



UNIVERSIDAD  
PRIVADA  
DEL NORTE

# FACULTAD DE INGENIERÍA

Carrera de Ingeniería de Sistemas Computacionales

“MINERÍA DE DATOS EN LA DETECCIÓN DE  
PATRONES DELICTIVOS: UNA REVISIÓN  
SISTEMÁTICA”

Trabajo de investigación para optar el grado de:

**Bachiller en Ingeniería de Sistemas Computacionales**

**Autor:**

Junior German Balcázar Gonzales

**Asesor:**

Mg. Ing. Pedro Gilmer Castillo Domínguez

Lima - Perú

2019



## DEDICATORIA

A mi madre, por ser la razón principal  
de mi crecimiento profesional y  
por ser el motor que me permite seguir adelante

A mi novia, por ser mi mayor motivación y  
la mejor compañera en la que siempre encuentro  
palabras de aliento.



## **AGRADECIMIENTO**

A mi asesor Ing. Pedro Gilmer Castillo Domínguez,  
quien me acompañó y asesoró durante todo el desarrollo de este proyecto de  
investigación.



## Tabla de contenido

<b>DEDICATORIA.....</b>	<b>2</b>
<b>AGRADECIMIENTO .....</b>	<b>3</b>
<b>CAPÍTULO I. INTRODUCCIÓN .....</b>	<b>5</b>
<b>CAPÍTULO II. METODOLOGÍA.....</b>	<b>7</b>
<b>CAPÍTULO III. RESULTADOS .....</b>	<b>10</b>
<b>CAPÍTULO IV. CONCLUSIONES .....</b>	<b>15</b>
<b>REFERENCIAS .....</b>	<b>18</b>

## CAPÍTULO I. INTRODUCCIÓN

El índice delincencial ha ido aumentando en los últimos años, evidenciándose un deterioro en el proceso de prevención del delito, ello representa una fuerte preocupación social por el incremento cuantitativo y cualitativo en su peligrosidad, ocasionando una fuerte sensación de inseguridad en los ciudadanos.

. En este sentido (Valera, 2012) asegura que “Es posible establecer patrones de comportamiento delictivo en base a registros históricos para así establecer diversas acciones preventivas en base a la acción policial”.

El análisis de los registros criminales es fundamental en la prevención del delito, porque permite el diseño de políticas y planes de prevención efectivos. En Perú este tipo de análisis se ha realizado históricamente mediante herramientas estadísticas descriptivas básicas, considerando fundamentalmente variables y relaciones primarias. Sin embargo, muchas veces la estadística descriptiva clásica no refleja la verdadera interrelación de las variables y, por lo tanto, el problema real. Este contexto requiere un tratamiento más complejo que obliga a evolucionar en el análisis de información criminal.

En este sentido, la cantidad de denuncias también ha crecido exponencialmente, sin embargo, la capacidad para procesarlos y utilizarlos no ha crecido en la misma dimensión,

Sobre ello (Jiménez, 2015), refiere que la falta de equipos tecnológicos y de sistemas de información impiden una labor más eficaz a los policías, para combatir el delito. Dentro de estas enormes masas de datos existe una gran cantidad de información "oculta", de gran



importancia estratégica, a la que solo se puede acceder por técnicas de recuperación de la información, como la minería de datos.

Según (Usama, 1996). “La minería de datos es el conjunto de técnicas y herramientas aplicadas al proceso trivial de extraer y presentar el conocimiento implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con el objeto de predecir de forma autorizada tendencias y comportamientos y/o descubrir de forma automatizada modelos previamente desconocidos”.

De acuerdo con lo expuesto surge la interrogante ¿De qué manera la minería de datos puede contribuir en la detección de patrones delictivos? El objetivo de esta investigación es poder determinar, mediante una revisión sistemática, la contribución de la minería de datos en la detección de patrones delictivos.



## CAPÍTULO II. METODOLOGÍA

### Diseño:

La presente investigación ha sido desarrollada en base a una revisión sistemática sobre la aplicación de la minería de datos para la detección de patrones delictivos con el fin de conocer su aporte en el mismo. La revisión sistemática trabaja sobre realidades de hecho y su característica fundamental es la de presentar una interpretación correcta.

### Selección de documentos y estrategias de búsqueda:

En base al objetivo de esta investigación, el cual es identificar el aporte de la minería de datos en la detección de patrones delictivos, se realizó una selección de documentos a través de la búsqueda automatizada en Google Scholar (Google académico), y a través de las bases Dialnet, World Wide Science, BASE, SciELO. Debido a su gran incremento anual de documentos científicos, estas búsquedas se realizaron en el mes de septiembre del 2019, considerando en su mayoría, investigaciones no mayores a 10 años de antigüedad.

Tabla 1. Cantidad de resultados obtenidos desde Google académico

2011	2012	2013	2014	2015	2016	2017	2018	2019
209	303	319	420	408	445	496	123	198

Fuente: (<https://scholar.google.com.pe>), Consultado: 22 de septiembre del 2019.

Para focalizar las búsquedas fue necesario especificar que la aplicación de la minería de datos haya sido implementada en la detección de patrones delictivos, considerando ello, se determinó incluir las técnicas de minería de datos el siguiente termino de búsqueda: “ o



“minería de datos como aporte en patrones delictivos” o “minería de datos para identificar patrones delincuenciales” o “descubrimiento de patrones delictivos con minería de datos” o “aplicando minería de datos por clasificación, regresión, agrupamiento, resumen, modelado de dependencias, análisis de secuencias en la detección de patrones delictivos”.

Como criterio adicional se estableció que el idioma de los documentos investigados sea castellano, a pesar de que la mayor cantidad de documentos referentes al tema investigado estaban redactados en inglés, sin embargo, de esta manera la recolección de información fue más concisa y en base a ello se pudo realizar una revisión más rápida y exacta de los contenidos

Se consideraron documentos de tipo Tesis y revistas científicas debido a su disponibilidad y fácil accesibilidad en los motores de búsquedas de académicos, obteniendo de las investigaciones la siguiente información:

- Año de publicación;
- Técnicas de minería de datos aplicadas en la investigación y;
- Implementación y resultados obtenidos.

El alcance de esta investigación consideró la revisión en profundidad de documentos de investigación que cumplieron con los criterios de selección, esta consideración se estableció posteriormente al filtrar un total de 121 documentos de investigación. La validez de los documentos seleccionados estuvo dada por el grado de evidencias demostrado, por las conclusiones y por la aplicabilidad a nuestro contexto. Con base en lo anterior se establecieron las siguientes interrogantes como objetivos específicos de la investigación:

- ¿Cuáles son las técnicas de minería de datos aplicadas para la detección de patrones delictivos?



- ¿Cuáles son las técnicas más efectivas y que han aportado más en la detección de patrones delictivos?
- ¿Cuáles son los patrones delictivos con más ocurrencia?

### CAPÍTULO III. RESULTADOS

*“Es posible establecer patrones de comportamiento delictivo en base a registros históricos para así establecer diversas acciones preventivas en base a la acción policial”.*

(Valera, 2012)

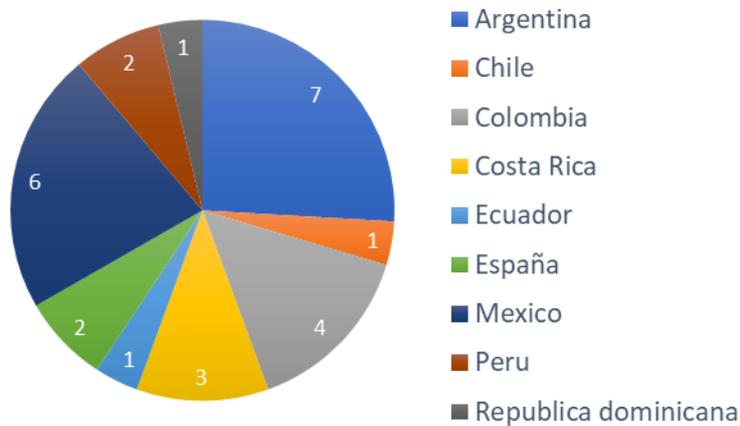
En base al análisis de datos considerando a los criterios de búsquedas establecidos, se obtuvo un total de 30 documentos de investigación, de los cuales se evidencia una mayor cantidad de documentos de investigación en el año 2015 y mayor énfasis de investigación argentina

Figura 1. Numero de investigaciones publicadas por año.



Figura 2. Distribución de investigaciones por país.

### Investigaciones por país



Fuente: Elaboración propia

Para la detección de los patrones delictivos, las investigaciones evidenciaron el uso del proceso KDD en donde la aplicación de minería de datos es un paso importante en este proceso del descubrimiento del conocimiento. Se identificaron las técnicas de minería de datos que fueron aplicadas en cada investigación resultando como la técnica más utilizada la de Agrupamiento con un 40% respecto a toda la muestra, la técnica de clasificación obtuvo un 29% seguida de la técnica Resumen con 26% y Análisis de secuencias con 6% de uso, así culmina la lista con la técnica de regresión y modelo de dependencias con 0% lo que no indica que estas últimas técnicas no sean efectivas, sin embargo, su aplicación para fines de detección de patrones delictivos, no son relevantes.

Figura 3. Proceso de descubrimiento del conocimiento en las bases de datos.



**Proceso de Knowledge Discovery in Databases (KDD)**

Recuperado de: <http://evaluandosoftware.com/tecnicas-data-mining/>

Tabla 2. Técnicas de minería de datos más utilizadas.

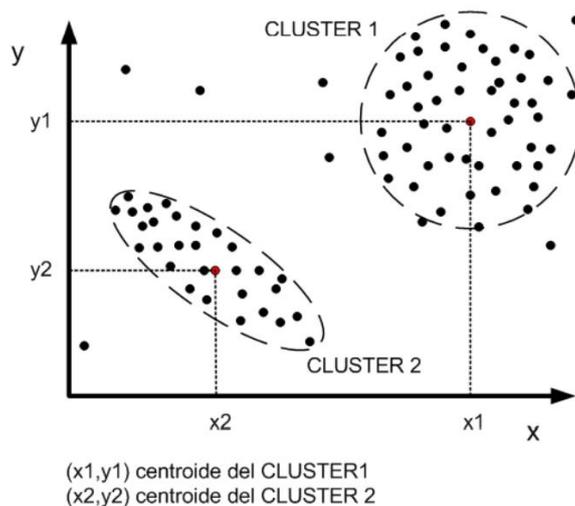
Técnicas de minería de datos	Argentina	Chile	Colombia	Costa Rica	Ecuador	España	México	Peru	Rep. Dominicana	TOTAL
Agrupamiento	7	1	1	2	0	0	1	2	0	14
Clasificación	2	0	4	0	1	2	0	0	1	10
Resumen	2	0	2	1	0	0	2	2	0	9
Análisis de secuencias	1	0	0	0	1	0	0	0	0	2
Regresión	0	0	0	0	0	0	0	0	0	0
Modelo de dependencias	0	0	0	0	0	0	0	0	0	0
	12	1	7	3	2	2	3	4	1	35

Fuente: Varios autores, elaboración propia

La principal función de la técnica de Agrupamiento es, identificar semejanzas entre objetos con el fin de agruparlos, esta técnica utiliza algoritmos de reconocimiento de patrones, de los cuales destaca *K-beans*, asignando a cada objeto una medida de semejanza, la cual se calcula con una función de distancia, siendo *Euclídea* la función más utilizada en técnicas de agrupamiento debido a que su distancia es la más corta entre objetos. Con base a lo expuesto, se utilizaron los atestados policiales con el fin de elaborar los modelos de bases de datos, con los cuales se implementaron y aplicaron las técnicas mencionadas en

búsqueda de hallazgos de patrones delictivos. Con un porcentaje promedio de 70% de acierto en las predicciones, se obtuvieron los siguientes resultados por cada cluster identificado:

Figura 4. Representación de la técnica de agrupamiento



Fuente: Benites, I.

Tabla 3. Cantidad de delitos por cluster edad.

EDAD	
DESCRIPCION	FRECUENCIA
0 - 18 AÑOS	16
19 - 20 AÑOS	26
21 - 30 AÑOS	152

Fuente: Varios autores, elaboración propia

Tabla 4. Cantidad de delitos por cluster estado civil.

ESTADO CIVIL	
DESCRIPCION	FRECUENCIA
Soltero	111
Casado	94
Divorciado	14
Conviviente	178
Exconviviente	74

Fuente: Varios autores, elaboración propia



Tabla 5. Cantidad de delitos por cluster reincidencia.

REINCIDENCIA	
DESCRIPCION	FRECUENCIA
Primario	194
Reiterante	171
Reincidencia múltiple	115

Fuente: Varios autores, elaboración propia

Tabla 6. Cantidad de delitos por cluster sexo.

SEXO	
DESCRIPCION	FRECUENCIA
Masculino	361
Femenino	119

Fuente: Varios autores, elaboración propia

Tabla 7. Cantidad de delitos por cluster instrucción.

INSTRUCCIÓN	
DESCRIPCION	FRECUENCIA
Sin instrucción	25
Primaria incompleta	54
Primaria completa	91
Secundaria incompleta	120
Secundaria completa	138
Técnica	38
Superior	14

Fuente: Varios autores, elaboración propia

Tabla 9. Cantidad de delitos por cluster condición anímica.

CONDICION ANIMICA	
DESCRIPCION	FRECUENCIA
Ebrio	85
Alterado	62
Drogado	45
Ecuanime	288

Fuente: Varios autores, elaboración propia

Tabla 10. Cantidad de delitos por cluster horario.

HORARIO	
DESCRIPCION	FRECUENCIA
8 HRS - 13 HRS	16
14 HRS - 20 HRS	26
21 HRS - 2 HRS	152
3 HRS - 7 HRS	114

Fuente: Varios autores, elaboración propia



## CAPÍTULO IV. CONCLUSIONES

De las investigaciones revisadas se encontraron distintas aplicaciones, así como diversas posiciones sobre las funcionalidades respecto a cada técnica aplicada, en esta sección encontraremos cuales técnicas de minería de datos que fueron utilizadas por los autores que han sido referencia en esta investigación con el fin de ampliar la perspectiva sobre la aplicación de minería de datos para la detección de patrones delictivos.

Jaulias, J., Vilcarromero J. (2015), Castillo, M. (2015), Aviles, R., Garcia, J., Gonzales, O., Gonzales, N. (2017) y Corso C., Garcia, A., Ciceri, L., Romero., F. (2014). afirman que las técnicas más efectivas para poder encontrar patrones delictivos que puedan contribuir con la prevención del delito, son las técnicas de agrupamiento (cluster) y clasificación teniendo en cuenta que para la técnica de clasificación rotula un dato dentro de una de las clases categóricas predefinidas responde a preguntas tales como, ¿Cuál es el riesgo de conceder un crédito a este cliente? ¿Dado este nuevo paciente qué estado de la enfermedad indican sus análisis?, así mismo la técnica de agrupamiento agrupación de registros, observaciones, o casos en clases de objetos similares. Un cluster es una colección de registros que son similares entre sí, y distintos a los registros de otro cluster. ¿Cuántos tipos de clientes vienen a mi negocio? ¿Qué perfiles de necesidades se dan en un cierto grupo de pacientes? Benites, I. (2005). también apoya las afirmaciones ya que, según lo expuesto, esta funcionalidad es la más idónea para garantizar la identificación de los patrones delictivos que servirán para los procesos de prevención del delito.

Lopez, N., Pollo, C., Britos, P., Garcia, M. (2011), Ruiz J. (2013), Britos, P., Fernández, E., Merlino, H., Pollo, M., Rodríguez, D., Procopio, C., Rancan, C., García, R. (2008) y Valenga, F., Fernández, E., Merlino, H., Rodríguez, D., Procopio, C., Britos, P., Garcia, R. (2008). en su defecto consideran que la técnica de análisis de consecuencias es la que puede aportar aún mas a este objetivo, sostiene que se modelan patrones secuenciales, como análisis de series temporales, secuencias de genes, etc. El objetivo es modelar los estados del proceso, o extraer e informar de la desviación y tendencias en el tiempo. ¿El



consumo de energía eléctrica de este mes es similar al del año pasado? Dados los niveles de contaminación atmosférica de la última semana cuál es la previsión para las próximas 24 horas, sostienen que se puede predecir actitudes y comportamientos en base los datos registrados anteriormente, sin embargo, Argüello, J., Malinowski, E. (2002), Peña, A. (2017), Clavero, A., Sepulveda S. (2009), Wanumen, L. (2010), Arruelas, S. (2012) y Jimenez, P., Oviedo, E. (2015), discrepan al considerar que para esa aplicación se debería contar con información histórica con lo cual no se podrían identificar a los posibles delincuentes primarios, en base a ello indican que aplicar la técnica de resumen es mas asertivo para obtener los resultados que realmente ayuden a determinar y adelantarnos a los actos delictivos, ellos sostienen que estos modelos proporcionan una descripción compacta de un subconjunto de datos. ¿Cuáles son las principales características de los delincuentes?

La presente investigación concluye después haber realizado una revisión detallada de documentos de investigación sobre la detección de patrones delictivos, de un total de 121 investigaciones se obtuvieron las siguientes conclusiones:

- Las técnicas aplicadas en este tipo de investigaciones son las de agrupamiento, clasificación, resumen y análisis de secuencias. Estas son las más usadas debido a su funcionalidad, el de tipo cluster o agrupamiento se refiere una colección de registros que son similares entre sí, y distintos a los registros de otro cluster, la técnica de clasificación rotula un dato dentro de una de las clases categóricas predefinidas, la de tipo resumen proporcionan una descripción compacta de un subconjunto de datos y en el caso del tipo de



análisis de secuencias, el objetivo es modelar los estados del proceso, o extraer e informar de la desviación y tendencias en el tiempo.

- Se identifico que las técnicas de agrupamiento y clasificación son las más utilizadas en este tipo de aplicación de minería de datos siendo la más efectiva la técnica de agrupamiento, otorgando un 70% de probabilidad de coincidencia en las detecciones de los patrones delictivos.
- Gracias a la minería de datos se identificaron los cluster más relevantes y de mayor reincidencia con respecto a patrones delictivos, obteniendo así información muy valiosa de cara a la prevención del delito. Entre los patrones más importantes encontramos que: Las personas que cometen mayor cantidad de actos delictivos, son mayores de 41 años, además son del sexo masculino, a ello se le suma que su estado civil es conviviente, en el caso de la reincidencia, la mayor cantidad de actos delictivos corresponde a delincuentes primerizos, sin dejar de observar a los reincidentes que ocupan el segundo lugar, el mayor grado de instrucción de la personas que han cometido mayores actos delictivos es la secundaria, también de identifico que el estado de personas que registran mayor índice de actos delincuenciales es cuando están ecuánimes seguido del estado ebriedad, finalmente se detecta que el intervalo horario en donde ocurren más actos delictivos es de 9 p.m. a 2 a.m.



## REFERENCIAS

- Lopez, N., Pollo, C., Britos, P., Garcia, M. (2011). *Un Protocolo de caracterización Empírica de Dominios para Uso en Explotación de Información*. XVII Congreso Argentino de Ciencias de la computación.
- Castillo, M. (2015). *Visualización de indicadores y tendencias delictivas a partir de informes gubernamentales*. Tesis de maestría. Instituto politécnico nacional de México.
- Ruiz J. (2013). *Análisis especial y supervisado de datos criminalísticos utilizando un sistema de información geográfica*. Tesis de maestría. Instituto politécnico nacional de México.
- Argüello, J., Malinowski, E. (2002). Un Modelo De Detección De Fraudes En Datos De Seguros. *Revista de la universidad de Costa Rica*, 12(1,2).
- Peña, A. (2017). *Modelo para la Caracterización del Delito en la Ciudad de Bogotá, Aplicando Técnicas de Minería de Datos Espaciales*. Tesis de maestría. Universidad distrital Francisco José de Caldas.
- Aviles, R., Garcia, J., Gonzales, O. y Gonzales, N. (2017). El uso de minería de datos. *Revista científica dominicana de ciencias*, 3(3). doi. <http://dx.doi.org/10.23857/dom.cien.pocaip.2017.3.3.jun.1308-1326>
- Clavero, A., Sepulveda S. (2009). *Aplicación de Minería de Datos para la Detección de Anomalías: Un Caso de Estudio*. Chile
- Jaulias, J., Vilcarromero J. (2015). *Sistema De Predicción De Hechos Delictivos Para La Mejora Del Proceso De Prevención Del Delito En El Distrito De La*



*Molina Utilizando Minería De Datos.* Tesis de título. Universidad de San Martín de Porres.

Wanumen, L. (2010). Minería de datos para la predicción de fraudes en tarjetas de crédito. *Vínculos*, 7(2).

Corso C., Garcia, A., Ciceri, L., Romero., F. (2014). *Minería de Datos aplicada a la Detección de factores para la prevención de incidentes informáticos.* Cordoba, Argentina.

Britos, P., Fernández, E., Merlino, H., Pollo, M., Rodríguez, D., Procopio, C., Rancan, C., García, R. (2008). *Explotación de información aplicada a inteligencia criminal.* XIV Congreso Argentino de Ciencias de la computación.

Flores, L., Mariño, S., Martins, S. (2018). *Propuesta de procedimiento para el análisis delictivo basado en la explotación de la información.* XX Workshop de Investigadores en Ciencias de la Computación.

Valenga, F., Fernández, E., Merlino, H., Rodríguez, D., Procopio, C., Britos, P. y Garcia, R. (2008). *Minería de datos aplicada a la detección de patrones delictivos en Argentina.* VII Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento

Jimenez, L. (2015). *Aplicación de un sistema de alerta temprana basada en minería de datos para identificar patrones delictivos en la ciudad de Chiclayo.* Tesis de título. Universidad católica Santo Toribio de Mogrovejo.

Acosta, A. (2016). *Definición de sistema tecnológico integrado de seguridad*



*pública*. Tesis de título. Universidad de la Sabana.

Bonilla, L. (2014). *Inteligencia artificial aplicada en la prevención y detección de lavado de activos y financiamiento del terrorismo*. San José, Costa Rica.

Bender, A., Mazza, N. (2017). *Análisis predictivo: Difusión e impacto en áreas de la sociedad*. Simposio Argentino sobre Tecnología y Sociedad (STS) - JAIIO 46.

Arruelas, S. (2012). *Sistema para la prevención de fraudes, gestión y monitoreo del negocio*. Tesis de título. Universidad Central del Ecuador.

Silva, S. (2017). *Diseño e implementación de un módulo para la detección de aplicaciones móviles maliciosas en mercados online*. Tesis de título. Universidad Politécnica de Madrid.

Aguilar, H., Quintana, S. (2012). *Sistema de relaciones criminales*. Tesis de título. Instituto Politécnico Nacional.

Salamea, D. (2014). *Modelo de aplicación laboral que incorpore la prevención del estado ecuatoriano respecto de la delincuencia juvenil y el buen vivir*. Tesis de magister. Universidad regional autónoma de los andes. Universidad católica de Cuenca.

Jimenez, P., Oviedo, E. (2015). *Diseño e implementación de una aplicación en dispositivos móviles para reducir la inseguridad de la población en el Municipio de Montelíbano*. Tesis de título. Universidad de Córdoba.

Morales, O. (2015). *Confianza en la política preventiva de la ciudad de México: Un análisis y propuesta a través de la minería de datos*. Tesis de maestría.



Centro de investigación y docencia económicas, A.C.

Lozano, A. (2008). El perfil financiero: una estrategia para detectar el lavado de activos. *Revista Criminalidad Policía Nacional – DIJIN*, 50(2).

Torres, F., Neira, L., Torres, L. (2015). Análisis criminal inteligente usando eventos geográficos y temporales con a la red neuronal de máxima sensibilidad. *Daena: International journal of good consciense*. 10(3).

Riquelme, J., Ruiz, R., Gilbert, C. (2006). Minería de datos: Conceptos y tendencias. *Inteligencia artificial, revista iberoamericana de inteligencia artificial*, 29.

Grajales, T. (2000), *Tipos de Investigación*, Revista de educación en línea.

Usama, F. (1996). El proceso KDD para extraer conocimiento útil de volúmenes de datos. *Revista comunicaciones del ACM*, 39(11). doi: <https://doi.org/10.1145/240455.240464>

Benites, I. (2005). *Técnicas de agrupamiento para el análisis de datos cuantitativos y cualitativos*.