

## **Análisis de Técnicas de Clasificación en Violencia de Género en Jujuy**

Mariela Rodríguez<sup>1</sup>, José H. Farfán<sup>1</sup>, Nazarena Laureano<sup>2</sup>, Gerardo L. Vargas<sup>1</sup>,  
Micaela A. Soria<sup>3</sup>, Norma G. Castro<sup>4</sup>.

Facultad de Ingeniería – Universidad Nacional de Jujuy<sup>1</sup>. Facultad de Humanidades y Ciencias Sociales – Universidad Nacional de Jujuy<sup>2</sup>. Universidad de Ciencias Empresariales Sociales<sup>3</sup>.  
Universidad Católica de Salta<sup>4</sup>

*mariela.rodriguez@fi.unju.edu.ar, jhfarfan@gmail.com, naza\_asis@hotmail.com,  
gherardo.gv@gmail.com, micaelasoria0@gmail.com, grisk307@gmail.com*

### **RESUMEN**

La Violencia de Género es una problemática que genera un gran flagelo en la comunidad, donde la víctima puede recibir desde daños morales hasta daños físicos que pueden ocasionar la muerte. Es importante que los organismos encargados de trabajar en esta problemática puedan contar con información adecuada para resolver las situaciones que se plantean diariamente. La minería de datos presenta un conjunto de técnicas que pueden tratar los datos y extraer información útil para los especialistas en Violencia de Género. Este trabajo aplica técnicas de clasificación en esta difícil problemática emergente en la sociedad actual. Se realizó un análisis situacional en el que se encuentran los datos a partir de la etapa del proceso KDD de Visualización, describiendo los principales atributos de dicha problemática. Para dar una respuesta a la optimización de recursos se utilizan técnicas no supervisada de datos tales como clustering identificando la cantidad de hechos por zona geográfica, empleando el algoritmo K-Means. Fue necesario también encontrar las relaciones entre los atributos disponibles mediante reglas de clasificación utilizando el algoritmo W-J48 para posteriormente predecir estos resultados y validarlos.

**Palabras Claves:** Violencia de Género, Minería de Datos, Técnicas de Clasificación, Clustering, KDD

### **INTRODUCCIÓN**

Según la Ley N° 26.485 de Protección Integral a las Mujeres, se entiende por violencia contra las mujeres toda conducta, acción u omisión, que de manera directa o indirecta, tanto en el ámbito público como en el privado, basada en una relación desigual de poder, afecte su vida, libertad, dignidad, integridad física, psicológica, sexual, económica o patrimonial, como así también su seguridad personal [1]. Quedan comprendidas las perpetradas desde el Estado o por sus agentes.

La ley además puntualiza que, se considera violencia indirecta, a los efectos de la presente ley, toda conducta, acción omisión, disposición, criterio o práctica discriminatoria que ponga a la mujer en desventaja con respecto al varón. La violencia de género puede adoptar diversas formas, lo que clasifica al delito, de acuerdo con el contexto en donde la ocurre: violencia física, psicológica, sexual, económica - patrimonial y simbólica; y según la modalidad que involucra: violencia doméstica, institucional, laboral, contra la libertad reproductiva, obstétrica y/o mediática. [2]

Según la ley, lo que diferencia a este tipo de violencia de otras formas de agresión y coerción, es que el factor de riesgo o de vulnerabilidad, acontece por el solo hecho de ser mujer. Sin embargo en vista a los datos obtenidos se consideró incluir en el análisis a las víctimas con diferentes orientaciones sexuales y el análisis de la violencia de la mujer hacia el hombre.

La necesidad de trabajar en la eliminación de la discriminación entre las mujeres y varones en todos los órdenes de la vida y desarrollar políticas públicas de carácter interinstitucional, es una tarea que compromete a todos los sectores, por lo tanto, es necesario que los organismos competentes, especialmente los vinculados a garantizar la seguridad personal contribuyen a: estudiar la complejidad de esta problemática, suministrar información confiable y relevante a la sociedad y a los organismos estatales competentes y, definir políticas públicas intersectoriales de largo plazo enfocadas hacia la sensibilización, prevención y empoderamiento por parte de la comunidad provincial.

A través del análisis de datos se pretende encontrar y descubrir los patrones que caracterizan los diferentes tipos y modalidades que la violencia de género adquiere en la vida cotidiana a partir de las llamadas recibidas a través de la línea de emergencia en el ámbito de las siguientes localidades de la Provincia de Jujuy: San Salvador de Jujuy, La Almona, Los Alisos, Río Blanco y Palpalá, contando para ello con un DataSet correspondiente al primer semestre de los años 2017, 2018 y 2019, y reconociendo, en base a él, la magnitud y extensión que el problema tiene a nivel local y provincial.

Esto permite construir un diagnóstico sobre cómo ésta problemática se presenta en la Provincia de Jujuy, a partir de un análisis inteligente de datos estadísticos que surgen del procesamiento de la información obtenida. Lo que posibilitará estudios con mayor nivel de profundidad que explique la situación existente y, finalmente ofrecer insumos para el diseño e implementación de políticas públicas intersectoriales que tiendan a la prevención, sanción y erradicación de esta problemática.

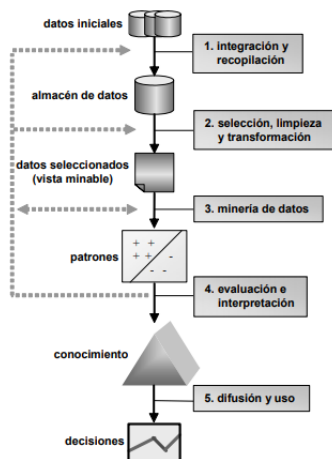
La minería de datos es la ciencia y tecnología de explorar datos en orden para descubrir patrones desconocidos, es una parte general del proceso de descubrimiento del conocimiento (KDD). Hoy en día se tiene masivas cantidades de información respecto a delitos que suceden en el ámbito provincial que pueden dar patrones que ayuden a diseñar trabajos operativos como estratégicos para el área de seguridad. [3]

## **METODOLOGÍA**

### **Proceso de obtención del conocimiento**

El descubrimiento de la información mencionada resulta posible por la Minería de Datos o Data Mining, pero es el descubrimiento del conocimiento (KDD, por sus siglas en inglés) el que se encarga de la preparación de los datos y de la interpretación de los resultados obtenidos, los cuales dan un significado a los patrones encontrados [4], es importante destacar que KDD es el producto del rápido desarrollo de la minería de datos y la gran aplicación de tecnologías de información y bases de datos.

Zhang et al [5] formulan el proceso basado en la extracción de conocimiento (KDD) en una secuencia iterativa de cuatro pasos: la definición del problema, el pre - procesamiento de datos (que incluye la preparación de datos), el data mining, y el post data mining. Otros autores como Hernández Orallo et al [6] profundizan estas fases, tal como puede observarse en la figura 1.



**Figura 1. Fases de KDD [6]**

## **Fase 1: Integración y recopilación**

Como se menciona precedentemente, el DataSet del presente trabajo se obtiene del Centro de Emergencias de la capital jujeña, la cual permite contar con una importante cantidad de datos confiables que brinda información válida acerca de las distintas formas que asume la violencia contra las mujeres en la provincia de Jujuy.

En dicho lugar se registran llamadas de diversas índoles, en la que se da intervención por Acción Policial, Emergencias, Servicio de Atención Médica, entre otras. A partir de esta información se realiza una depuración considerando solo aquellos registros relacionadas a la categoría de Violencia de Género. Se obtiene un total de 7728 registros que corresponden primer semestre de los años 2017, 2018 y 2019. Los mismos provienen particularmente de dos Departamentos de la Provincia de Jujuy: Dr. Manuel Belgrano y Palpalá, que según la proyección poblacional en el año 2019 ambas localidades cuentan con 364.310 habitantes. La cantidad de hechos de violencia de género que se registran, involucra aproximadamente el 1% de la población, lo que indica que los mismos sufrieron al menos un hecho de violencia en dicho semestre. Por último es importante mencionar que en relación a las personas que se comunican con dicho centro de operaciones, en el 11% del total reiteran las llamadas en distintas ocasiones.

Es importante mencionar que esta base de trabajo se obtuvo de forma coordinada con investigadores y responsables del área de seguridad correspondiente, e involucra procesos de integración y recopilación que no se detallan en el presente trabajo, teniendo en cuenta que el actual documento intenta abocarse al análisis de técnicas de minería de datos al problema en estudio.

## **Fase 2: Selección, limpieza y transformación**

La base de datos original posee una gran cantidad de tablas, registros y atributos, de la cual es importante saber que se encuentra normalizada y con diversos controles de integridad. A pesar de ello, se procede, primero a hacer un Pre procesamiento de Datos, realizando una limpieza de los mismos a través de la detección de valores fuera de rango u outlier.

De la misma, y luego de un gran esfuerzo, se obtuvo, como se menciona en la fase 1, una única tabla de trabajo con 7728 registros, con una cantidad también muy reducida de atributos. La cantidad de atributos se debe a que desde la fuente se cuenta con esta información y otros atributos de índole privada. A continuación se detalla el diccionario de datos correspondiente al DataSet en estudio del presente trabajo, considerando

solamente los atributos que se utilizan en este trabajo y que se extrajeron para el análisis a realizar:

- Día y hora de los hechos.
- Clasificador: amenaza con Arma de Fuego o Blanca, Daño Moral, Mujer Agredida.
- Medidas de Seguridad/Exclusión del Hogar.
- Sexo de la Víctima.
- Vínculo con el Agresor (Novio/Pareja, Ex Pareja, Padre/s, Hijo/s, Nieto/s, Hermano/s, Padrastro, Tío/s, Primo/s), Ideología de Género de la Víctima (F: femenino, M: Masculino, FF: Lesbiana, MM: Homosexual, T: Transexual).

Los atributos que se describieron corresponden a los hechos de violencia de género. En otros tipos de delitos se cuenta con más clasificadores.

### Fase 3: Minería de Datos

Orallo et al [6] y la mayoría de los principales autores de Minería de Datos coinciden en clasificar las técnicas de Minería de Datos en dos grandes grupos: Predictivas y Descriptivas, tal como puede observarse en la figura 2.

Nombre	PREDICTIVO		DESCRIPTIVO		
	Clasificación	Regresión	Agrupamiento	Reglas de asociación	Correlaciones / Factorizaciones
Redes neuronales	✓	✓	✓		
Árboles de decisión ID3, C4.5, C5.0	✓				
Árboles de decisión CART	✓	✓			
Otros árboles de decisión	✓	✓		✓	
Redes de Kohonen			✓		
Regresión lineal y logarítmica		✓			✓
Regresión logística	✓			✓	
Kmeans			✓		
Apriori				✓	
Naive Bayes	✓				
Vecinos más próximos	✓	✓	✓		
Análisis factorial y de comp. ppaes.					✓
Twostep, Cobweb			✓		
Algoritmos genéticos y evolutivos	✓	✓	✓	✓	✓
Máquinas de vectores soporte	✓	✓	✓		
CN2 rules (cobertura)	✓			✓	
Análisis discriminante multivariante	✓				

**Figura 2. Clasificación de las técnicas de Minería de Datos [6]**

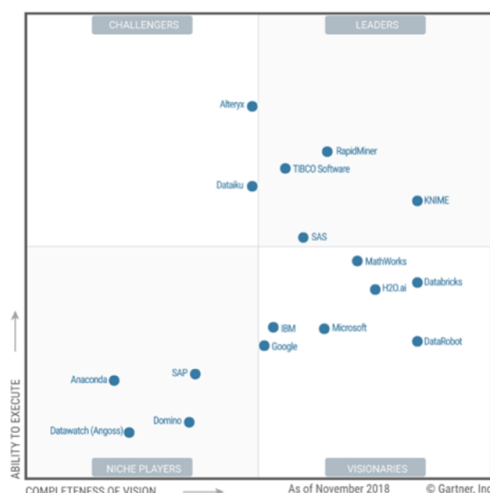
En forma general se puede afirmar que las técnicas Predictivas intentan encontrar un modelo que ayude a predecir un comportamiento como por ejemplo la devolución o no de un préstamo o la probabilidad de tener cierta enfermedad, etc. Mientras que las técnicas Descriptivas están centradas en encontrar una descripción de los datos de estudio, es decir la correlación que puede haber entre atributos o la importancia de ellos, como por ejemplo el análisis de los atributos que definen la tendencias de compras de los clientes de los supermercados o de los clientes de los bancos, etc.

Resulta muy importante, por lo tanto, para el actual análisis ahondar en las principales técnicas aplicables al caso de estudio, y que generen resultados significativos para aquellos organismos encargados de tomar decisiones a nivel prevención y seguridad.

### Herramienta utilizada para la minería de datos

Rapidminer es una de las principales herramientas disponibles en el mercado y la misma fue elegida por el grupo de investigación para realizar esta etapa de la Minería de Datos. Algo a destacar es, que, en el llamado “Cuadrante Mágico” para plataformas de Análisis Predictivo del año 2.019 (Figura 3), de la empresa Gartner (organización de investigación de las tecnologías de la información), en la cual se visualiza la comparación de las principales herramientas para Plataformas Analítica Avanzadas,

	CHALLENGERS		LEADERS	
--	-------------	--	---------	--



**Figura 3. “Cuadrante Mágico” para plataformas de Análisis Predictivo 2018 de Gartner. [7]**

Rapidminer es un programa informático para el análisis y minería de datos que permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores; esto significa que a través de una GUI (Interfaz Gráfica de Usuario) o de línea de comandos o a través de procesos batch (o de lotes) o desde otros programas a través de API de Java, permite realizar análisis estadístico, minería de datos y análisis predictivo.

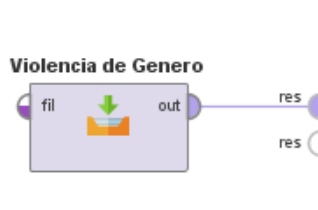
Es un software de tipo Open-Source con licencia GNU GPL, basado en java, que trabaja bajo las plataformas Windows y Linux, posee alrededor de 500 operadores que pueden ser combinados, utiliza el lenguaje de scripting XML para describir los operadores y su configuración y posee una gran cantidad de extensiones (plugins) [8]. Una de las mayores ventajas del Software Rapidminer son las herramientas que la misma dispone para la Exploración Visual, de los atributos de los que se dispone.

A continuación se describen algunos gráficos relevantes obtenidos (lo cual forma parte de una exploración visual) de la herramienta Rapidminer, para posteriormente analizar la aplicación de algunas técnicas específicas de minería de datos.

## DESARROLLO

## Exploración Gráfica

La exploración gráfica en minería de datos permite observar o describir atributos y relaciones de la base de datos de origen que son de gran importancia para los casos de estudio. Una vez que se obtienen los datos de la Base de Datos por parte de Rapidminer se pueden realizar una serie de gráficos de dicha herramienta para aplicar la Exploración Visual de los mismos. La conexión con la Base de Datos se efectúa en forma sencilla tal como se observa en la figura 4.



**Figura 4. Conexión de Rapidminer con la Base de Datos.**

A continuación se muestra un análisis de los gráficos más relevantes con los que se trabajaron:

### Gráficos por Periodo de Ocurrencia del Hecho (día y mes)

En primera instancia se visualiza la información concerniente a la cantidad de hechos registrados en el primer semestre de los años en cuestión por período de ocurrencia tal como se observa en la figura 5, la cual corresponde a los días del mes. Se observan tres picos representativos, el primero ocurre los días 1 de cada mes, luego entre 6, 7, tercer pico es entre 13, 14 cuarto entre 20, 21 y por último entre 28, 29. Estos datos tienen tendencia a que la mayor cantidad de hechos ocurren los fines de semana.

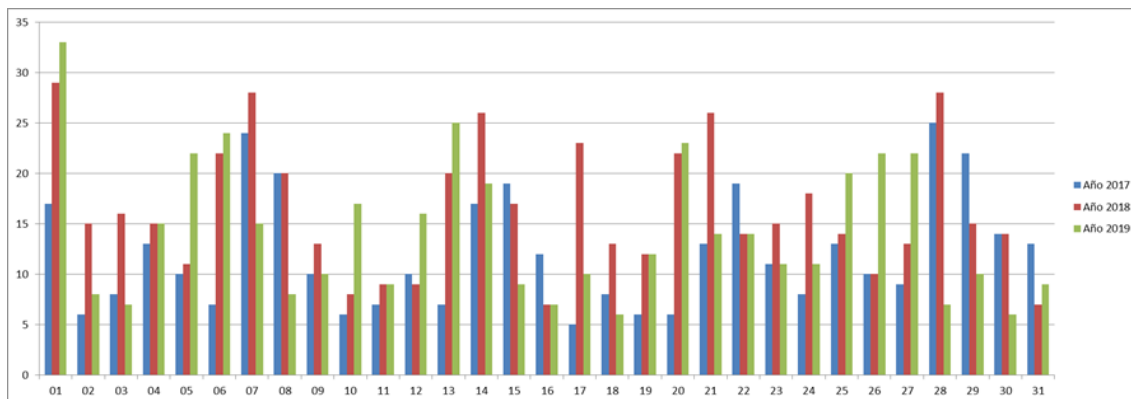


Figura 5. Ocurrencia de hechos delictivos por día del mes (1 a 31)

Discriminando los días, pero en los meses de los cuales se posee información (primer semestre 2017, 2018 y 2019), se observa en la figura 6 que se produjeron dos picos de relevancia que coinciden con fechas festivas tales como Año Nuevo en el mes de enero y el domingo de Pascuas, en abril y 1 de Mayo. La provincia de Jujuy se destaca por tener tradiciones fuertemente arraigadas por la religión católica, por ejemplo la festividad de Semana Santa, la cual exige ciertas prohibiciones como la abstinencia de consumo de carnes, escuchar música y el consumo de bebidas alcohólicas entre otras características. Si lo analizamos desde el círculo de la violencia podría decirse que el incremento de las prohibiciones incrementa el cúmulo de tensiones, a ello se le agrega el consumo problemático de alcohol que se encuentra naturalizado como práctica cultural en la provincia de Jujuy, ambos constituyen factores que incrementan la probabilidad de hechos de violencia de género en periodos festivos [9]. Este comportamiento se asemeja al estudio realizado por la Central Mexicana de Servicios Generales de Alcohólicos Anónimos [10] respecto a los altos consumos en Semana Santa.

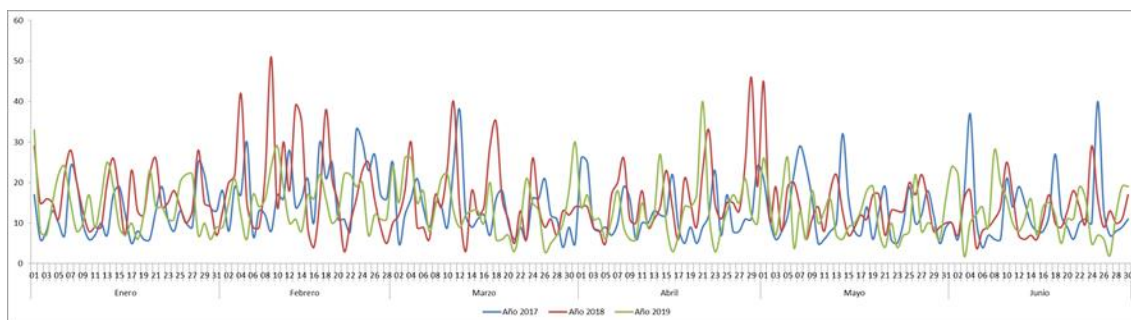


Figura 6. Ocurrencia de hechos delictivos por mes del año (de Enero a Febrero)

## Gráfico de Hechos Delictivos por Días de la Semana

En forma similar a lo analizado previamente, pero en relación al día de la semana, es importante remarcar que los fines de semana, son los días elegidos para cometer este tipo de delitos, tal como puede observarse en la figura .

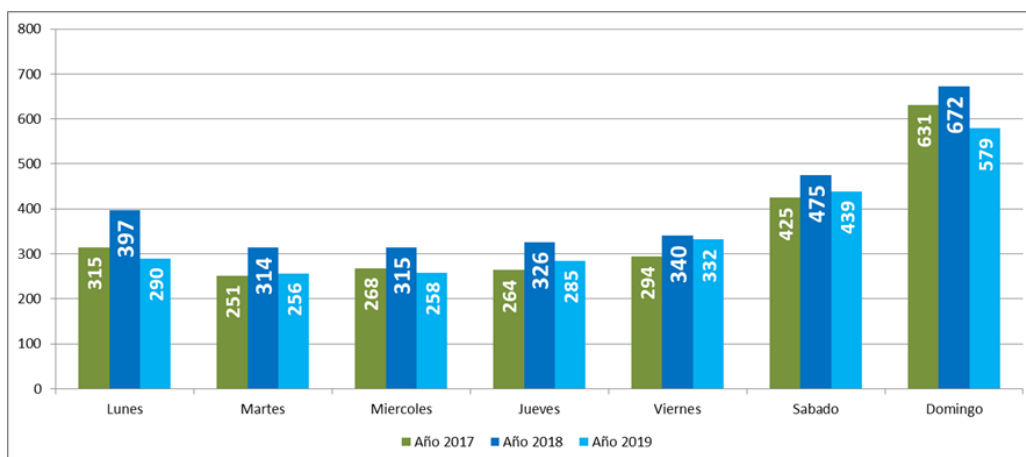


Figura 7. Ocurrencias de hechos delictivos día de la semana (de Lunes a Domingo)

Los llamados telefónicos, correspondientes a los días martes son relativamente bajos en todos los meses. Otra característica común, es que en todos los meses, el índice máximo ocurre los días domingos.

## Gráficos por Horas de Ocurrencia de Hechos Delictivos

Respecto a la hora en que ocurren los hechos predominan las horas nocturnas desde horas 20:00 a 01:00 donde ocurren la mayor cantidad de hechos, tal como se observa en la figura 8.

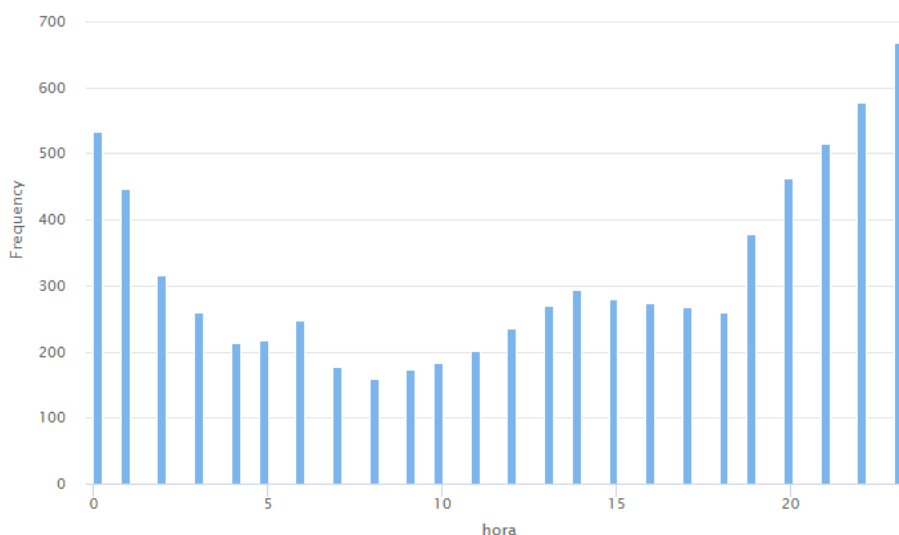


Figura 8. Ocurrencia de hechos delictivos por rango horario

## Aplicación de Técnicas de Minería de Datos

Como se mencionó anteriormente, luego de realizar una exploración visual se procede a aplicar técnicas de minería de datos, en donde para el caso de estudio, se detallan solo aquellas que permitieron obtener resultados relevantes al caso en estudio.

### Técnicas de Clustering para Ubicación de Hechos Delictuales

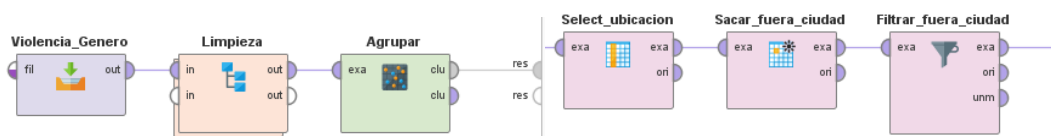
El Agrupamiento o Clustering tienen como finalidad obtener grupos entre los diferentes elementos de tal manera que los miembros de cada grupo tengan características similares, siendo importante considerar la factibilidad de contar o no con el número de grupos, y en este último caso analizar la posibilidad de aplicar algún algoritmo que permita obtenerlo.

Las técnicas de clustering hacen corresponder cada caso a una clase, con la peculiaridad de que las clases se obtienen directamente de los datos de entrada utilizando medidas de similitud. Es decir, agrupan a los datos bajo diferentes métodos y criterios [3].

K-medias es un método de clustering que tiene como objetivo generar una partición de un conjunto de observaciones en k grupos. Cada grupo está representado por el promedio de los puntos que lo componen. El representante de cada grupo se denomina centroide. La cantidad de grupos a descubrir, k, es un parámetro que se debe fijar a priori. [11]

La intervención a tiempo ante un hecho de violencia de género puede evitar consecuencias que resulten irreversibles. Por otro lado, las fuerzas de seguridad que se encuentran abocadas a medidas preventivas, resultan insuficientes debido a la cantidad de hechos crecientes. Los hechos de violencia de género resultan de índole privado hasta que se solicita la participación de la fuerza de seguridad, sin embargo no se podrían destinar recursos humanos para el cuidado de los domicilios y/o lugares privados, ante estas situaciones se puede utilizar el algoritmo de K-means ya que permite determinar los sectores de la ciudad donde se deben tener grupos de Comandos Policiales atento a estos hechos y brindar una intervención preventiva rápida.

Para analizar en qué lugares se pueden situar los comandos se estudiaron las posiciones geográficas del registro en cuestión. Se realizó la limpieza de información en los que se dejó fuera un 4% de los registros por encontrarse fuera de rango (lo cual mejora significativamente los resultados obtenidos), para este proceso se utilizó el operador Detect Outlier con búsqueda por distancia. En la figura 13 se visualiza el procedimiento realizado.



**Figura 13. Operadores de rapidminer para obtener Agrupamiento mediante K-means.**

Se hicieron varias iteraciones, a fin de determinar el parámetro de cantidad de clúster adecuado o K que permita obtener una medida equilibrada entre ellos con la cantidad de intervenciones adecuada a un Comando de Seguridad. Obteniéndose como resultado:

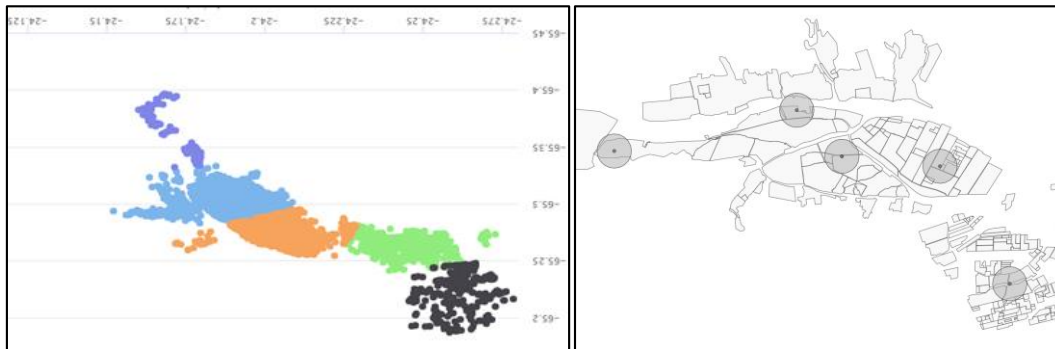
Modelo de Cluster: Cluster 0: 622 ítems, Cluster 1: 625 ítems, Cluster 2: 63 ítems, Cluster 3: 416 ítems, Cluster 4: 613 ítems. Total de número de ítems: 2339. Los centroides resultantes se observan en la Tabla 1.



**Tabla 1. Tabla de Centroides de Clúster**

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
latitud	-24.196	-24.247	-24.173	-24.179	-24.208
longitud	-65.297	-65.263	-65.359	-65.305	-65.271

En la figura 14 se obtuvo la ubicación de los centroides en el mapa de los dos departamentos involucrados Dr. Manuel Belgrano y Palpalá.



**Figura 14. Ubicación de los Centroides en el mapa. Obtenido con la herramienta Qgis.**

El resultado obtenido permite planificar estratégicamente los operativos que se realicen desde las fuerzas de seguridad.

### **Técnicas de algoritmos de clasificación**

En minería de datos estos algoritmos se utilizan como técnicas predictivas, en las que la variable puede distinguirse como dependiente e independiente. Las técnicas predictivas especifican el modelo para los datos en base a un conocimiento teórico previo. El modelo supuesto para los datos debe contrastarse después del proceso de minería de datos antes de aceptarlo como válido. Formalmente la aplicación de todo modelo debe superar las fases de identificación objetiva, estimación (proceso de cálculo de los parámetros del modelo elegido para los datos en la fase de identificación), diagnóstico (proceso de contraste de la validez del modelo estimado) y predicción (proceso de utilización del modelo identificado, estimado y validado para predecir valores futuros de las variables dependientes). [12]

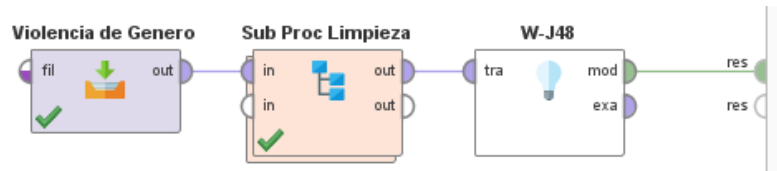
#### **Reglas de clasificación con Algoritmo W - J48**

El algoritmo J48, es uno de los algoritmos de minería de datos de clasificación más relevantes. Se destaca el nivel de confianza establecido para la poda del árbol generado, nivel de confianza, puesto que influye notoriamente en el tamaño y capacidad de predicción del árbol construido. El algoritmo J48 para tomar la decisión sobre el corte realizado en la iteración 'n', busca la variable predictora y el punto de corte exacto en el que el error cometido es el más bajo (tomando como criterio una variable preestablecida), siempre y cuando se encuentre en niveles de confianza superiores a los establecidos previamente. Una vez realizado el corte, el algoritmo vuelve a repetirse, hasta que ninguna de las variables predictoras alcance un nivel de confianza superior al establecido [13]. Se destaca la importancia de trabajar con el nivel de confianza, ya que, en caso de tener un gran número de sujetos y variables, este árbol puede resultar demasiado grande, característica que lo diferencia respecto a otros algoritmos de clasificación como ID3.

Se utiliza esta técnica de clasificación para encontrar patrones que puedan indicar las relaciones que existen entre la víctima, el sexo de ella, la relación que mantenía, la

hora del suceso y el tipo de agresión que propicia el victimario. El atributo que se utilizó como objetivo (label) es el tipo de agresión sufrida. El horario fue discretizado para obtener una mejor estimación: madrugada (0:00 a 06:00), mañana (06:00 a 12:00), tarde (12:00 a 18:00) y noche (18:00 a 24:00).

En la Figura 15 se observan los operadores aplicados para trabajar con esta técnica y en la Figura 16 los resultados obtenidos con la misma empleando Rapidminer.



**Figura 15. Operadores utilizados para realizar un Árbol de Decisión W-J48, obtenida con Rapidminer.**

vinculo.agresor-victima = EX PAREJA
hora = Madrugada
SEXO VICTIMA = F: MUJER AGREDIDA (284.62/172.08)
SEXO VICTIMA = M: MEDIDAS DE SEGURIDAD-EXCLUSIÓN DEL HOGAR (4.38/1.79)
hora = Mañana
SEXO VICTIMA = F: MUJER AGREDIDA (161.29/104.11)
SEXO VICTIMA = M: DAÑO MORAL (11.71/6.89)
hora = Tarde
SEXO VICTIMA = F: MUJER AGREDIDA (253.21/144.29)
SEXO VICTIMA = M: MEDIDAS DE SEGURIDAD-EXCLUSIÓN DEL HOGAR (9.79/5.75)
hora = Noche: MUJER AGREDIDA (387.0/235.0)
vinculo.agresor-victima = HERMANO: MUJER AGREDIDA (0.0)
vinculo.agresor-victima = HIJO: MUJER AGREDIDA (514.0/234.0)
vinculo.agresor-victima = NOVIO/ PAREJA: MUJER AGREDIDA (3598.0/818.0)
vinculo.agresor-victima = PADRE
hora = Madrugada: MUJER AGREDIDA (32.0/13.0)
hora = Mañana: DAÑO MORAL (10.0/5.0)
hora = Tarde: MUJER AGREDIDA (28.0/13.0)
hora = Noche: MUJER AGREDIDA (58.0/28.0)
vinculo.agresor-victima = SD: MUJER AGREDIDA (1918.0/141.0)

**Figura 16: Resultados Obtenidos con Algoritmo J48**

De acuerdo a los resultados obtenidos, en donde se puede observar que el clasificador que más predominó fue el de “Mujer Agredida” en los horarios de la madrugada, mañana y tarde. Los clasificadores que lo siguen son Exclusión del Hogar y Daño Moral.

Particularmente, las situaciones de violencia de género que se representan con el atributo “NOVIO/ PAREJA” suceden durante la juventud tienen características similares a las que se manifiestan luego en las edades adultas. La violencia durante el noviazgo puede manifestarse de diferentes maneras: a través de la violencia psicológica – hostigamiento-, la violencia sexual, la física e incluso la económica [14]. Las reglas de clasificación permite ilustrar características que el estado provincial debe analizar a fin de erradicar este tipo de hechos.

Resulta llamativo los casos de violencia de género en el vínculo madre e hijo. Al respecto Carrizo Antonio argumenta que en ocasiones, cuando la figura paterna está ausente, a uno de los hijos se le adjudica y/o asume el control de la casa, es decir, que

la figura de autoridad y de "macho" está representada en este hijo, quien ante una figura de mujer/madre tradicional y la figura machista que asume como "hombre", genera un vínculo violento con su madre. [15]

Los casos de violencia de la mujer hacia el hombre no se encontraron dentro de lo que la ley establece como violencia de género. Sin embargo es interesante destacar que este tipo de violencia física y psicológica es difícil de detectar, ya que los estereotipos contruidos sobre la figura del hombre "macho" actúan como un obstáculo para la visualización de esta problemática y sobre todo para pedir ayuda. Las mujeres por su constitución psíquica tiene más facilidad de expresar sus emociones, sin embargo lo hombre ven esta posibilidad como algo socialmente prohibido. De ahí la posibilidad de ocultamiento de la problemática y de la consideración de pedir ayuda al sistema de emergencia.

### Predicción de los datos con W - J 48 y Red Neuronal

El algoritmo J48, además de analizar la clasificación que existe entre los datos suministrados puede predecir, como se dijo anteriormente, el atributo que se estime como objetivo, atributo independiente.

Las redes neuronales permiten descubrir modelos complejos y afinarlos a medida que progresa la exploración de los datos. Gracias a su capacidad de aprendizaje permiten descubrir relaciones complejas entre variables sin ninguna intervención externa.

A continuación se realizará la predicción de información con el algoritmo W - J48 y el algoritmo Neural Net para estimar el tipo de daños que pueden ocasionar los agresores contra las víctimas de violencia de género. El primero de estos algoritmos se eligió porque es uno de los clasificadores que permite la poda de las ramas cuando se cuenta con varios estados y el segundo porque permite aprender cuando se tiene relaciones complejas. Para validar la información se utilizó el operador de validación cruzada.

La validación cruzada consiste en un método iterativo que aplica el procedimiento de entrenamiento prueba sobre diferentes particiones en cada iteración. La validación cruzada de k iteraciones toma como parámetro de entrada a un rango que puede ir de 0 a 1 y que corresponde al porcentaje de datos que se van a utilizar para testear el algoritmo en cada una de las k iteraciones. En la primera iteración se selecciona de manera aleatoria u ordenada un 100 (1- a) % de los datos como conjunto de entrenamiento, se ejecuta el correspondiente algoritmo de clasificación y se evalúa sobre el conjunto de prueba formado por el restante 100 %.

El procedimiento se repite durante las K iteraciones y al final se calcula la media de aciertos cometidos en términos de porcentaje. Este valor corresponde a la bondad del modelo de clasificación. [16]

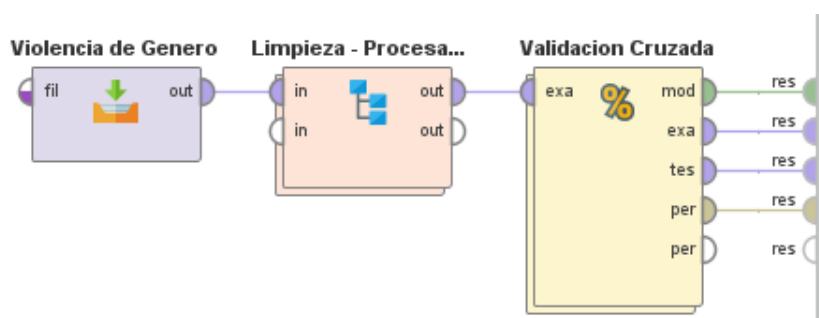


Figura 17. Modelo de Predicción con Validación Cruzada.

El modelo utilizado trabajo con los atributos nombrados con anterioridad sexo de víctima, relación con el agresor, hora y clasificador. Se utilizó el 80 % de los datos para entrenamiento y 20 % del dataSet para prueba con un modelo de performance de clasificador.

Según la figura 18a. el modelo de predicción J48 logró obtener un 73.14 % de precisión, es destacable la clase de precisión que se dio en el clasificador “mujer agredida” con un 74%, seguido de “medidas de exclusión del hogar” con un 33%.

accuracy: 73.14% +/- 0.50% (micro average: 73.14%)

	true AMENAZA C...	true AMENAZAS	true DAÑO MORAL	true DESCUIDO, ...	true MEDIDAS DE...	true MUJER AGR...	class precision
pred. AMENAZA C...	0	0	0	0	0	0	0.00%
pred. AMENAZAS	0	0	0	0	0	0	0.00%
pred. DAÑO MOR...	2	0	22	0	20	32	28.95%
pred. DESCUIDO,...	0	0	0	0	0	0	0.00%
pred. MEDIDAS D...	0	2	61	1	74	87	32.89%
pred. MUJER AG...	49	21	1142	1	535	5221	74.92%
class recall	0.00%	0.00%	1.80%	0.00%	11.76%	97.77%	

**Figura 18a. Predicción de Clasificador de Violencia de Género con J48**

accuracy: 73.45% +/- 0.00% (micro average: 73.45%)

	true AMENAZA C...	true AMENAZAS	true DAÑO MORAL	true DESCUIDO, ...	true MEDIDAS DE...	true MUJER AGR...	class precision
pred. AMENAZA C...	0	0	0	0	0	0	0.00%
pred. AMENAZAS	0	0	0	0	0	0	0.00%
pred. DAÑO MOR...	0	0	0	0	0	0	0.00%
pred. DESCUIDO,...	0	0	0	0	0	0	0.00%
pred. MEDIDAS D...	0	0	0	0	0	0	0.00%
pred. MUJER AG...	51	23	1225	2	629	5340	73.45%
class recall	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	

**Figura 18b. Predicción de Clasificador de Violencia de Género con Red Neuronal**

La red neuronal logra aprender respecto a modelos complejos, sin embargo la predicción que se obtiene es de 73,45%, sin embargo no logra identificar “Medidas de Exclusión” ni “Daño Moral” que son clasificadores que tienen una gran cantidad de registros.

## CONCLUSIONES

Las técnicas de minería de datos resultan adecuadas para ser aplicadas en cualquier ámbito: académico, privado y público; pero resultan sumamente útiles en el estudio de diversas problemáticas que afectan a la sociedad.

En este trabajo se realizó un análisis inteligente que conlleva un tiempo de aplicación y retroalimentación, para el estudio de la Violencia de Género. En primer instancia se identificaron patrones de los hechos delictuales del primer semestre de los años 2017, 2018 y 2019, de los que sobresalen los días del mes en el que ocurren los hechos, los días de la semana y horarios más frecuentes, circunstancias que permiten definir políticas de prevención. Es destacable la cantidad de hechos que ocurren los fines de semana como también en días festivos, situaciones que conllevan un estudio mayor de especialistas en Violencia de Género del Ministerio de Desarrollo Humano, quienes aportaron desde su conocimiento, que los hechos ocurren en esas fechas porque el agresor sale de la rutina y realiza actividades de esparcimiento, en las que, en muchas

ocasiones se acompaña con bebidas alcohólicas, lo cual es uno de los principales factores que pueden desembocar en la comisión de este tipo de delitos.

Al aplicar técnicas de agrupamiento se logró generar una propuesta preventiva en materia de seguridad a este tipo de hechos de índole privada, permitiendo brindar una orientación sobre cómo organizar los recursos disponibles de las fuerzas.

Para finalizar, las reglas de clasificación permitieron interpretar la relación que existe entre la víctima, el agresor, la hora en que ocurre el hecho y el medio de agresión (clasificador). Las reglas permitieron interpretar los factores que influyen en el medio de agresión. Posteriormente, a la información que se clasificó, se utilizaron técnicas para predecir y validar la información, que permitieron validar con un cierto grado de certeza la información suministrada. Se debe destacar que el Centro de Emergencias debe dar una respuesta rápida ante estas situaciones y no permite tener una variedad de información que pueda determinar con un mayor grado de certeza los hechos ocurridos.

El trabajo generó un estudio que permitirá la intervención de cada organismo, como las que tienen por objetivo el bienestar de la persona, salud e instituciones judiciales y, sobre todo, como materia investigativa que permitan conocer las causas que originan este tipo de delito.

Como trabajo a futuro, se pretende realizar un estudio interdisciplinario con la información que se obtiene del seguimiento de los casos.

## Agradecimientos

Se da un cordial agradecimiento a la Coordinación del Ministerio de Seguridad de Jujuy que a través del Centro de Monitoreo 911 brindaron acceso a la información y a la Dirección de Observatorio de Seguridad con quien, conjuntamente se realizó este trabajo en el proceso de Limpieza de Datos y Visualización.

## BIBLIOGRAFIA

- [1] “Ley 26.485. Protección Integral para Prevenir, Sancionar y Erradicar la Violencia contra las Mujeres en los Ámbitos en que Desarrollen sus Relaciones Interpersonales”. 2009. Consultado en Septiembre de 2.019.
- [2] Rico, N. “Violencia de género: un problema de derechos humanos”. Naciones Unidas, CEPAL. Santiago, Chile. 1.996. pp. 5-9.
- [3] Asencios, Violeta “Data Mining y el descubrimiento del conocimiento” Editor Industrial Data, vol. 7, núm. 2, 2004, pp. 83-86
- [4] Vallejos, S. J. “Minería de Datos”, Obtenido de <https://goo.gl/gHQAqK>, Universidad Nacional del Nordeste, 2006
- [5] Zhang, S., Zhang, C., & Yang, Q. , “Data preparation for data mining. Applied Artificial Intelligence”, San Francisco, 2003, págs. 375-381
- [6] Hernández Orallo J., Ramírez Quintana J., Ramírez C.F., “Introducción a la Minería de Datos”, Universidad Politécnica de Valencia, España, Ed.Pearson, 2004. pp 13-14, 19-42, 137-164.
- [7] Gartner, “Magic Quadrant for Advanced Analytics Platforms”. Accedido en Noviembre de 2.018.
- [8] Rapidminer, “Real Data Science, Fast and Simple”. Obtenido de <https://rapidminer.com/> en Septiembre de 2.016
- [9] Giménez C. Llopis, Rodríguez García M.I., e Hernández Mancha I., “Relación entre el consumo abusivo de alcohol y la violencia ejercida por el hombre contra su pareja en la unidad de valoración integral de violencia de género (UVIVG). 2014. de Sevilla” [http://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1135-76062014000300002](http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1135-76062014000300002). Accedido en Noviembre de 2.019.
- [10] Central Mexicana de Servicios Generales de Alcohólicos Anónimos. <http://aamexico.org.mx/>. Accedido en Noviembre de 2.019.
- [11] MacQueen, J. “Some methods for classification and analysis of multivariate observations”. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, 281 --297, University of California Press, Berkeley, Calif., 1967

- [12] Pérez López, César “Minería de datos: técnicas y herramientas” Thomson Ediciones, 2008. ppt 8
- [13] Martínez Abad, Fernando “Aplicación de Técnicas de Minería de Datos con Software Weka ”, Universidad de Salamanca. <https://knowledgesociety.usal.es/sites/default/files/Apuntes-curso-Weka.pdf>. Accedido en Noviembre de 2019.
- [14] Observatorio Nacional de Violencia contra la Mujer, “Linea 144: Violencia de Género en Jóvenes”. <https://www.argentina.gob.ar/sites/default/files/informejuveneslinea144.pdf>. Accedido en Noviembre de 2019.
- [15] Carrizo, Antonio Durbal, Ministerio de Desarrollo Humano de Prov. Jujuy. Entrevista realizada Pérez López, César “Minería de datos: técnicas y herramientas” Thomson Ediciones, 2008 pp 343.
- [16] Vicente Cestero, Eloy ; Caballero, Alfonso Mateos “Data science y redes complejas: Métodos y aplicaciones”, Editorial Universitaria Ramón Areces, 2017. ppt 168.