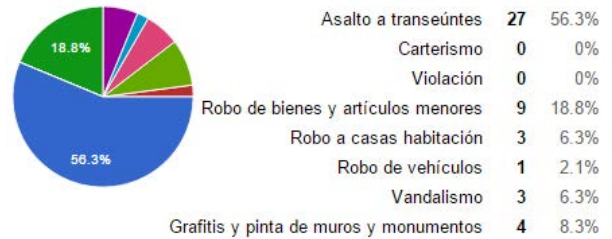


Figura 5. RTA sobre delincuencia.

Fuente: elaboración propia.

Para las franjas en que más se realizan este tipo de delitos se puede observar que con un porcentaje del 66% es considerado que se realiza de 6:00 pm a 12:00 pm, donde se opta por las horas de la tarde para realizar este tipo de delitos (Figura 7).

Considera usted que las franjas en las que más se comete este tipo de delitos es:

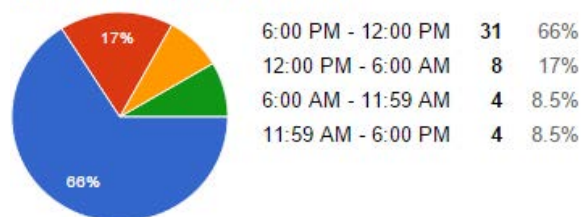


Figura 7. Respuesta sobre franjas horarias de delincuencia

Fuente: elaboración propia.

En el caso de dónde se realizan este tipo de delitos se encuentra que en un 68,9% es preferido realizar en vías públicas, seguido de las residencias con un 17,8% (Figura 8).

Para los encuestados es importante denunciar, aunque existe una percepción de 24,4%, que considera que estos actos se quedan impunes o que no existe tiempo suficiente para realizar las denuncias pertinentes (Figura 9).

Los sitios donde más se ven este tipo de delitos es:

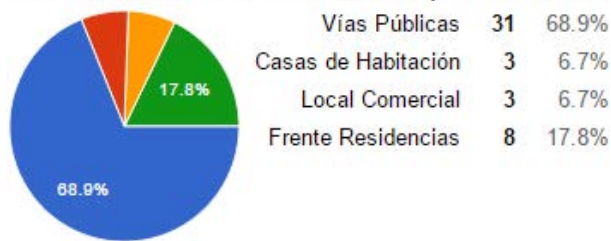


Figura 8. RTA sobre sitios donde se evidencia mayor la delincuencia

Fuente: elaboración propia.

Si usted es víctima de un delito, ¿Usted lo denunciaría?



Si la respuesta a la pregunta anterior fue "No", una de las razones por la que no lo haría es:

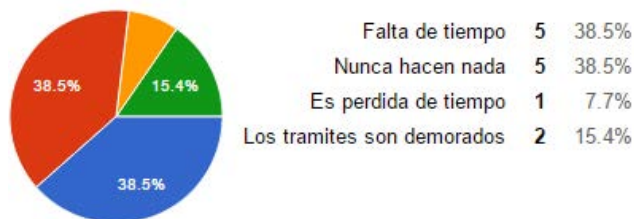


Figura 9. RTA Sobre las víctimas y su reacción de denuncia ante el robo

Fuente: elaboración propia.

Refinamiento y transformación de los datos

Después de obtener una serie de datos correspondientes a la delincuencia en la ciudad de Bogotá, se procedió reescribir los datos de forma cualitativa, para poder aplicar el algoritmo k-means. Para reescribir los datos se tomaron en cuenta las equivalencias presentadas a continuación.

- Atributo seguridad dada por gobierno: buena=3; regular=2; mala=1.
- Atributo localidades de Bogotá: Usaquén=1; Chapinero=2; Santa Fe=3; San Cristóbal=4; Usme=5; Tunjuelito=6; Bosa=7; Ciudad Kennedy=8; Fontibón=9; Engativá=10; Suba=11; Barrios Unidos=12; Teusaquillo=13; Los Mártires=14; Antonio Nariño=15; Puente Aranda=16; Candelaria=17; Rafael Uribe=18; Ciudad Bolívar=19; Sumapaz=20.
- Atributo percepción de seguridad en la localidad: mejorado=4; sigue igual=3; ha venido en decadencia=2; es pésima=1.
- Atributo delincuencia más común en la localidad: violación=1; robo de vehículos=2; asalto a transeúntes=3; robo a casas habitación=4; robo de bienes y artículos menores=5; grafitis y pinta de muros y monumentos=6; carterismo=7; vandalismo=8; otras=9.
- Atributo frecuencia de delincuencia: alta frecuencia=5; baja frecuencia=1.
- Atributo franja horaria (se comete delincuencia): 6:00 pm-12:00 pm=1; 11:59 am-6:00 pm=2; 12:00 pm-6:00 am=3; 6:00 am-11:59 am=4.
- Atributo lugares delictivos: vías públicas=1; casas de habitación=2; local comercial=3; frente residencias=4.
- Atributo causas de delitos: el desempleo=1; la falta de oportunidades educativas=2; la pobreza=3; la falta de castigo a quienes cometen delitos=4; la drogadicción=5; el mal funcionamiento de la justicia=6; falta de programas para los jóvenes=7; el narcotráfico=8; las diferencias entre ricos y pobres=9; la escasez de policías=10.

- Atributo denunciante: sí=1; no=0.
- Atributo razones de los no denunciantes: falta de tiempo=1; nunca hacen nada=2; es pérdida de tiempo=3; los tramites son demorados=4; no aplica=0;

Posterior a esta transformación, se registraron todos los datos en un archivo .arff (Figura 10), el cual es un tipo de formato propio de la herramienta Weka, y que será utilizado para la implementación del mismo en la herramienta.

```
@RELATION delincuencia_Bogota

@ATTRIBUTE seguridad_DadaPor_Gobierno numeric
@ATTRIBUTE localidades_Bogota numeric
@ATTRIBUTE percepcion_Seguridad_Localidad numeric
@ATTRIBUTE delincuencia_MasComun_Localidad numeric
@ATTRIBUTE frecuencia_Delincuencia numeric
@ATTRIBUTE franja_Horaria numeric
@ATTRIBUTE lugares_delictivos numeric
@ATTRIBUTE causas_delitos numeric
@ATTRIBUTE denunciantes numeric
@ATTRIBUTE razones_No_Denunciantes numeric

@DATA
2,19,2,3,5,1,1,4,0,2
1,19,1,4,4,3,2,4,1,0
2,7,3,5,3,1,1,4,0,1
1,11,3,3,4,1,1,6,1,0
1,11,3,3,4,1,1,9,1,0
1,11,2,3,4,1,1,3,1,0
1,11,2,3,4,1,1,3,1,0
1,11,2,3,4,1,1,3,1,0
2,10,3,5,2,1,1,6,1,0
2,10,3,5,3,1,1,4,1,0
1,11,3,5,3,1,4,5,0,2
2,6,3,5,3,1,2,7,1,0
1,7,2,3,5,1,1,4,0,2
2,5,3,3,4,1,1,4,1,0
3,10,3,8,4,3,1,5,1,0
1,11,2,6,4,3,4,5,0,2
2,18,2,3,2,1,1,4,0,1
2,7,4,3,3,1,1,7,1,0
2,8,2,3,3,1,4,9,1,0
2,8,2,3,3,1,4,9,1,0
2,8,2,3,5,1,4,6,1,0
1,7,2,5,4,2,1,1,0,1
1,11,2,3,4,1,1,4,1,0
1,16,2,3,4,1,1,6,1,0
1,7,2,5,4,4,1,4,1,0
1,19,3,3,4,1,1,6,0,4
2,10,3,3,3,2,4,5,1,3
1,1,2,3,4,1,1,4,1,0
1,8,1,3,3,4,1,2,1,0
2,7,4,3,4,1,1,5,1,0
1,7,2,5,4,2,1,1,0,1
3,14,1,8,3,1,3,6,0,2
2,1,3,3,3,3,3,8,1,0
2,1,3,3,3,3,3,8,1,0
2,19,3,6,4,1,1,2,1,0
1,14,3,4,4,4,4,4,1,0
1,5,2,9,5,1,1,10,1,0
2,10,3,3,3,1,1,2,1,0
1,4,2,3,3,1,1,4,1,0
1,16,1,3,5,3,1,4,0,4
2,9,2,2,3,1,2,3,1,1
1,5,2,3,4,2,1,6,1,0
2,8,4,5,2,1,4,5,1,0
3,18,3,8,3,4,1,4,1,0
2,7,3,3,5,1,1,4,1,0
```

Figura 10. Datos en Formato .arff.

Fuente: elaboración propia.

Estructura de clustering

Como ya se había mencionado anteriormente, se utilizó la herramienta Weka, junto al algoritmo k-means, para realizar *clustering* de los datos. Se seleccionaron cinco *clusters* para hacer la clasificación de los datos y obtener la información adecuada (Figura 11).

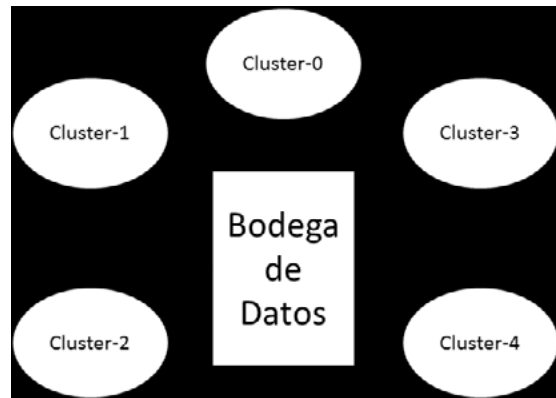


Figura 11. Esquema de clustering para los datos del modelo

Fuente: elaboración propia.

Resultados obtenidos por clustering

Los resultados de la simulación arrojaron la clasificación de los datos de la investigación realizada, observados en la Figura 1, en la cual se evidencia que los *clusters* que mayor información agrupó fueron los *clusters* 4, 3 y 0, respectivamente de mayor a menor. En la Figura 2 se puede analizar detenidamente cuál fue el valor resultante que se obtuvo con k-means para cada uno de los atributos analizados en el proceso de *clustering*.

Clustered Instances

0	9 (20%)
1	1 (2%)
2	3 (7%)
3	11 (24%)
4	21 (47%)

Figura 12. Clasificación de los datos en los cluster por parte de k-means y Weka

Fuente: elaboración propia.

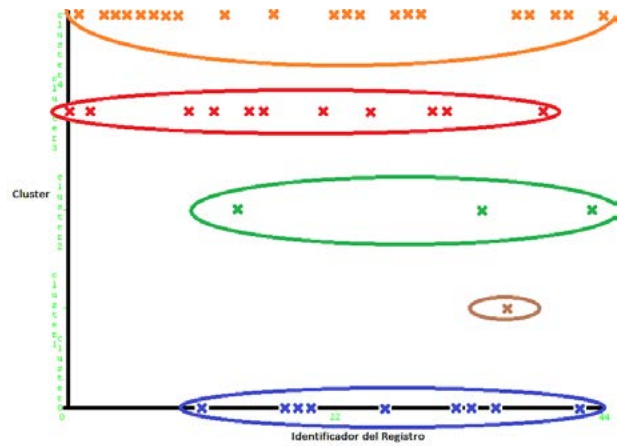


Figura 14. Gráfica de clustering por parte de k-means y Weka

Fuente: elaboración propia.

Number of iterations: 3
 Within cluster sum of squared errors: 21.087159432208562
 Missing values globally replaced with mean/mode

Cluster centroids:

Attribute	Full Data (45)	Cluster# 0 (9)	1 (1)	2 (3)	3 (11)	4 (21)
seguridad_DadaPor_Gobierno	1.5778	1.8889	1	2.6667	1.4545	1.381
localidades_Bogota	9.8444	7.1111	5	15.6667	12.3636	9.0952
percepcion_Seguridad_Localidad	2.4444	2.7778	2	3	2.0909	2.4286
delincuencia_MasComun_Localidad	4.0222	3.5556	9	7.3333	4.4545	3.2857
frecuencia_Delincuencia	3.6444	3.2222	5	3.6667	3.0102	3.6667
franja_Horaria	1.6222	1.8889	1	2.6667	1.5455	1.4286
lugares_delictivos	1.7333	3.5556	1	1	1.7273	1.0952
causas_delitos	4.8667	6.7778	10	3.6667	4	4.4286
denunciantes	0.7556	1	1	1	0	1
razones_No_Denunciantes	0.5778	0.3333	0	0	2	0.0476

Figura 13. Resultados de clustering por parte de k-means y Weka

Fuente: elaboración propia.

Una gráfica más representativa de los resultados obtenidos (Figura 12 y Figura 13) por medio de k-means se puede ver en el plano cartesiano de la Figura 14; el eje X corresponde al identificador (ID) de cada fila de datos analizados en la simulación, el eje Y corresponde a cada uno de los cinco clusters en los cuales la información fue repartida dado su comportamiento.

ANÁLISIS DE RESULTADOS

En la siguiente sección se analizarán los datos resultantes obtenidos atrás, con base al en el criterio de investigación de los autores; adicionalmente el énfasis estará centrado en los tres cluster con mayor cantidad de agrupamientos registrados en los resultados.

Para el cluster cero, del cual se obtuvieron nueve agrupaciones de registros (Figura 12), se pudo obtener la siguiente información relevante: para la ciudadanía no es suficiente la seguridad que se está brindando por parte del Estado, se percibe que no ha cambiado nada las políticas de seguridad ciudadana con respecto a otros gobiernos anteriores, a tal punto de evidenciarse una decadencia en vez de una mejoría. A partir de esto se podría decir que una solución para las problemáticas de delincuencia en la ciudad de Bogotá, podrían abordarse a través de una modificación de las actuales políticas de seguridad ciudadana, las cuales incluyen un gran número de factores económicos, sociales, culturales, entre otros, a los cuales se les puede dar cierta

importancia basándose en los datos obtenidos por los siguientes *clusters*.

Para el *cluster* tres, del cual se obtuvieron once agrupaciones de registros (Figura 12), se pudo obtener la siguiente información relevante: los patrones de robo de los delincuentes se evidencian en las horas de la noche (6:00 pm-12:00 pm), razón por la cual es más fácil para ellos cometer delitos como asalto a transeúntes, robo de locales, etc.; dada las condiciones en las que se genera el hurto, se puede descartar que dichas personas realicen estos actos delictivos por falta de educación, empleo o situaciones de pobreza, que aunque son factores que pueden promover estas conductas, no tienen el suficiente peso para ser el causante dado los resultados analizados hasta el momento. Una solución que se puede plantear siguiendo el análisis anterior es la de reforzar las zonas afectadas por estos actos delictivos en los horarios de mayor incidencia.

Por último, para el *cluster* cuatro, del cual se obtuvieron veintiún agrupaciones de registros (Figura 12), se pudo obtener la siguiente información relevante: existen diferentes circunstancias que pueden causar que una persona realice actos delictivos en una sociedad, como las ya mencionadas hasta el momento, pero basándose en los resultados de agrupamiento de este *cluster* se podrían atribuir directamente a la falta de un sistema judicial lo suficientemente rígido que castigue debidamente a los delincuentes. Un sistema de estas características le daría la confianza suficiente a una víctima de denunciar su caso, sin importar la demora en procesos y demás, dado que se está teniendo certeza de que realmente se lograra algo con esa denuncia. A partir de esto se puede sugerir un cambio en el sistema judicial que permita aumentar las penas y multas dadas a los delincuentes.

Con base en cada uno de los análisis realizados por separado para cada *cluster*, se puede obtener una solución común para la delincuencia en Bogotá, basándose en los datos recolectados por la encuesta realizada a los ciudadanos de esta.

Aunque existen diferentes factores que pueden fragmentar una sociedad a tal punto de generar actos delictivos como los que en este paper se analizaron, uno que repercutió más fue las deficiencias del sistema judicial del país. Una solución que posiblemente se podría plantear para este problema sería un cambio en las leyes que rigen el país, las cuales le permitan al ente judicial poder ser más rígidos con los castigos que apliquen a estos delincuentes, lo anterior permitiría que las víctimas fuesen capaces de denunciar los casos de los cuales ellos fueron partícipes y así no seguir dejando impune este tipo de acciones, pues sigue decayendo cada vez más el sistema de seguridad ciudadano.

CONCLUSIONES

La implementación de técnicas de la minería de datos, como el *clustering*, permite identificar focos que tienen características similares para los cuales se pueden aplicar criterios de la misma magnitud, afectando solo un grupo específico.

Weka, además de contar con el algoritmo básico para *clustering* (k-means) es una potente herramienta para la explotación de datos con características similares a otras herramientas de venta comercial, teniendo una ventaja al ser de acceso libre y completamente gratuita.

Al aplicar el modelo de *clustering* para hacer minería de datos, se pueden obtener diferentes patrones para cada conjunto de datos, esto dependiendo de cada *cluster*. Cada resultado en la mayoría de ocasiones converge a un mismo punto con respecto a los demás, este punto suele contener la información más relevante en el proceso de *datamining*.

El objetivo principal al realizar minería de datos es encontrar una serie de patrones no evidentes en la información recolectada, esto no implica la exclusión de información en la respuesta del modelo, como por ejemplo las causas que generan la delincuencia.

REFERENCIAS

- [1] Rivas, A. (s.f.). Una década de políticas de seguridad ciudadana en Colombia. Fundación Seguridad y Democracia. Recuperado de <http://pdba.georgetown.edu/Security/citizenssecurity/Colombia/evaluaciones/decadapoliticas.pdf>
- [2] Política Nacional de Seguridad y Convivencia Ciudadana. (2011). Alta Consejería Presidencial para la Convivencia y la Seguridad Ciudadana. Departamento Nacional de Planeación Dirección de Justicia, Seguridad y Gobierno. Recuperado de <https://colaboracion.dnp.gov.co/CDT/Justicia%20Seguridad%20y%20Gobierno/PNSCC%20FINAL%20AGO%202011.pdf>
- [3] Arriagada, I. y Godoy, L. (1999). Seguridad ciudadana y violencia en América Latina: diagnóstico y políticas en los años noventa. *Políticas Sociales*, 32. Recuperado de https://repositorio.cepal.org/bitstream/handle/11362/6263/S998552_es.pdf
- [4] El Tiempo (s.f.). Inseguridad en Bogotá. *El Tiempo*. Recuperado de <http://www.eltiempo.com/noticias/inseguridad-en-bogota>
- [5] Aluja, T. (2001). La minería de datos, entre la estadística y la inteligencia artificial. *Qüestió*, 25(3). 479-498.
- [6] Virseda, F. y Román, J. (s.f.). Minería de datos y aplicaciones. Madrid: Universidad Carlos III.
- [7] Vallejo, D. y Tenelanda, G. (2012). Minería de datos aplicada en detección de intrusos. *Ingenierías USB-Med*, 3(1), 50-61. Recuperado de <http://web.usb-med.edu.co/usbmed/fing/v3n1/v3n1a6.pdf>
- [8] Perversi, I. (2007) *Aplicación de minería de datos para la exploración y detección de patrones delictivos en Argentina*. (Trabajo de grado). Buenos Aires: Instituto Tecnológico de Buenos Aires.
- [9] Jiménez, J. (s.f.). Análisis de datos. Madrid: Universidad Carlos III.
- [10] Wanummen, L. (2010). Minería de datos para la predicción. *Vínculos*, 7(2), 44-57.