

EL ÁRBOL DE DECISIÓN EN EL ANÁLISIS DE LA CONVIVENCIA EN CASOS DE VÍCTIMAS MORTALES DE VIOLENCIA DE GÉNERO EN ESPAÑA

*THE DECISION TREE IN THE ANALYSIS OF COEXISTENCE IN
CASES OF FATAL VICTIMS OF GENDER VIOLENCE IN SPAIN*

<https://doi.org/10.5281/zenodo.4161179>

AUTORES: Farouk Ait Nasser^{1*}

DIRECCIÓN PARA CORRESPONDENCIA: farouk.aitnasser@hotmail.com

Fecha de recepción: 09 / 03 / 2020

Fecha de aceptación: 25 / 09 / 2020

RESUMEN

El concepto de violencia de género no explica una situación obvia de convivencia en un mismo domicilio. Cuando vinculamos el índice de muertes por violencia de género tenemos que diferenciar entre parejas que conviven juntas de las que no lo hacen. Entendemos que las mujeres son víctimas de violencia de género en varios contextos, sin que se cumpla la condición de convivencia. De ahí que, sabiendo la importancia de las consecuencias como la vulnerabilidad y el desamparo que surgen de una situación donde la víctima comparte vivienda con su agresor, nuestra atención se dirige hacia la eficacia de las medidas adoptadas por las autoridades. Para ello utilizamos el árbol de decisión como herramienta de investigación con el objeto de poder interpretar la utilidad de la información proporcionada en el Portal Estadístico de la Delegación del Gobierno contra la Violencia de Género (Ministerio de Igualdad de España) para ayudarnos a entender la causa del incremento o la disminución de la tasa de mujeres víctimas mortales de violencia de género en cohabitación, por medio de la evaluación de la relevancia o falta de eficacia de las medidas tomadas al respecto. Esta investigación se enmarca en el paradigma cuantitativo

^{1*}Doctor en Análisis de los Problemas Sociales, Máster en Cultura de Paz, Conflictos, Educación y Derechos Humanos, Máster en Cultura Hispánica y Comunicación, farouk.aitnasser@hotmail.com

tomando como objetivo la interpretación de los datos y la evaluación de las medidas llevadas a cabo, durante el periodo de estudio de 2009 a 2019, capaces de ayudar a situar el índice de muertes en cohabitación, en casos de víctimas mortales de violencia de género, por de debajo del 63,63%.

Palabras clave: minería de datos, árbol de decisión, inteligencia artificial, aprendizaje inductivo, muertes en convivencia

ABSTRACT

The concept of gender-based violence does not explain an obvious situation of living together in the same home. When we link the rate of deaths due to gender-based violence we have to differentiate between couples who live together and those who do not. We understand that women are victims of gender-based violence in various contexts, without the condition of living together being met. Therefore, knowing the importance of the consequences such as vulnerability and helplessness that arise from a situation where the victim shares housing with her aggressor, our attention is directed towards the effectiveness of the measures taken by the authorities. To this end, we use the decision tree as a research tool in order to be able to interpret the usefulness of the information provided in the Statistical Portal of the Government Delegation against Gender Violence (Spanish Ministry of Equality) to help us understand the cause of the increase or decrease in the rate of women who die of gender violence in cohabitation, by assessing the relevance or lack of effectiveness of the measures taken in this regard. This research is framed within the quantitative paradigm, taking as an objective the interpretation of the data and the evaluation of the measures carried out, during the study period from 2009 to 2019, capable of helping to place the rate of deaths in cohabitation, in cases of deaths from gender violence, below 63.63%.

Keywords: data mining, decision tree, artificial intelligence, inductive learning, deaths in coexistence

INTRODUCCIÓN

La violencia de género se entiende como una violencia que se ejerce por el varón contra la mujer, debido a la condición física o posición social que introduce el género (Ferrer-Pérez & Bosch-Fiol, 2019). Se entiende, asimismo, como instrumento y estrategia de dominación

para garantizar la sumisión de la mujer (Maqueda Abreu, 2006). No parece haber causas concretas, ni factores claves que inducen a los hombres a la violencia de género, existen, pues, distintas conductas de maltrato (López García, 2004). Cuando nos referimos a las víctimas mortales de violencia de género, señalamos un fenómeno que refleja la culminación de los actos de violencia cometidos por el varón contra la mujer. Existe una posibilidad real de limitar los casos de fallecimiento de las mujeres que cohabitan con su agresor si nos centramos en el análisis de las causas, medidas y subvenciones que se desarrollan. El asesinato de mujeres por parte de su agresor/pareja se entiende como uno de los crímenes más graves en materia de violencia de género, suele tener una gran repercusión mediática y es motivo de alarma social (Rodríguez Cárcela & López Vivas, 2020).

En este estudio, nos limitamos a hallar patrones que relacionan el comportamiento del índice de muertes en convivencia con otros ámbitos de análisis. Las clasificaciones que resultan de la exposición de atributos de mayor relevancia en relación con el comportamiento de la tasa de mortalidad por cohabitación se crean de forma automática empleando la inteligencia artificial.

Hipótesis

En lo que se refiere a la predicción de delitos, la interpretación de los resultados es tan importante como su precisión, por ello se debe presentar resultados fácilmente comprensibles para los usuarios que no son expertos en algoritmos de predicción (Wijenayake, Graham, & Christen, 2019). De todos modos, con el árbol de decisión contamos con un algoritmo que presenta resultados fácilmente interpretables (Baghdadi, Mallet, & Zribi, 2020), y es uno de los métodos más populares y frecuentes usados en minería de datos (Bouyer, Arasteh, & Movaghar, 2007).

Enfocándonos únicamente en las víctimas mortales de violencia de género en situación de convivencia, podemos analizar como el índice de fallecimientos puede bajar – por debajo de la media – si se puede discernir entre las medidas eficaces o si se puede establecer un patrón que determina las estrategias que se deban seguir. Como hipótesis, consideramos que esto es posible si evaluamos correctamente el resultado de la reacción política

gubernamental española, a través de las medidas que se toman en materia de violencia de género por medio del uso de la inteligencia artificial.

METODOLOGÍA

Con el propósito de alcanzar nuestro objetivo de investigación, consideramos apropiado recurrir a la inteligencia artificial y, más precisamente, el aprendizaje automático como análisis cuantitativo de datos.

El aprendizaje automático incluye varias herramientas, aun siendo muy útiles, generalmente se enfocan en predicciones de resultados. El interés que promueve esta investigación está vinculado con poder evaluar la eficacia de las medidas adoptadas, que ayudan a reducir el índice de muertes en convivencia por debajo de la media (2009-2019). Se opta por el uso del aprendizaje supervisado, más concretamente, el árbol de decisión, no para resolver un problema de regresión, sino de clasificación, con el objeto de descubrir y reconocer patrones, seleccionando y clasificando atributos proporcionados en los datos de entrada, para encontrar una estructura comprensible que ofrece una mejor interpretación de los datos que comprende la muestra. Sostenemos que de este modo podemos determinar la existencia de la correlación entre el conjunto de datos que poseemos (variables de entrada) y el índice de muertes en convivencia (variable de destino).

Gracias al aprendizaje inductivo, se desarrolla un proceso que permite inferir resultados que favorezcan la revelación de patrones que explican las fluctuaciones de la variable de destino, si es por encima o por debajo de la media, pudiendo observar dichos resultados por medio del árbol de decisión. El aprendizaje inductivo, que pertenece al área de la inteligencia artificial, es una técnica que transforma el conocimiento obtenido a un conjunto de reglas o en un árbol de decisión (Priore, De la Fuente, Pino, & Puente, 2004).

Los pasos seguidos en esta investigación

Para establecer la correspondiente correlación entre el conjunto de los ámbitos de análisis y el índice de muertes en convivencia, hemos utilizado el lenguaje de programación *python* (versión 3.7.3). En el proceso de obtención de resultados, se ha empleado la librería *sklearn* (<https://scikit-learn.org>). El árbol de decisión o clasificación es la herramienta empleada para realizar la selección y la clasificación de los atributos, funciona como un mecanismo deductivo a la hora de procesar los datos e inferir patrones.

Conviene señalar que por medio de la librería *sklearn* se utiliza el índice de Gini por defecto, en nuestro caso hemos mantenido este parámetro para medir la calidad de la clasificación. La fórmula matemática utilizada para medir la utilidad de las clasificaciones es la siguiente:

$$\text{Índice de Gini} = 1 - \sum_j P_j^2$$

Para poder cubrir un periodo de análisis de 10 años, desde 2009 a 2019, en este proceso hemos recogido únicamente los datos disponibles a partir del año 2009. Así pues, hemos tenido que descartar los siguientes ámbitos de análisis: “el sistema de seguimiento integral en los casos de violencia de género” y “usuario Atempro”.

El siguiente gráfico muestra el conjunto de datos recogidos en esta investigación, que clasificamos del siguiente modo:

Figura 1. Ámbitos de análisis recogidos para el análisis y la minería de datos

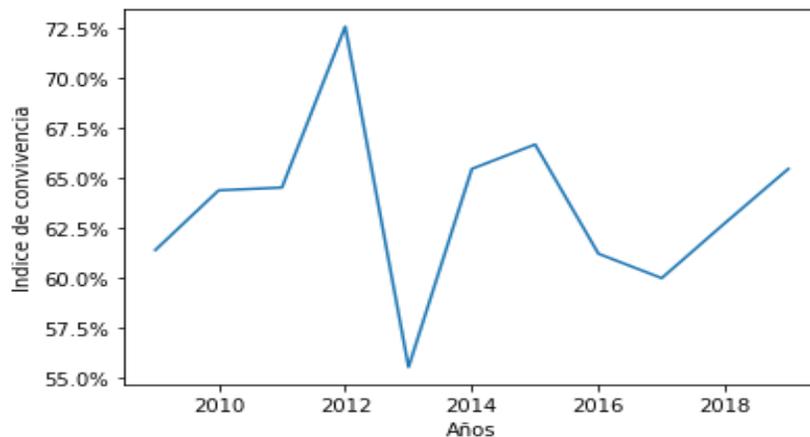
<u>Ayudas económicas</u>	<u>Medidas contra el agresor</u>	<u>Consecuencias de los malos tratos</u>
<ul style="list-style-type: none"> • Renta activa de inserción • Contratos bonificados • Contratos de sustitución • Ayudas concedidas Art 27 • Ayudas para cambio de residencia 	<ul style="list-style-type: none"> • Órdenes de protección • Dispositivos electrónicos activos 	<ul style="list-style-type: none"> • Denuncias por violencia de género • Llamadas pertinentes • Mujeres víctimas mortales

Fuente: Elaboración propia

Después de recabar los datos, nos hemos percatado de la dificultad de obtener resultados válidos para un análisis con distintas perspectivas de aplicación. Esto es, que la información recopilada no es suficiente si nuestro objetivo comprende, asimismo, el interés por la aplicación de varias herramientas de análisis.

De acuerdo con los datos recabados en el Portal Estadístico de la Delegación del Gobierno contra la Violencia de Género, observamos que desde el año 2009 a 2019, el índice de muertes en convivencia oscila entre el 55 y el 72,5%, dándonos una media de 63,63%.

Figura 2. Datos relativos a las muertes en convivencia expresados en porcentajes (2009-2019)



Fuente: Elaboración propia

Con el fin de poder tener información válida y manejable para la minería de datos, se ha optado por la conversión de la tasa de muertes en cohabitación a valores binarios, se ha asignado un “0” a los valores superiores a la media (0.6363), y “1” a los que están por debajo de esta, tal como se observa en la siguiente tabla:

Tabla 1. Índice de muertes en convivencia

Año	Mujeres víctimas mortales	Convivencia	Tasa media de convivencia	Valor binario
2009	57	35	0.6140	1
2010	73	47	0.6438	0
2011	62	40	0.6452	0
2012	51	37	0.7255	0
2013	54	30	0.5556	1
2014	55	36	0.6545	0
2015	60	40	0.6667	0
2016	49	30	0.6122	1
2017	50	30	0.6	1
2018	51	32	0.6275	1
2019	55	36	0.6545	0

Fuente: Elaboración propia

Análisis de los resultados

Tras introducir todos los atributos que incluye la muestra en forma de base de datos completa para su procesamiento – que incluye todos los ámbitos de análisis mostrados en la figura 1–, se ha procedido a llevar a cabo las clasificaciones correspondientes. Una vez hallado el atributo más útil, se crea una regla de decisión, y luego se opta por hallar el subconjunto de atributos más significativos.

A partir de ahí, hemos observado que, del conjunto de los registros de entrenamiento, el algoritmo nos ha devuelto varios resultados, cada resultado es un árbol de decisión distinto, que toma en consideración las variables que dividen de forma eficaz el conjunto de atributos. De este modo, en cada resultado de clasificación observamos las variables más significativas en la creación del árbol. Sin embargo, en el presente estudio, no podemos mostrar cada resultado con su correspondiente análisis, pero mostraremos el que presenta mayor precisión y revela un patrón más consistente.

Conviene aclarar que, al recopilar los datos en el mes de febrero de 2020, observábamos que en el año 2019 el número de órdenes de protección era de 30765, el algoritmo identificaba un patrón bastante claro. Cuando se superaba la media de 37933, el índice de muertes en convivencia siempre estaba debajo de la media (63,63%). Con la última actualización, realizada el 25/03/2020, se modificó a 40720, y las denuncias por violencia de género también, de 125936 a 168168. Los demás ámbitos de análisis no se modificaron.

Una vez actualizados los datos de la muestra, repetimos el proceso de minería de datos y deducimos del análisis de los resultados las variables de entrada más significativas, las que ejercen una clara influencia en el comportamiento de la variable de destino. Concluimos que las denuncias por violencia de género es el atributo más significativo. Ya que, presenta una clara asociación con la variable de destino. Es el que define el nodo raíz en el ejemplo que presentamos (figura 3). Observamos que en todos los casos cuando las denuncias son inferiores a 134822 no se alcanza el objetivo de minimizar la tasa de muertes en convivencia excepto en el año 2013, pero cuando es mayor, siempre se cumple el objetivo de tener una tasa de fallecimiento por cohabitación por debajo de la media, excepto en el año 2019.

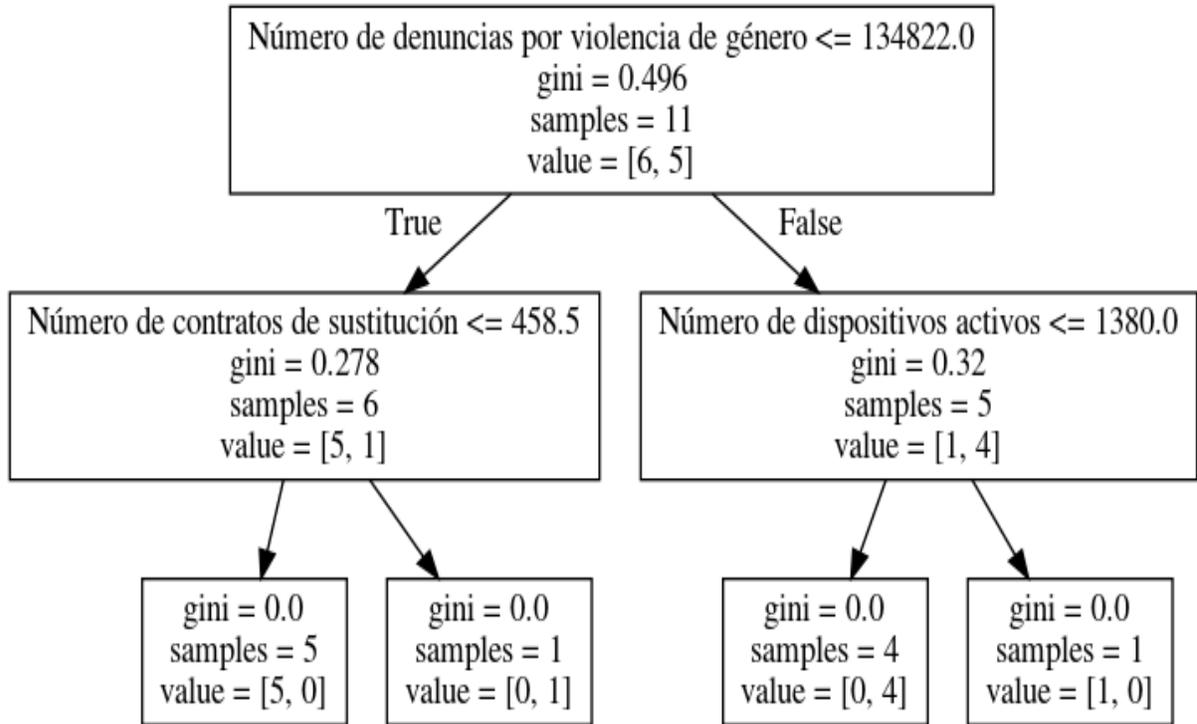
Tabla 2. Denuncias por violencia de género que superan la media anual de 134822

Año	Denuncias por violencia de género	Índice de convivencia	Valor binario
2009	135539	0.6140	1
2010	134105	0.6438	0
2011	134002	0.6452	0
2012	128477	0.7255	0
2013	124893	0.5556	1
2014	126742	0.6545	0
2015	129193	0.6667	0
2016	143535	0.6122	1
2017	166260	0.6000	1
2018	166961	0.6275	1
2019	168168	0.6545	0

Fuente: Elaboración propia

El árbol de clasificación recogido en este estudio responde a nuestro criterio de mostrar las reglas de decisión más coherentes y significativas. Cada nodo de hoja constituye una decisión, el nodo raíz está formado por la variable denuncias por violencia de género, en los nodos descendientes encontramos los contratos de sustitución y los dispositivos activos. El árbol en sí, mostrado en la figura 3, resulta bastante ilustrativo, ya que refleja una regla de decisión bastante sólida, y no se cumple debido a dos condiciones. Cuando las denuncias por violencia de género superan a 134822, la tasa de mortalidad en convivencia se sitúa por debajo de la media, excepto cuando los dispositivos activos superan a 1380, lo que nos hace pensar que se trata de una medida contraproducente. Sin embargo, cuando el número de denuncias por violencia de género baja de 134822, existe una sola condición de situar el índice de muertes por convivencia por encima de la media, aumentando los contratos de sustitución.

Figura 3. Árboles de clasificación elegido



Fuente: Elaboración propia

DISCUSIÓN DE LOS RESULTADOS

En nuestra revisión de la literatura, hemos hallado varios estudios afirmando que en España no existe una asociación entre las medidas penales o civiles adoptadas por las instituciones públicas y la tasa de mortalidad por la violencia de género (Sanz-Barbero, Heras-Mosterio, Otero-García, & Vives-Cases, 2016). En verdad, difícilmente podríamos hallar la relación, no obstante, en el presente estudio, al enfocarnos en el análisis de la tasa de mortalidad en cohabitación, hemos podido encontrar una asociación entre el índice de mortalidad en cohabitación y las denuncias por violencia de género. Consideramos que en cuanto se alcance una cantidad de datos ampliamente mayor, podríamos reconocer más patrones, formando ideas más claras sobre la influencia que ejerce el conjunto de las medidas cautelares, penales y civiles, así como las medidas en materia de servicios sociales implementadas por medio de las normativas de protección integral frente a la violencia contra las mujeres, en asociación con la tasa de muertes en convivencia.

CONCLUSIONES

Estimamos acertado la adopción de toda medida que ayude a las mujeres a tomar la decisión de denunciar a su agresor, y más precisamente si cohabitan con él. Ya que, de lo anteriormente expuesto, deducimos la existencia de un patrón que asocia las denuncias por violencia de género y el índice de muertes en convivencia. Concluimos que cuantas más denuncias, menos posibilidades tiene la mujer de fallecer en manos de su agresor.

Asimismo, sostenemos que no se deben infravalorar las órdenes de protección como medida cautelar, ni disminuir su media anual. Observamos que existe un aumento progresivo en su número desde el año 2015, pero, aun así, creemos que se le debe dar más importancia por parte de las instituciones públicas para aumentar su eficacia.

Por otra parte, en caso de manejar una muestra que comprende un volumen alto de información vinculada al tema tratado, ayudaría a mejorar los resultados de toda labor científica de investigación.

Finalmente, pensamos conveniente el desarrollo de varios estudios con relación a esta temática y precisamente aquellos que buscan la objetividad por medio del análisis de datos con el objeto de dar una explicación coherente del modo en cómo se debe actuar para minimizar las consecuencias de este fenómeno. Profundizar únicamente en el sufrimiento de la víctima no resuelve el problema. Una gestión eficaz y racional de los recursos públicos destinados a ayudar a la mujer que sufre de los malos tratos se presenta necesaria e imprescindible.

REFERENCIAS BIBLIOGRÁFICAS

Baghdadi, N., Mallet, C., & Zribi, M. (2020). *QGIS y sus aplicaciones en agua y en gestión del riesgo*. London: ISTE Group.

Bouyer, A., Arasteh, B., & Movaghar, A. (2007). A new hybrid model using case-based reasoning and decision tree methods for improving speedup and accuracy. *IADIS International Conference Applied Computing*, 787-789.

Ferrer-Pérez, V., & Bosch-Fiol, E. (2019). El género en el análisis de la violencia contra las mujeres en la pareja: de la “ceguera” de género a la investigación específica del mismo. *Anuario de Psicología Jurídica*, 69-76.

López García, E. (2004). La figura del agresor en la violencia de género: características personales e intervención. *Papeles del psicólogo*, 31-38.

Maqueda Abreu, M. (2006). La violencia de género: entre el concepto jurídico y la realidad social. *Revista electrónica de ciencia penal y criminología*, 1-13.

Priore, P., De la Fuente, D., Pino, R., & Puente, J. (2004). Utilización del aprendizaje inductivo en la toma de decisiones. Aplicación en un problema de secuenciación. *Investigaciones europeas de dirección y economía de la empresa*, 17-36.

Rodríguez Cárcela, R., & López Vivas, A. (2020). Tratamiento informativo de la violencia de género: asesinatos de mujeres. Análisis de la agencia EFE. *Ámbitos. Revista Internacional de Comunicación*, 23-60.

Sanz-Barbero, B., Heras-Mosterio, J., Otero-García, L., & Vives-Cases, C. (2016). Perfil sociodemográfico del feminicidio en España y su relación con las denuncias por violencia de pareja. *Gaceta sanitaria*, 272-278.

Wijenayake, S., Graham, T., & Christen, P. (2019). A decision tree approach to predicting recidivism in domestic violence. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 3-15.