

# **IDENTIFICACIÓN DE PATRONES DELICTIVOS EN COLOMBIA DURANTE EL PERIODO 2010-2016 MEDIANTE EL USO DE TÉCNICAS DE MINERÍA DE DATOS**

**Nicolás Aguirre Yacup, María Antonia Walteros Alcázar, Yimi Mauricio Noguera Pérez, Sandra Patricia Castillo Landínez**

**Corporación Universitaria Autónoma del Cauca  
Popayán, Colombia**

## **Resumen**

Los estudios realizados por diferentes organizaciones muestran que la percepción de inseguridad ha aumentado en gran parte de las ciudades del país, razón por la cual es de vital importancia identificar las zonas vulnerables donde mayor número de delitos se presentan y las características generales de éstos, a fin de que las autoridades gubernamentales planteen medidas de prevención y la fuerza pública optimice sus recursos humanos y financieros en la atención de situaciones que ponen en riesgo la seguridad de los ciudadanos.

Este trabajo está enmarcado en el proyecto de investigación "Identificación de patrones en datasets gubernamentales: caso de estudio hurtos y accidentes de tránsito en Colombia", el cual inició en septiembre de 2017, su propósito ha sido analizar los datos publicados por la Red Nacional de Observatorios del delito de la Policía Nacional en el periodo comprendido entre 2010 y 2016, que involucra el hurto de vehículos, celulares, motocicletas y abusos sexuales. Se utilizaron técnicas de minería de datos con el objetivo de identificar patrones característicos de estos actos delictivos en Colombia; de otra parte, se realizó un análisis visual que permitió conocer la situación de Popayán (Cauca) durante el mismo periodo en materia de hurtos.

**Palabras clave:** minería de datos; análisis visual; delitos

## Abstract

*The studies made by different organizations show that the perception of insecurity has increased in a large part of the country's cities, Therefore, it is vital to identify the vulnerable areas where the highest number of crimes are presented and the general characteristics of these, so that the government authorities can propose preventive measures and the law enforcements optimize their human resources and in attention to situations that they put in danger the safety of citizens.*

*This paper is embedded in the research project "Identification of patterns in government datasets: case of study thefts and traffic accidents in Colombia", which began in September 2017, its purpose has been to analyze the data published by the National Network of Observatories of the crime from the National Police in the period between 2010 and 2016, which involves the theft of vehicles, cell phones, motorcycles and sexual abuse.*

**Keywords:** *data mining; crime pattern; data analysis*

## 1. Introducción

La Encuesta de Convivencia y Seguridad Ciudadana de 2017 realizada por el DANE (DANE, 2017), basada en datos recopilados entre los años 2015 y 2016 en 28 ciudades en Colombia y en la cual se entrevistaron personas mayores de 15 años, presentó cifras bastante preocupantes:

- El 62.7% de la población se siente insegura.
- El 64.5% de las mujeres informaron que se consideran inseguras en su ciudad, frente al 60.7% de los hombres.
- Durante 2016, las víctimas de hurto reportaron que los objetos personales más robados fueron el teléfono celular (75.5%).
- El 24,2% indicó ser propietaria de al menos un vehículo (automóvil, motocicleta o bicicleta) durante 2016; de los cuales el 11.5% reveló haber sufrido hurto a la totalidad, una parte o accesorio de su vehículo.

En el caso de la Ciudad de Popayán, el Coronel Edgar Cárdenas encargado de la Policía Metropolitana de la ciudad, reveló que en el 2017 de acuerdo a las cifras de la Fiscalía, la inseguridad disminuyó en un 20% con respecto al año 2016, sin embargo la problemática aún está presente y afecta a un número significativo de ciudadanos.

En relación con delitos sexuales según cifras del Instituto de Medicina Legal, durante enero y abril de 2018 se han registrado 7173 casos de abuso sexual mientras que en el 2017 fueron 5831 para este mismo periodo; el grupo más afectado son los menores entre 10 a 13 años con 2.286 casos para el 2017 y 2.669 para el 2018 (Instituto Nacional de Medicina Legal y Forense, 2017).

De acuerdo con lo anterior, la percepción de inseguridad entre los ciudadanos presenta un nivel significativamente alto y es necesario que los entes gubernamentales y demás entidades

relacionadas con el tema, tomen cartas en el asunto y diseñen planes de contención para disminuir los índices de criminalidad y establecer mecanismos orientados a ofrecer una respuesta eficaz a la sociedad en general.

Para tal fin se hace necesario el análisis detallado y pormenorizado de los datos existentes y reportados por diferentes entidades relacionados con hechos delictivos con el propósito de identificar patrones de criminalidad. Para este caso, el volumen de datos a considerar es significativo y el procesamiento mediante técnicas tradicionales resulta ineficiente, razón por la cual se hace necesario el uso de técnicas de minería de datos para obtener productos relevantes y significativos que ayuden a las autoridades competentes a proponer, adoptar e implementar políticas y planes efectivos.

## **2. Minería de datos**

La sociedad moderna genera continuamente gran cantidad de información digital, se calcula que en promedio cada minuto se envían más de 200 millones de correos electrónicos, se realizan 2 millones de consultas a Google, se suben 48 horas de vídeo a YouTube, se escriben más de 100.000 mensajes en Twitter, se publican casi 30.000 nuevos artículos en sitios como Tumblr o WordPress y suben más de 6.000 fotografías a Instagram y Flickr (Beneyto, 2013).

Gran parte de estos datos simplemente se almacenan y no se aprovechan para crear productos de interés y que generen valor, en ocasiones, por el esfuerzo y el costo que implica el tratamiento de grandes datasets; la minería de datos surge entonces como un conjunto máquina-hombre que puede recibir información de cualquier tipo (texto, documentos, imágenes, videos, links) la cual será procesada usando algoritmos avanzados de análisis de datos con el fin de identificar nuevas relaciones, tendencias o patrones (Pérez, 2007); para tal fin se requiere el uso técnicas y tecnologías robustas que permitan procesar grandes volúmenes de datos de una forma relativamente sencilla y eficiente. En la actualidad, la minería de datos es una herramienta usada en diversos campos dada su gran utilidad para soportar la toma de decisiones, crear modelos de alta confiabilidad, ahorrar costos y abrir nuevas oportunidades de negocio (Yelitza & Rosalba, 2007). Es necesario mencionar que la minería de es una etapa dentro del proceso KDD.

El proceso KDD (por sus siglas en inglés Knowledge Discovery in Databases) se lleva a cabo justo con la minería de datos y permite generar conocimiento e información nueva que puede resultar potencialmente útil, cabe resaltar que este proceso es altamente arduo y complejo ya que los volúmenes de datos que se procesan tienden a ser bastantes robustos con el fin de extraer información de calidad que luego pueda ser usada (Ilanda, 2018). Las etapas del proceso KDD se muestran en la Figura 1.

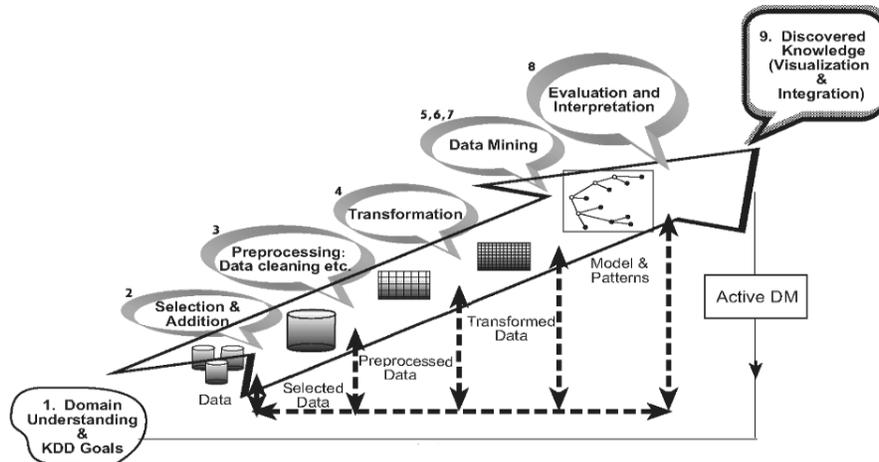


Figura 1. Proceso KDD. Fuente (Ianda, 2018)

Para el desarrollo del proyecto se utilizó la técnica de agrupamiento o clustering, la cual consiste en generar agrupamientos de tal forma que los objetos de un grupo sean similares entre sí y diferentes de los objetos de otros grupos (Benítez & Diez, 2005); en otros términos, un cluster es una colección de objetos de datos que son similares a otros dentro del mismo grupo y son distintos a los objetos de otros clusters (Farias, Durán, & Figueroa, 2008). Se buscó identificar características similares en los hurtos de teléfonos celulares, vehículos y motocicletas que fueron reportados durante el periodo 2010-2016.

### 3. Materiales y métodos

Para la realización del trabajo se usaron los datos reportados por el Observatorio del Delito de la Policía Nacional ([www.policia.gov.co/observatoriodeldelito](http://www.policia.gov.co/observatoriodeldelito)) referentes a hurtos a celulares, vehículos y motocicletas y delitos sexuales. El dataset inicial contaba con 21 atributos (Fecha, Departamento, Municipio, Día, Hora, Barrio, Zona, Clase Sitio, Arma Empleada, Móvil Agresor, Edad, Género, Móvil Víctima, Escolaridad, Profesión, Estado Civil, Clase de bien, Modelo, Color, Marca del bien, Línea) y 402.631 registros; uno de los mayores inconvenientes fue la estructuración y adecuación de los datos, ya que en muchas ocasiones para el mismo valor de un atributo se encontraron diferentes codificaciones, asociadas muy seguramente a problemas de digitación y falta de homogeneidad.

El procesamiento de los datos se ejecutó por fases, tomando como base la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) (Román, 2016). A continuación, se realiza una breve descripción:

- Fase 1: Alcance del problema

En esta fase se realiza una búsqueda de información para lograr apreciar mejor la problemática, teniendo como hallazgo la alta percepción de inseguridad que fue uno de los motivos por los cuales se tomó la decisión de realizar el presente proyecto. Se desarrollaron las siguientes actividades:

- Revisión bibliográfica.

- Valoración de la situación

- Fase 2: comprensión de los datos

En esta fase se da inicio a la búsqueda y recolección de los datos, también se aseguró la integridad de los mismos. Se desarrollaron las siguientes actividades:

- Recolección de datos iniciales
- Exploración de datos
- Verificación de la calidad de los datos

- Fase 3: preparación de los datos

En esta fase se aplicaron diferentes tipos de filtrados para la depuración de los datos y eliminación de atributos sin peso. Se desarrollaron las siguientes actividades:

- Selección de los datos
- Limpieza de los datos
- Integración de los datos
- Ajuste del formato de los datos

- Fase 4: análisis visual

En esta fase se realiza un análisis de los datos la cual se orienta a crear una representación gráfica de los mismos con el fin de facilitar su comprensión y facilitar la identificación de patrones. Se usó ArcGIS para la geo referenciación de puntos correspondientes a delitos en la ciudad de Popayán para realizar un análisis visual e identificar las zonas con mayor problemática. Se desarrollaron las siguientes actividades:

- Definir el tipo de visualizaciones a utilizar
- Construir la visualización de los datos
- Analizar las visualizaciones

- Fase 5: generación de cluster

El algoritmo de agrupamiento o clustering utilizado fue k-modes desarrollado por Zhexue Huang, el cual permite utilizar datos cualitativos, reemplazando las medias por modos y usa un método basado en frecuencia que actualiza los modos, teniendo como objetivo la sumatorio de todas las distancias de todos los objetos a todos los modos, habiendo tantos modos como clusters (HUANG, 1998). Las actividades que se realizaron fueron:

- Construcción del modelo
- Validación
- Análisis de resultados

## **4. Resultados**

- **Resultado del análisis visual**

Se identificaron los barrios de Popayán en los cuales se presentó el mayor número de hurtos de celulares, vehículos y motocicletas durante el periodo 2010 – 2016: El Centro, La Esmeralda, Alfonso López, El Barrio Bolívar y el Barrio Valencia (Figura 2).

**IDENTIFICACIÓN DE PATRONES DELICTIVOS EN COLOMBIA DURANTE EL PERIODO 2010-2016  
MEDIANTE EL USO DE TÉCNICAS DE MINERÍA DE DATOS**

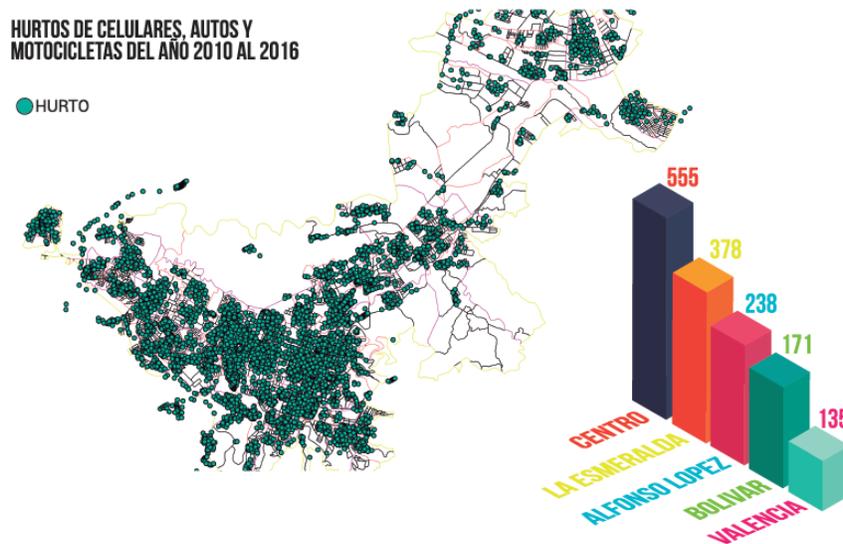


Figura 2. Hurtos de celulares, vehículos y motocicletas en Popayán en el periodo 2010 – 2016. Fuente propia

Se realizó un análisis diferenciando el sexo de las personas afectadas y se encontró que el número de mujeres que han sido víctimas de robos aumentó a lo largo del periodo de estudio; por otro lado, los hurtos a motocicletas y vehículos se presentan más en los hombres, mientras que a las mujeres les roban más el celular (ver Figura 3 y 4).

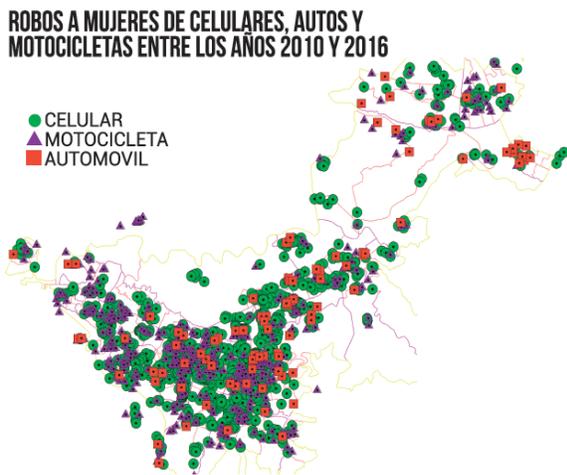


Figura 3. Robo de celulares, autos y motocicletas a mujeres en Popayán entre 2010 y 2016. Fuente propia

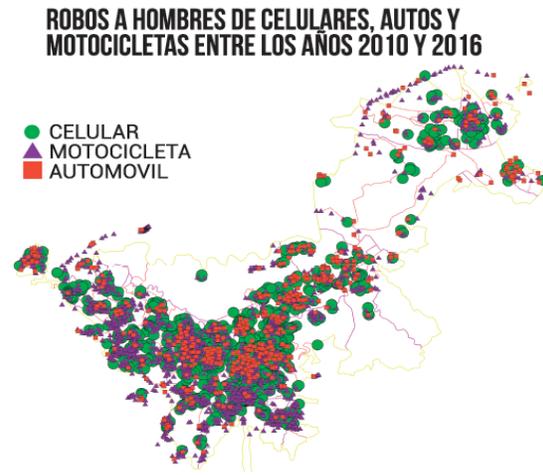


Figura 4. Robo de celulares, autos y motocicletas a hombres entre 2010 y 2016. Fuente propia

En cuanto a los delitos sexuales reportados entre 2010 y 2016, se contaba con 79.536 registros relacionados con acceso carnal violento, acto sexual violento, acceso carnal o acto sexual en persona puesta en incapacidad de resistir, acceso carnal abusivo con menor de 14 años, actos sexuales con menor de 14 años, acoso sexual, acceso carnal o acto sexual abusivo con incapaz de resistir, los anteriores con y sin agravantes, constreñimiento a la prostitución, inducción a la prostitución, demanda de explotación sexual comercial de persona menor de 18 años de edad,

estímulo a la prostitución de menores, pornografía con menores, utilización o facilitación de medios de comunicación para ofrecer servicios sexuales de menores y omisión de denuncia. Los departamentos con mayor incidencia de acuerdo al número de casos son Cundinamarca (10.720), Valle del Cauca (10.207), Santander (7.895) y Antioquia (5.809), en el mapa mostrado en la Figura 5 se aprecia como el tamaño de los puntos está directamente relacionado con el número de casos denunciados.

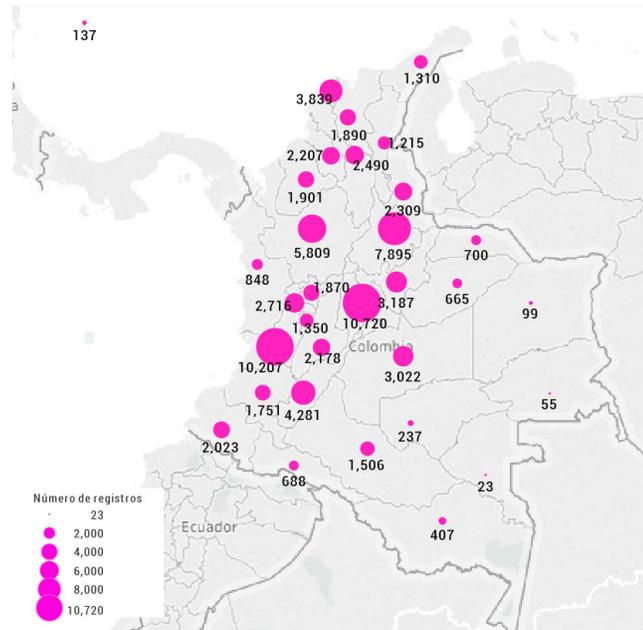


Figura 5. Incidencia de delitos sexuales en Colombia entre 2010 y 2016. Fuente propia

### • Resultados del análisis de los clusters

Se construyeron dos modelos, uno con tres grupos y otro con cinco. A continuación se describe cada uno:

#### 3 clusters ( $k=3$ )

- **Cluster 1:** formado por 151.383 registros; el mayor número de hurtos se presentó durante 2015 en el departamento de Cundinamarca los días viernes en horas de la tarde, en casa de habitaciones rentadas. Las víctimas fueron mujeres, solteras, con un promedio de edad de 34 años, escolaridad máxima hasta secundaria, en su mayoría empleadas independientes que se movilizaban a pie y a quienes les robaron móviles celulares, por su parte el asaltante no portaba armas de fuego y se movilizaba a pie.
- **Cluster 2:** constituido por 134.861 registros. La mayoría de estos delitos ocurrieron entre 2013 y 2014, en la ciudad de Bogotá el día miércoles en horas de la noche, en vías públicas. Las víctimas más comunes fueron hombres solteros, con edad promedio 31 años, máximo nivel educativo secundaria, se movilizaban a pie; sin embargo, el objeto de hurto más común fueron motocicletas. El asaltante se trasladaba a pie, no portaba armas, lo cual puede indicar que la mayoría de los robos se han presentado a motocicletas parqueadas o dejadas en vías públicas.

- **Cluster 3:** formado por 105.192 registros, ocurridos en 2012 y 2015 en el departamento de Antioquia. Los robos registrados en este cluster fueron realizado los día Domingo, el objeto de hurto fueron motocicletas en la vía pública conducidas por la víctima. El perpetrador se caracteriza por movilizarse a pie y el hurto es realizado intimidando o hiriendo a la víctima con arma de fuego.

*5 clusters (k=5)*

- **Cluster 1:** integrado por 98.476 registros. Los hurtos se presentaron durante el 2015 en el departamento de Antioquia, en especial en la ciudad de Medellín, durante los días jueves en la noche, El objeto de robo más común son las motocicletas; estas fueron hurtadas principalmente a hombres solteros de 31 años de edad en promedio con un máximo nivel escolar universitario; los cuales eran propietarios de la moto hurtada. El asaltante se trasladaba a pie portando un arma de fuego.
- **Cluster 2:** constituido por unos 66.729 registros, realizados en su mayoría en los años 2012 y 2015, en el departamento del Valle de Cauca y más específicamente en la ciudad de Santiago de Cali. Los días martes y sábados en horas de la mañana a hombres de 35 años en promedio con un nivel escolar máximo secundaria cuyos; los cuales son empleados particulares que viven en unión libre y eran propietarios de la motocicleta hurta. De forma similar, el asaltante se trasladaba a pie portando un arma de fuego; con la cual intimidaban o herían a la víctima.
- **Cluster 3:** Integrado por 71.430 registros de hurtos de celulares (Samsung y Blackberry), en su mayoría, presentados comunmente en el año 2013, en el departamento del Valle del Cauca, y en la zona del centro de la ciudad de Santiago de Cali. Los miercoles en la tarde es el momento más común para sufrir este tipo de hurto. Las víctimas se caracterizan por ser mujeres solteras. El asaltante se movilizaba a pie y no portaba ningún tipo de arma.
- **Cluster 4:** constituido por un total de 90.163 registros de hurtos de celulares (Samsung), durante 2013 y 2015 en el departamento de Cundinamarca y en especial en Bogotá, los días viernes en la tarde a hombres casados, empleados particulares, entre 51 y 52 años de edad en promedio y una escolaridad máxima en secundaria. El asaltante se trasladaba a pie y no portaba armas.
- **Cluster 5:** está formado por un total de 90.163 registros de hurtos a celulares, cuya víctima son mujeres solteras (empleadas particulares) con un promedio de edad de 26 años con máximo nivel escolar en secundaria. El perpetrador se movilizaba a pie y no portaba armas. Estos delitos ocurrieron durante el 2015, los miercoles en la tarde.

## **5. Conclusiones**

- Se requiere una mejor organización de los atributos o variables recopilados por parte de la Red Nacional de Observatorios del delito de la Policía Nacional lo cual permitirá obtener mejores resultados de análisis en futuros trabajos. Específicamente se habla de estandarización de los datos mediante el uso de formatos adecuados, herramientas eficientes de captura y validación de la información

- La minería de datos tiene un amplio campo de aplicación y pueden proponer soluciones eficientes para las organizaciones, sin embargo, la calidad de los datos es fundamental para generar resultados óptimos.
- Como trabajo futuro, se requiere incorporar información complementaria de tipo socio-económico para realizar un análisis más completo de las zonas más vulnerables en el país cuanto a hurtos de celulares, vehículos, motocicletas y delitos sexuales

## **6. Referencias**

### **Artículos de revistas**

- Benítez, I., & Diez, J. L. (septiembre de 2005). [www.researchgate.net](https://www.researchgate.net/profile/Ignacio_Benitez/publication/239526131_Tecnicas_de_Agrupamiento_para_el_Analisis_de_Datos_Cuantitativos_y_Cualitativos/links/00b7d51c15cca2cb1f000000/Tecnicas-de-Agrupamiento-para-el-Analisis-de-Datos-Cuantitativos-y-Cu). Obtenido de [https://www.researchgate.net/profile/Ignacio\\_Benitez/publication/239526131\\_Tecnicas\\_de\\_Agrupamiento\\_para\\_el\\_Analisis\\_de\\_Datos\\_Cuantitativos\\_y\\_Cualitativos/links/00b7d51c15cca2cb1f000000/Tecnicas-de-Agrupamiento-para-el-Analisis-de-Datos-Cuantitativos-y-Cu](https://www.researchgate.net/profile/Ignacio_Benitez/publication/239526131_Tecnicas_de_Agrupamiento_para_el_Analisis_de_Datos_Cuantitativos_y_Cualitativos/links/00b7d51c15cca2cb1f000000/Tecnicas-de-Agrupamiento-para-el-Analisis-de-Datos-Cuantitativos-y-Cu)
- Farias, R. A., Durán, E. B., & Figueroa, S. G. (2008). sedici. Obtenido de [http://sedici.unlp.edu.ar/bitstream/handle/10915/21990/Documento\\_completo.pdf?sequence=1](http://sedici.unlp.edu.ar/bitstream/handle/10915/21990/Documento_completo.pdf?sequence=1)
- Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering. Kluwer Academic Publishers., 22.
- Pérez, C. (2007). Minería de datos: técnicas y herramientas. Editorial Paraninfo, pp.789
- Román, J. V. (8 de agosto de 2016). singular data & analytics. Obtenido de <https://data.sngular.team/es/art/25/crisp-dm-la-metodologia-para-poner-orden-en-los-proyectos-de-data-science>
- Yelitza, J. M., & Rosalba, T. (enero de 2007). Minería de Datos como soporte a la toma de decisiones empresariales. Obtenido de [http://www.scielo.org.ve/scielo.php?pid=S1012-15872007000100008&script=sci\\_arttext](http://www.scielo.org.ve/scielo.php?pid=S1012-15872007000100008&script=sci_arttext)

### **Fuentes Electrónicas**

- Beneyto, R. G. (16 de septiembre de 2013). [documania20.wordpress.com](https://documania20.wordpress.com). Obtenido de <https://documania20.wordpress.com/2013/09/16/cuanta-informacion-se-genera-y-almacena-en-el-mundo/>
- DANE. (12 de 2017). [www.dane.gov.co](http://www.dane.gov.co). Obtenido de [https://www.dane.gov.co/files/investigaciones/poblacion/convivencia/2017/Presentacion\\_ECSC\\_2017.pdf](https://www.dane.gov.co/files/investigaciones/poblacion/convivencia/2017/Presentacion_ECSC_2017.pdf)
- Diario El Tiempo. (6 de mayo de 2018). Diario El Tiempo. Obtenido de <http://www.eltiempo.com/justicia/delitos/violacion-sexual-de-menores-en-colombia-214142>
- Landa, j. (2018). <http://fcojlanda.me>. Obtenido de <http://fcojlanda.me/es/sin-categorias/kdd-y-mineria-de-datos-espanol/>
- Instituto Nacional de Medicina Legal y Forense. (1 de junio de 2017). Instituto Nacional de Medicina Legal y Forense. Obtenido de [http://www.medicinalegal.gov.co/noticias/-/asset\\_publisher/vLcVEedo8qgD/content/inmlcf-da-a-conocer-cifras-sobre-violencia-en-contra-de-](http://www.medicinalegal.gov.co/noticias/-/asset_publisher/vLcVEedo8qgD/content/inmlcf-da-a-conocer-cifras-sobre-violencia-en-contra-de-)

lasmujeres?\_com\_liferay\_asset\_publisher\_web\_portlet\_AssetPublisherPortlet\_INSTANCE\_vLcVE  
edo8qgD\_redirect=http%3A%2F%2F

## Sobre los autores

- **Nicolás Aguirre Yacup:** Estudiante de 10 semestre de Ingeniería de Sistemas Informáticos, miembro del Semillero de Investigación en Minería de Datos (SIMD). [nicolas.aguirre.y@uniautonoma.edu.co](mailto:nicolas.aguirre.y@uniautonoma.edu.co)
- **María Antonia Walteros Alcázar:** Estudiante de 10 semestre de Ingeniería de Sistemas Informáticos, miembro del Semillero de Investigación en Minería de Datos (SIMD). [maria.walteros.a@uniautonoma.edu.co](mailto:maria.walteros.a@uniautonoma.edu.co)
- **Yimi Mauricio Noguera Pérez:** Estudiante de 9 semestre de Ingeniería de Sistemas Informáticos, miembro del Semillero de Investigación en Minería de Datos (SIMD). [yimi.noguera.p@uniautonoma.edu.co](mailto:yimi.noguera.p@uniautonoma.edu.co)
- **Sandra Patricia Castillo Landínez:** Ingeniera de Sistemas (Universidad Nacional de Colombia), Especialista en Administración de la Información y Bases de Datos (Colegio Mayor del Cauca), Certified Big Data Professional, Certified Big Data Scientist. Docente de la Facultad de Ingeniería, investigadora adscrita al Grupo de Investigación en Tecnología y Ambiente (GITA), coordinadora de la línea de Investigación en Ingeniería de Software y líder del Semillero de Investigación en Minería de Datos (SIMD). [sandra.castillo.l@uniautonoma.edu.co](mailto:sandra.castillo.l@uniautonoma.edu.co)

---

Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería.

Copyright © 2018 Asociación Colombiana de Facultades de Ingeniería (ACOFI)