

Celopatía y motivación sexual: Un análisis a través de *K-Means* en los establecimientos penitenciarios del Perú

Pilar Vanessa Hidalgo León
Universidad Andina del Cusco
Escuela Profesional de Ingeniería de Sistemas
Cusco Perú

RESUMEN

En el penal de Q'enqoro en la ciudad del Cusco, uno de cada cuatro internos están reclusos por delitos asociados a la violencia de género (INEI 2016), la ENDES (Encuesta Demográfica y de Salud Familiar) registra que las principales motivaciones para perpetrar estos crímenes están asociadas a la celopatía y al deseo sexual patológico. La OMS (2014) asegura que estos comportamientos pueden estar relacionados a «factores sociales, comunitarios, de reincidencia y situacionales en un determinado espacio geográfico». El objetivo de esta investigación es identificar el grado de cercanía de estos *factores* con la motivación de los perpetradores para cometer los delitos. Para esta investigación se utilizaron datos provenientes del «Censo Nacional Anual de la Población Penitenciaria en el Perú», que fue desarrollada por el Instituto Nacional de Estadística e Informática (INEI) y tomadas entre el 18 al 26 de abril de 2016, ejecutado en 66 establecimientos penitenciarios a nivel nacional, 77 500 internos encuestados en 24 departamentos del país y la Provincia Constitucional del Callao. La muestra cuenta con 1438 datos que incluyen delitos como feminicidios, asaltos sexuales e intento de homicidio. Para el procesamiento del *dataset* se usó el software de exploración de datos *Weka* versión 3.8, la experimentación de la búsqueda de patrones a través del algoritmo *K-means*, se reconocieron 2 grupos que definen las características cercanas involucradas con la motivación para perpetrar el delito: la celopatía y la motivación sexual.

Palabras clave: *K-means*, violencia de género, exploración de datos, violencia sexual, minería de datos.

ABSTRACT

Pathological jealousy and Sexual Motivation: An Analysis through K-Means in Peruvian Prison Establishments

In the Q'enqoro prison in the city of Cusco, one in four inmates are held for crimes associated with gender violence (INEI 2016), the ENDES (Demographic and Family Health Survey) records that the main motivations for perpetrating these Crimes are associated with Pathological jealousy and pathological sexual desire. OMS (2014) asserts that these behaviors may be related to social, community, recidivism and situational factors in a given geographic space. The objective of this investigation is to identify the degree of proximity of these factors with the motivation of the perpetrators to commit the crimes. For this research, data from the annual National Census of the prison population in Perú were used, which was developed by the National Institute of Statistics and Informatics (INEI) and taken between april 18 to 26, 2016, executed in 66 prisons nationwide, 77 500 interns surveyed in 24 departments of the country and the Constitutional Province of Callao. The sample counts 1438 data that includes crimes like feminicide, sexual assaults and attempt of homicide. For the dataset processing we used the Weka version 3.8 data exploration software, the experimentation of the search of patterns through the algorithm K-means, we recognized 2 groups that define the near characteristics involved with the motivation to perpetrate the crime: Pathological jealousy and sexual motivation.

Key words: K-means, gender violence, data exploration, sexual violence, data mining.

INTRODUCCIÓN

La «violencia de género» es considerada como un problema global, según las organizaciones internacionales (ONU, OMS, OPS), es parte de una serie de consecuencias relacionadas con la estabilidad de una sociedad [7, 8, 6]. Para poner fin a la violencia de género se requiere trabajo y esfuerzo conjunto de muchas instituciones. El Centro Internacional para la Prevención del Delito (2008) afirma que, el bienestar social debe priorizar la prevención, detección e intervención de los delitos a través de la cohesión social. La prevención, basada en el conocimiento, se relaciona con acciones que abordan la interpretación y diseminación de datos relacionados con la percepción de seguridad. La prevención del delito debe ser operada en los niveles: sociales, situacionales, comunitarios y de reincidencia, prevenir la delincuencia a través del conocimiento implica, principalmente, la aplicación de principios y métodos de investigación de calidad y evaluación adecuadas, que permitan a los usuarios tener cierto grado de confianza a las intervenciones. Nick Chrisman [15] también se refiere a la importancia de la inclusión del contexto social para que los sistemas de información puedan expresar adecuadamente, puntos que: 1) midan los aspectos de los fenómenos geográficos y los procesos; 2) representen estas mediciones, por lo general en forma de una base de datos informática, centrándose en cuestiones espaciales, entidades y relaciones; 3) operen sobre estas representaciones que producen más mediciones para descubrir nuevas relaciones integrando fuentes dispares y; 4) transformen estas representaciones que se ajustan a otros marcos de

entidades y relaciones. Estas actividades reflejan un contexto más amplio (instituciones y culturas) en el que estas personas realizan su trabajo.

Estos factores establecen relaciones con la motivación del delito como detonantes de la violencia de género y pueden ser estudiados cuantitativamente a través de las técnicas de *K-Means*. [8]

Bsoul, Q., Salim, J., & Zakaria, L. Q. [3], en un estupendo resumen acerca de las técnicas y algoritmos para procesar información y detectar crímenes afirman que *K-Means* es el más utilizado para estos tópicos. En la tabla adjunta se identifica el uso y limitaciones en investigaciones de criminalística que utilizan *K-Means* como herramienta principal: (Tabla A)

Tabla A: Resumen de los trabajos previos en algoritmos de *clustering*

<i>Data</i>	<i>Esta</i>	<i>Algoritmo</i>
<i>Used</i>	<i>do</i>	<i>superior</i>
	<i>del</i>	
	<i>Arte</i>	
	<i>del</i>	
	<i>algo</i>	
	<i>ritm</i>	
	<i>o</i>	
<i>Financi</i>	<i>AH</i>	<i>Enhance</i>
<i>al news</i>	<i>C</i>	<i>AHC</i>
<i>IRES</i>	<i>K-</i>	<i>SOM</i>
	<i>mea</i>	
	<i>ns</i>	
<i>Data</i>	<i>N/A</i>	<i>FSOM</i>
<i>series</i>		
<i>of</i>		
<i>crimes</i>		
<i>KDD</i>	<i>Esc</i>	<i>Kmeans</i>
	<i>alab</i>	
	<i>le,</i>	
	<i>k-</i>	
	<i>mea</i>	
	<i>ns</i>	
	<i>com</i>	
	<i>plet</i>	
	<i>e</i>	

<i>Crimes</i>	<i>Single pass, k-means</i>	<i>Enhance K-means</i>
<i>News</i>	<i>Singles pass</i>	<i>k-means</i>

Fuente: Bsoul, Q., Salim, J., & Zakaria, L. Q. (2013). An intelligent document clustering approach to detect crime patterns. *Procedia Technology*, 11, 1181-1187.

CLUSTERING

Los algoritmos de *clustering*, usualmente utilizan dos fases de búsqueda: un ciclo externo sobre el mayor número posible de grupos y un ciclo interno para adaptarse al mejor grupo posible. Estos métodos realizan tareas de agrupación de objetos de una base de datos en grupos llamados *clúster*, conformados por miembros tan similares como sea posible, es decir, que en los otros *clúster* se encuentran objetos diferentes. La minería de datos es un proceso que busca —en grandes conjuntos de datos— información relacionada y significativa. Varias técnicas para la agrupación o *clustering* son las siguientes: método de partición, método jerárquico, método basado en cuadrícula y método basado en la densidad. [10, 9, 1].

K-MEANS

El algoritmo *K-Means* es una técnica de partición muy conocida como método de agrupación de datos, usa como principio base la proximidad entre sí, según la distancia euclidiana. Se necesita un valor K_y como parámetro de entrada y la partición de un conjunto de ‘n’ objetos de *clusters* K_y ; El valor medio del objeto se toma como similitud para formar *clusters*. La media del

grupo o centro se forma mediante la selección aleatoria del objeto K_y . Por comparación, la mayoría tiene similitud con otros objetos y se asignan al grupo. Para cada vector de datos este algoritmo calcula la distancia entre el vector de datos y cada *centroide* de agrupamiento utilizando la ecuación siguiente [1].

$$D(Z_p, M_j) = \sqrt{\sum (Z_p, K_y - M_j, K_y)}$$

Donde Z_p : es el dato inicial elegido al azar antes de la construcción del modelo, M_j es el *centroide de clúster "j"*.

El centroide es recalculado cada vez, respectivamente, antes de agregar un nuevo dato al clúster "j". La iteración final para la construcción del modelo sigue la siguiente ecuación:

$$M_j = \frac{1}{N_j} \sum Z_p, Z_p \in C_j$$

Donde N_j es el número de datos que pertenecen al *clúster*.

Esta es la manera en la que el algoritmo de *K-Means* trabaja. [4,6].

MATERIALES Y MÉTODOS

Si hemos de hablar acerca de un método de investigación, el análisis y exploración de datos son parte de este trabajo —por ello es de alcance descriptivo—, al incluir los métodos de *clustering* la observación nos permite analizar conocimiento implícito que no habría sido identificado sin el apoyo de estas metodologías.

La encuesta que realizó el INEI tiene por objetivo conocer las condiciones por las cuales el interno procedió a delinquir, además de la situación carcelaria y su interacción con las autoridades. Los datos se dividen en 4 temas: las condiciones sociales y familiares del interno, la tipificación del delito, las condiciones de vida en el penal y, la relación e intervención de las autoridades.

Respecto a lo descrito anteriormente la ‘violencia de género’ se puede ver influenciada por diversos factores: sociales (educación, nivel de alfabetismo, deserción estudiantil, edad, estado civil), comunitarios (lugar de procedencia), reincidencia de delitos y situacionales (consumo de drogas, enfermedades mentales) aparentemente aisladas entre sí.

Es entonces que tomamos estos valores de la encuesta que corresponden a 16 atributos respectivamente:

1. Ciudad (46 ciudades del país para la muestra)
2. Estado civil (soltero, casado, viudo, divorciado, conviviente, separado, etc.)
3. Religión (católica, adventista, mormona, testigo de Jehová, otra, ninguna)
4. Edad (entre 18 a 79 años)
5. Estudios (primaria, secundaria, universitaria, superior no universitaria)
6. Deserción estudiantil (motivo de deserción)
7. Mala influencia en la escuela (sí lo tuvo o no)
8. Depresión (sí sufre y sí está diagnosticado)
9. Ansiedad (sí sufre y sí está diagnosticado)
10. Adicción (sí sufre y sí está diagnosticado)
11. Consumo de drogas 6 horas del delito (sí o no)
12. Tipo de drogas (marihuana, alcohol, cocaína, heroína, otras)
13. Centro juvenil (sí estuvo o no recluso antes)
14. Reincidencia del delito (sí o no)
15. Reincidencia de otros (lesiones, secuestro, violaciones sexuales, homicidio culposos)
16. Motivación del delito (celos, motivación sexual)

Siguiendo la metodología para descubrir conocimiento de *Fayyad*, se tienen los siguientes procesos para la evaluación del *dataset*: [10, 5, 2].

- **Selección de datos**

Los elementos de entrada que se establecieron durante el proceso de aplicación del algoritmo fueron 3 parámetros: el cálculo de distancia entre puntos, el número de *cluster* y el número de iteraciones. Es importante mencionar que no son los únicos disponibles, en *Weka*, existen otros tantos pero, que en esta investigación no se abordarán, puesto que los experimentos se llevaron a cabo con algoritmos naturales, es decir, sin alteración de

parámetros. El número de *cluster* definidos para este primer experimento fueron 2 en un conjunto de 1438 datos, inicialmente.

- **Análisis de datos**

Luego de depurar manualmente los datos disponibles en la encuesta y utilizar *All Filter* para los espacios vacíos, los histogramas muestran el conjunto de datos con afinidad al punto representativo de la muestra.

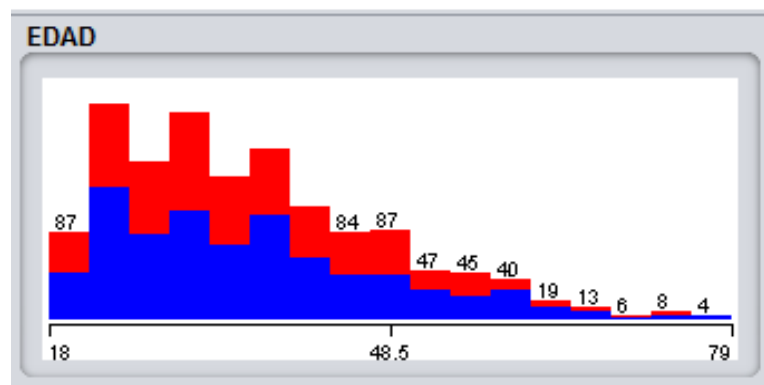


Figura 1: Histograma edad promedio de la muestra
Fuente de elaboración propia

<i>Statistic</i>	<i>Value</i>
<i>Minimum</i>	<u>18</u>
<i>Maximum</i>	<u>79</u>
<i>Mean</i>	<u>35.997</u>
<i>StdDev</i>	<u>11.819</u>

Este histograma (figura 1) muestra que la población tomada en la investigación fluctúa entre los 18 y los 79 años, se identifica por ejemplo que la ‘celopatía’ (color azul) es mayor que la motivación sexual (color rojo) en perpetradores de todas las edades. La definición de estos clústeres se realizó según los siguientes parámetros:

- 1) La elección aleatoria del número de objetos representativos iniciales para cada dato;

2) *Ky* reúne y define cada objeto según los requisitos de pertenencia al *clúster* basado en el valor medio del objeto en él.

Luego de un bucle finito y después de que se calcula para cada *clúster* el valor medio del objeto, se repite esta metodología desde el paso inicial hasta que no se haya asignado ningún punto más.

Los criterios de satisfacción del bucle pueden ser el número de iteración o el cambio de posición del *centroide*³ en sucesivas iteraciones.

- **Transformación del conjunto de datos**

Los datos, inicialmente, estaban dispuestos por capítulos en la encuesta del INEI, además, no contaban con todas las preguntas (hubo preguntas no obligatorias) por privacidad del interno; la información en los campos tales como nombre, dirección, situación procesal; fueron espacios ausentes en las bases de datos.

Se eligieron los campos que están incluidos en los factores que influyen en la violencia de género según la OMS. Definiendo de este modo la «motivación para perpetrar el delito» presente en la pregunta 203 de la encuesta del INEI, como la característica relevante de los *clúster*, transformando así, las variables del fichero por variables no numéricas, para ser reconocidas por *Weka*.

Ya que existen preguntas no obligatorias en la encuesta —pero son necesarias para la investigación—, se aplicó la herramienta *All Filter* para filtrar los espacios vacíos.

Una vez que los datos fueron transformados y declarados aptos para la experimentación, se ha escogido el algoritmo *K-Means*, que según los

³ En geometría, el centroide o baricentro de un objeto X perteneciente a un espacio n -dimensional es la intersección de todos los hiperplanos que dividen a X en dos partes de igual n -volumen con respecto al hiperplano.

criterios mencionados, cumple con las características necesarias para el análisis de datos.

RESULTADOS

Para presentar los resultados y siguiendo la metodología *Knowledge Discover Data*, de *Fayyad*, se realizó el proceso de ‘selección y aplicación de la técnica de minería de datos’ de forma objetiva, sugerida por las investigaciones mencionadas en este trabajo pues cumplen el rigor científico de revisión.

En la ‘evaluación e interpretación de los resultados’ se realizaron 3 experimentos en distintas situaciones para validar el modelo y perfil construido a partir de la motivación del perpetrador para cometer el delito.

Primer experimento

En el experimento se utilizó como distancia entre los puntos la función de «distancia euclidiana», donde el máximo de candidatos fue 1000 y el mínimo 100. El principio de ordenamiento fue *R first-last* eligiendo los *centroides* de manera aleatoria. El número de instancias totales fue 1427 luego de la depuración, analizando 16 atributos y evaluando la muestra de entrenamiento a través del test full training set. El número de iteraciones para la búsqueda patrones fue 4, con un error cuadrático de 6433,092553143473.

En la primera iteración el punto escogido para dar inicio a la construcción del modelo fue:

Cluster 0: el parámetro es la celopatía, provenientes del penal de 'La Esperanza', con situación civil conviviente, religión católica, 48 años de edad promedio, con secundaria completa, no especifica la razón por la cual desistió de los estudios, no sufre de depresión, ansiedad o adicción descartadas por un especialista, no consumió drogas 6 horas previas al delito, consume alcohol, no proviene de reclusorios juveniles, es la primera vez que comete el delito, no ha reincidido en ningún otro delito anteriormente.

Clúster 1: el parámetro es motivación sexual, provenientes del ‘Penal de Socabaya’, estado civil: soltero, religión evangélica, 37 años de edad promedio, con educación primaria incompleta, la deserción estudiantil fue a causa de la pobreza de la familia, no tuvo mala influencia en la escuela, no sufre de depresión, ansiedad, adicción pues ha sido descartado por especialista, no consumió drogas 6 horas previas al delito, no consume drogas, no tiene antecedentes penales juveniles, es la primera vez que comete el delito y no ha reincidido en otros.

Este es el resultado final del primer experimento, finalmente los *centroides* encontrados (tabla 1).

Tabla 1: Resultado de los *centroides* significativos en la última iteración del algoritmo.

Las instancias agrupadas son: 0: 657 (46 %), 1: 770 (54 %)

	Clúster #	0=celopatía	1= motivación sexual
Atributos	(1427.0)	(657.0)	(770.0)
PENAL	SAN JUAN DE LURIGANCHO	SAN JUAN DE LURIGANCHO	SAN JUAN DE LURIGANCHO
Estado civil	Soltero	Conviviente	Soltero
Religión	Católica	Católica	Adventista
Edad	37,99972	37,1537	35,0104
Estudios	Secundaria incompleta	Secundaria completa	Secundaria completa
Razón de deserción	No especifica	No especifica	No Especifica
Influencias en la escuela	No	No	No
Depresión	No	No	No
Ansiedad	No	No	No
Adicción	No	No	No
Drogas previas a delinquir	No	Si	No
Tipo de drogas	No consume	Alcohol	No consume
Reincidencia del delito	No	No	No
Centro juvenil anteriormente	No	No	No
Reincidencia de otros delitos	No	No	No

Fuente de elaboración propia

Segundo experimento

Para este caso, el experimento utilizó como distancia entre los puntos la función de ‘distancia Manhattan’, donde el máximo de candidatos será 10 000 y el mínimo 100. El principio de ordenamiento será *R first-last* eligiendo los *centroides* de manera aleatoria. El número de instancias totales es 1427, analizando 16 atributos, evaluando la muestra de entrenamiento a través del test al 66 % de la muestra del *training set*. El número de iteraciones para la búsqueda patrones fue 4, con una suma de las distancias entre *clúster* de 6505,96721. La primera iteración: el punto escogido para dar inicio a la construcción del modelo fue el mismo del primer experimento.

Este es el resultado de la primera iteración del experimento, *centroides* aleatorios (tabla 2).

Tabla 2: Resultado de los *centroides* aleatorios en la primera iteración del algoritmo.

Atributos	Clúster # Celopatía (1427.0)	O=celopatía (664.0)	1= motivación sexual (770.0)
PENAL	SAN JUAN DE LURIGANCHO	SAN JUAN DE LURIGANCHO	SAN JUAN DE LURIGANCHO
Estado civil	Soltero	Conviviente	Soltero
Religión	Católica	Católica	Católico
Edad	34	36	32
Estudios	Secundaria incompleta	Secundaria completa	Secundaria incompleta
Influencias en la escuela	No	No	No
Depresión	No	No	No
Ansiedad	No	Sí	No
Adicción	No	No	No
Drogas previas a delinquir	No	Si	No
Tipo de drogas	No consume	Alcohol	No consume
Reincidencia del delito	No	No	No
Centro juvenil anteriormente	No	No	No
Reincidencia de otros delitos	No	No	No

Fuente de elaboración propia

El modelo y evaluación del test en *Split* con 3 iteraciones y la suma de las distancias entre *clusters* fue 4337,68333. En la segunda iteración, las instancias agrupadas de manera aleatoria fueron:

Clúster 0: Celopatía, 'Instituto Penitenciario de Moyobamba', estado civil: separado, religión: evangélica, 62 años de edad, secundaria completa, no especifica el motivo de la deserción a los estudios, no padece de ansiedad, adicción o de depresión, consume alcohol, consumió drogas 6 horas previas al delito, es la primera vez que comete el delito, no ha reincidido anteriormente en otros delitos, refiere no haber pertenecido a establecimiento penitenciario alguno.

Clúster 1: Motivación sexual, 'Establecimiento Penitenciario Nuevo Imperial', soltero, católico, 50 años de edad, educación superior incompleta universitaria, no especifica la razón de su deserción escolar, no sufre depresión, ni ansiedad, ni adicción, no consume drogas, no las consumió 6 horas antes del crimen. Es el primer delito y no ha reincidido en otros delitos, ni haber pertenecido a un reclusorio juvenil anteriormente.

Para finalmente encontrar los siguientes *centroides* (tabla 3).

Tabla 3: Resultado de los *centroides* representativos después de la última iteración del algoritmo.

Atributos	Clúster # Celopatía (1427.0)	O=celopatía (664.0)	1=motivación sexual (770.0)
PENAL	SAN JUAN DE LURIGANCHO	SAN JUAN DE LURIGANCHO	SAN JUAN DE LURIGANCHO
Estado civil	Soltero	Conviviente	Soltero
Religión	Católica	Católica	Católico
Edad	34	36	32
Estudios	Secundaria incompleta	Secundaria completa	Secundaria incompleta
Razón de deserción	No especifica	No especifica	Factor económico

Influencias en la escuela	No	No	No Especifica
Depresión	No	No	No
Ansiedad	No	No	No
Adicción	No	No	No
Drogas previas a delinquir	No	Si	No
Tipo de drogas	No consume	Alcohol	No consume
Reincidencia del delito	No	No	No
Centro juvenil anteriormente	No	No	No
Reincidencia de otros delitos	No	No	No

Fuente de elaboración propia

Las instancias agrupadas son: 0: 198 (41 %), 1: 288 (59 %)

Se realizó un tercer experimento, y los resultados fueron similares al primero. Los factores sociales (educación, nivel de alfabetismo, deserción estudiantil, edad, estado civil), factores comunitarios (lugar de procedencia), reincidencia de delitos y factores situacionales (consumo de drogas, enfermedades mentales), en apariencia aisladas entre sí, están incluidos en los experimentos.

Un punto realmente interesante es la interpretación de otros grupos si los contrastamos con fenómenos que parecen no estar relacionados con la violencia —como es el consumo de drogas—, pero son: el grado de educación y la pertenencia a un grupo gnóstico, características resaltantes en los perfiles de los victimarios, al igual que el estado civil. Por ejemplo, en la Figura 2: el uso de drogas y su relación con la motivación para perpetrar el delito, teniendo en cuenta que las pudo utilizar 6 horas antes de cometer el crimen (eje "y", valores: sí o no); en la Figura 2, la ingesta de alcohol (eje "x", valores: tipo de droga que ingiere) tiene influencia en la celopatía (región azul), sin embargo, en quienes no ingirieron sustancia alguna la motivación fue sexual (región roja), en un 46 %, y la celopatía 54 %. [11, 12, 13, 14, 15].

Figura 2: Xplot del versus de los atributos Tipo de droga que consume y si las ingirió 6 horas antes del delito



Fuente de elaboración propia. Weka 3.8 etapa de visualización de datos.

Se afirma, entonces, que la ingesta de alcohol comúnmente asociada con los asaltos sexuales, aparece de forma más frecuente en los celópatas. Es así que se hallaron 2 grupos, identificando 2 perfiles; dependiendo de la motivación del delito y otros problemas de fondo del victimario.

Discusión

Este trabajo fue realizado por medio de *clúster* naturales, no han sufrido alteración en su estructura, de igual modo, en los resultados de la construcción del modelo. Como refiere Arora, P., & Varshney, S. (2016). Es muy probable que por esta razón, en la actualidad, hay muy pocos trabajos que investiguen las características de los métodos en términos características del *clúster* que podrían ser de interés en diferentes aplicaciones de la agrupación como método. También es probable que muchos investigadores estén preocupados por el fuerte impacto subjetivo pueda sesgar los análisis y conclusiones de las investigaciones, ya que darán resultados que dependen claramente del observador. [1, 2, 3].

El objetivo científico de las investigaciones debe ser el de apoyar el reconocimiento de las diferencias individuales y el impacto del punto de vista de los individuos, con lo que se comparte la afirmación de que las decisiones individuales y su racionalidad se hagan transparentes y sean de tal manera que la «realidad fuera de nuestro alcance» pueda transmitir un mensaje *per sé*. [4, 5].

Por ejemplo —citando a Bsoul, Q., Salim, J., & Zakaria—, las variables deben ser elegidas por su relevancia en la investigación y no para producir agrupaciones específicas motivadas por el investigador. Hay una serie de razones para tomar decisiones que dependen de los datos, en particular si el análisis inicial revela que los investigadores no formalizaron correctamente sus objetivos, en cuyo caso, una confirmación de nuevos datos fuera de la validación, normalmente sería necesaria para convencer a la audiencia de que los resultados son significativos. [3, 4, 6].

Los trabajos recopilados acerca de *clustering* específicamente en *K-means*, refieren que el análisis de *clúster* depende de contextos, objetivos y decisiones del investigador, pero, también puede ser científico, transparente y claro con respecto a sus conceptos, objetivos subyacentes y abiertos al impacto de la realidad que está fuera de nuestra observación. El análisis de *k-means* está siendo rescatado para darle a la investigación de las ciencias sociales mayor soporte desde el enfoque estadístico a gran escala. [14, 15].

Conclusiones

En los resultados, el 88,7 % de los internos no muestran signos de problemas mentales como la depresión y la ansiedad, de quien, naturalmente, puede verse afectado por atentar contra la vida y la libertad de otra persona. [15].

Los algoritmos de *clustering*, nos permiten llevar la interpretación estadística simple más allá de la descripción de los problemas. La motivación de los crímenes de género, entonces, no es singular sino más bien; llega a ser estructural, puesto que tiene influencia en factores ligados a la sociopatía, producto del medio en que se vive [11, 12, 13], como los encontrados en la visualización de los grupos o *clúster* luego de la construcción del modelo. (Figura 2)

Es concluyente que la violencia, en general, tiene muchos factores. Los seres humanos hemos tratado de explicarla para prevenirla de algún modo o, en algunos casos, controlarla. Sin duda, existen sucesos que implican problemas sociales como: la falta de educación, la deserción escolar y el consumo de drogas, que hacen más difícil la erradicación de la violencia [14, 15]. El modelo construido a partir de los 2 *clúster* determinados, nos muestran el perfil de perpetradores «motivados por los celos», con las siguientes características: estado civil: conviviente, entre 35 a 37 años, secundaria completa, no sufre depresión, ansiedad a pesar de haber pasado por situaciones traumáticas, consume alcohol, al delinquir estaba bajo efectos de las drogas y sin antecedentes penales; y, los que tuvieron «motivación sexual», con el siguiente perfil: estado civil soltero, entre 32 a 35

años, secundaria completa, no sufre depresión, ansiedad a pesar de haber pasado por situaciones traumáticas, no consume drogas y no tiene antecedentes penales.

La definición del perfil del victimario es importante para la construcción de estrategias de intervención criminalística en el nivel preventivo y, el debido tratamiento de estas personas con perfiles similares al de un potencial agresor.

Trabajos futuros

Se recomienda realizar el estudio también con *K-medoids*, para comparar resultados más densos. La data del INEI también cuenta con información acerca de la niñez de los internos, tanto de mujeres como varones, se podrían utilizar para realizar un análisis de la relación entre estos factores y la reincidencia de sus delitos.

De igual modo, la minería de datos espacio-temporal nos ayudaría a identificar los perfiles en lugares y situaciones para apoyar a las estrategias de intervención de la violencia.

Agradecimientos

Nuestro agradecimiento al Dr. Hugo Alatrística, por su guía y soporte; al Dr. Cesar Beltrán, por brindar apoyo al conocimiento científico y forma correcta que contribuye a la mejora de nuestras publicaciones. Indirectamente, quisiera agradecer al Estado, porque la apertura de los datos que gestionan los organismos públicos aporta —de manera sustancial— a la toma de decisiones, tanto en entidades privadas como públicas afines.

Contribuciones de autoría

A la Dra. Yolanda León, por su contribución del punto de vista salubrista y de género.

Fuentes de financiamiento

Ninguna, actualmente.

Conflictos de interés

La autora declara no tener conflictos de interés en la publicación de este artículo.

BIBLIOGRAFÍA

- [1] T. Velmurugan, and T. Santhanam, (2011) “A Survey of Partition based Clustering Algorithms in Data Mining: A n Experimental Approach” An experimental approach Information. Technology. Journal, Vol, 10, No .3, pp 478-484
- [2] Arora, P., & Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm For Big Data. *Procedia Computer Science*, 78, 507-512.(7)
- [3] Bsoul, Q., Salim, J., & Zakaria, L. Q. (2013). An intelligent document clustering approach to detect crime patterns. *Procedia Technology*, 11, 1181-1187.
- [4] Tar, J., Bitó, J. F., Rudas, I., & Várkonyi, T. A. (2011). International Conference on Emerging Trends in Engineering and Technology, ICETET. In 4th International Conference on Emerging Trends in Engineering and Technology, ICETET 2011.
- [5] Kusriani, K. (2015). Grouping of Retail Items by Using K-Means Clustering. *Procedia Computer Science*, 72, 495-502.
- [6] Bardales, Olga; Vásquez Hernán, (2013) *Feminicidio bajo la lupa*. Lima: MIMP.
- [7] MIMP, (2013), Estado de las investigaciones sobre violencia familiar y sexual en el Perú: 2006-2010, Lima: MIMP.
- [8] MIMP, (2013), Intervención profesional frente al feminicidio. Aportes desde los CEM para la atención y prevención, Lima: MIMP.
- [9] KANTARDZIC, Mehmed (2001). Data Mining: concepts, models, methods, and algorithms. Wiley – Interscience, USA: 2001, p. 2, 117, 165.
- [10] HAND, D., MANNILA, H., and SMYTH, P. (2001) Principles of Data Mining. The MIT Press, USA: 2001, pp. 1.
- [11], Alberto Ochoa, José-Alberto Hernández, Francisco Alvares, (2006). *Más allá del razonamiento basado en casos y una aproximación al modelo de sociedades utilizando minería de datos*. Universidad Autónoma de Zacatecas, México CP.9800.
- [12] E. Fernández, H. Merlino, D. Rodríguez, García-Martínez, I. Perversi, (S/F) Aplicación de minería de datos para la exploración y detección de patrones delictivos, ARGENTINA, tesis Licenciatura en Informática. Universidad de Morón.
- [13] Gutiérrez Rüegg, P., Merlino, H., Rancan, C., Procopio, C., Rodríguez, D., Britos, P., García-Martínez, R. (S/F) Identificación de patrones característicos de la población carcelaria mediante minería de datos. Departamento de Ingeniería Industrial. Instituto Tecnológico de Buenos Aires.
- [14] Jorge Enrique Rodríguez Rodríguez, (2008). Minería de datos para la determinación del grado de exclusión social.
- [15] Sección VII, violencia de varones: intervención, Hombres violentos, hallazgos de investigación y reflexiones para una política pública de prevención y atención, Martha Ramírez Solórzano, pág. 349.