

# Algoritmo de minería de datos para la clasificación de violencia en pareja.

## Data Mining Algorithm for Classification of Intimate Partner Violence.

Miriam Vargas Reyes (1).  
Tecnológico Nacional de México/TES Valle de Bravo, Estado de México, Valle de Bravo.  
[L201607039@vbravo.tecnm.mx](mailto:L201607039@vbravo.tecnm.mx).

Mariana Carolyn Cruz Mendoza\* (2). Tecnológico Nacional de México/TES Valle de Bravo, Estado de México, Valle de Bravo. [mariana.cm@vbravo.tecnm.mx](mailto:mariana.cm@vbravo.tecnm.mx).

Cesar Primero Huerta (3). Tecnológico Nacional de México/TES Valle de Bravo, Estado de México, Valle de Bravo.  
[cesar.ph@vbravo.tecnm.mx](mailto:cesar.ph@vbravo.tecnm.mx).

---

\*corresponding author.

Artículo recibido en diciembre 18, 2020; aceptado en febrero 18, 2021.

### Resumen.

*Desde hace muchos años, la violencia, en todas sus clasificaciones, ha sido considerada como parte de nuestra cultura, y de esta forma, se acepta como un integrante en la formación de la sociedad, donde se daña, principalmente, la integridad de la mujer, causando daños, no solo físicos, sino también psicológicos, por lo tanto, es necesario atender el problema, para que de algún modo se pueda frenar o minimizar. La presente investigación aborda la violencia en pareja, donde aplicando instrumentos de recolección de datos se obtendrá información de diferentes participantes mujeres que sufren o han sufrido algún tipo de violencia, teniendo como propósito principal implementar algoritmos de minería de datos para clasificarla en 4 grupos; violencia física, psicológica, sexual y verbal y con ello generar planes de acción recomendados por un especialista.*

**Palabras clave:** Violencia de pareja, algoritmos de clasificación, minería de datos, planes de acción.

### Abstract.

*For many years, violence, in all its classifications, has been considered as part of our culture, and in this way, it is accepted as an integral part of the formation of society, where the integrity of women is mainly damaged, causing damage, not only physical, but also psychological, therefore, it is necessary to address the problem, so that it can somehow be stopped or minimized. This research addresses intimate partner violence, where, by applying data collection instruments, information will be obtained from different female participants who suffer or have suffered some type of violence, with the main purpose of implementing data mining algorithms to classify it into 4 groups; physical, psychological, sexual and verbal violence and thereby generate action plans recommended by a specialist.*

**Keywords:** Intimate partner violence, classification algorithms, data mining, action plans

## 1. Introducción.

La violencia ha existido desde tiempos muy remotos hasta llegarse a considerar como parte de nuestra cultura. Los factores que la caracterizan, como el comportamiento agresivo, ha estado presente a través de todos los tiempos, y ha quedado plasmada en documentos que van desde las antiguas escrituras hasta las estadísticas actuales. Gran parte de estas cifras se centran en la violencia de género, ya que, organismos internacionales como ONUMUJERES, estima que el 35 % de las mujeres de todo el mundo ha sufrido algún tipo de violencia ya sea física y/o sexual, por parte de una persona con lazos sentimentales. Es por esto, que es necesario atender el problema, para que de algún modo se pueda frenar o minimizar buscando soluciones con ayuda de las tecnologías de la información. El siguiente documento presenta una investigación acerca de la violencia en pareja, la información será utilizada para implementar algoritmos de clasificación como J48 o árboles de decisión y Bayes Net haciendo uso de minería de datos ya que ésta ayuda a analizar grandes volúmenes de información, para clasificar, agrupar o crear modelos predictivos.

Dichos datos serán recolectados mediante una encuesta, que incluye datos personales y características matrimoniales o de relación de las participantes. La clasificación que se desea obtener es verbal, sexual, psicológica y física para recomendar planes de acción a las víctimas de acuerdo con la violencia que están viviendo, dichos planes serán recomendados por un especialista.

En la actualidad existen diversas aplicaciones web, donde las víctimas de violencia pueden ingresar de manera anónima y contestar encuestas, como, por ejemplo, “Encuesta sobre violencia contra la mujer” por parte del Instituto Mexicano de Seguro Social, con el propósito de conocer si la víctima sufre de violencia, estos datos son enviados y posteriormente colocan enlaces donde se puede pedir ayuda o simplemente se envían los datos y no se tiene ninguna otra función. Por otra parte, se han desarrollado diversas investigaciones dentro del marco de minería de datos, por ejemplo, el proyecto “Aplicación de minería de datos para determinar las causas de la violencia de género contra las mujeres en Ecuador” donde se analiza información para determinar cuáles son las causas principales de la violencia hacia la mujer, se enfoca en la violencia de pareja, noviazgo, en la niñez y en el ámbito social, educativo y laboral.

## 2. Métodos.

### Metodología.

KDD (Knowledge discovery in databases): Es una de las metodologías dominantes para el proceso de la minería de datos.

#### Fases

- Selección o búsqueda de información: Investigaciones que proporcionen información sobre el problema, lineamientos para la obtención de bases de datos
- Preprocesamiento: obtención de datos, cumplimiento de lineamientos establecidos
- Transformación: depuración de base de datos, selección de información concreta para la resolución del problema
- Minería de datos: Aplicación de algoritmos de clasificación orientados a cada caso
- Evaluación e implantación: interpretación de los datos: presentación de resultados simplificados y comprensibles.

### Algoritmos de clasificación.

- Redes Bayesianas: Son una representación gráfica donde se modelan dependencias para el razonamiento probabilístico, consiste en una red que contiene nodos que representan las variables aleatorias y los arcos que representan las relaciones de dependencia directa entre dichas variables.
- Árboles de clasificación, algoritmo J48 (Clasificador C4.5): El algoritmo J48 de WEKA es una implementación del algoritmo C4.5, algoritmo de minería de datos más utilizado. (Sierra, s.f.).
- El algoritmo C4.5 permite generar un árbol de decisión a partir de datos obtenidos.

- Árboles de decisión: son una representación de los procesos que se involucran en las tareas de clasificación.

**Software de entrenamiento Weka.**

WEKA (Waikato Environment for Knowledge Analysis) es un software gratuito de aprendizaje automático y minería de datos distribuido bajo licencia GNU-GPL. Tiene una gran colección de algoritmos para tareas de minería de datos y modelado predictivo. También incluye herramientas para el preprocesamiento de los datos (filtros), clasificación (árboles, tablas), clustering, reglas de asociación, y adicionalmente, diversas formas de visualización de los datos, tanto en el inicio del proceso de carga de datos, como después de haber aplicado un algoritmo (Pineda, 2017).

**Muestra.**

Para poder armar el banco de datos es necesario tener al menos 100 datos, por lo tanto, fue necesario elegir tres comunidades; Loma de Rodríguez, La Rinconada y El castellano, pertenecientes al municipio de Valle de Bravo, Estado de México. Teniendo un total de 142 parejas en las tres comunidades.

Se utilizó el software SURVEYMONKEY, donde se ingresa el tamaño de la población, que sería el total de parejas, con un nivel de confianza del 95% y con un margen de error del 5%, dando un total de 104 mujeres para encuestar.

**Recolección de datos.**

El siguiente instrumento de recolección de datos se creó con ayuda de una encuesta validada del autor Javier Gustavo Fernández Teruelo, la cual recolecta información de violencia que sufre la víctima, en situación de pareja o expareja, para ayudar a predecir y prevenir feminicidios. Así mismo se utilizaron algunas características del instrumento que creó la doctora Martha Alicia Tronco Rosas directora de la Unidad Politécnica de Gestión con Perspectiva de Género del Instituto Politécnico Nacional, denominado “Violentometro”. Por otro lado, para obtener los datos personales se acudió a un trabajo de investigación denominado Escala de Evaluación del Tipo y Fase de Violencia de Género (EETFVG). Así mismo se pidió ayuda para la revisión de la encuesta, a un científico calificado.

Proyecto “Algoritmo de minería de datos para la clasificación de violencia en pareja”

Encuesta

**Objetivo:** Obtener información de mujeres que viven en pareja de las comunidades de Loma de Rodríguez, La Rinconada y El Castellano para conocer si sufren violencia y con ello aplicar técnicas de minería de datos para clasificarla en física, verbal, sexual y psicológica.

**Instrucciones.** Indique de forma voluntaria sus características de:

A. Datos Personales		
Edad:		
Estado civil:	<input type="checkbox"/> Casado	<input type="checkbox"/> Unión libre
Situación económica:	<input type="checkbox"/> Clase alta	<input type="checkbox"/> Clase media <input type="checkbox"/> Clase baja
Trabaja:	<input type="checkbox"/> Si	<input type="checkbox"/> No
Número de hijos:		
Lugar de nacimiento:		
Lugar de residencia:		
B. Religión		
<input type="checkbox"/> Católica	<input type="checkbox"/> Bíblicas diferentes a evangélicas	
<input type="checkbox"/> Protestantes y evangélicas	<input type="checkbox"/> Otro	
C. Formación integral de la salud		
¿Cómo considera su estado de salud?		
<input type="checkbox"/> Malo	<input type="checkbox"/> Regular	
<input type="checkbox"/> Bueno	<input type="checkbox"/> Excelente	
¿Tiene alguna dependencia sobre las siguientes sustancias?		
<input type="checkbox"/> Alcohol	<input type="checkbox"/> Tabaco	<input type="checkbox"/> Sustancias ilegales
D. Nivel de educación y analfabetismo		
<input type="checkbox"/> No sabe leer ni escribir	<input type="checkbox"/> Lee y escribe	<input type="checkbox"/> Primaria
<input type="checkbox"/> Secundaria	<input type="checkbox"/> Preparatoria	<input type="checkbox"/> Estudios universitarios
E. Factores de vulnerabilidad de la víctima		
Discapacidad	<input type="checkbox"/> Si	<input type="checkbox"/> No
En periodo de gestación	<input type="checkbox"/> Si	<input type="checkbox"/> No
Victima extranjera	<input type="checkbox"/> Si	<input type="checkbox"/> No

**Figura 1.** Encuesta. Datos personales. Elaboración propia

## F. Situación matrimonial o de relación

¿Ha existido algún tipo de violencia por parte del agresor?	<input type="checkbox"/> Si	<input type="checkbox"/> No
¿Ha existido reacción defensiva de la víctima ante la agresión?	<input type="checkbox"/> Si	<input type="checkbox"/> No
¿Su pareja ha empleado armas u objetos contra usted?	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja empleó:	<input type="checkbox"/> Arma blanca	<input type="checkbox"/> Arma de fuego
¿Su pareja tiene acceso a armas de fuego?	<input type="checkbox"/> Si	<input type="checkbox"/> No
Señale marcando un Si o un No si su pareja le ha hecho alguna de las siguientes acciones:		
Su pareja la ha acosado	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja ha intentado aislarla de su familia y/o amigos	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja le ha hecho bromas hirientes	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja la ha chantajeado	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja intenta controlarla	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja la culpabiliza	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja la ha descalificado	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja ha disminuido su persona	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja no muestra interés o tiene falta de atención hacia usted	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja constantemente le grita	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja le dice groserías	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja la ha humillado	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja la ignora	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja la insulta	<input type="checkbox"/> Si	<input type="checkbox"/> No
Se siente intimidada por su pareja	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja le ha dejado de hablar o le ha aplicado la ley de hielo	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja la ha maldecido	<input type="checkbox"/> Si	<input type="checkbox"/> No
Su pareja la ha manipulado	<input type="checkbox"/> Si	<input type="checkbox"/> No

Figura 2. Fragmento de encuesta datos matrimoniales. Elaboración propia

## 3. Desarrollo.

## Análisis de datos.

Una vez que se diseñó el instrumento de recolección de datos se localizaron 86 variables para el entrenamiento de los datos, las cuales se describen a continuación.

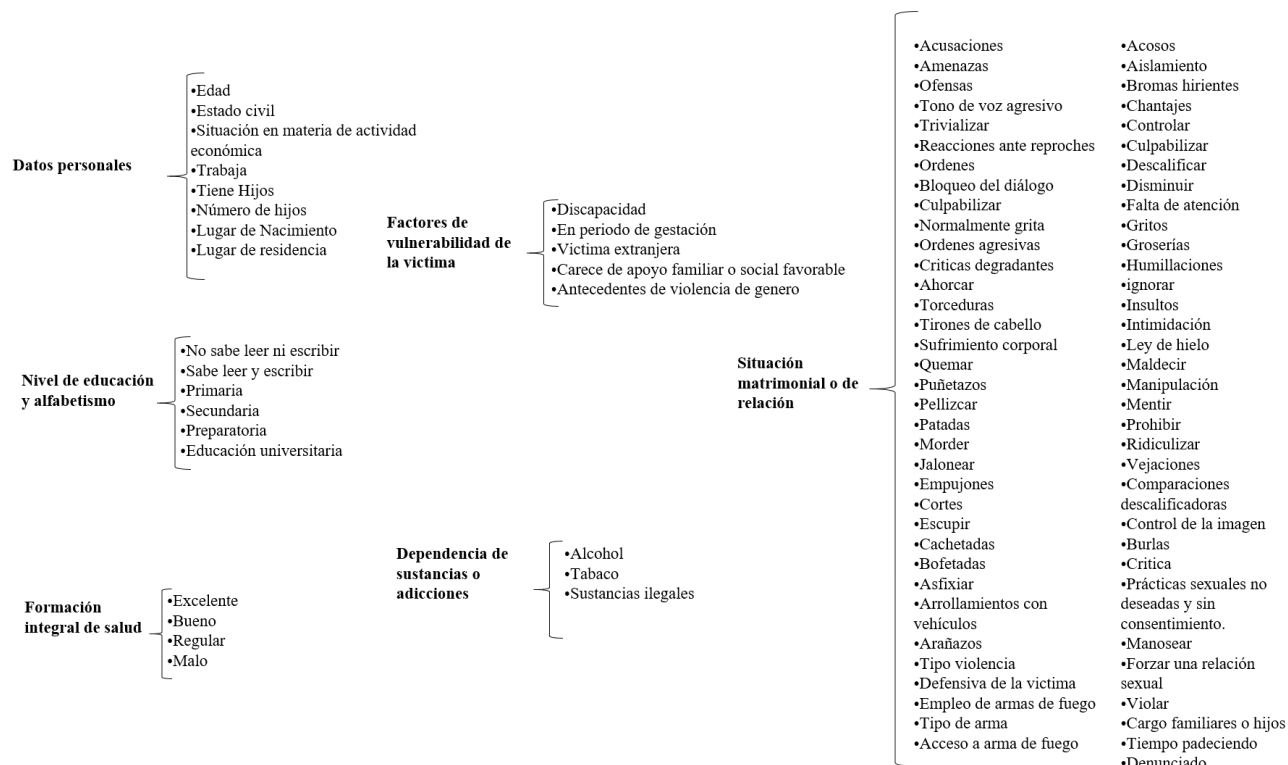


Figura 3. Variables localizadas

**Aplicación de la encuesta y creación del dataset.**

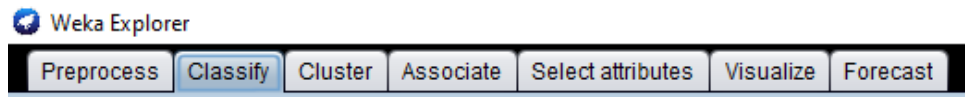
Las entrevistas se aplicaron de manera presencial, debido a que las comunidades seleccionadas son rurales, mucha de las mujeres no tiene acceso a internet o teléfonos celulares, para que se haya aplicado de manera digital. En un análisis previo las encuestas contestadas arrojaron que la edad de las participantes se repite más de los 29 a 50 años y la mayoría están casadas por el civil y por la iglesia, todas las mujeres practican la religión católica y la mitad trabajan, la mayoría de las mujeres arriba de 50 años no lee ni escribe y muy pocas saben escribir y leer, así mismo debajo de la edad de 49 años sabe leer y escribir, asistieron a la educación básica y superior y muy pocas mujeres tienen estudios de nivel medio superior y superior. Su estado de salud se encuentra en el nivel “Bueno”, no tienen discapacidad y son originarias de Valle de Bravo, así mismo solo tres o cuatro mujeres están en periodo de gestación. Los resultados obtenidos se colocaron en una hoja de cálculo de Excel para crear el dataset para el entrenamiento de los datos, a continuación, se muestra un fragmento.

edad	estado civil	situacion_economica	trabaja	hijos	numero_hijos	lugar_nacimiento	lugar_residencia	religion	formacion_integral_salud	adcciones	grado_estudios	discapacidad	periodo_gestacion	victima_extranjera	carece_apoyo	antecedentes_violencia	tipo_violencia_agresor	defensiva_victima
42	Casado	Clase media	No	Si	3	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Primaria	No	No	No	No	No	No	No
30	Casado	Clase media	Si	Si	2	Cerro Colora	Loma de Rod	Catolica	Bueno	Ninguna	Estudios univ	No	No	No	No	No	No	No
60	Casado	Clase baja	No	Si	18	La Rinconada	Loma de Rod	Catolica	Regular	Ninguna	No sabe leer	No	No	No	No	No	No	No
30	Casado	Clase media	Si	Si	2	La Mesa	Loma de Rod	Catolica	Bueno	Ninguna	Primaria	No	No	No	No	No	Si	No
21	Union libre	Clase media	No	Si	1	El Castellano	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
47	Casado	Clase media	Si	Si	3	Rincon de Es	Loma de Rod	Catolica	Bueno	Ninguna	Lee y escribe	No	No	No	No	No	Si	No
48	Casado	Clase media	No	Si	5	Loma de Chil	Loma de Rod	Catolica	Bueno	Ninguna	No sabe leer	No	No	No	No	No	No	No
41	Casado	Clase media	Si	Si	5	Rincon de Es	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	Si	No	No	No
45	Union libre	Clase media	Si	Si	4	Rincon de Es	Loma de Rod	Catolica	Bueno	Ninguna	Primaria	No	No	No	No	No	Si	No
34	Casado	Clase media	Si	Si	2	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
56	Union libre	Clase baja	Si	Si	6	Loma de Rod	Loma de Rod	Catolica	Regular	Ninguna	Lee y escribe	No	No	No	Si	No	Si	Si
49	Casado	Clase media	Si	Si	5	San Mateo	Loma de Rod	Catolica	Regular	Ninguna	Lee y escribe	No	No	No	No	No	Si	Si
28	Union libre	Clase media	Si	Si	2	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	Si	No	No	No	No	No
65	Casado	Clase baja	No	Si	2	El Fresno	Loma de Rodriguez	Regular	Alcohol	No sabe leer	No	No	No	No	Si	No	Si	No
30	Union libre	Clase media	Si	Si	2	Loma de Chil	Loma de Rod	Catolica	Regular	Ninguna	Secundaria	No	No	No	No	No	Si	No
27	Union libre	Clase media	Si	Si	1	Valle de Brav	Loma de Rod	Catolica	Bueno	Ninguna	Estudios univ	No	No	No	No	No	No	No
33	Union libre	Clase media	Si	Si	3	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	Si	Si
21	Union libre	Clase media	No	Si	2	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Estudios univ	No	No	No	No	No	No	No
49	Union libre	Clase media	Si	Si	3	Avandaro	Loma de Rod	Catolica	Regular	Ninguna	No sabe leer	No	No	No	No	No	No	No
24	Union libre	Clase media	Si	No	0	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
25	Union libre	Clase media	Si	Si	1	Polvillos Am	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
27	Union libre	Clase media	Si	Si	1	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
40	Casado	Clase baja	No	Si	2	Loma de Chil	Loma de Rod	Catolica	Bueno	Ninguna	Lee y escribe	No	No	No	No	No	No	No
70	Casado	Clase media	No	Si	9	Los Alamos	Loma de Rod	Catolica	Bueno	Ninguna	No sabe leer	No	No	No	No	No	Si	No
36	Union libre	Clase media	Si	Si	3	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
34	Casado	Clase media	Si	Si	2	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
30	Union libre	Clase media	Si	No	0	Rincon de Es	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	Si	No
40	Casado	Clase media	Si	Si	3	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
32	Union libre	Clase media	Si	Si	2	Barrio de Gu	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
89	Casado	Clase media	No	Si	8	El Castellano	Loma de Rod	Catolica	Regular	Ninguna	Lee y escribe	No	No	No	No	No	Si	No
46	Casado	Clase media	Si	Si	3	Los Alamos	Loma de Rod	Catolica	Bueno	Ninguna	Lee y escribe	No	No	No	No	No	No	No
65	Casado	Clase media	No	Si	5	San Pedro	Loma de Rod	Catolica	Regular	Tabaco	Lee y escribe	No	No	No	Si	No	Si	Si
52	Casado	Clase media	No	Si	13	El Castellano	Loma de Rod	Catolica	Regular	Ninguna	No sabe leer	No	No	No	Si	No	Si	No
53	Casado	Clase media	No	Si	5	La Rinconada	Loma de Rod	Catolica	Regular	Ninguna	Lee y escribe	No	No	No	No	No	Si	No
26	Casado	Clase media	Si	Si	2	Cerro Colora	Loma de Rod	Catolica	Bueno	Ninguna	Preparatoria	No	No	No	No	No	Si	Si
48	Union libre	Clase media	No	Si	6	San Juan Ate	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
56	Casado	Clase media	No	No	0	Loma de Rod	Loma de Rod	Catolica	Regular	Ninguna	Primaria	No	No	No	Si	No	No	No
49	Casado	Clase media	Si	Si	6	Loma de Rod	Loma de Rod	Catolica	Regular	Ninguna	Secundaria	No	No	No	Si	No	No	No
55	Casado	Clase media	No	Si	5	Rincon de Es	Loma de Rod	Catolica	Bueno	Ninguna	Lee y escribe	No	No	No	No	No	No	No
32	Casado	Clase media	Si	Si	1	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
68	Casado	Clase media	No	Si	4	San Mateo	Loma de Rod	Catolica	Regular	Ninguna	Lee y escribe	No	No	No	No	No	Si	No
50	Casado	Clase media	No	Si	4	San Mateo	Loma de Rod	Catolica	Bueno	Ninguna	Estudios univ	No	No	No	No	No	No	No
28	Union libre	Clase media	Si	Si	3	San Mateo	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
38	Union libre	Clase media	Si	Si	3	San Mateo	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
28	Union libre	Clase media	Si	Si	1	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	No	No
34	Casado	Clase baja	No	Si	6	Loma de Rod	Loma de Rod	Catolica	Bueno	Ninguna	Lee y escribe	No	No	No	Si	Si	Si	Si
26	Union libre	Clase media	No	Si	1	Rincon de Es	Loma de Rod	Catolica	Bueno	Ninguna	Secundaria	No	No	No	No	No	Si	No
32	Union libre	Clase media	No	Si	2	San Juan Ate	Loma de Rod	Catolica	Bueno	Ninguna	Estudios univ	No	No	No	No	No	No	No
51	Casado	Clase media	No	Si	5	Ciudad de M	Loma de Rod	Catolica	Regular	Ninguna	Primaria	No	No	No	No	No	No	No

Figura 4. Fragmento de dataset. Elaboración propia.

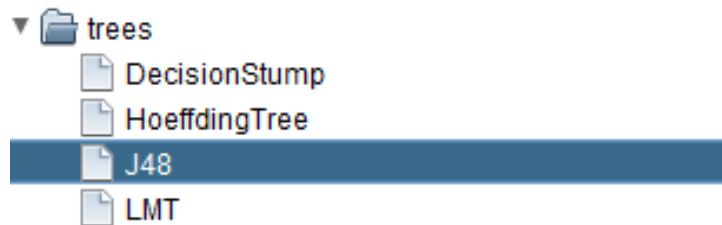
**Implementación del algoritmo J48.**

Los archivos contenidos en la hoja de cálculo Excel, serán guardados en archivo CSV, para que el software de entrenamiento de datos pueda leerlos. La implementación del algoritmo J48 genera arboles de clasificación para la toma de decisiones, utiliza únicamente variables de tipo cualitativo. Para poder utilizarlo es necesario abrir al archivo de tipo CSV con los datos que se desean entrenar mediante este algoritmo. En la parte superior izquierda localizar CLASSIFY y dar clic.



**Figura 5.** Menú. Software Weka.

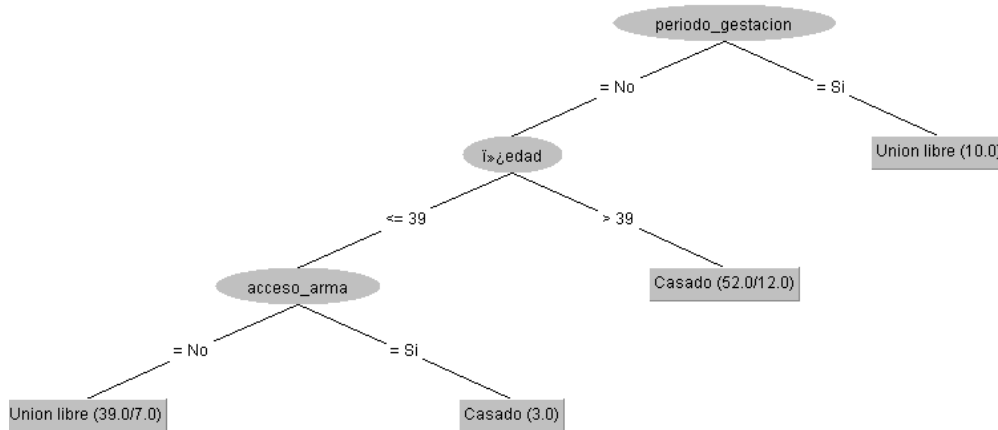
Localizar el apartado de TREES y elegir el algoritmo J48



**Figura 6.** Seleccionar algoritmo J48. Software Weka.

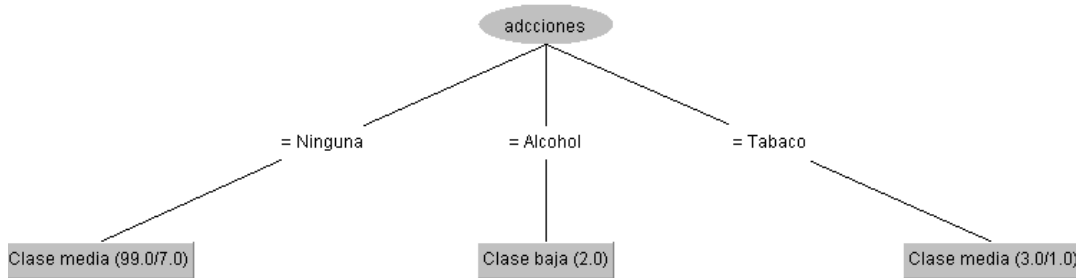
Este algoritmo se aplica a cada una de las variables que se localizaron, generando arboles para la toma de decisiones. A continuación, se presentan algunos ejemplos y su análisis.

Variable estado\_civil: Este árbol tiene un tamaño de 11, y se muestra la relación entre la variable estado\_civil con las variables periodo\_gestacion, edad y acceso\_arma. Las mujeres que están en periodo de gestación están en unión libre, las mujeres que no están en periodo de gestación tienen una edad mayor a 39 años y están casadas, y las mujeres menores o igual a 39 años y que no tienen acceso a arma están en unión libre, así mismo las parejas de las mujeres que tienen acceso a arma y son menores de o igual a 39 años están casadas.



**Figura 7.** Árbol J48. Variable estado\_civil

Variable situación\_economica: El tamaño de este árbol es de 4 y relaciona las variables de situación\_economica y adicciones.



**Figura 8.** Árbol J48. Variable situacion\_economica

**Resultados obtenidos del análisis de los árboles de decisión.**

Al hacer el análisis de todos los árboles generados, se encuentra que algunas variables no tienen relación entre si y se obtienen dos grupos. El primer grupo con características de violencia física y el segundo con características de violencia verbal, sexual y psicológica.

Ahorcar, arañazos, tirones de cabello, patadas, jalonear, sufrimiento\_corporal, trabaja, carece\_apoyo, maldecir, puñetazos, sufrimiento\_corporal, pellizcar, patadas, tirones\_de\_cabello, sufrimiento\_corporal, morder, jalonear, empujones, cortes, escupir, cachetadas, bofetadas, asfixiar, arrollamineto\_con\_vehiculos, arañazos.

**Figura 9.** Primer grupo obtenido.

estado\_civil, lugar\_de\_nacimiento, situación\_economica, grado\_estudios, carece\_apoyo, trabaja, comparaciones\_descalificadoras, tiempo\_padeciendo, formacion\_integral\_salud, edad, antecedentes\_violencia, culpabilizar, intimidación, ley\_hielo, manipulación, periodo\_gestacion, acusaciones, tipo\_violencia\_agresor, acceso\_arma, ignorar, acoso, armas\_contra\_la\_victima, tipo\_arma, asfixiar, defensiva\_victima, bromas\_hirientes, falta\_atencion, chantajear, maldecir, insultos, descalificar, humillaciones, gritos, groserías, insultos, control\_imagen, disminuir, maldecir, tono\_de\_voz\_agresivo, pellizcar, tiempo\_padeciendo, asifixiar, mentir, burlas, trivilizar, prohibir, control\_imagen, controlar, criticas, ridiculizar, vejaciones, comparaciones\_descalificadoras, practicas\_sexuales\_no\_deseadas, violar, forzar\_relacion, manosear, amenazas, reacciones\_ante\_reproches, ofender, trivializar, criticas\_degradantes, vejaciones, bloqueo\_del\_dialogo, normalmente\_grita, ordenes\_agresivas,

**Figura 10.** Segundo grupo obtenido.

**Conclusiones.**

La aplicación de nuevas tecnologías a problemas que son bastante graves ha ido aumentando de manera considerable, la minería de datos ha ayudado a analizar grandes volúmenes de información para crear modelos de clasificación, agrupamiento o modelos predictivos.



La elaboración y aplicación del instrumento de recolección de datos o encuesta es bastante difícil ya que se deben buscar encuestas que ya están validadas o crear una con ayuda de un especialista, en el momento de aplicar la encuesta se deben de tomar en cuenta varios aspectos para que las participantes no se sientan intimidadas o con miedo, por ejemplo proporcionar datos personales, debido a la seguridad, incomodidad debido al tema, las emociones que pueden surgir debido a las preguntas y no tener conocimiento para manejar la situación, por lo que es necesario pedir ayuda de un psicólogo para que la situación esté un poco más controlada. Weka es un software con una interfaz sencilla de entender y se pueden elegir los algoritmos para aplicar de una manera rápida. Se localizaron 86 variables de la encuesta que se aplicó, se acomodaron en la hoja de cálculo de Excel con las respuestas que se obtuvieron de las participantes y se exportó el archivo a formato CSV, que es el formato que admite Weka. Con esto se pudo observar cómo se visualizan cada una de las variables y el porcentaje de respuestas que tiene cada una. El algoritmo J48 crea árboles para la toma de decisiones donde se puede observar la relación entre las variables y se pueden llegar a agrupar. Algunas no presentan relación y pueden ser descartadas.

### Referencias bibliográficas.

- Ayala Salgado, L. y. (1 de Mayo de 2012).** *Contribuciones a las Ciencias Sociales*. Obtenido de "La violencia hacia la mujer. Antecedentes y aspectos teóricos": <http://www.eumed.net/rev/cccss/20/ashm.html>
- Campos, A. (Mayo de 2003).** *La violencia Familiar (Tesis de grado. "Master en derecho penal") Universidad Autonoma de Nuevo Leon*. Obtenido de <http://eprints.uanl.mx/1412/1/1020148989.PDF>
- Esnel Montero Hechavarria, M. (Abril de 2011).** *Scielo*. Obtenido de Realidades de la violencia familiar en el mundo contemporaneo: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1029-30192011000400016](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192011000400016)
- Garnica, N. J. (15 de diciembre de 2011).** *Scielo*. Obtenido de Exploración de la violencia sexual en la ciudad de Bogotá: una aplicación de técnicas de minería de datos: <http://www.scielo.org.co/pdf/crim/v53n2/v53n2a07.pdf>
- INEGI. (12 de Mayo de 2017).** Obtenido de Tiempo digital: <https://tiempodigital.mx/2017/05/12/los-datos-duros-de-la-violencia-intrafamiliar-en-mexico/>
- López, C. P. (2007).** *Minería de datos: Técnicas y herramientas*. Madrid, España: Paraninfo Cengage Learning.
- María García Jiménez, A. Á. (s.f.).** *Universidad Carlos III*. Obtenido de Análisis de Datos en WEKA – Pruebas de selectividad: <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/28.pdf>
- Orallo, J. H. (s.f.).** *Introducción a la minería de datos*. España: Pearson .
- Pilar Jara Romero, A. R. (s.f.).** *Universitat Jaume-I*. Obtenido de Escala de evaluación del tipo y fase de la violencia de género (EETFVG): [http://repositori.uji.es/xmlui/bitstream/handle/10234/77672/forum\\_2009\\_18.pdf?sequence=1&fbclid=IwAR02HfEfQdfEUkwz6O5ttD05nmoZq7GlkXe8bDYnZM9m70vsI0kBaK1JvWc](http://repositori.uji.es/xmlui/bitstream/handle/10234/77672/forum_2009_18.pdf?sequence=1&fbclid=IwAR02HfEfQdfEUkwz6O5ttD05nmoZq7GlkXe8bDYnZM9m70vsI0kBaK1JvWc)
- Pineda, O. M. (julio de 2017).** *ResearchGate*. Recuperado el 12 de octubre de 2020, de [https://www.researchgate.net/publication/325766285\\_Categorizacion\\_automatizada\\_de\\_tweets\\_sobre\\_el\\_tem\\_a\\_politico\\_electoral\\_aplicando\\_algoritmos\\_de\\_clasificacion\\_supervisada#pfc](https://www.researchgate.net/publication/325766285_Categorizacion_automatizada_de_tweets_sobre_el_tem_a_politico_electoral_aplicando_algoritmos_de_clasificacion_supervisada#pfc)
- SurveyMonkey. (12 de junio de 2020).** *SurveyMonkey*. Obtenido de SurveyMonkey: <https://www.surveymonkey.com/mp/sample-size-calculator/>

**Teruelo, J. G. (2017).** *Revista Electrónica de Ciencia Penal y Criminología*. Obtenido de Diagnóstico del sistema de protección y propuestas de intervención para la predicción y prevención de feminicidios en contexto de pareja o expareja: <http://criminet.ugr.es/recpc/19/recpc19-23.pdf>

### Información de los autores.



**Miriam Vargas Reyes** estudiante del noveno semestre de la carrera de Ingeniería en Sistemas Computacionales en el Tecnológico de Estudios Superiores de Valle de Bravo, Estado de México.



**Mariana Carolyn Cruz Mendoza** es Ingeniero en Sistemas Computacionales por el Tecnológico de Estudios Superiores de Valle de Bravo, Estado de México, cuenta con Maestría en Ingeniería en Sistemas Computacionales por el Tecnológico de Estudios Superiores de Ecatepec. Tiene la distinción de Perfil Deseable ante PRODEP y es Miembro del Cuerpo Académico en Formación “Procesamiento de datos con industria 4.0” del Tecnológico de Estudios Superiores de Valle de Bravo.



**César Primero Huerta** es Ingeniero en Sistemas Computacionales por el Tecnológico de Estudios Superiores de Valle de Bravo, Estado de México, cuenta con Maestría en Ingeniería en Sistemas Computacionales por el Tecnológico de Estudios Superiores de Ecatepec. Tiene la distinción de Perfil Deseable ante PRODEP y es Líder del Cuerpo Académico en Formación “Procesamiento de datos con industria 4.0” del Tecnológico de Estudios Superiores de Valle de Bravo.