



UNIVERSIDAD NACIONAL DE COLOMBIA

Modelo para la caracterización y clasificación de los tipos de violencia intrafamiliar desde los registros del sistema de salud

Andrea del Pilar Montaña Ramírez

Universidad Nacional de Colombia

Facultad de Ciencias

Departamento de Estadística

Bogotá, Colombia

2022

Modelo para la caracterización y clasificación de los tipos de violencia intrafamiliar desde los registros del sistema de salud

Andrea del Pilar Montaña Ramírez

Trabajo de grado presentado como requisito parcial para optar al título de:
Magíster en Ciencias Estadística

Directora:

B. Piedad Urdinola Ph.D.

Codirector:

Rafael Ricardo Rentería Ph.D.

Línea de Investigación:

Análisis de redes

Universidad Nacional de Colombia

Facultad de Ciencias

Departamento de estadística

Bogotá, Colombia

2022

Dedicatoria

Dedicado a mi Tita, a mi madre Pilar Ramírez y a mi hermanita Valentina, son todo en mi vida.

“La aventura más grande que puedes llevar a cabo es la de vivir tus sueños”

Oprah Winfrey

Agradecimientos

Quiero agradecer a mi mamá María Pilar por amarme, cuidarme, protegerme y enseñarme a amar; por ser padre y madre, por mostrarme que puedo comerme el mundo y que no existen barreras o límites que no pueda superar pero sobretodo, por haberme dado todas las herramientas necesarias para convertirme en la mujer que soy.

A mi madrina Beatriz, por ser mi apoyo; mi refugio, mi segunda madre, mi persona favorita, por haberme entregado tanto amor durante toda mi vida y por hacerme sentir como su propia hija. A mi tío Alfonso, por aconsejarme el estudiar esta carrera, que me ha llenado de alegrías y éxitos, por siempre darme una palabra de aliento y por estar ahí para mí como un padre. A mi hermana Valentina que es luz en mi vida, que me ha llenado de fuerza y valentía en los momentos más oscuros de mi existencia, por enseñarme a luchar por mis sueños y a no rendirme jamás y a mi Tita, por criarme, por enseñarme de qué se trata la vida, por amarme profundamente, por ser la cabeza de mi familia y darme ejemplo de rectitud, integridad y fortaleza. A mi tío Pepito, le agradezco por sus consejos, amor y apoyo, a mi tío James, a mi tía Rosis y a todas mis tías políticas les agradezco por su apoyo incondicional, por su amor y por todos los momentos que hemos compartido juntas. Soy muy afortunada de tenerlos en mi vida y les agradezco profundamente.

A Laura Cañón, la mejor amiga que he podido tener, por su apoyo y amor y por darme los empujones necesarios para seguir adelante. A mis amigas Paula, Karol, Sharin y Angélica, por enseñarme el poder que hay en mi interior, por sus risas, abrazos y besos y por todos los hermosos momentos juntas.

Agradezco a mis directores de tesis Piedad y Rafael por creer en mi y darme la oportunidad de trabajar a su lado, por su guía y sus recomendaciones, sin las cuales no hubiera podido terminar este proyecto. Son seres humanos excelentes y me han permitido con este trabajo aportar a una problemática que me importa y conmueve.

Por último, le agradezco a la Universidad Nacional de Colombia por ser mi alma máter, por brindarme todas las herramientas necesarias para triunfar como profesional; por aportar a mi desarrollo como un ser humano íntegro y exitoso. Soy lo que soy por esta universidad, me lo ha dado todo y le estaré eternamente agradecida.

Resumen

Modelo para la caracterización y clasificación de los tipos de violencia intrafamiliar desde los registros del sistema de salud

En este trabajo se presenta un modelo de identificación de población víctima basado en diagnósticos clínicos de individuos víctimas de violencia intrafamiliar. Este modelo, determina los patrones de diagnóstico del individuo, identifica sus multi-morbilidades y los usa para identificar correctamente a la población víctima, de la población general. El análisis se realiza considerando dos tipologías específicas de violencia intrafamiliar y los datos corresponden a víctimas del Departamento de Risaralda, Colombia. Los conceptos y métodos utilizados vienen de la fundamentación del análisis de redes y del aprendizaje relacional estadístico.

Palabras clave: Análisis de redes, aprendizaje relacional estadístico, violencia intrafamiliar, tipologías de violencia intrafamiliar, multi-morbilidades.

Abstract

Model for the characterization and classification of the types of domestic violence from the records of the health system

This work presents an identification model based on clinical diagnoses of individual victims of domestic violence, which allows the identification of multi-morbidities in the clinical history of the individual and, based on these diagnostic patterns, correctly identify the victim population from the general population. The analysis is carried out considering two specific typologies of domestic violence and the data correspond to victims from the Department of Risaralda, Colombia. The concepts and methods used are based on network analysis and statistical relational learning.

Keywords: Network analysis, statistical relational learning, domestic violence, typologies of domestic violence, multi-morbidities.

Lista de Figuras

3-1. Diagrama de barras del género de las víctimas por grupos etarios vs la población general	21
3-2. Histogramas de la edad de las víctimas por género	21
3-3. Diagrama de cajas de la edad de las víctimas por género	22
3-4. Diagrama de cajas de la edad de las víctimas por género y tipo de violencia intrafamiliar	22
3-5. Fuente: Tomado de Leone et al. (2019) Red parte A	26
3-6. Fuente: Tomado de Leone et al. (2019) Red parte B	27
4-1. Entradas y salidas del algoritmo propuesto	36
4-2. Partes del algoritmo propuesto	36
4-3. División en grupos de víctimas y no víctimas para las bases RIPS y SIANIESP	37
4-4. Selección de las poblaciones de interés para las bases RIPS y SIANIESP . .	38
4-5. División poblaciones de interés de RIPS y SIANIESP en grupos etarios . . .	39
4-6. División en entrenamiento y prueba menores de edad	40
4-7. División en entrenamiento y prueba mayores de edad	40
4-8. División en entrenamiento y prueba población general	41
4-9. Resumen proceso de entrenamiento en la población de menores de edad . . .	46
4-10. Resumen proceso de entrenamiento en la población de mayores de edad . . .	46
4-11. Resumen proceso de entrenamiento en la población general	47
5-1. Red de diagnósticos clínicos de todas las víctimas	52
5-2. Red de diagnósticos clínicos de las víctimas menores de edad	54
5-3. Red de diagnósticos clínicos de las víctimas mayores de edad	57
5-4. Diagrama de los preliminares del algoritmo propuesto	61
5-5. Diagrama del algoritmo de separación del algoritmo propuesto	62
5-6. Diagrama del entrenamiento del algoritmo propuesto	63
5-7. Diagrama de la prueba del algoritmo propuesto	67

Lista de Tablas

4-1. Ejemplo de tabla con valores del grado nodal para cada corrida del algoritmo central	42
4-2. Ejemplo de tabla con valores del grado nodal para el capítulo traumatismos en la red de entrenamiento de víctimas	42
4-3. Ejemplo de tabla con valores del grado nodal para el capítulo traumatismos en la red de entrenamiento de no víctimas	42
4-4. Ejemplo de tabla con valores del DKL para el capítulo traumatismos comparando las redes entrenamiento víctimas y entrenamiento no víctimas	42
4-5. Ejemplo de tabla con valores de la Divergencia de Kullback Leibler para la corrida 1 del algoritmo central	43
4-6. Ejemplo de tabla con valores de la Divergencia de Kullback Leibler para las n corridas del algoritmo central	43
4-7. Ejemplo de tabla con valores de la Divergencia de Kullback Leibler para cada corrida del algoritmo central donde el test corresponde a víctimas	44
4-8. Ejemplo de tabla con valores de la Divergencia de Kullback Leibler para cada corrida del algoritmo central donde el test corresponde a no víctimas	44
4-9. Ejemplo de tabla con valores de la Divergencia de Kullback Leibler para la corrida 1 del algoritmo central donde el test corresponde a víctimas y el capítulo es Traumatismos	44
5-1. Tabla con el resultado del algoritmo propuesto para las víctimas de la población general	68
5-2. Tabla con el resultado del algoritmo propuesto para las víctimas mayores de edad	68
5-3. Tabla con el resultado del algoritmo propuesto para las víctimas menores de edad	69

1. Introducción

La violencia intrafamiliar corresponde a una subcategoría de la violencia interpersonal en la que se agrupan todas las conductas que se dan dentro del hogar y producen daño físico o psicológico a los miembros del mismo. En Colombia, ésta es una problemática compleja y vigente que se ha estudiado desde diferentes áreas del conocimiento con el fin de analizar sus causas, consecuencias sociales y relación con otras problemáticas como los trastornos mentales e intentos suicidas [15] [9] [5] [13].

En la literatura se encuentran múltiples investigaciones donde se ha estudiado la relación entre el intento suicida y las agresiones por violencia doméstica. En algunas de estas investigaciones se ha podido concluir que el intento suicida es consecuencia de haber sido víctima de violencia intrafamiliar y/o violencia sexual [30] [5]. El Departamento de Risaralda es uno de los departamentos de Colombia donde la tasa de intentos suicidas es más alta que el promedio nacional y considerando el aumento del intento suicida así como el aumento de denuncias por violencia intrafamiliar como consecuencia de la pandemia por Covid-19, se propone estudiar la problemática de la violencia intrafamiliar desde el análisis de diagnósticos clínicos, con el fin de desarrollar un modelo que permita identificar población víctima de violencia intrafamiliar, lo cual contribuiría a reducir de alguna manera el intento suicida.

El análisis de redes se ha usado en diferentes áreas del conocimiento para estudiar sistemas donde es importante entender y considerar las relaciones existentes entre los actores del mismo. En análisis relacionados con la violencia intrafamiliar, se han usado ampliamente las redes como metodología de trabajo, ejemplos de estos son: Uso del ARS (Análisis de redes sociales) para estudiar el comportamiento de los diagnósticos clínicos de la población víctima y brindar herramientas a los servicios de salud para una atención más efectiva [21] y uso de redes y algoritmos de aprendizaje profundo para análisis textuales de publicaciones en redes sociales con el fin de identificar automáticamente víctimas de violencia intrafamiliar en condiciones críticas.[37]

Considerando que el presente trabajo se concentra en los diagnósticos clínicos de los individuos reportados en los sistemas de salud colombianos, el análisis de redes es la metodología de trabajo propuesta para identificar patrones entre los diagnósticos y usarlos como insumos en un modelo de detección de población víctima que permita la caracterización y clasificación de los tipos de violencia intrafamiliar. Se espera que, el modelo pueda ser usado en la práctica como una herramienta de identificación en las instituciones de salud, que contribuya a identificar, apoyar y/o tratar a las víctimas de violencia doméstica. Así, considerando la

relación de causalidad entre la violencia intrafamiliar y el intento suicida, se espera que, el usar el modelo de identificación como herramienta de apoyo, puede contribuir a la reducción del intento suicida.

El presente trabajo está organizado de la siguiente manera: El segundo capítulo, corresponde al marco teórico donde se introduce toda la teoría necesaria para comprender el texto. El tercer capítulo, presenta un análisis del contexto de la violencia intrafamiliar en Colombia, se describen los datos, se expone el análisis demográfico de los mismos y se mencionan algunas investigaciones donde se utilizó el análisis estadístico de redes para analizar casos particulares de la violencia intrafamiliar. En el cuarto capítulo, se define la metodología de trabajo, se describen las fuentes de información, las bases de datos y se especifica el algoritmo de identificación propuesto. En el quinto capítulo, se presentan y discuten los resultados de las redes y los resultados del modelo de identificación propuesto. Por último, se presentan las conclusiones junto con las recomendaciones de trabajo futuro.

2. Marco Teórico

Considerando las limitaciones que presenta realizar un análisis estadístico de los diagnósticos de la población víctima sin considerar multi-morbilidades para la caracterización y clasificación de las víctimas, y las diferentes investigaciones en las que se han usado las redes como herramientas para analizar diagnósticos clínicos y multi-morbilidades en pacientes que han sufrido algún tipo de agresión; el presente capítulo introduce las bases conceptuales y teóricas sobre las cuales se establece la metodología de trabajo.

La teoría presentada en este capítulo corresponde a conceptos y definiciones relacionadas con el análisis estadístico de redes y el aprendizaje relacional. Específicamente, conceptos de teoría de grafos, análisis de redes, y aprendizaje relacional estadístico, los cuales son necesarios para entender los capítulos posteriores. Los conceptos y definiciones corresponden a una recopilación teórica cuya bibliografía principal se encuentra en [1], [20] y [36].

2.1. Teoría de Grafos

2.1.1. Definiciones

Las siguientes definiciones son tomadas de [1]:

Grafo: un grafo \mathcal{G} está definido como una pareja ordenada $(\mathcal{V}, \mathcal{E})$, donde \mathcal{V} es un conjunto finito no vacío de vértices o nodos y \mathcal{E} es el conjunto de aristas o enlaces entre los vértices $\mathcal{E} \subseteq \{(u, v) | u, v \in \mathcal{V}\}$. Algunos grafos especiales se definen de la siguiente manera:

- **Grafo sin auto-bucles:** cuando la relación \mathcal{E} es irreflexiva, lo que significa que $\forall v \in \mathcal{V}, (v, v) \notin \mathcal{E}$ se dice que el grafo está libre de auto-bucles. Esto significa que no hay forma de viajar al mismo vértice en una sola transición.
- **Grafo con auto-bucles:** cuando la relación \mathcal{E} satisface la siguiente restricción $\exists v \in \mathcal{V}, (v, v) \in \mathcal{E}$, se dice que el grafo tiene auto-bucles. Esto significa que uno puede viajar de regreso al mismo vértice a través de un borde sin dejarlo.

Además, denotamos por $V = |\mathcal{V}|$ y $E = |\mathcal{E}|$ el número de vértices y aristas, respectivamente, del grafo.

Grafo completo: un grafo completo es un grafo en el que existen vínculos entre cada par

de vértices. El grafo completo con vértices \mathbf{V} se denota con \mathcal{K}_V . Un grafo completo también se puede clasificar en con o sin bucles automáticos, de acuerdo con la definición de grafo.

Grafo no dirigido: cuando la relación \mathcal{E} es simétrica, lo que significa que $\forall(u, v) \in \mathcal{E} \rightarrow (v, u) \in \mathcal{E}$, se dice que el grafo no está dirigido. En otros términos, cuando hay una arista que une los vértices u a v , habrá un vínculo de v a u .

Cabe resaltar que, para la construcción de las redes complejas vamos a hacer uso justamente de grafos no dirigidos.

Grafo dirigido (dígrafo): cuando la relación \mathcal{E} satisface la siguiente restricción: $\exists(u, v) \in \mathcal{E} | (v, u) \notin \mathcal{E}$, se dice que la gráfica está dirigida (dígrafo). En otros términos, este tipo de grafo debe tener al menos un borde arbitrario que vincule a u con v , sin el vínculo opuesto.

Grafo ponderado: un grafo ponderado \mathcal{G} se define como un triple $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathbf{W} \rangle$, donde \mathcal{V} y \mathcal{E} son los conjuntos de vértices y aristas, respectivamente, y \mathbf{W} es una matriz que lleva los pesos de las aristas. Por ejemplo, la entrada $\mathbf{W}_{uv} = w$, $(u, v) \in \mathcal{E}$, fijando como $w > 0$ el peso de la arista que une los vértices u con v . Si $(u, v) \notin \mathcal{E} \rightarrow \mathbf{W}_{uv} = 0$

Cuando \mathbf{W} es una matriz binaria, entonces el grafo ponderado se reduce a un grafo no ponderado, que es el grafo especial proporcionado en la definición de grafo.

2.1.2. Conectividad

Vértices adyacentes: dos vértices $u \in \mathcal{V}$ y $v \in \mathcal{V}$ se denominan adyacentes si comparten un borde común, en cuyo caso se dice que el borde común une los dos vértices.

En grafos no dirigidos, si u es adyacente a v , entonces v también debe ser adyacente a u .

En los dígrafos, u adyacente a v no implica que v sea adyacente a u . Específicamente, si $(u, v) \in \mathcal{E}$ y $(v, u) \notin \mathcal{E}$, entonces v es adyacente a u , pero lo contrario no se cumple.

Vecindad de un vértice: la vecindad de un vértice $v \in \mathcal{V}$, en un grafo \mathcal{G} es el conjunto de vértices adyacentes a v . La vecindad se denota por $\mathcal{N}(v)$ y está formalmente dada por $\mathcal{N}(v) = \{u : (v, u) \in \mathcal{E}\}$

Grado (valencia o conectividad) de un vértice: en un grafo no dirigido, el grado de un vértice v es el número total de vértices adyacentes a v . El grado de un vértice v se denota por k_v . Se puede definir de manera equivalente el grado de un vértice como la cardinalidad de su conjunto vecino y decir que, para cualquier vértice v , $k_v = |\mathcal{N}(v)|$, es decir,

$$k_v = |\mathcal{N}(v)| = |\{u : (v, u) \in \mathcal{E}\}| = \sum_{u \in \mathcal{V}} \mathbf{1}_{[(v, u) \in \mathcal{E}]}$$

En el que $\mathbf{1}_{[K]}$ representa la función delta o indicadora de Kronecker que da 1 si la expresión lógica \mathbf{K} es verdadera; de lo contrario, devuelve 0.

Los valores factibles de k_v están dentro del intervalo de valores discretos $\{0, \dots, V - 1\}$ si no se permiten bucles automáticos y en $\{0, \dots, V\}$ si se permiten bucles automáticos.

Cuando $k_v = 0$, entonces se dice que v es un vértice único o aislado.

Cuando k_v asume valores relativamente grandes que el resto de los vértices de la red, decimos que v es un centro.

En este caso, como las redes a construir están basadas en grafos no dirigidos, ésta va a ser la definición de grado nodal con la que se va a trabajar. Además, esta métrica, es la métrica de interés que se usa en el algoritmo propuesto.

Grado medio de la red: el grado medio de la red, o conectividad de red, viene dado por:

$$\bar{k} = \frac{1}{V} \sum_{v \in \mathcal{V}} k_v = \frac{1}{V} \sum_{(v,u) \in \mathcal{V}^2} \mathbb{1}_{[(v,u) \in \mathcal{E}]}$$

Fuerza: en un grafo ponderado no dirigido, la fuerza de un vértice $v \in \mathcal{V}$, indicado por s_v , representa la suma total de conexiones ponderadas de v hacia sus vecinos

$$s_v = \sum_{u \in \mathcal{V}} \mathbf{W}_{vu}$$

En la que \mathbf{W}_{vu} es el peso del borde de v a u .

Grafo regular: un grafo es regular si todos los vértices del grafo tienen el mismo grado. En particular, si el grado de cada vértice es k , se dice que \mathcal{G} es k -regular.

En el caso de las redes que tenemos no se encuentra esta característica particular, dado que si esto pasara no se podría usar la medida del grado de la red como métrica de discriminación entre grupos.

Si \mathcal{G} es un grafo completo con vértices V , entonces es $(V - 1)$ regular.

2.1.3. Caminos y Ciclos

Caminata: Sea $v_1, \dots, v_K \in \mathcal{V}$, $K \geq 2$. Una caminata \mathcal{W} es una secuencia ordenada de aristas: $\mathcal{W} = \{(v_1, v_2), (v_2, v_3), \dots, (v_{K-1}, v_K)\}$ tal que $\forall k \in \{2, \dots, K\} : (v_{k-1}, v_k) \in \mathcal{E}$. En este caso, v_1 y v_k se denominan origen y destino de la caminata, respectivamente. Nótese que los vértices se pueden volver a visitar en el mismo recorrido.

Una caminata se considera cerrada si $v_1 = v_K$ y abierto en caso contrario.

Una caminata que consta de un solo vértice se denomina caminata trivial.

Sendero: un sendero es una caminata en la que no se repite ningún borde. Los senderos también se pueden clasificar en senderos abiertos y cerrados.

Tour o circuito: un tour es un sendero cerrado.

Longitud de la caminata: la longitud de una caminata $\mathscr{W} = \{(v_1, v_2), (v_2, v_3), \dots, (v_{K-1}, v_K)\}$, $K \geq 2$, es el número de bordes que atraviesa la caminata, es decir, $|\mathscr{W}| = K - 1 \geq 1$

Ruta: Una ruta \mathscr{P} es una caminata no trivial en la que todos los vértices (excepto posiblemente el primero y el último) son distintos.

Una ruta es siempre una caminata.

Ciclo: un ciclo es una ruta cerrada.

Distancia de caminata o ruta: la distancia d de la caminata

$$\mathscr{W} = \{(v_1, v_2), (v_2, v_3), \dots, (v_{K-1}, v_K)\}, \quad K \geq 2$$

está dada por:

$$d(\mathscr{W}) = \sum_{k=2}^K |(v_{k-1}, v_k)| = \sum_{k=2}^K W_{k-1,k}$$

en el cual $|(v_{k-1}, v_k)|$ es el peso del borde que une el vértice v_{k-1} con v_k .

Ruta más corta (ruta geodésica) entre vértices: la ruta más corta entre $u \in \mathscr{V}$ y $v \in \mathscr{V}$, denotada aquí como d_{uv} , viene dada por la ruta que comienza en u y termina en v con la menor distancia. Matemáticamente,

$$d_{uv} = \min_{\mathscr{W}_{u \rightarrow v}} d(\mathscr{W}_{u \rightarrow v})$$

donde $\mathscr{W}_{u \rightarrow v}$ representa caminatas que comienzan en u y terminan en v .

Para medidas que requieren dos entradas, como la ruta más corta entre vértices, usamos d_{uv} cuando los subíndices son variables y $d_{1,2}$ cuando son números. Es decir, mantenemos la notación lo más sucinta posible. La coma se utiliza para mayor claridad cuando se indexan los números.

Distancia entre vértices: la distancia d_{uv} entre dos vértices u y v es siempre su distancia de trayectoria más corta.

Nótese que, d_{uv} siempre se evalúa a partir de una ruta. Es decir, la distancia entre u y v no puede ser una ruta que no sea un camino.

La distancia entre cualquier vértice y él mismo es 0.

Si no hay una ruta de u a v , entonces $d_{uv} = \infty$

2.1.4. Subgrafos

Accesibilidad: Decimos que $v_2 \in \mathscr{V}$ es accesible desde $v_1 \in \mathscr{V}$ si $d_{v_1 v_2}$ es finito. Alternativamente, v_1 llega a v_2 si hay al menos una caminata que comienza en v_1 y termina en v_2 .

Conectividad: el grafo \mathscr{G} está conectado si, para cada par de vértices v_1 y v_2 , se puede

acceder a v_2 desde v_1 o si se puede acceder a v_1 desde v_2 .

Fuerte conexión: el grafo \mathcal{G} está fuertemente conectado si, para cada par de vértices v_1 y v_2 , v_2 es accesible desde v_1 y v_1 es accesible desde v_2 .

Una conexión fuerte implica conexión.

En grafos no dirigidos, la conexión implica una fuerte conexión. Esto es cierto porque si v_1 llega a v_2 , entonces lo contrario debe ser cierto, ya que los bordes son bidireccionales en gráficos no dirigidos.

En los gráficos dirigidos, la conexión no implica una conexión fuerte.

Componente gráfico: El subgrafo \mathcal{G}_C de \mathcal{G} es un componente si:

- \mathcal{G}_C está conectado;
- Todos los subconjuntos adecuados de \mathcal{G}_C no están conectados.

Alternativamente, \mathcal{G}_C es un componente de grafo si dos de sus vértices son alcanzables al menos de uno a otro, y si sus miembros de vértice no están conectados a vértices adicionales en el resto del gráfico.

Un grafo conectado siempre tiene un solo componente.

Clique: un clique en un grafo no dirigido es un subconjunto de vértices de modo que cada dos vértices del subconjunto están conectados por un borde. Por lo tanto, las camarillas son subgrafos o grafos completos.

2.1.5. Representación gráfica

Matemáticamente, un grafo no ponderado $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ o un grafo ponderado $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{W} \rangle$ se representan con frecuencia mediante una matriz de adyacencia \mathbf{A} que se construye a partir de los conjuntos de vértices y aristas.

Matriz de adyacencia: Sea $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathcal{W} \rangle$ un grafo ponderado. Entonces, la matriz de adyacencia \mathbf{A} se define así:

- El número de vértices $|\mathcal{V}| = V$ sirve para establecer la dimensión de la matriz de adyacencia, que siempre es $V \times V$
- El conjunto de aristas contribuye a definir los valores de entrada de la matriz de adyacencia de la siguiente manera. La (i, j) -ésima entrada de \mathbf{A} se denota como $\mathbf{A}_{ij} = a_{ij} = \mathbf{W}_{ij}$, donde \mathbf{W}_{ij} es el peso del borde que une i con j . Formalmente, $\forall (i, j) \in \mathcal{E} : a_{ij} \neq 0$ y $\forall (i, j) \notin \mathcal{E} : a_{ij} = 0$

2.2. Análisis de redes

2.2.1. Definición

Una red se define como una colección de puntos conectados por enlaces, los puntos se denominan vértices o nodos y los enlaces se conocen como aristas. Esta definición es bastante flexible dependiendo de las diferentes conexiones y relaciones que especifiquen los enlaces. Existen en la literatura múltiples definiciones del concepto de red, pero esta es la más sencilla. [4], [25], [11]

2.2.2. Redes complejas

Las redes complejas son usadas para describir múltiples sistemas tecnológicos, físicos, biológicos e intelectuales. La célula, por ejemplo, se describe usando una red compleja cuyos nodos corresponden a sustancias químicas conectadas por reacciones químicas, el Internet es otro sistema que puede modelarse usando una red compleja de enrutadores y computadoras conectadas por varios enlaces físicos o inalámbricos. Estos son solo algunos ejemplos de los sistemas donde se implementa una red compleja para analizar su comportamiento. [1]

Desde la década de 1950, aquellas redes que no exhiben un diseño determinado se han catalogado como grafos aleatorios, los cuales son considerados como la representación más simple de una red compleja. Los matemáticos Paul Erdős y Alfred Rényi fueron los primeros en estudiar los grafos aleatorios en el modelo Erdos-Renyi, también conocido como modelo de redes aleatorias. Este modelo ha sido la guía para el análisis de redes complejas desde su publicación. Sin embargo, en ejemplos como los de la célula y el internet surge un interrogante interesante, ¿efectivamente, las redes reales detrás de estos sistemas complejos son fundamentalmente aleatorias?. Asumiendo que estos sistemas deben tener algún tipo de organización intrínseca en su topología, se han estado estudiando y desarrollando nuevas herramientas para analizar las características topológicas de las redes complejas cuyas características difieren a las de un grafo aleatorio.[1][36]

2.2.3. Redes aleatorias

En el artículo que data de 1959, Erdős y Rényi desarrollaron un modelo que genera redes aleatorias que consisten en vértices V y aristas E . Partiendo de vértices V completamente desconectados (sin bordes en la red), la red se construye a partir de la adición gradual de bordes L creados aleatoriamente, de tal manera que se evita el auto bucle. Otro modelo similar establece V vértices en una red, y existe una probabilidad $p > 0$ de conectar cada posible par de vértices. Este último modelo es ampliamente reconocido como el modelo de Erdős y Rényi. Cabe resaltar, que no se utiliza ninguna relación espacial entre los vértices. En esta formación de red, simplemente se crean bordes de una manera probabilística uniforme, independientemente de la similitud entre los vértices.

Dado que, para cada vértice $i \in \mathcal{V}$ de la red (un total de V), existen $V-1$ posibilidades diferentes de conexiones con otros vértices, se deduce que la cardinalidad del espacio muestral, $|\Omega|$, que cuantifica el máximo número teórico de aristas entre los vértices, viene dado por:

$$|\Omega| = \frac{V(V-1)}{2}$$

en el que la división por 2 proviene del hecho de que se está considerando que el grafo es no dirigido, es decir, las aristas siempre son bidireccionales en relación a ambos vértices enlazados. En general, la presencia de estos dos bordes representa la ocurrencia del mismo evento probabilístico, debido al acoplamiento inherente (bidireccional). Teniendo en cuenta que un borde arbitrario está presente en una red aleatoria con probabilidad p y está ausente con probabilidad $1-p$, y recordando que hay $\binom{V-1}{k}$ formas de elegir k vértices sobre $V-1$ en total, y p^k denota la probabilidad conjunta de estos k vértices de poseer exactamente k vértices conectados, entonces $\binom{V-1}{k}p^k$ proporciona la probabilidad de que estos k vértices tengan exactamente k otros vértices interconectados. Sin embargo, en este análisis, se debe imponer que no existen más aristas más allá de estos k , es decir, para la cantidad de vértices reminiscentes, $V-1-k$, el evento probabilístico complementario de aristas existentes, es decir, $(1-p)^{(V-1-k)}$, debe suceder. Ante este razonamiento, la distribución de grados sigue una distribución Binomial con parámetros Binomial($V-1, p$), cuya ecuación se rige por la siguiente expresión

$$P(k) = \binom{V-1}{k} p^k (1-p)^{V-1-k}$$

Dado que $V \rightarrow \infty$ y $p \ll 1$, se puede demostrar que una distribución Binomial parametrizada con Binomial($V-1, p$) se aproxima asintóticamente a una distribución de Poisson con parámetro Poisson(λ), con la siguiente condición de vinculación:

$$(V-1, p) = \lambda$$

Nótese que, la media μ y la varianza, σ^2 , de un Poisson(λ) están dadas por $\mu = \sigma^2 = \lambda$. Además, la ruta media más corta $\langle d \rangle$ es pequeña en redes aleatorias. Esta cantidad aumenta proporcionalmente al logaritmo del tamaño de la red, es decir, $\langle d \rangle \sim \frac{\ln(V)}{\ln(\langle k \rangle)}$, donde $\langle k \rangle$ viene dado por el valor promedio de la distribución de Poisson (grado medio), lo que significa que $\langle k \rangle = \lambda = (V-1)p$.

El gran descubrimiento de Erdős y Rényi fue que muchas propiedades importantes de una red aleatoria pueden revelarse a medida que se modifican los parámetros de una Binomial($V-1, p$). En su estudio, demostraron que, para valores de la probabilidad de conexión p mayores que una probabilidad crítica p_c , casi todas las redes aleatorias presentan una propiedad específica \mathbf{Q} con probabilidad 1. Esa misma propiedad no se verifica siempre que $p \leq p_c$. Por ejemplo, si p es mayor que un cierto valor de p_c , las redes aleatorias pueden presentar un solo componente conectado. Pero, para valores por debajo de este umbral crítico, las redes aleatorias ya no presentan un solo componente, sino varios subgrafos no conectados.

2.2.4. Redes de mundo pequeño

Varias redes del mundo real exhiben la propiedad de mundo pequeño, es decir, la mayoría de los vértices pueden ser alcanzados por otros, por medio de una pequeña cantidad de pasos intermedios (bordes). Esta característica se encuentra, por ejemplo, en las redes sociales, donde prácticamente todas las personas del mundo pueden ser contactadas por una pequeña cadena de personas.

Para construir una red que presente la propiedad del mundo pequeño, se puede utilizar el siguiente proceso de formación de redes:

- Inicialmente, la red es regular, con V vértices, en la que cada vértice se conecta a sus k vecinos más cercanos en cada dirección, totalizando $2k$ conexiones
- Luego, cada borde se reubica al azar, es decir, dado un vértice arbitrario $i \in \mathcal{V}$, elegimos al azar una de sus $2k$ conexiones originales. La arista seleccionada, digamos que enlazan los vértices i y $j \in \mathcal{V}$, se reubica aleatoriamente, de modo que el destino de j se cambia a otro vértice $u \in \mathcal{V}$, $j \neq u$, con probabilidad p .

Cuando $p = 0$, no se realizan reordenamientos y, por tanto, la red sigue siendo regular. Por el contrario, cuando $p \rightarrow 1$, todos los bordes están efectivamente reubicados. Cabe resaltar que, para $p = 0$, la red resultante es virtualmente regular. A medida que p aumenta (pero sigue siendo pequeño), la propiedad del mundo pequeño se hace evidente. Cuando $p = 1$, la red resulta ser aleatoria. En este caso, el pico de la distribución de grados, siguiendo este enfoque, se sitúa cerca de $2k$.

La implicación inmediata para las redes que tienen la propiedad de mundo pequeño es que la difusión de cualquier información, dado que se generó en cualquier vértice arbitrario de la red, es muy rápida. Por ejemplo, en las redes de contagio viral con la propiedad del pequeño mundo, dado que una persona ha contraído algún virus, entonces se espera que, en poco tiempo, muchas personas se infecten por este virus debido a la topología de la red que favorece la rapidez propagación.

2.2.5. Redes agrupadas aleatorias

Algunas redes del mundo real, como las sociales y biológicas, presentan estructuras modulares llamadas comunidades. Estas comunidades consisten en conjuntos de vértices que satisfacen una regla simple: los vértices que pertenecen a la misma comunidad tienen muchos bordes interconectados, mientras que diferentes comunidades comparten relativamente pocos bordes interconectados entre sí. Girvan & Newman (2002) propusieron un modelo para generar tales comunidades. Este método aglomerativo agrupa V vértices inicialmente aislados en M comunidades. Esto se gestiona creando un vínculo entre dos vértices con probabilidad p_{in} , si pertenecen a la misma comunidad, o con probabilidad p_{out} , si pertenecen a comunidades

distintas. Los valores de p_{in} y p_{out} se pueden elegir arbitrariamente para controlar el número de enlaces intracomunitarios e intercomunitarios, z_{in} y z_{out} , respectivamente, para un grado de red medio arbitrario $\langle k \rangle$. La red usada en [21], es una red social donde justamente se usan las comunidades de la red para realizar la caracterización de los diagnósticos de la población víctima.

Los valores altos de p_{in} y los valores bajos de p_{out} se refieren a redes con comunidades bien definidas, es decir, hay una alta concentración de bordes confinados dentro de cada comunidad y muy pocos bordes que interconectan diferentes comunidades. Por el contrario, valores bajos de p_{in} y valores altos de p_{out} contribuyen a la aparición de comunidades muy mezcladas entre sí. En base a estos parámetros, se puede definir la fracción de enlaces intracomunitarios $z_{in}/\langle k \rangle$ y, de igual forma, la fracción de enlaces intercomunitarios $z_{out}/\langle k \rangle$. La cantidad $z_{out}/\langle k \rangle$ define la mezcla entre diferentes comunidades. Esencialmente, a medida que aumenta $z_{out}/\langle k \rangle$, las comunidades se vuelven más mixtas y más difíciles de identificar.

Empíricamente, el $p_{out} \ll p_{in}$ debe cumplirse para garantizar la presencia de comunidades en la red.

2.2.6. Distribución del grado

Erdos y Renyi fueron los primeros en estudiar la distribución del grado mínimo y máximo en un grafo aleatorio. Sea p la probabilidad de conexión, el grado k_i del nodo i sigue una distribución binomial de parámetros $N - 1$ y p :

$$P(k_i = k) C_{N-1}^k p^k (1-p)^{N-1-k}$$

Esta probabilidad representa el número de formas en las que k enlaces se pueden trazar desde un determinado nodo. Así, la probabilidad de k enlaces es p^k y la probabilidad de ausencia de aristas adicionales es $(1-p)^{N-1-k}$, y existirían C_{N-1}^k formas de seleccionar los k puntos para esos enlaces. Con base en esta probabilidad se realiza un desarrollo teórico que permite concluir que para N suficientemente grande la distribución del grado en un grafo aleatorio sigue una distribución Poisson: $P(k) \approx \exp - \langle k \rangle \frac{\langle k \rangle^k}{k!}$. Adicional, se evidenció que así la posición de los enlaces sea aleatoria, los grafos aleatorios son muy homogéneos, dado que la mayoría de nodos tienen el mismo número de enlaces.

2.2.7. Medidas de redes complejas

Densidad: La densidad de red denotada por D mide qué tan fuertes están conectados los vértices de un grafo. Se define como la fracción de conexiones reales sobre el total de conexiones posibles.

Para una red dirigida, la densidad D se define como:

$$D = \frac{E}{2\binom{V}{2}} = \frac{2E}{2V(V-1)} = \frac{E}{V(V-1)}$$

donde $2\binom{V}{2}$ denota el número total de conexiones posibles en un gráfico dirigido. En especial, el binomio cuenta para obtener el número total de combinaciones por pares entre dos vértices en la red. Multiplicamos por dos porque el orden (vértices de inicio y destino) de esas conexiones por pares importa en un gráfico dirigido.

Para una red no dirigida, la densidad D es:

$$D = \frac{E}{\binom{V}{2}} = \frac{2E}{V(V-1)}$$

en el que, en este caso, no importa el orden de las conexiones por pares. La densidad asume valores en el intervalo $[0,1]$. Cuando $D = 0$, decimos que \mathcal{G} es un gráfico vacío. Por el contrario, cuando $D = 1$, se dice que \mathcal{G} es un grafo clique completo o máximo.

Asortatividad: Esta medida captura, en un sentido estructural, la preferencia de los vértices para unirse a otros que son similares o diferentes en términos del grado. La asociatividad a menudo se operacionaliza como un grado de correlación entre vértices. El coeficiente de asortatividad r es esencialmente el coeficiente de correlación de Pearson de grado entre pares de vértices vinculados. Por tanto, los valores positivos de r indican una correlación entre vértices de grado similar, mientras que los valores negativos indican relaciones entre vértices de diferentes grados. En general, r se encuentra entre -1 y 1 . Cuando $r = 1$, se dice que la red tiene patrones de mezcla de ordenación perfecta, mientras que en $r = -1$ la red es completamente desasortativa.

2.3. Divergencia de Kullback-Leibler

La divergencia de Kullback-Leibler también conocida como entropía relativa entre dos funciones de probabilidad $f(x)$ y $g(x)$ se define como:

$$D(f||g) = \int f(x) \log \frac{f(x)}{g(x)} dx$$

La divergencia satisface tres propiedades:

- Auto-similitud: $D(f||f) = 0$
- Auto-identificación: $D(f||g) = 0$ solo si $f = g$
- Es positiva

$$D(f||g) \geq 0$$

para todos f, g

Se utiliza principalmente para medir la similitud entre dos funciones de densidad.

[17] [16]

Existen múltiples problemas donde puede aplicarse como en reconocimiento de imágenes, análisis del discurso, análisis textual, entre otros. En este caso, se va a utilizar para evaluar la similitud entre las distribuciones del grado nodal para las víctimas y no víctimas de violencia intrafamiliar.

2.4. Aprendizaje relacional estadístico

A continuación, se presenta una breve descripción del área de trabajo del aprendizaje relacional estadístico tomado de [20].

El aprendizaje relacional estadístico (SRL por sus siglas en inglés) es una área emergente de investigación. La investigación de SRL intenta representar, razonar y aprender en dominios con una estructura relacional compleja y rica probabilísticamente. Otros términos que se han utilizado recientemente incluyen aprendizaje de lógica probabilística y minería de datos multirrelacional. Muchas de las tareas conocidas como problemas de predicción estructurada también se superponen en gran medida con los problemas abordados por la investigación de SRL.

La mayoría de los sistemas de SRL propuestos se pueden distinguir en varias dimensiones. Los formalismos de representación más comunes se basan en formalismos lógicos (por ejemplo, formalismos basados en reglas) o basados en marcos (por ejemplo, orientados a objetos).

La semántica probabilística se basa principalmente en modelos gráficos o gramáticas estocásticas; Los primeros enfoques de SRL a menudo se definían en términos de modelos gráficos dirigidos (por ejemplo, redes bayesianas), mientras que recientemente ha habido un interés creciente en modelos no dirigidos (por ejemplo, redes de Markov). Los modelos dirigidos pueden representar modelos generativos complejos, mientras que los modelos no dirigidos pueden representar dependencias no causales. También son posibles otras alternativas, como las redes de dependencia y los modelos mixtos dirigidos y no dirigidos.

La semántica de muchos de los sistemas SRL se da en términos de un modelo gráfico desarrollado o en tierra. Por lo tanto, un enfoque para hacer inferencia en estos modelos es realizar la inferencia probabilística apropiada en el modelo de nivel base. Una simple optimización al estilo de KBMC es hacer uso de la consulta en la construcción de la red. En lugar de construir todo el modelo de nivel base, la construcción puede hacerse más eficiente construyendo solo la parte de la red requerida para responder a la consulta. Pero esto no explota ninguna de las estructuras inherentes al modelo probabilístico. Nótese que, en muchos casos los modelos se pueden descomponer en sistemas débilmente acoplados, y muestran cómo las interfaces entre los componentes pueden usarse para encapsular la inferencia dentro de los componentes. Esto permite la reutilización y el almacenamiento en caché de inferencias y

puede conducir a mejoras significativas en la eficiencia durante la inferencia. Los enfoques más generales, como la eliminación de variables de primer orden, combinan la eliminación de variables con la unificación y permiten realizar una inferencia elevada.

No es sorprendente que el aprendizaje sea un componente fundamental en cualquier enfoque de SRL. El poder de la representación estructurada es la naturaleza jerárquica de los modelos estadísticos. La ventaja de los modelos jerárquicos, y lo que los distingue de los modelos estadísticos "planos", es compartir o vincular parámetros. El intercambio de parámetros ocurre cuando los parámetros potencialmente distintos del modelo están restringidos a ser los mismos. Un ejemplo simple ocurre en un modelo de Markov oculto: debido a la suposición de Markov, los parámetros que determinan el siguiente estado son los mismos en cada instancia de tiempo, por lo tanto, no requerimos parámetros distintos indexados por valores específicos de t , simplemente se tiene un conjunto de parámetros $\theta_{t+1|t}$.

Esta vinculación de parámetros no solo nos proporciona un modelo compacto para clases ricas de distribuciones, sino que también permite que la estimación de parámetros robustos sea factible. A diferencia de los escenarios tradicionales de aprendizaje automático, donde el sistema de aprendizaje recibe como entrada una secuencia de i.i.d. observaciones, la entrada a un algoritmo de aprendizaje de SRL suele ser solo una instancia única, ricamente conectada. Si no hubiera compartición de parámetros, esta instancia sería de poca utilidad para realizar inferencias estadísticas. Pero, debido a que se usan los mismos parámetros en varios lugares del modelo, aún se pueden extraer estadísticas significativas de los datos para usar en nuestros procedimientos de inferencia estadística.

En particular, para dar respuesta a nuestra pregunta de investigación, se hará uso del análisis de redes complejas construyendo un algoritmo que detecte población víctima usando la distribución del grado nodal y la divergencia de Kullback-Leibler para identificar los capítulos de diagnósticos que permiten distinguir entre víctimas y no víctimas y así, poder construir un modelo de detección que identifique población víctima de violencia intrafamiliar y permita caracterizar y clasificar los tipos de violencia intrafamiliar.

3. Contexto de la violencia intrafamiliar

3.1. Violencia intrafamiliar en Colombia

En el informe mundial sobre la violencia y la salud traducido por la Organización Panamericana de la Salud, la OMS plantea que no existe país ni comunidad en el mundo que sea ajeno a la violencia. “Cada año, más de 1,6 millones de personas en el mundo pierden la vida violentamente” [8]. Además, clasifica a la violencia en tres categorías según el autor del acto violento: violencia dirigida contra uno mismo (también denominada violencia autoinfligida), violencia interpersonal y violencia colectiva. Estas categorías a su vez se subdividen en tipos de violencia más específicos: la violencia dirigida contra uno mismo se subdivide en comportamiento suicida y autolesiones, la violencia interpersonal en violencia de familia o pareja y comunidad, y la violencia colectiva en social, política y económica. De esta manera, según el informe, la violencia intrafamiliar o de pareja es un tipo de violencia dentro de la categoría de violencia interpersonal que se define como aquel acto violento que en la mayoría de los casos se produce entre miembros de la familia o compañeros sentimentales, y suele acontecer en el hogar. A su vez, abarca formas de violencia como el maltrato de los niños, la violencia contra la pareja y el maltrato de los ancianos.

En Colombia, en el concepto 123 de 2017 el Instituto Colombiano de Bienestar Familiar (ICBF) define la violencia intrafamiliar de la siguiente manera:

“La violencia intrafamiliar es una problemática que atenta y destruye la unidad y armonía familiar consagrada en la Constitución Política y que es abordada por la Ley desde la perspectiva de protección a las víctimas y punitiva respecto de los agresores. Se entiende como toda conducta que se da dentro del contexto familiar, que produzca daño físico, psíquico o a su integridad sexual, o amenaza, agravio, ofensa o cualquier otra forma de agresión por parte de un miembro del grupo familiar.” [18]

Considerando lo anterior, se puede ver que en la literatura, la violencia intrafamiliar tiene varias definiciones. Sin embargo, en relación con lo postulado por la OMS, por el ICBF y con el fin de unificar criterios, se definirá la violencia intrafamiliar como: todo acto de violencia física, psicológica, sexual y/o económica que produce un miembro de la familia a otro. [35] [18] [19] [28] [40] [8]

En el informe Violencia de género e intrafamiliar publicado por el Instituto Nacional de Salud se presenta la metodología para la vigilancia en salud pública del evento violencia de género

e intrafamiliar en Colombia, a través del Sistema Nacional de Vigilancia en Salud Pública - Sivigila y se postula que “las violencias de género y entre ellas, las violencias sexuales, son violaciones sistemáticas y masivas de los derechos humanos que afectan la vida, la salud física, mental y social, la integridad, la libertad e igualdad de niñas, niños, adolescentes, jóvenes, mujeres, hombres y personas con orientaciones diferentes a la heterosexual, en particular a las de la comunidad de lesbianas, gays, bisexuales, transgénero e intersexuales (LGBTI)” [32].

De igual manera, el informe [32] señala que “en 2018 se notificaron al Sivigila 109 699 casos de violencia de género e intrafamiliar; 52 548 (47,9%) de violencia física, 29 804 (27,2%) de violencia sexual, 18 647 (17,0%) de negligencia y abandono y 8 700 (7,9%) de violencia psicológica. En el 79,9% (87 699) de los casos la víctima fue una mujer, según la edad de la víctima se observaron mayores incidencias de violencia física en personas entre los 15 y 44 años, de violencia psicológica entre los 25 y 49 años, en violencia sexual de 5 a 14 años, (el grupo de edad de 10 a 14 años es el más afectado) y de negligencia y abandono en los menores de cinco años” y sostiene que “la exposición a la violencia en la niñez (como víctima o como testigo) aumenta el riesgo de sufrir otras formas de violencia en etapas posteriores de la vida”. Además, señala que parece existir algún tipo de acuerdo generalizado que persuade a las mujeres de buscar ayuda cuando son víctimas de violencia o de aceptar el apoyo de terceros en caso de maltrato por parte de su pareja.

Por su parte, el Instituto Nacional de Medicina Legal y Ciencias Forenses en su reporte anual mantiene que “durante el año 2018 se conocieron 28 645 casos de violencia intrafamiliar, correspondiente a la población adulta mayor de 2 261 casos (7,8%), población de niños, niñas y adolescentes con 10 794 casos (37,6%) y por último violencia entre otros familiares con 15 590 casos (54,4%)” [19]. De los casos reportados tanto en población adulta mayor como en niños, niñas y adolescentes la mayoría de las víctimas son de sexo femenino.

Si bien al evaluar estos reportes e informes, se evidencia una medición anual de los casos de violencia intrafamiliar y se observa que hay reportes en múltiples fuentes como medicina legal y Sivigila, en Colombia y el mundo se ha evidenciado un subregistro de los casos reportados de diferentes tipos de violencia dado que un porcentaje de la población no denuncia las agresiones por miedo, vergüenza, temor a represalias, entre otros. [22].

Adicional a lo anterior, en múltiples investigaciones se ha estudiado la relación entre el intento suicida y el haber sido víctima de violencia intrafamiliar [26] [10] [15] [13]. En Colombia, esta relación en particular es de vital interés. Se destaca el caso del Departamento de Risaralda, donde el suicidio se considera una problemática de salud pública que tiene una estrecha relación con la violencia intrafamiliar. Se han realizado múltiples estudios al respecto y se evidenció que el intento suicida en algunos casos, se genera como resultado de situaciones adversas como la violencia intrafamiliar y violencia sexual. [30] [5]

De igual manera, el Departamento se encuentra en el top 10 de los departamentos con más suicidios en Colombia, ha sufrido un aumento del intento suicida y para el año 2022,

según el Instituto Nacional de Salud (INS) el promedio de intentos de suicidio por cada 100 000 habitantes es de aproximadamente 15,8 mientras que el promedio nacional está en 13,5 aproximadamente.

Considerando lo anterior, es importante analizar la información de las víctimas de violencia intrafamiliar en el Departamento de Risaralda, para definir si es posible implementar el análisis de redes dentro de un algoritmo de aprendizaje relacional, con el fin de construir un modelo de identificación de población víctima que permita caracterizar las tipologías de violencia intrafamiliar.

3.2. Análisis descriptivo de las víctimas de violencia intrafamiliar en Risaralda

En esta sección se realizará un análisis descriptivo de los individuos del Departamento de Risaralda que han reportado por lo menos una agresión por violencia intrafamiliar en la ficha 875. Se analizarán la cantidad de agresiones que han reportado, la cantidad de atenciones en urgencias, hospitalizaciones y consultas programadas, junto con una caracterización demográfica de los individuos.

3.2.1. Medidas de tendencia central de las agresiones y atenciones médicas de los individuos víctimas de violencia intrafamiliar

Evento 875 de la ficha para Vigilancia en salud pública de la violencia de género e intrafamiliar código INS: 875

Se evidenció que los 12 722 registros en la base de datos corresponden a 11 593 individuos. 688 de ellos reportaron por lo menos dos agresiones por violencia intrafamiliar. Se calcularon medidas de tendencia central para aquellos individuos que reportaron haber sufrido por lo menos dos agresiones por violencia intrafamiliar:

- Media: 2,118 agresiones
- Mediana: 2 agresiones
- Máximo: 5 agresiones

Urgencias

Se evidenció que de los 340 843 registros en la base de datos sólo 6 495 corresponden a 1 810 individuos con por lo menos un reporte en la base del evento 875. Los cinco diagnósticos más recurrentes de estos 1 810 individuos a la salida de urgencias son: Diarrea y gastroenteritis de presunto origen infeccioso, abuso sexual, rinofaringitis aguda (resfriado común), infección viral, no especificada y asma, no especificada.

Se calcularon medidas de tendencia central del número de atenciones en urgencias de los 1 810 individuos que reportaron haber sufrido por lo menos una agresión por violencia intrafamiliar:

- Media: 3,445 atenciones
- Mediana: 2 atenciones
- Máximo: 63 atenciones

Consultas

Se evidenció que de los 8 000 378 registros en la base de datos sólo 99 137 corresponden a 8 817 individuos con por lo menos un reporte en la base del evento 875. Los cinco diagnósticos principales de consulta más recurrentes en estos individuos son: Caries de la dentina, supervisión de embarazo de alto riesgo (sin otra especificación), control de salud de rutina del niño, diarrea y gastroenteritis de presunto origen infeccioso y rinofaringitis aguda (resfriado común)

Se calcularon medidas de tendencia central del número de atenciones en consultas de los individuos que reportaron haber sufrido por lo menos una agresión por violencia intrafamiliar:

- Media: 11,23 consultas
- Mediana: 7 consultas
- Máximo: 252 consultas

Hospitalizaciones

Se evidenció que de los 214 419 registros en la base de datos sólo 5 155 corresponden a 1 927 individuos con por lo menos un reporte en la base del evento 875. Los 5 diagnósticos más recurrentes al ingreso de la hospitalización de estos individuos son: Abuso sexual, neumonía (no especificada), parto único espontáneo (presentación cefálica de vértice), neumonía bacteriana (no especificada) y diarrea y gastroenteritis de presunto origen infeccioso. Los 5 diagnósticos más recurrentes al egreso de la hospitalización de estos individuos son: Parto único espontáneo (presentación cefálica de vértice), abuso sexual, neumonía (no especificada), parto único espontáneo (sin otra especificación) y neumonía bacteriana (no especificada).

Se calcularon medidas de tendencia central del número de hospitalizaciones de los individuos que reportaron haber sufrido por lo menos una agresión por violencia intrafamiliar:

- Media: 2,654 atenciones
- Mediana: 2 atenciones
- Máximo: 47 atenciones

RIPS

Se evidenció que de los 10 401 910 registros en la base de datos sólo 159 429 corresponden a 9 725 individuos con por lo menos un reporte en la base del evento 875. Los 5 diagnósticos más recurrentes en las consultas de estos individuos son: Control de salud de rutina del niño, caries de la dentina, examen médico general, supervisión de embarazo de alto riesgo, sin otra especificación y rinofaringitis aguda (resfriado común)

Se calcularon medidas de tendencia central del número de consultas de los individuos que reportaron haber sufrido por lo menos una agresión por violencia intrafamiliar:

- Media: 16,39 consultas

- Mediana: 12 consultas

- Máximo: 157 consultas

3.2.2. Caracterización de las víctimas

Evento 875 de la ficha para Vigilancia en salud pública de la violencia de género e intrafamiliar código INS: 875

Se define la población víctima, como todos aquellos individuos que hayan reportado algún tipo de agresión en la ficha 875, donde la modalidad de violencia sea física o psicológica. La base de datos de todos los individuos que han reportado un acto de violencia en la ficha 875 cuenta con 12 722 registros los cuales corresponden a 11 593 individuos. De estos, sólo 5 734 individuos reportaron un acto de violencia física o psicológica. De ellos, 262 reportaron por lo menos 2 actos de violencia.

Análisis demográfico

A continuación se presenta la distribución por género y edad de las 5 734 víctimas, diferenciando por dos grupos etarios: menores de edad y mayores de edad:

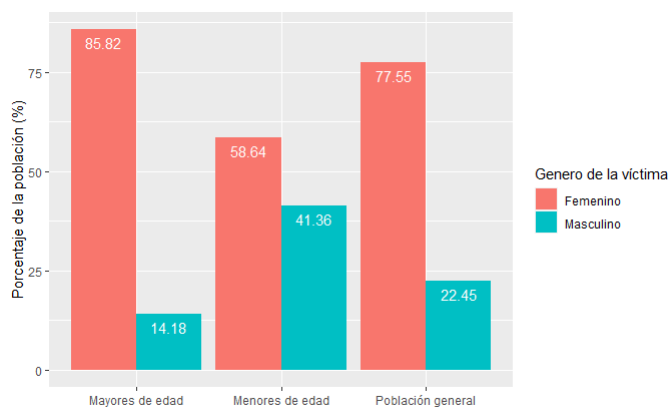


Figura 3-1.: Diagrama de barras del género de las víctimas por grupos etarios vs la población general

En la figura **3-1** se observa el porcentaje de víctimas por género, en cada una de las tipologías de violencia intrafamiliar definidas y en la población general sin diferenciar por grupo etario. La figura muestra que en todos los casos la proporción de mujeres es mayor que la de hombres. Sin embargo, el comportamiento difiere entre los dos grupos etarios dado que para las víctimas menores de edad la relación hombre-mujer es casi 1 a 6 mientras que para las víctimas mayores de edad la relación es aproximadamente 1 a 1.

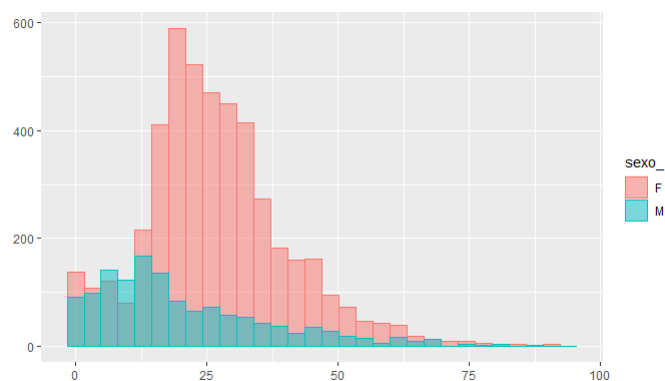


Figura 3-2.: Histogramas de la edad de las víctimas por género

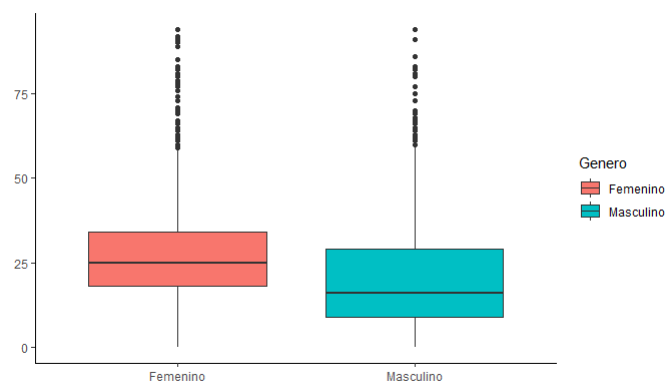


Figura 3-3.: Diagrama de cajas de la edad de las víctimas por género

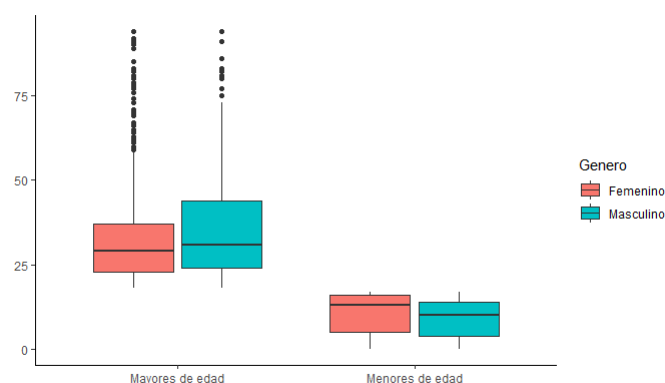


Figura 3-4.: Diagrama de cajas de la edad de las víctimas por género y tipo de violencia intrafamiliar

En las figuras 3-2 3-3 3-4 se observa que en la población general de víctimas, las mujeres tienen una edad mediana mayor que los hombres. Sin embargo, al evaluar el comportamiento de la edad dentro de cada grupo etario se puede observar que para las víctimas mayores de edad la edad mediana de los hombres es mayor a la de las mujeres.

Como se mencionó anteriormente, el trabajo está enfocado en la caracterización de las tipologías de violencia intrafamiliar, en este caso, viendo las diferencias que presentan las víctimas dependiendo del grupo etario al que pertenecen, se propone basar la clasificación en la edad de la víctima al momento de la agresión, por tanto, la primera tipología corresponderá a menores de edad y la segunda tipología corresponderá a mayores de edad.

Cruce con las bases RIPS y SIVIGILA

Una vez determinada la población víctima se realiza un cruce con las bases del RIPS y de Sivigila. Así, al realizar el cruce con los datos del RIPS sólo se encuentra información de 587 víctimas y al hacer el cruce con los datos de Sivigila, se encuentra información de 4 075 individuos, al consolidar ambas bases y eliminar duplicados se tiene información de 4 147 individuos en total.

De los 587 individuos con información en el RIPS, se evidenció que los 5 diagnósticos principales con mayor número de registros son:

- Supervisión de embarazo de alto riesgo, sin otra especificación (178 registros)
- Consejo y asesoramiento general sobre la anticoncepción (124 registros)
- Examen médico general (84 registros)
- Hipertensión esencial (primaria) (76 registros)
- Caries de la dentina (71 registros)

De los 4 056 individuos con información en el Sivigila, se evidenció que los 5 diagnósticos principales con mayor número de registros son:

- Caries de la dentina (431)
- Hipertensión esencial (primaria) (257)
- Gingivitis crónica (232)
- Supervisión de embarazo de alto riesgo, sin otra especificación (189)
- Infección de vías urinarias, sitio no especificado (182)

Considerando lo anterior, se observa que los diagnósticos principales de la población víctima corresponden a diagnósticos comunes que al analizarse de forma individual no aportan mucha información acerca de la víctima y en general no podrían discriminar entre la población víctima y la no víctima. Además, se realizaron diferentes pruebas usando técnicas de estadística clásica, pero no fue posible encontrar una que permitiera identificar y/o caracterizar los patrones de diagnóstico de ambas poblaciones y diferenciarlas. Por tal motivo, a continuación, se presentan algunas investigaciones donde se hizo uso de redes para trabajar el análisis de diagnósticos clínicos y otros tipos de datos relacionados con la violencia intrafamiliar.

3.3. Uso de redes en el análisis de casos de violencia

Se han desarrollado diferentes metodologías matemáticas y estadísticas de análisis para examinar el problema de la violencia en el mundo. Dentro de las metodologías propuestas se han utilizado técnicas como: aprendizaje automático, aprendizaje profundo y análisis de redes sociales. Los modelos desarrollados han permitido identificar potenciales víctimas que no han reportado el hecho violento.

Dentro de los modelos de redes sociales se destaca el paper de Leone et al. [21] donde hacen uso del análisis de redes sociales (ARS) en los datos del Departamento de Emergencias (DE) de la región de Lazio, Italia, con el objetivo de identificar patrones de diagnósticos dentro de todos los accesos a urgencias de mujeres en situación de agresión y así, poder desarrollar una herramienta de evaluación de riesgos para los profesionales de urgencias que permita fortalecer su capacidad para gestionar víctimas de violencia.

Para el estudio se tomó una cohorte de 124 691 mujeres entre los 15 y los 70 años de edad con por lo menos una visita al Departamento de Emergencias por agresión entre 2003 y 2015 y, para cada una de ellas, se reconstruyó el historial de diagnósticos y traumas del DE por medio del Sistema de Información de Emergencias Sanitarias (SIES). La información histórica se definió como todos los diagnósticos y traumas reportados en todas las visitas al DE que ocurrieron durante los 3 años anteriores a la última visita por agresión. Los diagnósticos fueron divididos dentro de las categorías de violencia y no violencia. Cabe resaltar que la categoría de diagnósticos de violencia incluía a aquellos códigos de diagnósticos relacionados con síntomas y señales de psico-abuso lógico, físico o sexual, por ejemplo: “Maltrato de adultos no especificado” (ICD9-CM: código 995 80); “abuso físico para adultos” (ICD9-MC: código 995 81), entre otros. La presencia de uno de estos diagnósticos en el historial de urgencias identificó al grupo de mujeres víctimas de violencia (MVV) dentro de la población de estudio.

El análisis estadístico se realizó de la siguiente manera: La población seleccionada de 124 691 mujeres se dividió aleatoriamente en una submuestra de desarrollo de 93 519 mujeres correspondiente al 75 % de la población total y una submuestra de validación de 31 172. Para retener el porcentaje de MVV observado en la población de estudio (0,47 %) se realizó una estratificación proporcional. En la submuestra de desarrollo, se realizó un enfoque ARS para identificar patrones de diagnóstico de violencia a partir de la historia del DE. Luego, sobre la base de los resultados del ARS, se desarrolló un modelo predictivo para evaluar el impacto de estos patrones de diagnóstico en las probabilidades de una mujer de pertenecer al grupo MVV en la submuestra de validación. Por último, se propuso una tabla de riesgos como un marco potencial para que el personal de los servicios de urgencias priorice y gestione rápidamente los riesgos de salud potencialmente relacionados con la violencia de género.

De esta manera, se realizó el análisis con base en 29 datos de información clínica: nueve diagnósticos de MVV, otros 15 grupos diagnósticos y cinco grupos de traumatismos. Adicio-

nal, para cada mujer, se recopiló información demográfica como la edad al momento de la visita por agresión, lugar de residencia y nacionalidad. Una vez creada esta red, se aplicaron algoritmos de detección de comunidades para utilizar las interacciones más densas entre los diagnósticos asociadas con las medidas de centralidad de los mismos para obtener solamente aquellos nodos que están fuertemente asociados.

Para el análisis de redes sociales, se definió como nodo cada diagnóstico y traumatismo informado en cada visita al DE. Para construir el gráfico de la red social, se organizó la información en una matriz X simétrica $N \times N$ donde la celda genérica x_{ij} expresa una frecuencia conjunta. Este último valor viene dado por el número de veces que se recuperó el nodo c_i (diagnóstico o trauma i) junto con el nodo c_j (diagnóstico o trauma j), es decir, el número de veces que este par de diagnósticos se informó en el historial de urgencias de mujeres incluidas en la submuestra de desarrollo. Luego se normalizó la matriz reduciendo todos los nodos a la misma frecuencia marginal utilizando el método iterativo de ajuste proporcional (IAP). El método IAP es un procedimiento matemático mediante el cual los valores originales se ajustan iterativamente para ajustarse a los valores marginales de fila y columna dejando sin cambios las relaciones entre los nodos y la simetría de la matriz. Restringimos todos los nodos a tener frecuencias marginales iguales a 1. Se utilizó la normalización para evitar las relaciones artificiales de los diagnósticos/traumas de urgencias más comunes en los datos brutos.

Para identificar una comunidad dentro de la red se realizaron tres algoritmos de agrupamiento diferentes: Óptimo, Intermedio de arista y agrupamiento rápido y codicioso. Se utilizó el coeficiente de modularidad que indica qué tan bien se descompone una red en comunidades para definir el mejor algoritmo de agrupamiento. Con respecto a las medidas de centralidad, se utilizó el grado, la cercanía y la intermediación para cuantificar la fuerza de la conexión entre un diagnóstico dado o diagnósticos de trauma y violencia. En concreto, la centralidad del grado cuantifica los nodos con un gran número de conexiones. La centralidad de proximidad cuantifica los nodos fácilmente accesibles desde otros nodos en función de la distancia geodésica entre ellos. Finalmente, la centralidad de intermediación cuantifica los nodos que ocurren con frecuencia en las rutas más cortas entre otros nodos.

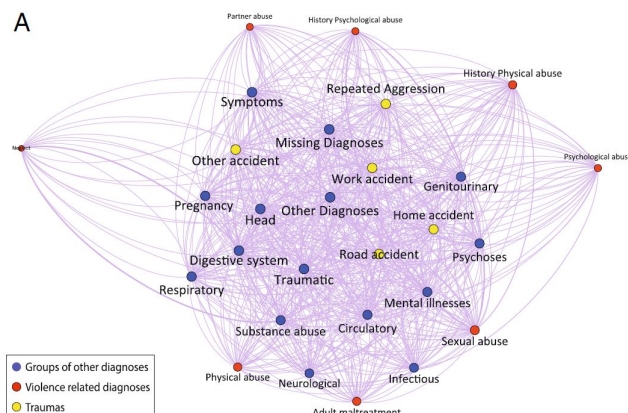


Figura 3-5.: Fuente: Tomado de Leone et al. (2019) Red parte A

Con respecto a los resultados, la inspección visual del gráfico de la red social muestra una red muy densa (densidad igual al 0,85 %), dado que cada nodo está conectado con la mayoría de los otros nodos al menos una vez (Fig. 1 Red parte A). Cada nodo representa una categoría específica de diagnóstico: violencia (color rojo), diagnósticos no violentos (color azul) y grupos de trauma (color amarillo).

Algunos nodos de la red eran más grandes que otros, lo que refleja el mayor peso de las entidades dentro de la red. Se puede observar que los diagnósticos reportados con mayor frecuencia durante las visitas a urgencias por mujeres que experimentan una agresión fueron: síntomas no especificados, diagnósticos del sistema traumático y musculoesquelético, diagnósticos de cabeza y rostro y enfermedades del aparato digestivo. El trauma reportado con más frecuencia fue un trauma no intencional que ocurrió en un lugar diferente al del trabajo, la escuela, la carretera o el hogar. Además, los diagnósticos de violencia mostraron una menor frecuencia de denuncia y, por lo tanto, se retiraron a la periferia de la red. Sin embargo, al destacar las asociaciones más fuertes (frecuencias conjuntas superiores al percentil 95 de la distribución de frecuencias conjuntas normalizadas FCN) en el gráfico, surgió un grupo complejo de conexiones con el diagnóstico de violencia.

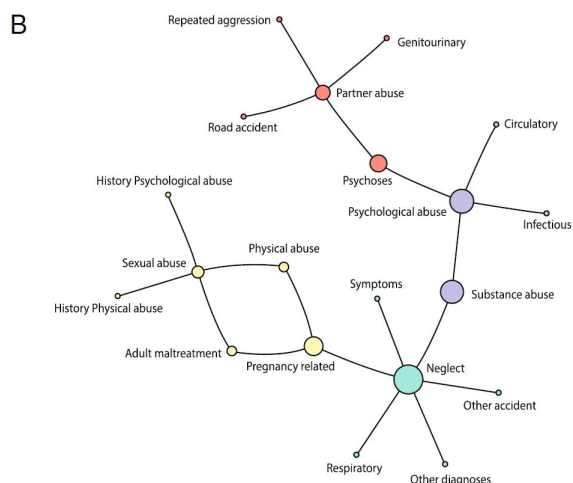


Figura 3-6.: Fuente: Tomado de Leone et al. (2019) Red parte B

Se identificaron cuatro comunidades, (Fig. 2 Red parte B) cada una representando unos patrones específicos de potenciales diagnósticos relacionados con un código de violencia. La comunidad violeta se caracteriza por un diagnóstico de abuso psicológico (adulto emocional / psico-abuso lógico) y mostró el nivel más alto de conectividad con alcohol o abuso de sustancias, sistema circulatorio y enfermedades infecciosas. En la comunidad naranja está involucrado un diagnóstico de abuso de pareja, los vínculos más fuertes fueron con accidentes de tráfico, trastornos de psicosis y enfermedades genitourinarias. Esta comunidad se caracteriza por un fuerte vínculo con la agresión repetida dentro de la historia en el DE. La comunidad verde se centra en un mal diagnóstico de tratamiento (negligencia de adultos, abuso de adultos y negligencia, maltrato de adultos no especificado) y está fuertemente conectado con enfermedades respiratorias y síntomas no especificados, otros traumas u otros diagnósticos sin asociación previamente reportada con violencia de género. La comunidad amarilla se caracteriza por un fuerte vínculo entre el abuso sexual y los diagnósticos de violencia de maltrato de adultos, abuso físico y antecedentes de abuso físico o emocional. Las condiciones relacionadas con el embarazo son el único diagnóstico no violento conectado dentro de esta comunidad.

Por último, dentro del mismo estudio se desarrolló un modelo logístico para validar la capacidad de estos patrones para predecir las probabilidades de que una mujer tenga antecedentes de violencia. Así, el análisis de redes demostró ser un instrumento adecuado para el estudio de información correspondiente a conjuntos de datos que presentan múltiples relaciones entre sí, en este caso particular, para diagnósticos comúnmente reportados por víctimas de violencia. A partir de un denso sistema de relaciones en el que cada diagnóstico se menciona a menudo entre sí, el método ARS es capaz de detectar patrones diagnósticos claros dentro de los cuales las referencias cruzadas son más frecuentes, mientras que los diagnósticos externos al grupo son raros. El método también aclaró los roles particulares que juegan algunos

diagnósticos; hay algunos diagnósticos que ocurren juntos, mientras que otros se encuentran en diferentes patrones de violencia, que conectan diferentes procesos de violencia. Así, el estudio de Leone [21] es el primero que aplica ARS de forma predictiva en Italia.

Por otro lado, se ha abordado el problema de la violencia doméstica a través de los modelos de aprendizaje automático, en [3] se propone un modelo de aprendizaje automático para predecir un arresto por violencia doméstica dentro de los dos años siguientes a la lectura de cargos. Los datos corresponden a 28 646 cargos por violencia doméstica que condujeron a cargos oficiales y una liberación entre el 1 de enero de 2007 y el 31 de octubre de 2011. Cabe aclarar que los datos corresponden únicamente a casos liberados dado que solo estos casos tienen la oportunidad de reincidir. El estudio definió un seguimiento de dos años para cada caso procesado hasta finales de octubre de 2013 considerando que este lapso de tiempo logra capturar la mayoría de los casos que les preocupan a las autoridades.

Para el estudio se consideraron tres posibles resultados: ningún arresto por un delito de violencia doméstica, un arresto por violencia doméstica que no involucra daño físico, un intento de causar daño físico o la amenaza de daño físico y un arresto por violencia doméstica que involucra lesiones físicas (incluida la violación), un intento de causar lesiones físicas o la amenaza de lesiones físicas. Es de aclarar, que en el modelo sólo se consideran casos en los que efectivamente se haya efectuado un arresto y se aclara que el hecho de no haberlo no implica que no se haya cometido el delito.

Se seleccionaron 36 variables de entrada para el modelo, las cuales en su mayoría corresponden a conteos de cargos y delitos anteriores. Es de resaltar que no se hizo ningún esfuerzo por reducir la lista de variables de entrada y que éstas están altamente correlacionadas. Los bosques aleatorios son el método de aprendizaje automático escogido puesto que este procedimiento no presenta ninguno de los problemas que tendría el análisis de regresión convencional al tener tantas variables correlacionadas e incluso puede funcionar con más entradas de pronóstico que observaciones. De igual manera, el objetivo principal no es determinar qué entradas son importantes y cuáles no, sino utilizar todas las entradas como grupo para llegar a pronósticos precisos. Además, posee formas fáciles de ponderar los errores de previsión por su relación relativa de costos y proporciona visualizaciones del desempeño de pronóstico que pueden ser muy instructivas.

Se obtiene que los casos que presentan poco riesgo se pueden pronosticar con precisión casi el 90 por ciento de las veces.

En [37] adoptan como enfoque el aprendizaje profundo para identificar automáticamente víctimas de violencia doméstica en condiciones críticas basados en publicaciones de redes sociales. La metodología de trabajo propuesta se dividió en 5 etapas: extracción de datos, etiquetado de datos, extracción de características, construcción del modelo y evaluación del desempeño. Para la extracción de los datos se usaron como principal fuente de información las redes sociales, entre ellas Facebook haciendo uso de su API. Para el etiquetado de la

información usaron calificadores humanos que leían cada publicación y determinaban si era crítica o no. Para la extracción de características usaron la técnica Word2vec, la cual, consiste en un algoritmo para procesamiento de lenguaje natural que usa una red neuronal para aprender asociaciones entre las palabras del corpus. Para la construcción del modelo usaron 5 modelos de aprendizaje profundo (Redes neuronales convolucionales CNN, redes neuronales recurrentes RNN, redes recurrentes artificiales de largo y corto plazo LSTM, unidades recurrentes cerradas GRU, y redes recurrentes artificiales bidireccionales de largo y corto plazo BLSTM). Por último, para evaluar el desempeño del modelo las métricas usadas fueron: Precisión, recuperación, medida F y exactitud.

Actualmente, no se ha encontrado en la literatura el uso de un modelo estadístico o de aprendizaje profundo basado en redes para caracterizar y/o clasificar los tipos de violencia intrafamiliar en Colombia desde los registros de morbi-mortalidad del Sistema de Salud Colombiano.

4. Metodología

En este capítulo se define la metodología de trabajo para abordar el tema de la violencia doméstica desde la construcción de redes no direccionadas creadas a partir de los diagnósticos clínicos de los individuos. En la primera parte del capítulo, se describen las fuentes de información y las bases de datos utilizadas para el análisis; posteriormente, se detalla la manera cómo se limpian, depuran y consolidan las bases de datos, de tal manera tengan el formato necesario para construir las redes; a continuación, se describe la selección de la metodología de trabajo y por último, se presenta el algoritmo propuesto para la construcción de las redes y la forma cómo se evaluará su ajuste.

A continuación, se usarán los términos: red de víctimas y red de no víctimas, cada resaltar que estos términos corresponden a red de diagnósticos clínicos de víctimas y red de diagnósticos clínicos de no víctimas respectivamente. Todas las redes aquí presentadas corresponden a redes construidas con los diagnósticos clínicos de los individuos de interés.

4.1. Fuentes y sistemas de información

A continuación, se presentan las fuentes y sistemas de información usadas para realizar el análisis propuesto y construir las redes de los individuos. Inicialmente, se detalla qué información contiene cada sistema/fuente de información junto con el decreto y/o resolución que lo reglamenta y posteriormente, se describen con mayor detalle:

4.1.1. Decretos y/o resoluciones que reglamentan a las fuentes y sistemas de información

- **Sistema de Vigilancia en Salud Pública SIVIGILA:**

Por medio del Decreto 3518 de 2006 el Presidente de la República crea y reglamenta el Sistema de Vigilancia en Salud Pública (SIVIGILA) [7]. Cuyo objeto corresponde a: “provisión en forma sistemática y oportuna, de información sobre la dinámica de los eventos que afecten o puedan afectar la salud de la población, con el fin de orientar las políticas y la planificación en salud pública; tomar las decisiones para la prevención y control de enfermedades y factores de riesgo en salud; optimizar el seguimiento y evaluación de las intervenciones; racionalizar y optimizar los recursos disponibles y lograr la efectividad de las acciones en esta materia, propendiendo por la protección

de la salud individual y colectiva.”.

Cabe resaltar que, según lo estipulado en el decreto [7] “Todas las acciones que componen el Sistema de Vigilancia en Salud Pública, SIVIGILA, tendrán el carácter de prioritarias en salud pública.”. Así, la aplicación del sistema de vigilancia tiene como una de sus principales finalidades estimar la magnitud de los eventos de interés en salud pública, entendiendo como eventos de interés, “Aquellos eventos considerados como importantes o trascendentes para la salud colectiva por parte del Ministerio de la Protección Social, teniendo en cuenta criterios de frecuencia, gravedad, comportamiento epidemiológico, posibilidades de prevención, costo–efectividad de las intervenciones, e interés público; que además, requieren ser enfrentados con medidas de salud pública.”.

Por otro lado, en términos de responsables se estipula que: “El Ministerio de la Protección Social establecerá los mecanismos para la integración y operación del Sistema de Vigilancia en Salud Pública, así como los lineamientos para la planeación, organización, dirección, operación, seguimiento y evaluación del Sistema en todos los niveles”. [7]

En particular, son de interés los datos correspondientes al reporte por violencia intrafamiliar que se encuentran reportados en la Ficha 875 del sistema de información.

■ **Registro Individual de Prestación de Servicios de Salud RIPS:**

Según la resolución 3374 de 2000 publicada por el Ministerio de Salud [33], el Registro Individual de Prestación de Servicios de Salud (RIPS) “Es el conjunto de datos mínimos y básicos que el Sistema General de Seguridad Social en Salud requiere para los procesos de dirección, regulación y control, y como soporte de la venta de servicio, cuya denominación, estructura y características se ha unificado y estandarizado para todas las entidades a que hace referencia el artículo 2o de la presente Resolución. Los datos de este registro se refieren a la identificación del prestador del servicio de salud, del usuario que lo recibe, de la prestación del servicio propiamente dicho y del motivo que originó su prestación, diagnóstico y causa externa.”.

Así, RIPS se encuentra conformado por 3 tipos de datos: Datos correspondientes a la identificación, al servicio de salud y al motivo que originó su prestación.

■ **Sistema de información de apoyo a la notificación e investigación de eventos de interés en salud pública SIANIESP:**

Según [12] SIANIESP es un aplicativo que “fue desarrollado e integrado a SIVIGILA dentro del marco del Proyecto piloto de viabilidad y factibilidad del uso de los RIPS como fuente de información para la vigilancia en salud pública, liderado por el Grupo de Investigación en Epidemiología Aplicada – GEA del Instituto Nacional de Salud, en cooperación con la Universidad de los Andes y la Fundación Universitaria Juan N. Corpas”.

- **Base de Datos Única de Afiliados BDUA:**

Por medio de la resolución 4622 de 2016 [27] el Ministro de Salud y Protección social resuelve “establecer el reporte de datos de afiliación al Sistema General de Seguridad Social en Salud, fijar los lineamientos aplicables al Régimen Contributivo, al Régimen Subsidiado, a los Regímenes Especiales y de Excepción, a las entidades que ofertan Planes Voluntarios de Salud, al Instituto Nacional Penitenciario y Carcelario —INPEC, y actualizar en lo pertinente, la Base de Datos Única de Afiliados - BDUA.”.

Así, según ADRES [31] “La Base de Datos Única de Afiliados – BDUA, es la base que contiene la información de los afiliados plenamente identificados, de los distintos regímenes del Sistema General de Seguridad Social en Salud. (Régimen Contributivo, Régimen Subsidiado, Regímenes de Excepción y Especiales y entidades prestadoras de Planes Voluntarios de Salud).”.

4.1.2. Descripción de las bases de datos

EVENTO 875 - SIVIGILA

La base de datos del Evento 875 de SIVIGILA es el principal insumo para el análisis. En esta base se van a identificar a los individuos que reportaron algún tipo de agresión producto de violencia intrafamiliar. La base cuenta con 119 campos, que caracterizan tanto a la víctima como al agresor y presentan características de la agresión. El campo más importante corresponde a la identificación de la víctima, dado que es con base en este campo que se realiza el cruce con las demás bases de datos. En este caso, cabe resaltar que un mismo individuo puede haber reportado en más de una ocasión el haber sufrido alguna agresión producto de violencia intrafamiliar y por tanto, en la base cada registro no corresponde a una víctima sino a un reporte de agresión.

La base de datos actual corresponde a información de individuos del Departamento de Risaralda, cuenta con 12.722 registros y contiene información desde el año 2012 hasta el año 2018.

SIANIESP

Dentro del aplicativo SIANIESP se tiene acceso a los meses de febrero a diciembre de 2016, enero a noviembre de 2017 y enero a marzo de 2018. Para cada mes se tiene acceso a diferentes archivos como son: AP = Archivo de procedimientos, AH = Archivo de hospitalización, AU = Archivo de urgencias, entre otros. **revisar Anexo A**

En este caso, son de interés solamente 3 de estos archivos: archivo de consulta, archivo de urgencias y archivo de hospitalización. Para ver la variables dentro de cada archivo **revisar Anexo A**

RIPS

La base de datos RIPS contiene la información de todas las consultas, urgencias y hospitalizaciones, que no son eventos de interés reportados en SIANIESP. En este caso, la base de datos tiene todas las consultas médicas, hospitalizaciones y urgencias consolidadas de los individuos; cuenta con 9 campos que incluyen el código del diagnóstico principal y los 3 diagnósticos relacionados al mismo y tiene 10.681.210 registros correspondientes a los años 2011 a 2016.

BDUA

La base de datos BDUa está dividida en 2 bases, una para el régimen contributivo y otra para el régimen subsidiado. Se tienen actualmente, 2 bases para cada régimen, una que contiene información desde el año 2011 al año 2016 y otra que contiene toda la información del año 2017. Para el régimen contributivo las bases tienen 7.429.872 y 522.698 registros respectivamente; para el régimen subsidiado las bases tienen 4.984.328 y 373.411 registros respectivamente. Las bases tienen diferentes campos, pero en general contienen información relacionada con el estado de afiliación, fecha de nacimiento, grupo poblacional, zona, entre otros.

4.1.3. Consolidación de las bases de trabajo

SIANIESP

La estructura básica de los archivos mensualizados de las bases de SIANIESP se presenta a continuación:

Urgencias:

PRESTADOR	N_FACTURA	tip_ide	num_ide	fec_ing	hora_ing	num_auto	causa_ext	diag_sal	diag_sal1	diag_sal2	diag_sal3	destino_usu	estado_sal	causa_muerte_ur	fec_sal	hora_sal	mes
HSJP0001647935	660010076201.00	CC	xxxx	24/11/2015	30300.00	13	I209					3.00	1.00		24/11/2015	34980.00	Febrero
HSJP0001648565	660010076201.00	CC	xxx	27/11/2015	34680.00	13	O321					1.00	1.00		27/11/2015	40740.00	Febrero

Consultas:

PRESTADOR	N_FACTURA	tip_ide	num_ide	fec_cons	num_auto	cod_cons	fin_cons	cau_ext	diag_prin	diag_prin1	diag_prin2	diag_prin3	tipo_diag_prin	valor_cons	valor_cuota_mod	valor_neto	mes
DN1374260	660010178101.00	CC	xxx	01/10/2015	236757	800202.00	10	13	F328				1.00	35000.00	19500.00	19500.00	Febrero
DN1374324	660010178101.00	CC	xxxxx	01/10/2015	236760	800302.00	10	13	F330				1.00	35000.00	19500.00	19500.00	Febrero

Hospitalizaciones:

N_FACTURA	PRESTADOR	tip_ide	num_ide	via_ingreso	fec_ing	hora_ing	num_auto	cau_ext	diag_prin1	diag_prin2	diag_prin3	diag_prin4	diag_prin5	diag_prin6	diag_prin7	diag_prin8	diag_prin9	diag_prin10	estado_salida	diag_causa_muerte	fec_sal	hora_sal	mes
HSJP0001647157	660010076201.00	CC	108838128	3.00	09/11/2015	77340.00	13	C821	Z782	F599									1.00		15/11/2015	11:30	Febrero
HSJP0001647167	660010076201.00	CC	42062844	3.00	09/11/2015	37140.00	13	C349	C343	E440									2.00	C343	21/11/2015	11:54	Febrero

Como se mencionó anteriormente, las bases mensualizadas de SIANIESP corresponden a urgencias, hospitalizaciones y consultas. Cada una de las bases de cada tipología tiene las mismas columnas, por tanto, para hacer la unificación solo es necesario adicionar cada archivo como nuevas filas a la base de datos inicial (sin pérdida de generalidad, se toma como

base inicial el archivo con fecha más antigua de cada tipología). Así, partiendo de dichas bases mensualizadas, se consolidan 3 bases de datos: urgencias, consultas y hospitalizaciones. Estas bases abarcan el periodo febrero de 2016 a marzo de 2018.

Una vez consolidadas las 3 bases, se identifican dentro de las mismas, los campos de interés: identificación del individuo, diagnóstico principal y diagnósticos relacionados. Estos campos se renombran de tal manera que en cada base de datos los nombres sean exactamente iguales. De cada base se filtran las columnas de interés mencionadas anteriormente, tomando como base inicial la base de datos de urgencias, se adicionan las bases de consultas y hospitalizaciones como nuevos registros. Cabe resaltar que no se va a discriminar por el tipo de atención, es decir, no es de interés saber si el diagnóstico del individuo fue resultado de una consulta programada, una urgencia o una hospitalización, por tal motivo, no es necesario saber a qué base inicial pertenece dicho registro. De esta manera, se llega a una única base de datos de todas las consultas, hospitalizaciones y urgencias de SIANIESP.

Análogo a lo realizado con las bases de SIANIESP, se hace una consolidación de las bases del BDUA para cada uno de los régimen, de tal manera que se obtiene una base con toda la información desde el año 2011 al año 2017 para el régimen contributivo y otra base con toda la información desde el año 2011 al año 2017 para el régimen subsidiado. Cabe resaltar, que en este caso, las bases originales también tienen los mismos campos para cada régimen, así que, solo es necesario adicionar las bases como nuevas filas y es por esto que el procedimiento es análogo al usado en SIANIESP.

Para las bases consolidadas de cada régimen de la BDUA, SIANIESP y RIPS se identifican los registros que correspondan a individuos que tienen por lo menos un reporte en la base de datos del evento 875. Con esta información, se realiza un análisis descriptivo del comportamiento clínico y sociodemográfico de estos individuos.

Por otro lado, para la base del Evento 875 se realiza un filtro dependiendo la naturaleza del evento reportado, considerando que en este caso sólo son de interés, los registros cuya naturaleza corresponda a violencia física o violencia psicológica, de esta manera, se obtiene una base de datos con 5.780 registros.

4.2. Selección de la metodología de trabajo

Como se puede observar en el capítulo de contexto, en los análisis descriptivos y exploratorios de las bases de trabajo se observa que al evaluar los diagnósticos de cada individuo de forma individual, sin considerar la relación entre diagnósticos, no se logra encontrar una diferencia significativa entre los grupos de individuos víctimas y no víctimas, no es posible observar una correlación directa entre el diagnóstico y el ser víctima o no, razón por la cual, considerando la revisión bibliográfica y los estudios realizados por múltiples autores sobre las multi-morbilidades [10] [15] [13], se decide utilizar redes para el análisis de información.

Para el análisis de la información se hará uso de redes complejas para encontrar patrones dentro de la población de víctimas y no víctimas, se compararán las medidas descritas en el marco teórico de cada población y con base en los patrones encontrados, se construirá un modelo que permita distinguir basado en el patrón de diagnósticos del individuo la probabilidad de que un grupo de individuos pertenezca al grupo de víctimas.

4.3. Algoritmo propuesto

El propósito principal de esta investigación es desarrollar un modelo de identificación de población víctima que permita caracterizar los tipos de violencia intrafamiliar basado en los registros clínicos reportados en los sistemas de salud colombianos. Al tener los datos de la población víctima del departamento de Risaralda es posible construir la red de los diagnósticos clínicos de estos individuos. Sin embargo, para poder construir el algoritmo de identificación, es necesario construir redes para los individuos que no han reportado ser víctimas de violencia intrafamiliar, de tal manera, que se pueda identificar patrones dentro de las redes que discriminen a ambas poblaciones.

Dado que el número de individuos que no han reportado haber sufrido algún tipo de agresión por violencia intrafamiliar es 4 000 a 1 aproximadamente, es necesario hacer muestreos de los individuos no víctima (individuos que no han reportado haber sufrido algún tipo de agresión por violencia intrafamiliar) para que las redes se construyan en poblaciones balanceadas y los resultados sean comparables. Además, es importante resaltar que el costo computacional de construir una red para toda la población no víctima es muy alto y tomaría meses. De igual manera, es necesario considerar que el hecho de que un individuo no haya reportado haber sufrido una agresión por violencia doméstica no implica que no la haya sufrido, por tanto, es necesario minimizar al máximo la posibilidad de usar individuos mal clasificados. A continuación, se explica cada parte del algoritmo detalladamente. En las figuras **4-1** **4-2** se presentan el diagrama con las entradas y salidas del algoritmo y el diagrama del algoritmo propuesto, respectivamente:



Figura 4-1.: Entradas y salidas del algoritmo propuesto

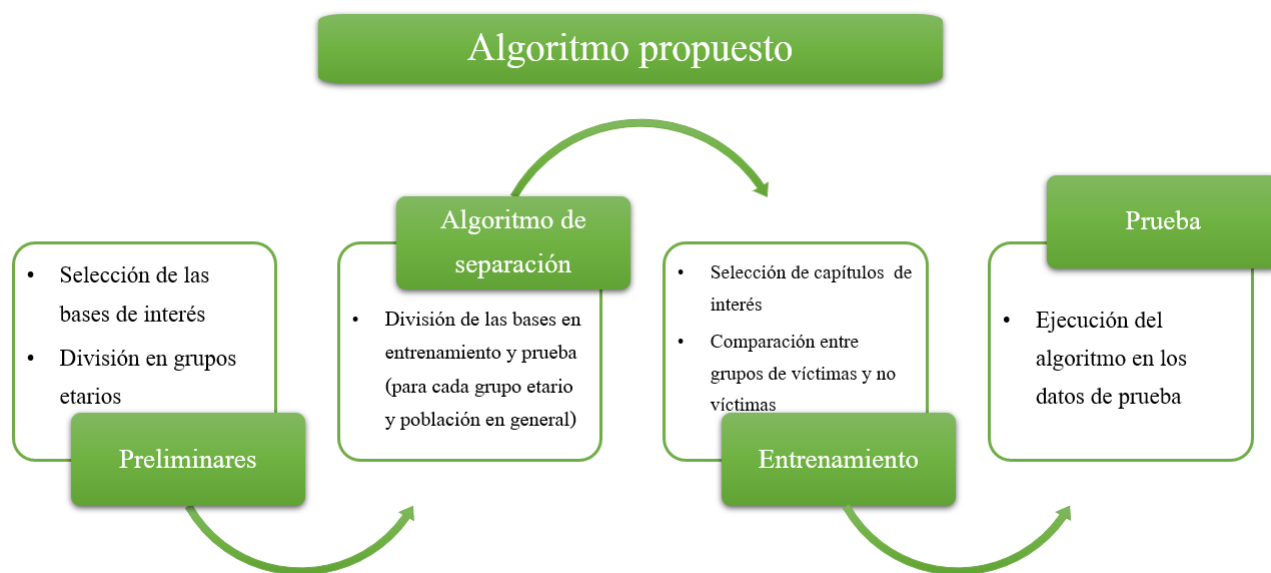


Figura 4-2.: Partes del algoritmo propuesto

4.3.1. Preliminares

Una vez consolidadas las bases de trabajo (RIPS y SIANIESP) cada una es separada en dos nuevas bases de trabajo: víctimas y no víctimas. Para el caso de las bases de víctimas, se realiza la selección identificando a los individuos que se encuentran en la base del evento 875 y las bases de no víctimas corresponden a los individuos restantes de las bases de datos originales. De esta manera, se obtienen 4 bases de datos de trabajo. (Ver figura 4-3)



Figura 4-3.: División en grupos de víctimas y no víctimas para las bases RIPS y SIANIESP

Considerando que dentro de las bases de no víctimas se pueden encontrar muchas víctimas que no han reportado la agresión y con el fin de minimizar el riesgo de no encontrar patrones claros que permitan identificar a la población víctima se define usar solamente la información de las víctimas cuyos diagnósticos clínicos hayan incluido por lo menos uno de los diagnósticos que caracterizan a las víctimas mencionados en [21] (**ver anexo B**). De igual manera, para la población no víctima solo se van a incluir los individuos que en todo su historial clínico no hayan presentado ninguno de los diagnósticos que caracterizan a las víctimas mencionados en [21]. De esta manera, se minimiza la posibilidad de estar entrenando el algoritmo con una mala clasificación entre víctimas y no víctimas.

En el diagrama 4-4 se resume el proceso mencionado anteriormente, para seleccionar los individuos de trabajo tanto de víctimas como de no víctimas y se resaltan las 4 bases de trabajo finales.

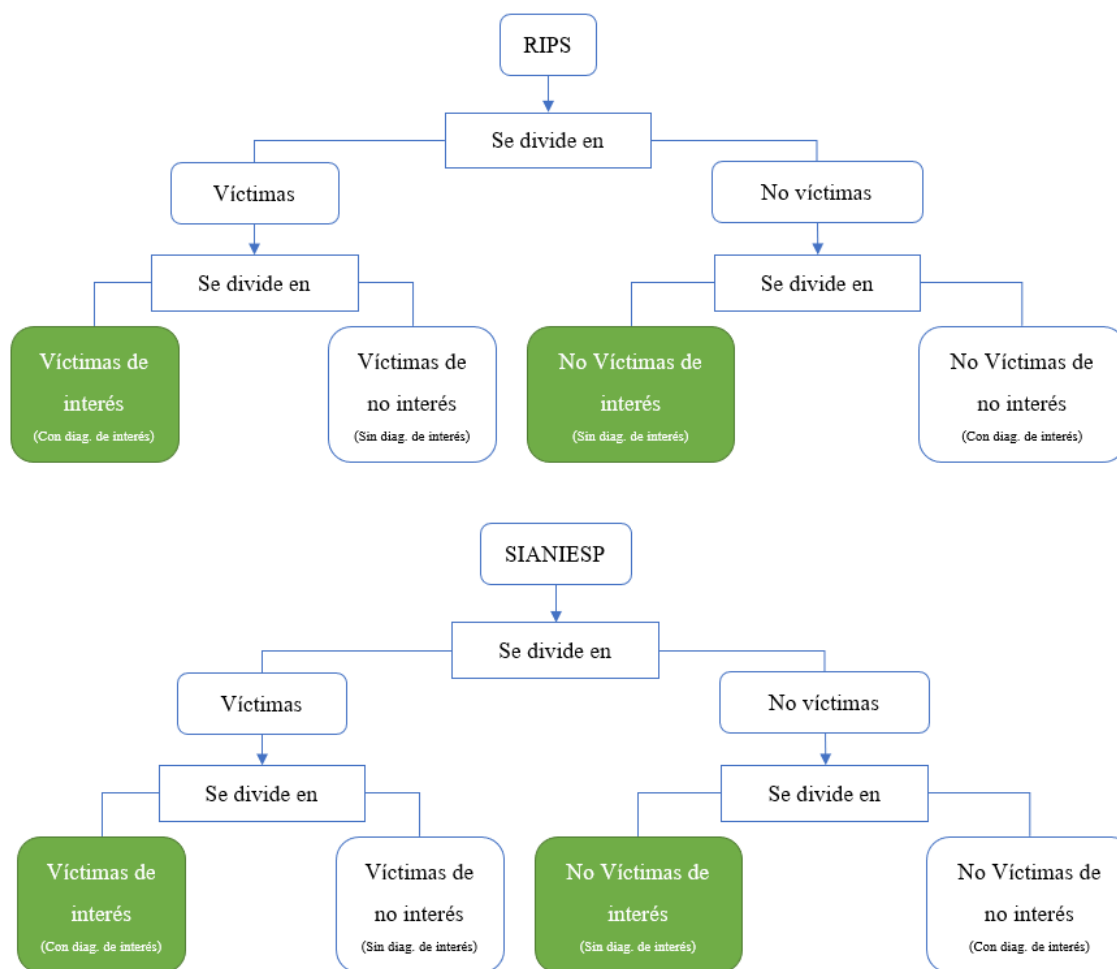


Figura 4-4.: Selección de las poblaciones de interés para las bases RIPS y SIANIESP

Con respecto a las tipologías de violencia intrafamiliar, existen múltiples clasificaciones que van desde el género de la víctima, su edad y orientación de género, hasta la relación que tiene con el agresor, el género del agresor y la edad del mismo. [35] [19] [28] [40]. En este caso se evaluarán dos tipologías de violencia doméstica: la violencia hacia menores de edad y la violencia hacia mayores de edad. Es importante resaltar que no es posible trabajar sobre tipologías de violencia más específicas dadas las limitaciones de los datos, puesto que, en términos de las características de las víctimas no hay suficiente información para discriminar a los adultos mayores y en términos del agresor no es posible identificar en la mayoría de los casos sus características socio-demográficas ni su relación con la víctima.

De esta manera y considerando que el interés de la investigación es caracterizar las tipologías de violencia, se divide la población en dos grupos etarios: Menores de edad y mayores de edad. Así, se podrán evaluar las diferencias entre los patrones encontrados en cada uno de los grupos etarios y los patrones encontrados en la población en general.

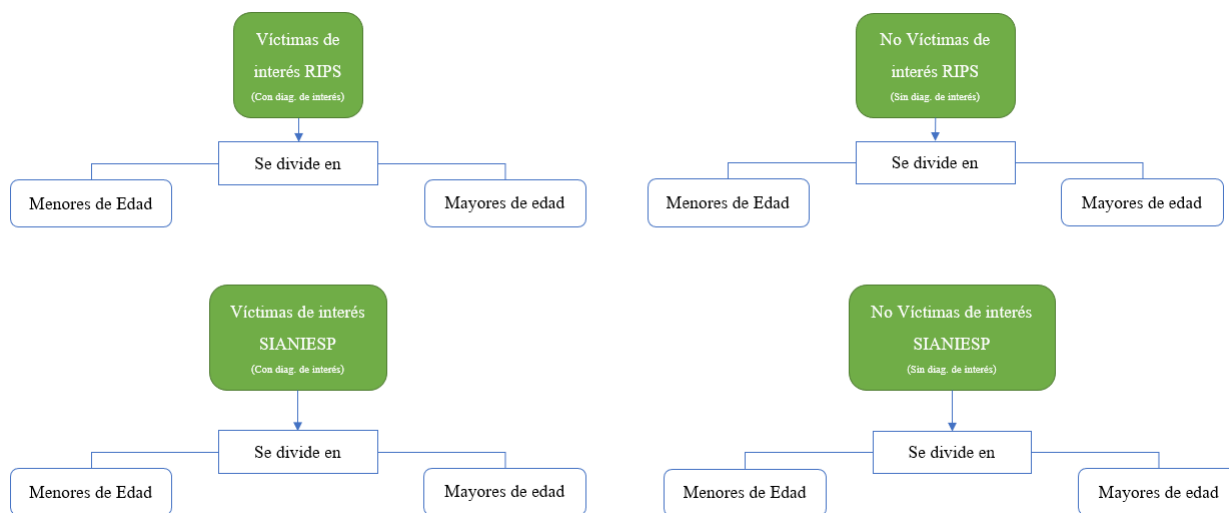


Figura 4-5.: División poblaciones de interés de RIPS y SIANIESP en grupos etarios

4.3.2. Algoritmo de separación

Para realizar la separación de los datos en entrenamiento y prueba es necesario extraer de cada base de datos la identificación de los individuos y consolidar un vector con estas identificaciones. Es importante aclarar, que es sobre el vector de identificaciones que se hace la división en entrenamiento y prueba dado que lo que se busca es dividir a los individuos que se van a usar en entrenamiento y de los individuos que se van a usar en prueba.

Considerando la división preliminar en dos grupos etarios de acuerdo con las tipologías de violencia, se divide cada uno de estos grupos en dos: entrenamiento y prueba, la proporción es 80 % – 20 %. Adicional, se hace una separación de la población general siguiendo la misma proporción y estratificando por la proporción de individuos que hay de cada grupo etario. En términos generales tanto en la base de RIPS como en la base de SIANIESP, el 25 % aproximadamente corresponde a menores de edad y el 75 % restante corresponde a mayores de edad.

De esta manera, se obtienen las bases de trabajo de entrenamiento para menores de edad, mayores de edad y población general. Ver figuras 4-6, 4-7, 4-8

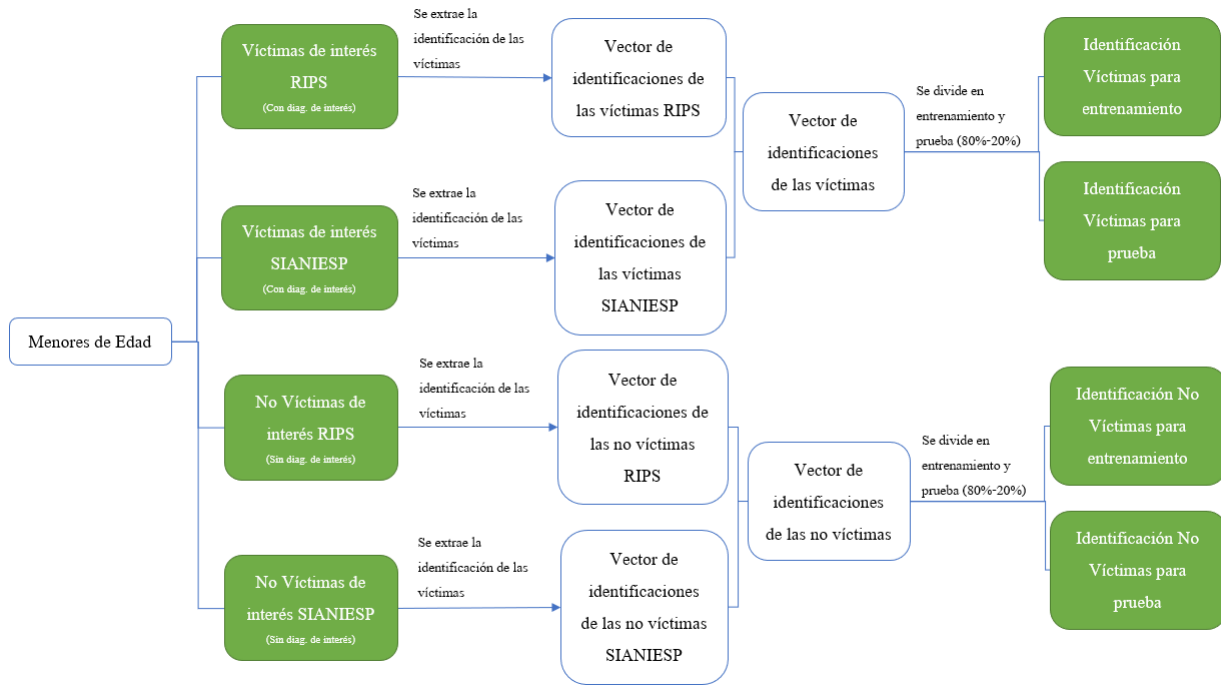


Figura 4-6.: División en entrenamiento y prueba menores de edad

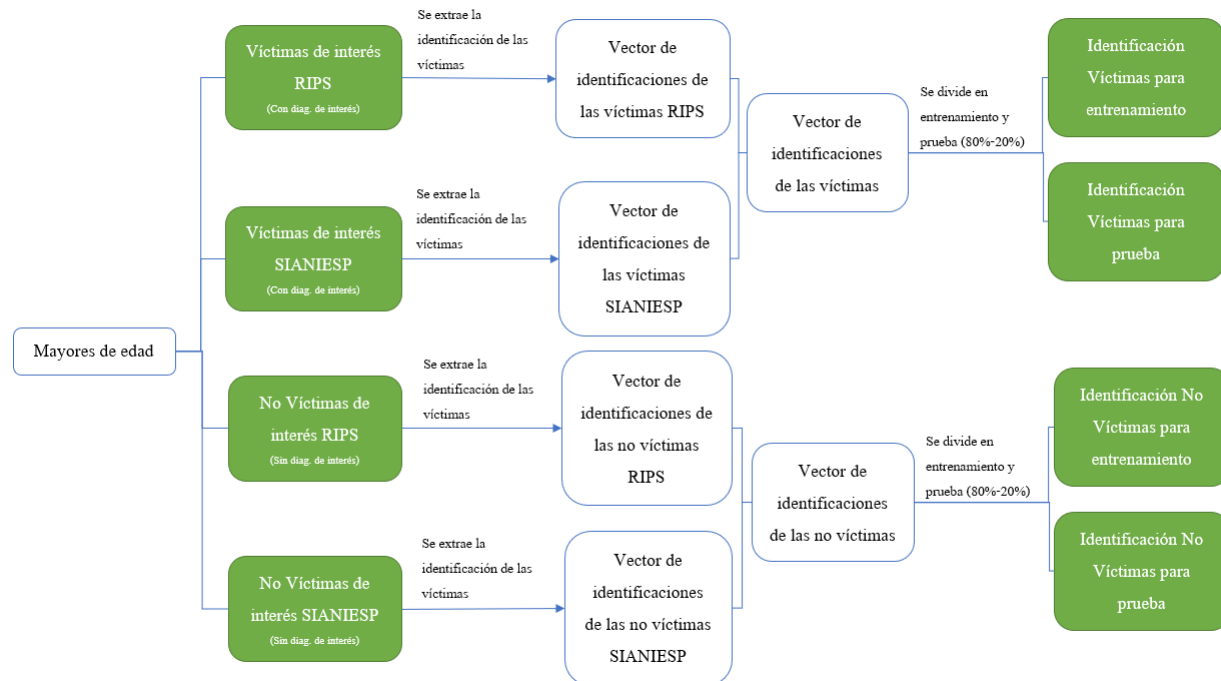


Figura 4-7.: División en entrenamiento y prueba mayores de edad

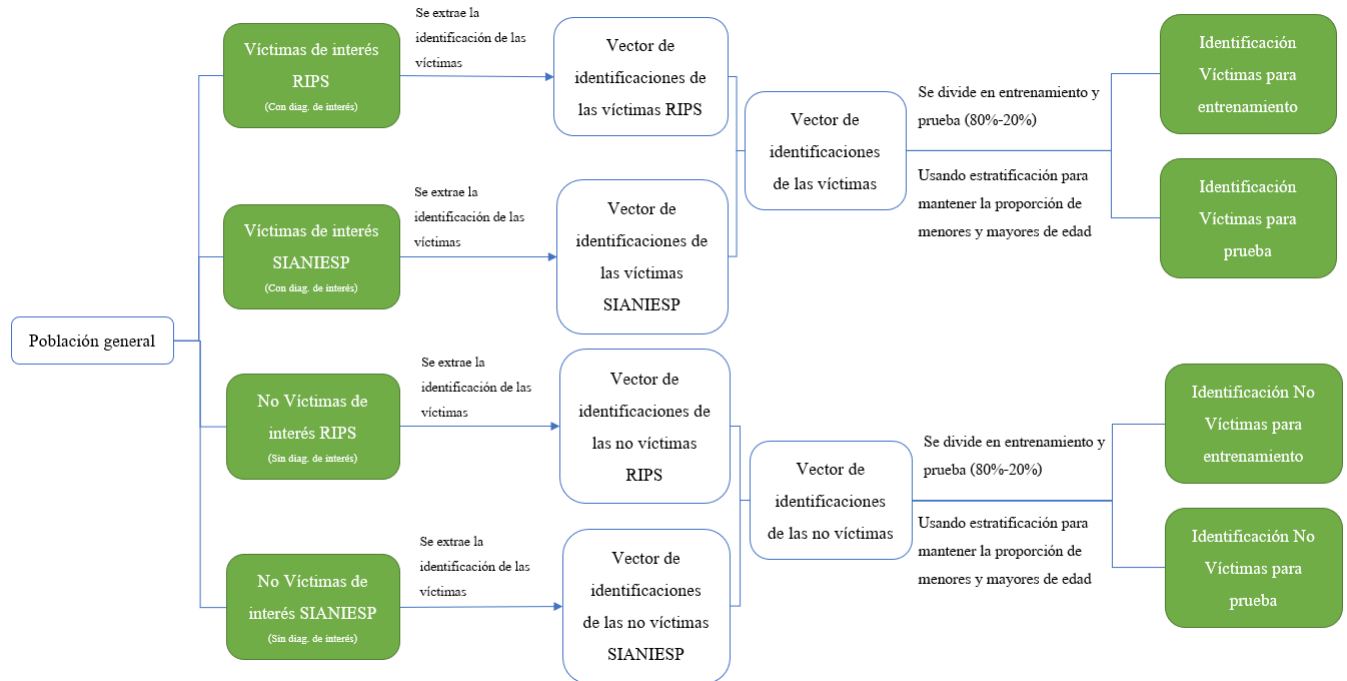


Figura 4-8.: División en entrenamiento y prueba población general

4.3.3. Entrenamiento

Algoritmo inicial para la selección de capítulos de interés

Los siguientes pasos se realizan en los datos de entrenamiento para cada uno de los grupos de interés, es decir, menores de edad, mayores de edad y población general:

1. Determinar la cantidad de víctimas en la base de datos n_v
2. Sacar una muestra de tamaño n_v de las personas no víctimas
3. Los siguientes pasos son repetidos n veces:
 - a) Dividir la población víctima en entrenamiento y prueba siguiendo una proporción 80 % – 20 %
 - b) Dividir la población no víctima en entrenamiento y prueba siguiendo una proporción 80 % – 20 %
 - c) Construir la red para la población víctima de entrenamiento
 - d) Construir la red para la población víctima de prueba
 - e) Construir la red para la población no víctima de entrenamiento
 - f) Construir la red para la población no víctima de prueba

- g) Consolidar una base con los valores de los grados nodales para cada red donde se identifique el valor del grado nodal para cada nodo y el capítulo al que pertenece el nodo.

Código diagnóstico	Grado Nodal	Descripción diagnóstico	Código Capítulo	Capítulo
A661	24	lesiones papilomatosas múltiples y frambesia con paso de cangrejo	A66	Ciertas enfermedades infecciosas y parasitarias
D509	111	anemia por deficiencia de hierro sin otra especificación	D50	Enfermedades de la sangre y de los órganos hematopoyéticos, y ciertos trastornos que afectan el mecanismo de la inmunidad

Tabla 4-1.: Ejemplo de tabla con valores del grado nodal para cada corrida del algoritmo central

- h) Considerando que cada nodo es un diagnóstico y que cada diagnóstico pertenece a un capítulo según la clasificación CIE-10, se filtra por capítulo en cada grupo (ejemplo, Train víctimas) y se compara la distribución del grado nodal de los nodos pertenecientes a ese capítulo calculando la divergencia de Kullback-Leibler entre los grupos. Las siguientes tablas ilustran un ejemplo del procedimiento:

Repetición	Grupo	Código diagnóstico	Grado Nodal	Capítulo
1	Train Víctimas	S000	15	Traumatismos
1	Train Víctimas	S001	20	Traumatismos
1	Train Víctimas	S002	18	Traumatismos

Tabla 4-2.: Ejemplo de tabla con valores del grado nodal para el capítulo traumatismos en la red de entrenamiento de víctimas

Repetición	Grupo	Código diagnóstico	Grado Nodal	Capítulo
1	Train No Víctimas	S000	10	Traumatismos
1	Train No Víctimas	S001	8	Traumatismos
1	Train No Víctimas	S002	11	Traumatismos

Tabla 4-3.: Ejemplo de tabla con valores del grado nodal para el capítulo traumatismos en la red de entrenamiento de no víctimas

Repetición	Capítulo	Comparación	DKL
1	Traumatismos	Train Víctimas - Train No Víctimas	0.0095

Tabla 4-4.: Ejemplo de tabla con valores del DKL para el capítulo traumatismos comparando las redes entrenamiento víctimas y entrenamiento no víctimas

La comparación se realiza entre las siguientes redes:

- Red víctimas de entrenamiento vs Red víctimas prueba

- Red no víctimas de entrenamiento vs Red víctimas prueba
 - Red víctimas de entrenamiento vs Red no víctimas prueba
 - Red no víctimas de entrenamiento vs Red no víctimas prueba
- i)* Consolidar en una base de datos los valores de la divergencia por capítulo en cada una de las comparaciones

Repetición	Comparación	Capítulo	DKL
1	Train Víctimas - Test Víctimas	Traumatismos	0.5
⋮	⋮	⋮	⋮
1	Train No Víctimas - Test Víctimas	Traumatismos	0.9
⋮	⋮	⋮	⋮
1	Train Víctimas - Test No Víctimas	Traumatismos	0.5
⋮	⋮	⋮	⋮
1	Train No Víctimas - Test No Víctimas	Traumatismos	0.2

Tabla 4-5.: Ejemplo de tabla con valores de la Divergencia de Kullback Leibler para la corrida 1 del algoritmo central

De esta manera, al correr el algoritmo n veces se obtiene una base de datos con la siguiente estructura:

Repetición	Comparación	Capítulo	DKL
1	Train Víctimas - Test Víctimas	Traumatismos	0.5
⋮	⋮	⋮	⋮
1	Train No Víctimas - Test Víctimas	Traumatismos	0.9
⋮	⋮	⋮	⋮
1	Train Víctimas - Test No Víctimas	Traumatismos	0.5
⋮	⋮	⋮	⋮
1	Train No Víctimas - Test No Víctimas	Traumatismos	0.2
⋮	⋮	⋮	⋮
n	Train Víctimas - Test Víctimas	Traumatismos	0.3
⋮	⋮	⋮	⋮
n	Train No Víctimas - Test Víctimas	Traumatismos	0.85
⋮	⋮	⋮	⋮
n	Train Víctimas - Test No Víctimas	Traumatismos	0.74
⋮	⋮	⋮	⋮
n	Train No Víctimas - Test No Víctimas	Traumatismos	0.18

Tabla 4-6.: Ejemplo de tabla con valores de la Divergencia de Kullback Leibler para las n corridas del algoritmo central

Selección de capítulos de interés

La base de datos resultante del algoritmo central **4-6** se divide dependiendo el grupo de test en dos nuevas bases de datos:

Repetición	Comparación	Capítulo	DKL
1	Train Víctimas - Test Víctimas	Traumatismos	0.5
⋮	⋮	⋮	⋮
1	Train No Víctimas - Test Víctimas	Traumatismos	0.9
⋮	⋮	⋮	⋮

Tabla 4-7.: Ejemplo de tabla con valores de la Divergencia de Kullback Leibler para cada corrida del algoritmo central donde el test corresponde a víctimas

Repetición	Comparación	Capítulo	DKL
1	Train Víctimas - Test No Víctimas	Traumatismos	0.5
⋮	⋮	⋮	⋮
1	Train No Víctimas - Test No Víctimas	Traumatismos	0.2
⋮	⋮	⋮	⋮

Tabla 4-8.: Ejemplo de tabla con valores de la Divergencia de Kullback Leibler para cada corrida del algoritmo central donde el test corresponde a no víctimas

Algoritmo final para la selección capítulos

1. En la base de datos **4-7** se identifican aquellos capítulos cuyo valor DKL es menor al comparar con el mismo grupo y mayor al comparar con grupos cruzados, es decir, se seleccionan aquellos capítulos donde el DKL es menor entre Train victimas - Test víctimas y mayor entre Train no victimas - Test Víctimas.

Repetición	Comparación	Capítulo	DKL
1	Train Víctimas - Test Víctimas	Traumatismos	0.5
1	Train No Víctimas - Test Víctimas	Traumatismos	0.9

Tabla 4-9.: Ejemplo de tabla con valores de la Divergencia de Kullback Leibler para la corrida 1 del algoritmo central donde el test corresponde a víctimas y el capítulo es Traumatismos

Por ejemplo, como se observa en la tabla **4-9** para la repetición 1, el DKL del capítulo traumatismos en la comparación Train victimas - Test víctimas es 0.5 y para la comparación Train no victimas - Test Víctimas es 0.9, en este caso como el DKL es menor en la comparación del mismo grupo Víctimas - Víctimas. Por lo tanto, el capítulo Traumatismos se convierte en un capítulo de interés.

2. Una vez identificados los capítulos de interés en el paso 1, se realiza un conteo del número de veces que dicho capítulo funcionó en el total de remuestreos realizados
3. Los capítulos identificados en el paso 1, son consolidados como un conjunto, del cual se halla su conjunto potencia excluyendo el conjunto vacío, así se determinan todas las

combinaciones posibles de capítulos de interés.

Por ejemplo, si se identificaron 3 capítulos: Cap_1 , Cap_2 y Cap_3 , se genera un listado con todas las posibles combinaciones de los capítulos de la siguiente manera:

$$\{Cap_1, Cap_2, Cap_3, (Cap_1, Cap_2), (Cap_1, Cap_3), (Cap_2, Cap_3), (Cap_1, Cap_2, Cap_3)\}$$

4. Para cada subconjunto i del conjunto potencia hallado en el paso 3, se implementa el siguiente algoritmo:
 0. Seleccionar el elemento i del conjunto potencia
 1. Filtrar de la base de víctimas **4-7** resultante del algoritmo central las filas que corresponden a capítulos contenidos en el elemento i del conjunto potencia. Ejemplo, si $i = (Cap_1, Cap_2)$, entonces se filtrarán todos los registros que correspondan a información del Cap_1 o del Cap_2
 2. Se agrupa por repetición y comparación y se resume el máximo del DKL
 3. Se define una variable indicadora que toma el valor 1 cuando el valor del DKL es menor en la comparación del mismo grupo (ejemplo, Train Víctimas - Test Víctimas) comparado con el valor del DKL cuando se comparan grupos diferentes (ejemplo, Train No Víctimas - Test Víctimas) y 0 en otro caso
 4. Se suma el valor de la variable indicadora y se divide por la cantidad de corridas realizadas para encontrar la proporción de acierto.
5. Se consolida una base con los resultados del algoritmo del paso 5 y se determina la combinación de capítulos que arroja un mayor porcentaje de acierto.

Volviendo al ejemplo anterior, la tabla resultante quedaría de la siguiente manera:

Capítulo	Porcentaje de acierto
Cap_1	20 %
Cap_2	40 %
Cap_3	50 %
(Cap_1, Cap_2)	75 %
(Cap_1, Cap_3)	15 %
(Cap_2, Cap_3)	35 %
(Cap_1, Cap_2, Cap_3)	80 %

En las figuras **4-9**, **4-10** y **4-11** se resume lo mencionado anteriormente, sobre el proceso de entrenamiento del algoritmo central.

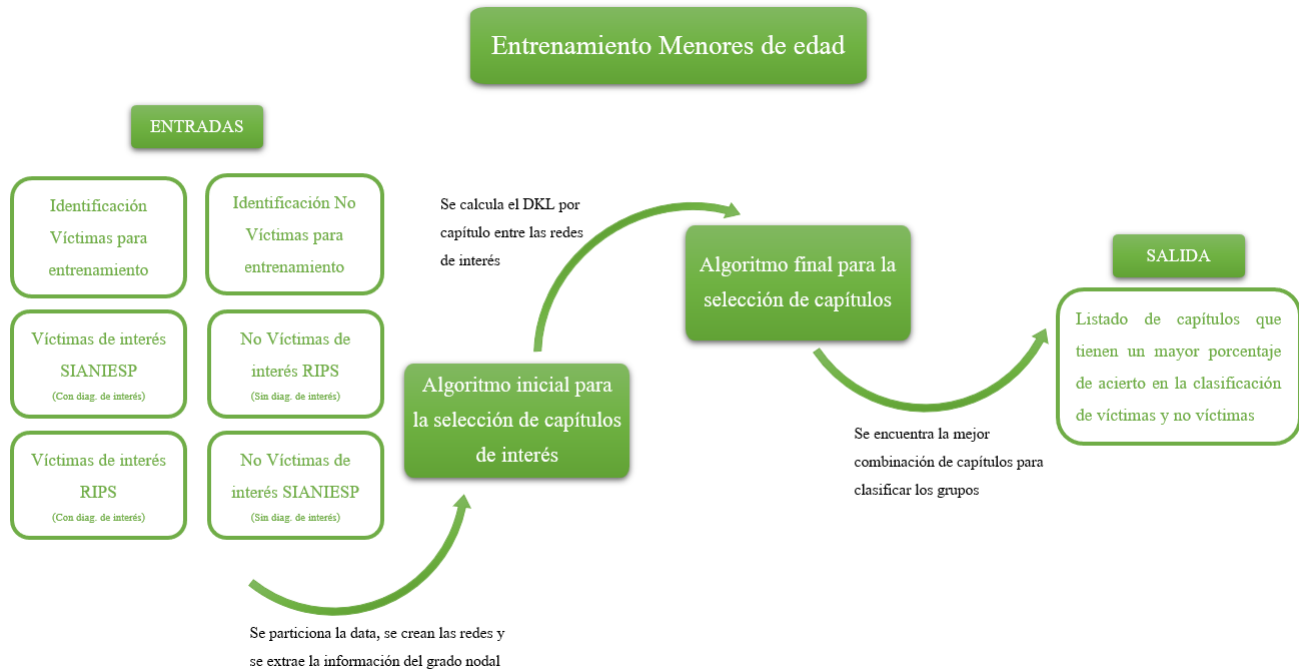


Figura 4-9.: Resumen proceso de entrenamiento en la población de menores de edad

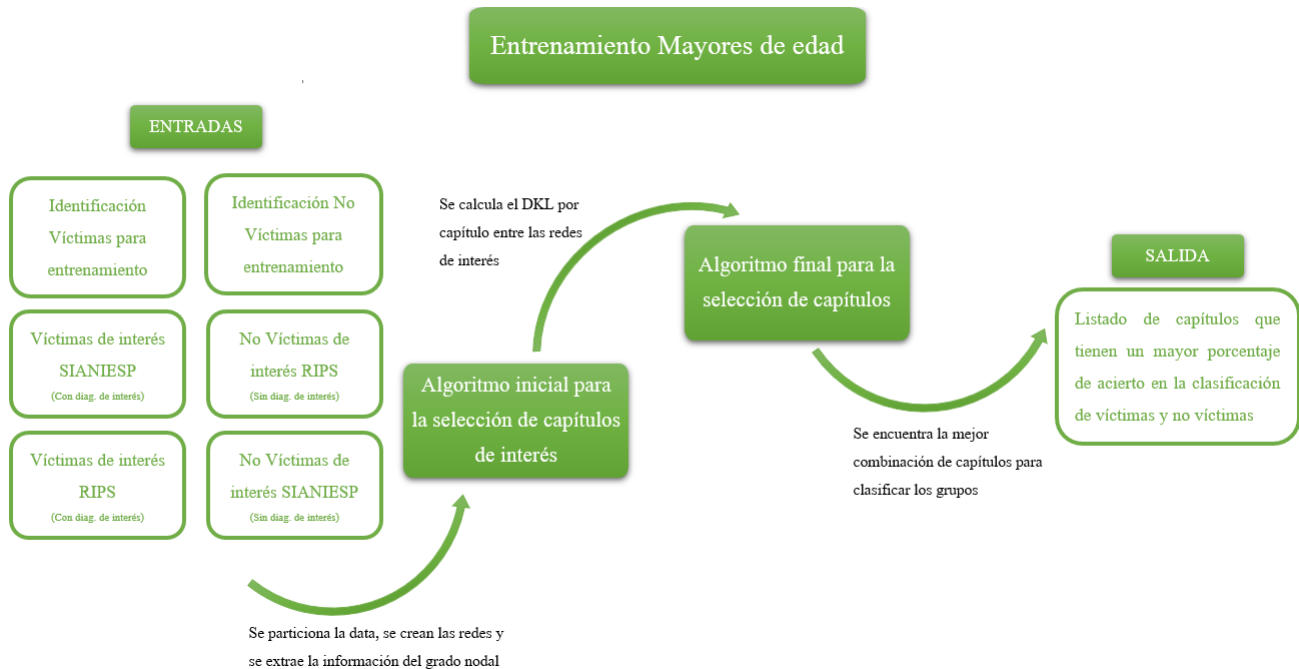


Figura 4-10.: Resumen proceso de entrenamiento en la población de mayores de edad



Figura 4-11.: Resumen proceso de entrenamiento en la población general

4.3.4. Prueba

Considerando que el objetivo es identificar aquellos capítulos de diagnósticos cuya distribución del grado nodal difiere entre víctimas y no víctimas, esta última parte del algoritmo propuesto busca evaluar si al usar los capítulos identificados en el entrenamiento es posible acertar en cuanto a la clasificación de los datos de prueba. Por tanto, una vez determinados los capítulos de interés usando los datos de entrenamiento, se aplica el algoritmo central a los datos de prueba y se determina si el algoritmo clasifica correctamente los nuevos datos en víctimas y no víctimas.

Es importante resaltar que, para los datos de prueba, también se realizan múltiples corridas de los datos con el fin de no tener resultados espurios.

5. Resultados y discusión

Como se había mencionado anteriormente, el objetivo principal del trabajo es analizar el comportamiento de los diagnósticos clínicos de la población víctima de violencia intrafamiliar del Departamento de Risaralda de tal manera que, sea posible construir un modelo de identificación de población víctima que permita caracterizar las tipologías de violencia usando únicamente sus diagnósticos clínicos. Por tanto, en este capítulo se presentan y discuten los resultados del análisis propuesto.

Al inicio del capítulo se presenta la manera como se construyeron las redes y se discuten sus medidas globales. Posteriormente se presentan los resultados del algoritmo propuesto para los datos de entrenamiento y por último, se presentan los resultados para los datos del grupo de prueba. Al final, se discuten los resultados obtenidos y se plantean algunos interrogantes que surgen a partir de los mismos.

5.1. Análisis de redes

5.1.1. Construcción de las redes

Las redes, como se mencionó anteriormente, se pueden definir de muchas maneras y aplican a diferentes contextos donde se tenga un conjunto de elementos conectados. En este caso, los nodos de la red corresponden a diagnósticos clínicos de los individuos, conectados por enlaces cuya relación está dada por la presencia de los diagnósticos en la historia clínica del paciente, es decir, los enlaces están dados por la presencia de los diagnósticos en la historia clínica de un mismo paciente. La red se construye como un grafo no dirigido donde los nodos son los diagnósticos y los enlaces son la relación entre éstos. Cabe resaltar que sólo se construyen las redes completas para las víctimas dado lo mencionado en la metodología sobre las limitaciones para construir las redes de los individuos no víctimas de violencia intrafamiliar.

5.1.2. Víctimas de violencia intrafamiliar física o psicológica

Medidas globales de la red

Se construye la red de víctimas con los 5 734 individuos cuya naturaleza de la agresión corresponde a violencia física o psicológica y se sacan sus principales medidas globales:

- **Cantidad de nodos y enlaces:** La red tiene 3 723 nodos y 310 331 enlaces.

- **Densidad:** La densidad de la red es 0,045.

Considerando que la red corresponde a un grafo no dirigido y que la densidad de una red indica la conectividad de la misma y siendo ésta, la propiedad de la red que mide la proporción de las relaciones presentes sobre el máximo número de relaciones que pueden existir en la red, una densidad de 4,5 % indica que la relación entre los nodos de la red es mínima, es decir, los diagnósticos no están muy relacionados entre si, lo que implica que, no es muy homogénea la distribución de relaciones que existen entre los diagnósticos de la población víctima.

De esta manera, lo anterior implica que no todos los diagnósticos están conectados unos con otros en la red de víctimas, sugiriendo que pueden haber diagnósticos que nunca se relacionan con otros, lo cual podría ser de utilidad al reducir el número de diagnósticos que se deberían considerar en el análisis, dado que podrían ser más importantes aquellos diagnósticos que están más conectados.

- **Asortatividad:** La assortatividad de la red es -0,336.

Considerando que la assortatividad es una medida que determina la preferencia de los nodos de la red para relacionarse con nodos de características similares, el hecho de que en este caso el coeficiente de assortatividad sea negativo implica una red disortativa.

Una red disortativa sugiere que los nodos se relacionan entre sí con otros de características diferentes, es decir, en la red de víctimas la relación se presenta entre vértices de diferentes grados, implicando que, diagnósticos con grado nodal grande se relacionan con diagnósticos que tienen un grado nodal pequeño.

Con base en lo anterior, al examinar el coeficiente de assortatividad en la red de víctimas se puede inferir que en la red los diagnósticos no tienden a relacionarse con diagnósticos que tengan un grado nodal similar, sino con diagnósticos con un grado nodal diferente. Esto sugiere que, los diagnósticos con muy baja interacción con otros diagnósticos, en general, se relacionan con diagnósticos que tienen muy alta interacción con otros diagnósticos.

- **Grado promedio:** El grado promedio de la red es 166,75.

Considerando que el grado promedio corresponde a la media aritmética de los grados de cada diagnóstico, es una medida más fácil de interpretar que la densidad de la red [24]. En este caso, al tener un grafo no dirigido es posible relacionar ambos valores y dado que el grado promedio da 167 aproximadamente, que la densidad da 4,5 % aproximadamente y que hay 957 diagnósticos cuyo grado es igual o mayor al grado promedio se puede concluir que la distribución del grado nodal no es muy homogénea, que la red no es muy densa y que posiblemente es necesario realizar el filtro de los diagnósticos de interés para el análisis de la población víctima ya que al considerar

todas las atenciones y todos los diagnósticos de la población víctima no se evidencia un patrón claro dentro de la red y la distribución del grado nodal no es homogénea. A continuación se listan los diagnósticos con mayor grado nodal:

- Caries de la dentina (2 835)
- Examen médico general (2 748)
- Infección de vías urinarias, sitio no especificado (2 640)
- Rinofaringitis aguda (resfriado común) (2 446)
- Otros dolores abdominales y los no especificados (2 378)

Como se puede observar, los diagnósticos con mayor grado nodal, en general son diagnósticos comunes que puede tener cualquier individuo sea víctima de violencia intrafamiliar o no.

Por lo tanto, un análisis individual de los diagnósticos puede no conducir a una caracterización de la población víctima, lo cual tiene sentido, dado que al analizar los diagnósticos de forma individual se pierde la posibilidad de ver y evaluar multi-morbilidades, las cuales como ya se mencionó en capítulos anteriores, son las que se quieren estudiar y usar para realizar una caracterización de los patrones de diagnóstico en las poblaciones de víctimas y no víctimas [14] [29]

Los cinco capítulos con mayores diagnósticos dentro de estos 957 diagnósticos con grado mayor o igual al grado promedio son:

- Traumatismos, envenenamientos y algunas otras consecuencias de causas externas (134 diagnósticos)
- Factores que influyen en el estado de salud y contacto con los servicios de salud (92 diagnósticos)
- Trastornos mentales y del comportamiento (77 diagnósticos)
- Enfermedades del sistema digestivo (75 diagnósticos)
- Enfermedades del sistema osteomuscular y del tejido conjuntivo (67 diagnósticos)

Es importante resaltar que el capítulo de Trastornos mentales y del comportamiento es el tercer capítulo con más diagnósticos que superan el valor del grado promedio, esto apoya la teoría de una relación entre los trastornos mentales y sufrir violencia intrafamiliar.

Por otro lado, estos capítulos pueden ser aquellos que al aplicar el algoritmo propuesto evidencien algún tipo de diferencia entre la distribución del grado nodal de víctimas y la distribución del grado nodal de no víctimas. Además, es de resaltar que a estos capítulos pertenecen la mayoría de diagnósticos que se definen como de interés en la población víctima según [21] (Ver anexo B).

Los cinco subgrupos de diagnósticos a tres dígitos con mayores diagnósticos dentro de estos 957 diagnósticos con grado mayor o igual al grado promedio son:

- Trastorno afectivo bipolar (10 diagnósticos)
- Dorsalgia (8 diagnósticos)
- Atención para la anticoncepción (8 diagnósticos)
- Gastritis y duodenitis (8 diagnósticos)
- Supervisión de embarazo de alto riesgo (8 diagnósticos)

Es interesante que el subgrupo de diagnósticos a tres dígitos Trastorno afectivo bipolar, sea el subgrupo con mayor cantidad de diagnósticos cuyos grado nodal sea superior al grado nodal promedio, ya que esto puede sugerir que en la población víctima predominan diagnósticos con grado nodal mayor al grado promedio que corresponden a diagnósticos de enfermedades mentales (Nótese que el capítulo de Trastornos mentales fue el 3er capítulo con más diagnósticos cuyo grado nodal era mayor al grado promedio).

De igual manera, comparar los cinco diagnósticos, los cinco capítulos y los cinco subgrupos de diagnósticos a tres dígitos con mayores valores de grado nodal, sugiere que probablemente es más útil analizar los diagnósticos por subgrupo o capítulo que de forma individual. Además, es interesante ver como diagnósticos que corresponden a enfermedades mentales predominan en la red de víctimas, puesto que se refuerza el por qué hay que considerar las multi-morbilidades del individuo en el análisis, además, de analizar la relación entre la violencia intrafamiliar y el intento suicida [10] [13] [15].

5.1.3. Víctimas de interés

Considerando lo definido a nivel metodológico, se construyen las redes para las víctimas de interés, es decir, para aquellas víctimas cuyos diagnósticos clínicos incluyen por lo menos uno de los diagnósticos del artículo de [21] (Ver anexo B). Como se mencionó anteriormente, el objetivo es identificar diferencias y similitudes en el comportamiento de los diagnósticos clínicos de la población víctima en sus diferentes tipologías, por tanto, se construyen tres redes diferentes: una red para la población general de víctimas, otra para víctimas menores de edad y la última para víctimas mayores de edad.

Medidas globales de las redes

Población general

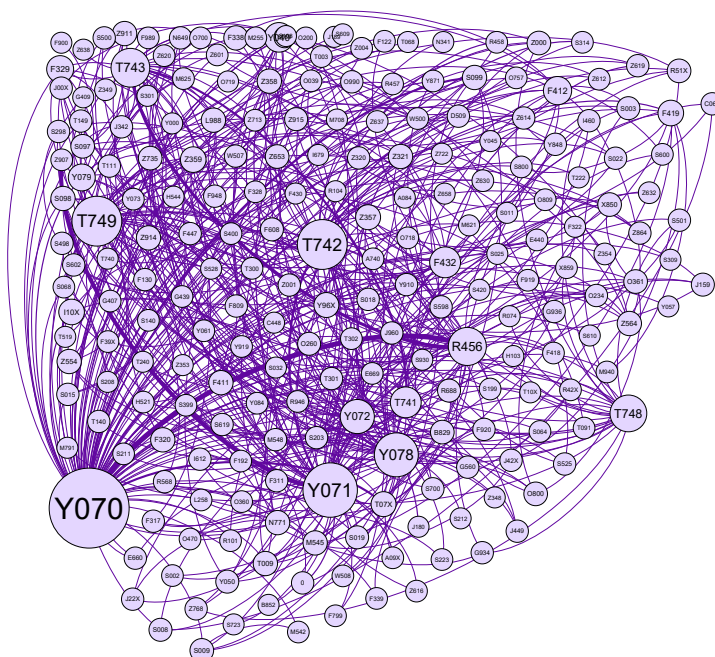


Figura 5-1.: Red de diagnósticos clínicos de todas las víctimas

En la figura **5-1** el tamaño de la etiqueta del nodo está dado por el grado nodal del mismo. Se observa que el diagnóstico con el mayor grado nodal es Y070: “Otros síndromes de maltrato: por esposo o pareja”, seguido por los diagnósticos: Y071: “Otros síndromes de maltrato: por padre o madre” y T749: “Síndrome de maltrato, no especificado”. La relación más fuerte entre diagnósticos se da entre Y070: “Otros síndromes de maltrato: por esposo o pareja”, R456: “Violencia física” y T743: “Abuso psicológico”.

Lo anterior, muestra que en términos generales, los diagnósticos con mayor grado nodal dentro de la red de víctimas de interés corresponden a diagnósticos relacionados con el maltrato, que pertenecen al capítulo de Causas externas de morbilidad y de mortalidad y que a su vez pertenecen al subgrupo de diagnóstico a tres dígitos Otros síndromes de maltrato. Estos

diagnósticos están fuertemente relacionados con diagnósticos que en su mayoría pertenecen al subgrupo de diagnóstico a tres dígitos Síndromes de maltrato, los cuales corresponden al capítulo Traumatismos, envenenamientos y algunas otras consecuencias de causas externas.

Se puede ver que la red no parece ser muy densa, no todos los diagnósticos están conectados unos con otros, además, se observa que hay nodos más fuertemente relacionados con otros (líneas más oscuras en la figura 5-1) y que el grado nodal de éstos es muy diferente, por ejemplo, Y070: “Otros síndromes de maltrato: por esposo o pareja” con S099: “traumatismo de los vasos sanguíneos de la cabeza no clasificados en otra parte”. En general, diagnósticos con grado nodal alto se relacionan fuertemente con diagnósticos que tienen grado nodal bajo, lo cual indica una posible red disortativa, como la red de todas las víctimas de violencia física o psicológica presentada anteriormente.

Las medidas globales de la red se describen a continuación:

- **Cantidad de nodos y enlaces:** La red tiene 226 nodos y 508 enlaces.

- **Densidad:**

La densidad de la red es 0,02.

Como se mencionó anteriormente, la densidad de una red es aquella propiedad que mide la proporción de las relaciones presentes sobre el máximo número de relaciones que pueden existir en la red, una densidad de 2 % indica que los diagnósticos no están muy relacionados entre sí, lo que implica que, no es muy homogénea la distribución de relaciones que existen entre los diagnósticos de la población víctima de interés.

- **Asortatividad:**

La asortatividad de la red es -0,346.

Como se mencionó anteriormente, la asortatividad es una medida que determina la preferencia de los nodos de la red para relacionarse con nodos de características similares, el hecho de que en este caso el coeficiente de asortatividad sea negativo implica una red disortativa. Lo cual corrobora lo observado en la figura 5-1, donde diagnósticos con grado nodal grande se relacionan con diagnósticos que tienen un grado nodal pequeño.

- **Grado medio:**

El grado medio de la red es: 4,496

- **Grado medio con pesos:**

El grado medio con pesos de la red es: 4,867

Como se mencionó anteriormente, el grado promedio corresponde a la media aritmética de los grados de cada diagnóstico. En este caso, al tener un grado promedio y un grado promedio con pesos de 5 aproximadamente, y con una densidad de 2 % se puede concluir

que la distribución del grado nodal no es muy homogénea y que la red de víctimas de interés no es muy densa.

- **Diámetro de la red:**

El diámetro de la red es: Cuatro (4)

Considerando que el diámetro de la red es una medida que representa la máxima distancia existente entre dos nodos en toda la red, se puede concluir que la distancia máxima entre dos diagnósticos de las víctimas de interés es cuatro (4). Esto indica que los diagnósticos no están muy distantes entre sí y en general, se sitúan a los alrededores del centro de la red.

Población menores de edad

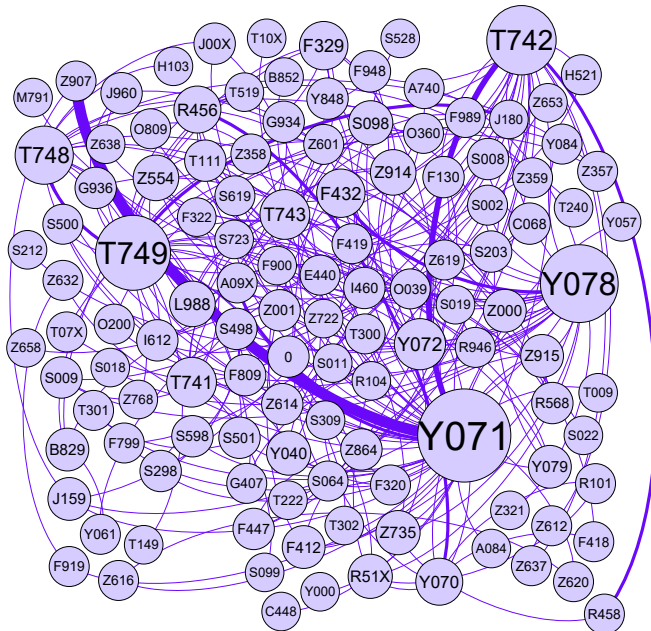


Figura 5-2.: Red de diagnósticos clínicos de las víctimas menores de edad

En la figura 5-2 el tamaño de la etiqueta del nodo está dado por el grado nodal del mismo. Se observa que el diagnóstico con el mayor grado nodal es Y071: “Otros síndromes de maltrato: por padre o madre”, seguido por los diagnósticos: Y078: “Otros síndromes de maltrato: por otra persona especificada”, T749: “Síndrome de maltrato, no especificado” y T742: “Abuso sexual”. La relación más fuerte entre diagnósticos se da entre Y071: “Otros síndromes de maltrato: por padre o madre”, Z907: “Ausencia adquirida de órganos genitales” y T742: “Abuso sexual”.

El hecho de que el diagnóstico con mayor grado nodal sea el Y071: “Otros síndromes de maltrato: por padre o madre” es muy interesante, dado que evidencia lo presentado por múltiples autores y estudios acerca del principal agresor de los menores de edad, puesto que, según la literatura, en la mayoría de los casos el agresor corresponde al padre y/o a la madre [2] [19] [38]. De esta manera, el tener dicho diagnóstico como principal, en términos del grado nodal muestra que la red permite vislumbrar comportamientos y patrones del agresor desde los diagnósticos clínicos de la víctima.

De igual manera, la red de víctimas de interés de menores de edad muestra que en términos generales, los diagnósticos con mayor grado nodal corresponden a diagnósticos relacionados con el maltrato, que pertenecen al capítulo de Causas externas de morbilidad y de mortalidad y que a su vez pertenecen al subgrupo de diagnóstico a tres dígitos Otros síndromes de maltrato. Estos diagnósticos están fuertemente relacionados con diagnósticos que en su mayoría pertenecen al subgrupo de diagnóstico a tres dígitos Síndromes de maltrato, los cuales corresponden al capítulo Traumatismos, envenenamientos y algunas otras consecuencias de causas externas. Es interesante ver, que este mismo comportamiento se observa en la población general, por lo que se puede inferir que aún cuando los menores de edad representan solo el 25 % del total de víctimas de la población de interés, el comportamiento de los diagnósticos se mantiene.

Por otro lado, se puede ver que la red no parece ser muy densa, no todos los diagnósticos están conectados unos con otros, además, se observa que hay nodos más fuertemente relacionados con otros (líneas más oscuras en la figura 5-2) y que el grado nodal de éstos es muy diferente, por ejemplo, Y071: “Otros síndromes de maltrato: por padre o madre” con Z907: “Ausencia adquirida de órganos genitales”. En general, diagnósticos con grado nodal alto se relacionan fuertemente con diagnósticos que tienen grado nodal bajo, lo cual indica una posible red disortativa, como las dos redes de víctimas presentadas anteriormente.

Las medidas globales de la red se describen a continuación:

- **Cantidad de nodos y enlaces:** La red tiene 121 nodos y 219 enlaces.

- **Densidad:**

La densidad de la red es 0,03.

Como se mencionó anteriormente, la densidad de una red es aquella propiedad que

mide la proporción de las relaciones presentes sobre el máximo número de relaciones que pueden existir en la red, una densidad de 3% indica que los diagnósticos no están muy relacionados entre si, lo que implica que, no es muy homogénea la distribución de relaciones que existen entre los diagnósticos de la población víctima de interés menor de edad.

Cabe resaltar que en este caso, la red de víctimas menores de edad es un poco más densa que la red de la población general de víctimas de interés (3% menores de edad vs 2% población general). Sin embargo, no es una diferencia muy grande y la conclusión sigue siendo que en ambos casos ninguna red es muy densa.

- **Asortatividad:**

La asortatividad de la red es -0,3414.

Como se mencionó anteriormente, la asortatividad es una medida que determina la preferencia de los nodos de la red para relacionarse con nodos de características similares, el hecho de que en este caso el coeficiente de asortatividad sea negativo implica una red disortativa. Lo cual corrobora lo observado en la figura 5-2, donde diagnósticos con grado nodal grande se relacionan con diagnósticos que tienen un grado nodal pequeño.

Análogo a lo mencionado con respecto a la densidad, en este caso, el coeficiente de asortatividad de la red de víctimas menores de edad tiene un valor casi igual al coeficiente de asortatividad de la red de la población general de víctimas de interés (-0,3414 menores de edad vs -0,346 población general). De esta manera, se concluye que en ambos casos las redes son disortativas.

- **Grado medio:**

El grado medio de la red es: 3,62

- **Grado medio con pesos:**

El grado medio con pesos de la red es: 3,835

Como se mencionó anteriormente, el grado promedio corresponde a la media aritmética de los grados de cada diagnóstico. En este caso, al tener un grado promedio y un grado promedio con pesos de 4 aproximadamente, y con una densidad de 3% se puede concluir que la distribución del grado nodal no es muy homogénea y que la red de víctimas de interés menores de edad no es muy densa.

Como se mencionó con las medidas anteriores, este resultado coincide con la conclusión obtenida al analizar el comportamiento del grado promedio en la red de víctimas de la población general, lo cual reafirma la similitud entre ambas redes.

- **Diámetro de la red:**

El diámetro de la red es: Seis (6)

Considerando que el diámetro de la red es una medida que representa la máxima distancia existente entre dos nodos en toda la red, se puede concluir que la distancia máxima entre dos diagnósticos de las víctimas de interés menores de edad es seis (6). Esto indica que los diagnósticos no están muy distantes entre sí y en general, se sitúan a los alrededores del centro de la red. Si bien en el caso de la red de la población general el valor del diámetro es cuatro (4) la diferencia no es muy grande, por lo cual, se puede asumir una similitud en el diámetro de ambas redes.

Población mayores de edad

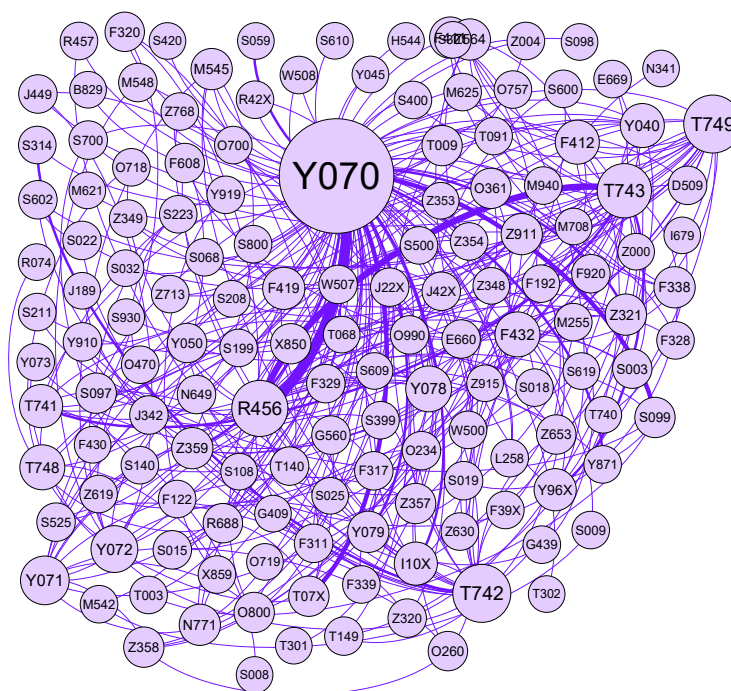


Figura 5-3.: Red de diagnósticos clínicos de las víctimas mayores de edad

En la figura 5-3 el tamaño de la etiqueta del nodo está dado por el grado nodal del mismo. Se observa que el diagnóstico con el mayor grado nodal es Y070: “Otros síndromes de

maltrato: por esposo o pareja”, seguido por los diagnósticos: T742: “Abuso sexualz T749: “Síndrome de maltrato, no especificado”. La relación más fuerte entre diagnósticos se da entre Y070: “Otros síndromes de maltrato: por esposo o pareja”, R456: “Violencia físicaz T743: “Abuso psicológico”.

Lo anterior, muestra que en términos generales, los diagnósticos con mayor grado nodal dentro de la red de víctimas de interés mayores de edad, corresponden a diagnósticos relacionados con el maltrato, que pertenecen al capítulo de Causas externas de morbilidad y de mortalidad y que a su vez pertenecen al subgrupo de diagnóstico a tres dígitos Otros síndromes de maltrato. Estos diagnósticos están fuertemente relacionados con diagnósticos que en su mayoría pertenecen al subgrupo de diagnóstico a tres dígitos Síndromes de maltrato, los cuales corresponden al capítulo Traumatismos, envenenamientos y algunas otras consecuencias de causas externas. Es interesante ver, que este mismo comportamiento se observa en la población general y en la población de menores de edad, por lo que se puede inferir que los diagnósticos clínicos tanto de los mayores de edad como los de menores de edad tienen el mismo comportamiento.

Por otro lado, se puede ver que la red no parece ser muy densa, no todos los diagnósticos están conectados unos con otros, además, se observa que hay nodos más fuertemente relacionados con otros (líneas más oscuras en la figura 5-3) y que el grado nodal de éstos es muy diferente, por ejemplo, Y070: “Otros síndromes de maltrato: por esposo o pareja” con R456: “Violencia física”. En general, diagnósticos con grado nodal alto se relacionan fuertemente con diagnósticos que tienen grado nodal bajo, lo cual indica una posible red disortativa, como las tres redes de víctimas presentadas anteriormente.

Las medidas globales de la red se describen a continuación:

- **Cantidad de nodos y enlaces:** La red tiene 146 nodos y 305 enlaces.
- **Densidad:**

La densidad de la red es 0,029.

Como se mencionó anteriormente, la densidad de una red es aquella propiedad que mide la proporción de las relaciones presentes sobre el máximo número de relaciones que pueden existir en la red, una densidad de 2,9 % indica que los diagnósticos no están muy relacionados entre si, lo que implica que, no es muy homogénea la distribución de relaciones que existen entre los diagnósticos de la población víctima de interés mayor de edad.

Cabe resaltar que en este caso, la red de víctimas mayores de edad es un poco más densa que la red de la población general de víctimas de interés (2,9 % mayores de edad vs 2 % población general). Sin embargo, no es una diferencia muy grande y la conclusión sigue siendo que en ambos casos ninguna red es muy densa. Además, el valor de la densidad es aproximadamente igual al valor de la densidad de la red de menores

de edad, lo cual indica una consistencia en las conclusiones para las 3 poblaciones.

■ **Asortatividad:**

La assortatividad de la red es -0.2823.

Como se mencionó anteriormente, la assortatividad es una medida que determina la preferencia de los nodos de la red para relacionarse con nodos de características similares, el hecho de que en este caso el coeficiente de assortatividad sea negativo implica una red disortativa. Lo cual corrobora lo observado en la figura **5-3**, donde diagnósticos con grado nodal grande se relacionan con diagnósticos que tienen un grado nodal pequeño.

Análogo a lo mencionado con respecto a la densidad, en este caso, el coeficiente de assortatividad de la red de víctimas mayores de edad es un poco inferior al coeficiente de assortatividad de la red de la población general de víctimas de interés y al coeficiente de assortatividad de la red de menores de edad (-0,2823 mayores de edad vs -0,346 población general vs -0,3114 menores de edad). Sin embargo, la conclusión es la misma, las redes son disortativas.

■ **Grado medio:**

El grado medio de la red es: 4,178

■ **Grado medio con pesos:**

El grado medio con pesos de la red es: 4,466

Como se mencionó anteriormente, el grado promedio corresponde a la media aritmética de los grados de cada diagnóstico. En este caso, al tener un grado promedio y un grado promedio con pesos de 4 aproximadamente, y con una densidad de 2,9% se puede concluir que la distribución del grado nodal no es muy homogénea y que la red de víctimas de interés mayores de edad no es muy densa.

Como se mencionó con las medidas anteriores, este resultado coincide con la conclusión obtenida al analizar el comportamiento del grado promedio en la red de víctimas de la población general y en la red de menores edad, lo cual reafirma la similitud entre todas las redes.

■ **Diámetro de la red:**

El diámetro de la red es: Cinco (5)

Considerando que el diámetro de la red es una medida que representa la máxima distancia existente entre dos nodos en toda la red, se puede concluir que la distancia máxima entre dos diagnósticos de las víctimas de interés menores de edad es cinco (5). Esto indica que los diagnósticos no están muy distantes entre sí y en general, se sitúan a los alrededores del centro de la red. Si bien en el caso de la red de la población

general el valor del diámetro es cuatro (4) la diferencia no es muy grande, por lo cual, se puede asumir una similitud en el diámetro de ambas redes.

5.2. Algoritmo de propuesto

5.2.1. Preliminares

Como se mencionó en la metodología, la primera parte del algoritmo propuesto tiene como nombre “preliminares” y consiste en:

1. consolidación de las bases de trabajo,
2. separación de las bases con respecto a las tipologías de violencia a trabajar,
3. construcción de las redes para cada tipología de violencia, y
4. análisis preliminar de las redes de cada tipología de violencia.

En la figura 5-4 se presenta el resumen de lo que se realiza en la sección de preliminares del algoritmo y se puede identificar cuáles son los datos de entrada y qué salidas arroja esta sección particular del algoritmo. Cabe resaltar que como se observa en la figura 5-4, es en esta parte del algoritmo donde se dividen los datos en las tipologías de violencia definidas: víctimas de violencia intrafamiliar con diagnósticos de interés menores de edad y víctimas de violencia intrafamiliar con diagnósticos de interés mayores de edad.

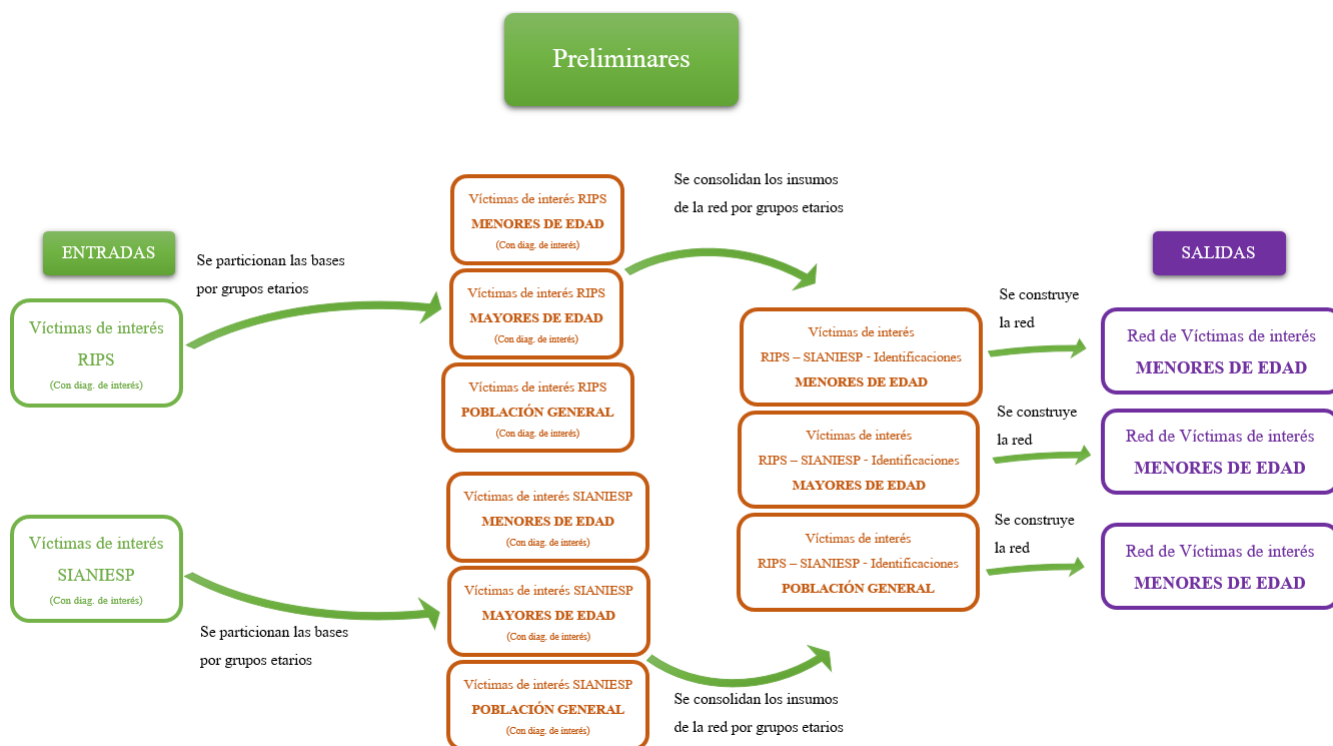


Figura 5-4.: Diagrama de los preliminares del algoritmo propuesto

5.2.2. Algoritmo de separación

Una vez que se ha implementado la sección de preliminares del algoritmo propuesto y habiendo analizado las características, medidas globales y patrones de comportamiento dentro de cada red de los diagnósticos, se implementa la segunda sección del algoritmo propuesto denominada: “algoritmo de separación”.

En el algoritmo de separación se toman las bases obtenidas en la sección de preliminares para víctimas y no víctimas de cada grupo etario y se extraen las identificaciones de los individuos para crear dos grupos de individuos, los que van a usarse en el entrenamiento del algoritmo y los que van a usarse para probar el algoritmo. La figura 5-5 resume lo que hace esta sección del algoritmo y presenta las entradas y salidas del mismo.

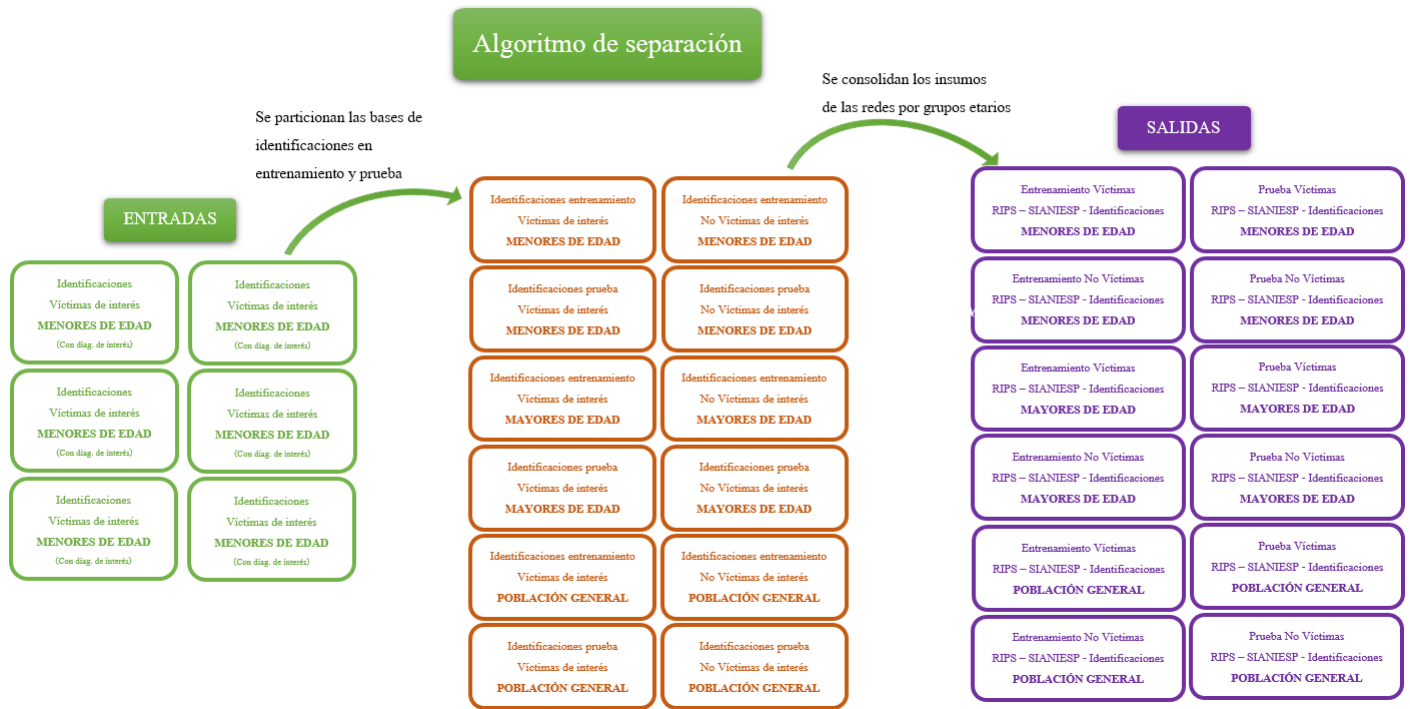


Figura 5-5.: Diagrama del algoritmo de separación del algoritmo propuesto

5.2.3. Entrenamiento

Una vez que se han implementado las secciones uno y dos del algoritmo propuesto, se implementa la tercera sección del algoritmo denominada: “Entrenamiento”. En este caso, el entrenamiento se va a realizar de forma independiente en cada una de las poblaciones de interés: menores de edad, mayores de edad y población general. En la figura 5-6 se presenta el resumen de lo que hace esta sección del algoritmo junto con sus entradas y salidas:

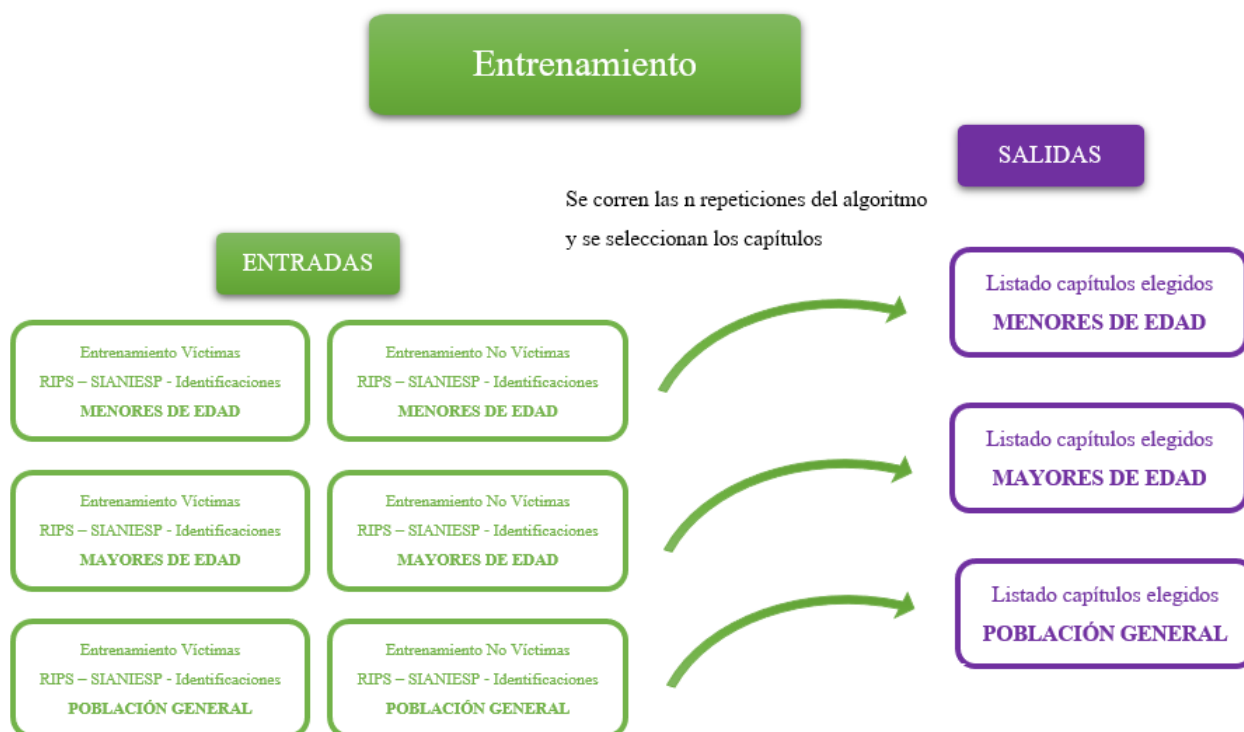


Figura 5-6.: Diagrama del entrenamiento del algoritmo propuesto

A continuación, se presentan los resultados para cada una de las poblaciones.

Población General

Para el entrenamiento se tomaron 373 remuestreos de las poblaciones de entrenamiento de víctimas y no víctimas, usando la función de agrupación del máximo se eligen los siguientes capítulos para la prueba del modelo:

- Embarazo, parto y puerperio.
- Enfermedades del ojo y sus anexos.
- Enfermedades del sistema digestivo.
- Enfermedades del sistema nervioso.
- Enfermedades del sistema osteomuscular y del tejido conjuntivo.
- Enfermedades endocrinas, nutricionales y metabólicas.
- Factores que influyen en el estado de salud y contacto con los servicios de salud.
- Síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte.

- Tumores [neoplasias].

De esta manera, al implementar el entrenamiento en los datos de las víctimas se identifican nueve (9) capítulos de diagnósticos cuya distribución del grado nodal difiere entre las víctimas y no víctimas de entrenamiento.

Es interesante ver que uno de los capítulos elegidos en el entrenamiento sea el relacionado con el embarazo, dado que en la literatura se encuentran varias investigaciones que han estudiado la relación entre el embarazo y la violencia de género y violencia doméstica. [6] [23] [34] [39].

De igual manera, se destaca la presencia del capítulo Factores que influyen en el estado de salud y contacto con los servicios de salud en la lista de capítulos seleccionados, dado que éste incluye diagnósticos relacionados con el embarazo, problemas de crianza y trastornos mentales. Lo anterior, muestra una consistencia con respecto a los resultados obtenidos al analizar la red de víctimas de la población general y es interesante evidenciar que algunos de los capítulos que se habían identificado como candidatos a capítulos caracterizadores del comportamiento de los diagnósticos clínicos de las víctimas, efectivamente lo son.

Mayores de edad

Para el entrenamiento se tomaron 840 remuestreos de las poblaciones de entrenamiento de víctimas y no víctimas, usando la función de agrupación del máximo se eligen los siguientes capítulos para la prueba del modelo:

- Causas externas de morbilidad y de mortalidad.
- Embarazo, parto y puerperio.
- Enfermedades de la piel y del tejido subcutáneo.
- Enfermedades de la sangre y de los órganos hematopoyéticos, y ciertos trastornos que afectan el mecanismo de la inmunidad.
- Enfermedades del ojo y sus anexos.
- Enfermedades del oído y de la apófisis mastoides.
- Enfermedades del sistema circulatorio.
- Enfermedades del sistema digestivo.
- Enfermedades del sistema genitourinario.
- Enfermedades del sistema nervioso.
- Enfermedades del sistema osteomuscular y del tejido conjuntivo.

- Enfermedades del sistema respiratorio.
- Enfermedades endocrinas, nutricionales y metabólicas.
- Factores que influyen en el estado de salud y contacto con los servicios de salud.
- Síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte.
- Trastornos mentales y del comportamiento.
- Traumatismos, envenenamientos y algunas otras consecuencias de causas externas.
- Tumores [neoplasias].

De esta manera, al implementar el entrenamiento en los datos de las víctimas menores de edad se identifican nueve (18) capítulos de diagnósticos cuya distribución del grado nodal difiere entre las víctimas y no víctimas de entrenamiento menores de edad.

Es interesante ver que uno de los capítulos elegidos en el entrenamiento sea el relacionado con el embarazo, dado que como se mencionó anteriormente, en la literatura se encuentran varias investigaciones que han estudiado la relación entre el embarazo y la violencia de género y violencia doméstica. [6] [23] [34] [39]. Además, también se encuentran dentro de la lista de capítulos seleccionados los capítulos Trastornos mentales y del comportamiento, y Traumatismos, envenenamientos y algunas otras consecuencias de causas externas. Estos capítulos, como se evidenció en el análisis de la red de víctimas mayores de edad, son los capítulos a los que pertenecen los diagnósticos con mayor grado nodal y los que tienen más diagnósticos relacionados fuertemente entre ellos. Por lo cual, desde el análisis de la red eran candidatos a ser buenos caracterizadores del comportamiento de los diagnósticos clínicos de las víctimas mayores de edad.

De igual manera, se destaca la presencia del capítulo Factores que influyen en el estado de salud y contacto con los servicios de salud en la lista de capítulos seleccionados, dado que éste incluye diagnósticos relacionados con el embarazo, problemas de crianza y trastornos mentales. Lo anterior, muestra una consistencia con respecto a los resultados obtenidos al analizar la red de víctimas mayores de edad y es interesante evidenciar que, como se mencionó anteriormente, la mayoría de los capítulos que se habían identificado como candidatos a capítulos caracterizadores del comportamiento de los diagnósticos clínicos de las víctimas, efectivamente lo son.

Menores de edad

Considerando que la cantidad de individuos víctima menores de edad es considerablemente menor a la cantidad de individuos víctima mayores de edad se tomaron 6 512 remuestreos de

las poblaciones de entrenamiento de víctimas y no víctimas, usando la función de agrupación del máximo se eligen los siguientes capítulos para la prueba del modelo:

- Causas externas de morbilidad y de mortalidad.
- Ciertas enfermedades infecciosas y parasitarias.
- Enfermedades de la piel y del tejido subcutáneo.
- Enfermedades del ojo y sus anexos.
- Enfermedades del oído y de la apófisis mastoides.
- Malformaciones congénitas, deformidades y anomalías cromosómicas.
- Enfermedades del sistema digestivo.
- Enfermedades del sistema genitourinario.
- Enfermedades del sistema nervioso.
- Enfermedades del sistema osteomuscular y del tejido conjuntivo.
- Enfermedades del sistema respiratorio.
- Enfermedades endocrinas, nutricionales y metabólicas.
- Factores que influyen en el estado de salud y contacto con los servicios de salud.
- Síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte.
- Trastornos mentales y del comportamiento.
- Traumatismos, envenenamientos y algunas otras consecuencias de causas externas.
- Tumores [neoplasias].

De esta manera, al implementar el entrenamiento en los datos de las víctimas menores de edad se identifican nueve (17) capítulos de diagnósticos cuya distribución del grado nodal difiere entre las víctimas y no víctimas de entrenamiento menores de edad.

De igual a manera a lo mencionado en el caso de la población general y las víctimas mayores de edad, es interesante ver que uno de los capítulos elegidos en el entrenamiento sea el relacionado con el embarazo, dado que como se mencionó anteriormente, en la literatura se encuentran varias investigaciones que han estudiado la relación entre el embarazo y la violencia de género y violencia doméstica. [6] [23] [34] [39]. Además, también se encuentran dentro de la lista de capítulos seleccionados los capítulos Trastornos mentales y del comportamiento,

y Traumatismos, envenenamientos y algunas otras consecuencias de causas externas. Estos capítulos, como se evidenció en el análisis de la red de víctimas mayores de edad, son los capítulos a los que pertenecen los diagnósticos con mayor grado nodal y los que tienen más diagnósticos relacionados fuertemente entre ellos. Por lo cual, desde el análisis de la red eran candidatos a ser buenos caracterizadores del comportamiento de los diagnósticos clínicos de las víctimas mayores de edad.

De igual manera, se destaca la presencia del capítulo Factores que influyen en el estado de salud y contacto con los servicios de salud en la lista de capítulos seleccionados, dado que éste incluye diagnósticos relacionados con el embarazo, problemas de crianza y trastornos mentales. Lo anterior, muestra una consistencia con respecto a los resultados obtenidos al analizar la red de víctimas mayores de edad y es interesante evidenciar que, como se mencionó anteriormente, la mayoría de los capítulos que se habían identificado como candidatos a capítulos caracterizadores del comportamiento de los diagnósticos clínicos de las víctimas, efectivamente lo son.

5.2.4. Prueba

Una vez que se han implementado las secciones uno, dos y tres del algoritmo propuesto, se implementa la última sección del algoritmo denominada: “Prueba”. En este caso, la prueba se va a realizar de forma independiente en cada una de las poblaciones de interés: menores de edad, mayores de edad y población general. En la figura 5-7 se presenta el resumen de lo que hace esta sección del algoritmo junto con sus entradas y salidas:

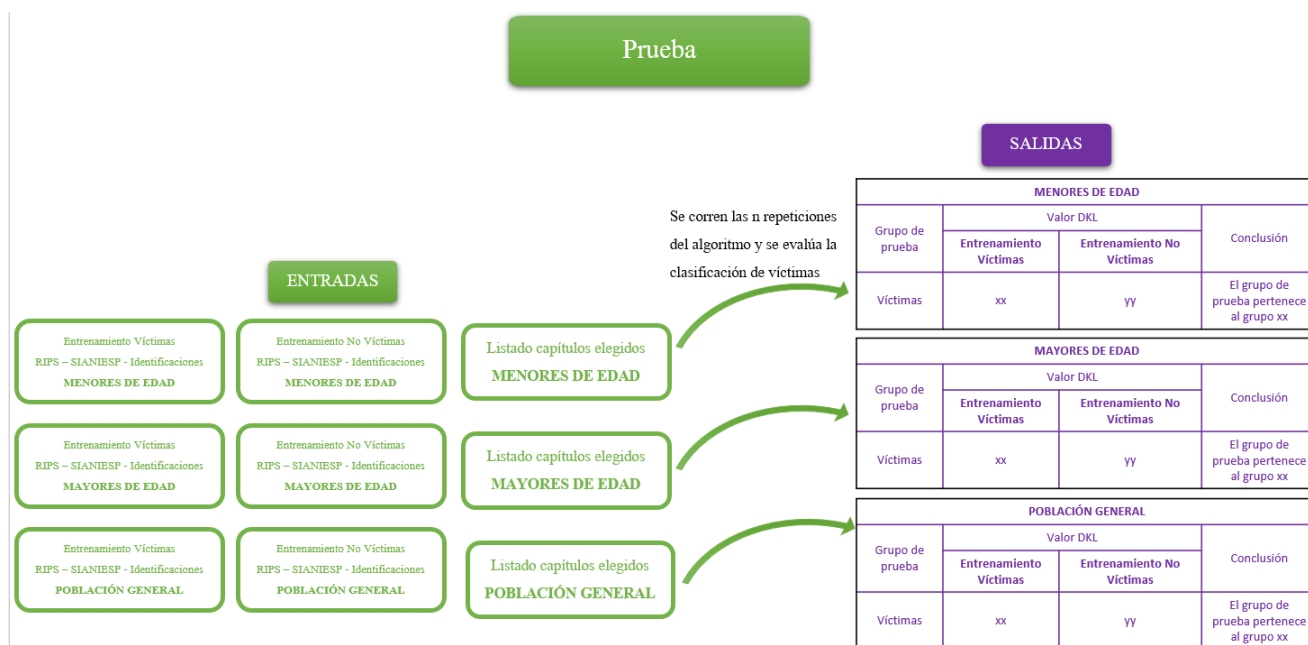


Figura 5-7.: Diagrama de la prueba del algoritmo propuesto

A continuación, se presentan los resultados para cada una de las poblaciones.

Población general

Se realizan 1 000 remuestreos de las poblaciones de prueba de víctimas y no víctimas y usando los capítulos identificados en el entrenamiento, el algoritmo logra identificar los datos de las víctimas de prueba como víctimas.

Grupo de Prueba	Valor DKL		Conclusión
	Train Víctimas	Train No Víctimas	
Test Víctimas	0.100703	0.108579	Los datos de prueba corresponden a población víctima de violencia intrafamiliar

Tabla 5-1.: Tabla con el resultado del algoritmo propuesto para las víctimas de la población general

De esta manera, en la tabla **5-1** se puede observar que usando los capítulos elegidos en el entrenamiento, para las víctimas de la población general, el algoritmo logra clasificar correctamente a la población de víctimas de violencia intrafamiliar con diagnósticos de interés.

Cabe resaltar que como se mencionó en la metodología, el algoritmo basa su decisión en el valor DKL más pequeño, es decir, el algoritmo asigna la población de prueba a la población de entrenamiento cuyo valor DKL sea menor. En este caso, al calcular el DKL entre las víctimas de prueba contra las víctimas de entrenamiento y entre las víctimas de prueba contra las no víctimas de entrenamiento, se obtiene un DKL más pequeño cuando los datos de entrenamiento corresponden a víctimas, por consiguiente el algoritmo asigna la población de prueba al grupo de víctimas.

Considerando que el grupo de prueba efectivamente corresponde a víctimas, el algoritmo hace la clasificación correctamente.

Mayores de edad

Se realizan 1 000 remuestreos de las poblaciones de prueba de víctimas y no víctimas y usando los capítulos identificados en el entrenamiento, el algoritmo logra identificar los datos de las víctimas de prueba como víctimas.

Grupo de Prueba	Valor DKL		Conclusión
	Train Víctimas	Train No Víctimas	
Test Víctimas	1.115029	1.56357	Los datos de prueba corresponden a población víctima de violencia intrafamiliar

Tabla 5-2.: Tabla con el resultado del algoritmo propuesto para las víctimas mayores de edad

De esta manera, en la tabla **5-2** se puede observar que usando los capítulos elegidos en el entrenamiento, para las víctimas mayores de edad, el algoritmo logra clasificar correctamente a la población de víctimas de violencia intrafamiliar con diagnósticos de interés.

Cabe resaltar que como se mencionó anteriormente, el algoritmo basa su decisión en el valor DKL más pequeño, es decir, el algoritmo asigna la población de prueba a la población de entrenamiento cuyo valor DKL sea menor. En este caso, al calcular el DKL entre las víctimas de prueba contra las víctimas de entrenamiento y entre las víctimas de prueba contra las no víctimas de entrenamiento, se obtiene un DKL más pequeño cuando los datos de entrenamiento corresponden a víctimas, por consiguiente el algoritmo asigna la población de prueba al grupo de víctimas.

Considerando que el grupo de prueba efectivamente corresponde a víctimas mayores de edad, el algoritmo hace la clasificación correctamente.

Menores de edad

Se realizan 1 000 remuestreos de las poblaciones de prueba de víctimas y no víctimas y usando los capítulos identificados en el entrenamiento, el algoritmo logra identificar los datos de las víctimas de prueba como víctimas.

Grupo de Prueba	Valor DKL		Conclusión
	Train Víctimas	Train No Víctimas	
Test Víctimas	1.189857	1.679836	Los datos de prueba corresponden a población víctima de violencia intrafamiliar

Tabla 5-3.: Tabla con el resultado del algoritmo propuesto para las víctimas menores de edad

De esta manera, en la tabla **5-3** se puede observar que usando los capítulos elegidos en el entrenamiento, para las víctimas menores de edad, el algoritmo logra clasificar correctamente a la población de víctimas de violencia intrafamiliar con diagnósticos de interés.

Cabe resaltar que como se mencionó anteriormente, el algoritmo basa su decisión en el valor DKL más pequeño, es decir, el algoritmo asigna la población de prueba a la población de entrenamiento cuyo valor DKL sea menor. En este caso, al calcular el DKL entre las víctimas de prueba contra las víctimas de entrenamiento y entre las víctimas de prueba contra las no víctimas de entrenamiento, se obtiene un DKL más pequeño cuando los datos de entrenamiento corresponden a víctimas, por consiguiente el algoritmo asigna la población de prueba al grupo de víctimas.

Considerando que el grupo de prueba efectivamente corresponde a víctimas menores de edad, el algoritmo hace la clasificación correctamente.

6. Conclusiones y trabajo futuro

6.1. Conclusiones

El análisis descriptivo de los datos sugiere que no es adecuado trabajar con técnicas clásicas de análisis estadístico para caracterizar a la población víctima desde sus registros de diagnósticos clínicos dado que, al no considerar las relaciones entre los diagnósticos y los factores de multimorbilidad que presentan los individuos no es posible hacer una caracterización que distinga a individuos víctimas de violencia intrafamiliar de individuos que no han sido víctimas de violencia intrafamiliar y/o no lo han reportado.

De acuerdo con el análisis descriptivo de los datos tiene sentido no trabajar con toda la información de los diagnósticos clínicos de las víctimas de violencia intrafamiliar dado que, en la mayoría de los casos, estas atenciones nunca evidencian un diagnóstico relacionado con violencia lo cual dificulta poder hacer una caracterización de la población víctima usando los diagnósticos clínicos.

Es importante resaltar que, las tipologías de violencia definidas en este trabajo responden principalmente a las limitaciones en cuanto a información, dado que solo se cuenta con la edad de la víctima para realizar la clasificación de los tipos de violencia intrafamiliar. Considerando lo anterior, el análisis descriptivo de los datos muestra que tanto en las características demográficas como en los diagnósticos clínicos las dos tipologías de violencia parecen tener ciertas diferencias, tales como: la proporción de mujeres dentro de cada grupo etario, los diagnósticos con mayor grado nodal en las redes y las aristas con mayor peso.

En términos generales, se puede observar que en los tres casos de estudio: población en general, víctimas menores de edad y víctimas mayores de edad; las redes son disortativas, es decir, la relación se da entre vértices de mayor grado con vértices de menor grado como se puede observar gráficamente; su densidad es pequeña, lo cual sugiere que la distribución de las relaciones que existen entre los diagnósticos de la población víctima no es homogénea y su diámetro es pequeño y muy similar entre redes.

El algoritmo de aprendizaje relacional desarrollado en este trabajo alcanza su objetivo de identificar a la población víctima en cada uno de los casos de estudio: población en general, víctimas menores de edad y víctimas mayores de edad.

Los resultados presentados en este trabajo abren la puerta a una nueva metodología de análisis de información clínica. Actualmente, en Colombia no existe ninguna investigación donde

se usen redes complejas para el análisis de diagnósticos clínicos y como se evidencia en los datos, el uso de esta metodología permite no solo encontrar factores de multimorbilidad en los pacientes sino a su vez permite usar características de la red como el grado nodal para crear algoritmos de clasificación e identificación de población víctima.

6.2. Trabajo futuro

Considerando los resultados obtenidos en este trabajo, se proponen los siguientes aspectos como trabajo futuro:

- Trabajar con tipologías de violencia intrafamiliar diferentes como clasificaciones basadas en: relación con el agresor, edad del agresor, género de la víctima, género del agresor, entre otras.
- Aplicar la metodología y el algoritmo de identificación en una base de datos que contenga información hasta el año 2021 para evaluar el efecto de la pandemia por COVID-19 en los patrones de diagnóstico de la población víctima y no víctima de violencia doméstica.
- Aplicar la metodología propuesta, tomando toda la población de individuos que no han sufrido y/o reportado algún hecho de violencia doméstica y comparar los resultados obtenidos con los de este trabajo.

A. Anexo A: Archivos SIANIESP

Listado de archivos:

- CT = Archivo de control
- AF = Archivo de las transacciones
- US = Archivo de usuarios de los servicios de salud
- AD = Archivo de descripción agrupada de los servicios de salud prestados
- AC = Archivo de consulta
- AP = Archivo de procedimientos
- AH = Archivo de hospitalización
- AU = Archivo de urgencias
- AN = Archivo de recién nacidos
- AM = Archivo de medicamentos
- AT = Archivo de otros servicios

Listado de variables para cada archivo de interés:

Archivo de consulta (AC)

- Número de la factura
- Código del prestador de servicios de salud
- Tipo de identificación del usuario
- Número de identificación del usuario en el sistema
- Fecha de la consulta
- Número de autorización
- Código de la consulta

- Finalidad de la consulta
- Causa externa
- Código del diagnóstico principal
- Diagnóstico relacionado Nro. 1
- Diagnóstico relacionado Nro. 2
- Diagnóstico relacionado Nro. 3
- Tipo de diagnóstico principal
- Valor de la consulta
- Valor de la cuota moderadora
- Valor neto a pagar

Archivo de urgencias (AU)

- Número de la factura
- Código del prestador de servicios de salud
- Tipo de identificación del usuario
- Número de identificación del usuario en el sistema
- Fecha de ingreso del usuario a observación
- Hora de ingreso del usuario a observación
- Número de autorización
- Causa externa
- Diagnóstico a la salida
- Diagnóstico relacionado Nro. 1 a la salida
- Diagnóstico relacionado Nro. 2 a la salida
- Diagnóstico relacionado Nro. 3 a la salida
- Destino del usuario a la salida de observación
- Estado a la salida

- Causa básica de muerte en urgencias
- Fecha de la salida del usuario en observación
- Hora de la salida del usuario en observación

Archivo de hospitalización (AH)

- Número de la factura
- Código del prestador de servicios de salud
- Tipo de identificación del usuario
- Número de identificación del usuario en el sistema
- Vía de ingreso a la institución
- Fecha de ingreso del usuario a la institución
- Hora de ingreso del usuario a la institución
- Número de autorización
- Causa externa
- Diagnóstico principal de ingreso
- Diagnóstico principal de egreso
- Diagnóstico relacionado Nro. 1 a egreso
- Diagnóstico relacionado Nro. 2 a egreso
- Diagnóstico relacionado Nro. 3 a egreso
- Diagnóstico de la complicación
- Estado a la salida
- Diagnóstico de la causa básica de muerte
- Fecha de la salida del usuario a la institución
- Hora de la salida del usuario a la institución

B. Anexo B: Listado de diagnósticos según el paper de [21]

Código	Descripción
T740	negligencia o abandono
T741	abuso físico
T742	abuso sexual
T743	abuso psicológico
T748	otros síndromes de maltrato
T749	síndrome de maltrato, no especificado
Y060	negligencia y abandono: por esposo o pareja
Y061	negligencia y abandono: por padre o madre
Y062	negligencia y abandono: por conocido o amigo
Y068	negligencia y abandono: por otra persona especificada
Y069	negligencia y abandono: por persona no especificada
Y070	otros síndromes de maltrato: por esposo o pareja
Y071	otros síndromes de maltrato: por padre o madre
Y072	otros síndromes de maltrato: por conocido o amigo
Y073	otros síndromes de maltrato: por autoridades oficiales
Y078	otros síndromes de maltrato: por otra persona especificada
Y079	otros síndromes de maltrato: por persona no especificada
Z610	problemas relacionados con la pérdida de relación afectiva en la infancia
Z611	problemas relacionados con el alejamiento del hogar en la infancia
Z612	problemas relacionados con alteración en el patrón de la relación familiar en la infancia
Z613	problemas relacionados con eventos que llevaron a la pérdida de la autoestima en la infancia
Z614	problemas relacionados con el abuso sexual del niño por persona dentro del grupo de apoyo primario
Z615	problemas relacionados con el abuso sexual del niño por persona ajena al grupo de apoyo primario
Z616	problemas relacionados con abuso físico del niño
Z617	problemas relacionados con experiencias personales atemorizantes en la infancia
Z618	problemas relacionados con otras experiencias negativas en la infancia
Z619	problemas relacionados con experiencia negativa no especificada en la infancia

Bibliografía

- [1] ALBERT, Réka ; BARABÁSI, Albert-László: Statistical mechanics of complex networks. En: *Reviews of modern physics* 74 (2002), Nr. 1, p. 47
- [2] APOSTÓLICO, Maíra Rosa ; NÓBREGA, Caroline R. ; GUEDES, Rebeca N. ; FONSECA, Rosa Maria Godoy Serpa d. ; EGRY, Emiko Y.: Características de la violencia contra los niños en una capital brasileña. En: *Revista Latino-Americana de Enfermagem* 20 (2012), Nr. 2, p. 266–273
- [3] BERK, Richard A. ; SORENSON, Susan B. ; BARNES, Geoffrey: Forecasting domestic violence: A machine learning approach to help inform arraignment decisions. En: *Journal of Empirical Legal Studies* 13 (2016), Nr. 1, p. 94–115
- [4] BREDE, Markus. *Networks—An Introduction. Mark EJ Newman.*(2010, Oxford University Press.) \$65,38,£ 35.96 (hardcover), 772 pages. ISBN-978-0-19-920665-0. 2012
- [5] CAÑÓN BUITRAGO, Sandra Constanza M. ; CASTAÑO CASTRILLÓN, José J. ; MEDINA JIMÉNEZ, Natalia ; MOSQUERA ESCOBAR, Kower K. ; QUINTERO CARDONA, Carlos A. ; VARÓN CAICEDO, Leidy L.: Caracterización del intento suicida en pacientes que ingresaron al Instituto del Sistema Nervioso en la ciudad de Pereira, Risaralda, Colombia, 2013-2014. (2015)
- [6] CASTRO, Roberto ; RUÍZ, Agustín: Prevalencia y severidad de la violencia contra mujeres embarazadas, México. En: *Revista de Saúde Pública* 38 (2004), p. 62–70
- [7] DE PROTECCIÓN SOCIAL PRESIDENCIA DE LA REPÚBLICA DE COLOMBIA, Ministerio: Decreto 3518 de 2006. (2006)
- [8] CONCHA-EASTMAN, Alberto ; KRUG, Etienne. *Informe mundial sobre la salud y la violencia de la OMS: una herramienta de trabajo.* 2002
- [9] CONTRERAS, María Isabel Niño: Estado del arte de la violencia intrafamiliar en Colombia: estudios socio-jurídicos. En: *Derecho Público* (2014), Nr. 33, p. 1
- [10] DÍAZ-MARTÍNEZ, Alejandro ; ESTEBAN-JIMÉNEZ, Ramón: I. Violencia intrafamiliar. En: *Gaceta médica de México* 139 (2003), Nr. 4, p. 353–355

- [11] EASLEY, David ; KLEINBERG, Jon: *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge university press, 2010
- [12] DE SALUD INS DIRECCIÓN DE VIGILANCIA Y ANÁLISIS DEL RIESGO EN SALUD PÚBLICA GRUPO DE INVESTIGACIÓN EN EPIDEMIOLOGÍA APLICADA, Ministerio: Uso de RIPS para vigilancia en Salud Pública, Una alternativa para sustitución de fuentes. SIANIESP, Colombia, 2013. (2013)
- [13] ESPINOZA-GÓMEZ, Francisco ; ZEPEDA-PAMPLONA, Víctor ; BAUTISTA-HERNÁNDEZ, Víctor ; HERNÁNDEZ-SUÁREZ, Carlos M. ; NEWTON-SÁNCHEZ, Oscar A. ; PLASENCIA-GARCÍA, Guadalupe R.: Violencia doméstica y riesgo de conducta suicida en universitarios adolescentes. En: *salud pública de méxico* 52 (2010), Nr. 3, p. 213–219
- [14] FERNÁNDEZ-NIÑO, Julián A ; BUSTOS-VÁZQUEZ, Eduardo: Multimorbilidad: bases conceptuales, modelos epidemiológicos y retos de su medición. En: *Biomédica* 36 (2016), Nr. 2, p. 188–203
- [15] GONZÁLEZ, Reinier M. ; GARCÍA, Laidy M. ; LOZANO, Dunia Mercedes F.: Funcionamiento familiar e intento suicida en escolares. En: *Revista Cubana de Medicina General Integral* 33 (2017), Nr. 3, p. 281–295
- [16] HASTIE, Trevor: A closer look at the deviance. En: *The American Statistician* 41 (1987), Nr. 1, p. 16–20
- [17] HERSHEY, John R. ; OLSEN, Peder A.: Approximating the Kullback Leibler divergence between Gaussian mixture models. En: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07* Vol. 4 IEEE, 2007, p. IV–317
- [18] ICBF: Instituto Colombiano de Bienestar Familiar https://www.icbf.gov.co/cargues/avance/docs/concepto_cbf_000123_2017.ht. (2017)
- [19] INMLCF: Instituto Nacional de Medicina Legal y Ciencias Forenses FORENSIS 2018 DATOS PARA LA VIDA. (2019)
- [20] KOLLER, Daphne ; FRIEDMAN, Nir ; DŽEROSKI, Sašo ; SUTTON, Charles ; MCCALLUM, Andrew ; PFEFFER, Avi ; ABBEEL, Pieter ; WONG, Ming-Fai ; MEEK, Chris ; NEVILLE, Jennifer [u. a.]: *Introduction to statistical relational learning*. MIT press, 2007
- [21] LEONE, Michela ; LAPUCCI, Enrica ; DE SARIO, Manuela ; DAVOLI, Marina ; FARCHI, Sara ; MICHELOZZI, Paola: Social network analysis to characterize women victims of violence. En: *BMC public health* 19 (2019), Nr. 1, p. 494
- [22] MENCHÓN, P ; ARTACHO, AI ; CASTELLANOS, P ; HOLGADO, M ; BELLIDO, J ; MUÑOZ, N. *Sobre la inhibición a denunciar de las víctimas de violencia de género*. 2015

- [23] MONTERROSA-CASTRO, Álvaro ; ARTETA-ACOSTA, Cindy ; ULLOQUE-CAAMAÑO, Liezel: Violencia doméstica en adolescentes embarazadas: caracterización de la pareja y prevalencia de las formas de expresión. En: *Iatreia* 30 (2017), Nr. 1, p. 34–46
- [24] NEAL, Zachary. *Borgatti, SP, Everett, MG, & Johnson, JC (2013). Analyzing Social Networks: Thousand Oaks, CA: Sage. 296 pp., 54.00(paper), 130.00 (cloth).* 2015
- [25] NEWMAN, Mark E.: The structure and function of complex networks. En: *SIAM review* 45 (2003), Nr. 2, p. 167–256
- [26] PINZÓN, Mildred Alexandra V. ; MOÑETÓN, Marly Johana B. ; ALARCÓN, Linda Lilliana A.: Variables psicosociales asociadas al intento suicida, ideación suicida y suicidio en jóvenes. En: *Tesis psicológica* 8 (2013), Nr. 1
- [27] DE SALUD Y PROTECCIÓN SOCIAL MINISTRO DE SALUD Y PROTECCIÓN SOCIAL, Ministerio: Resolución número 4622 de 2016. (2016)
- [28] RIBERO, Rocío ; FABIO, SÁ [u. a.]: Determinantes, efectos y costos de la violencia intrafamiliar en Colombia / Universidad de los Andes-CEDE. 2004. – Informe de Investigación
- [29] RODRÍGUEZ, Jorge J. ; KOHN, Robert ; AGUILAR-GAXIOLA, Sergio: *Epidemiología de los trastornos mentales en América Latina y el Caribe*. Pan American Health Org, 2009 (632)
- [30] RODRÍGUEZ-ESCOBAR, José A. ; MEDINA-PÉREZ, Óscar A. ; CARDONA-DUQUE, Deisy V.: Caracterización del suicidio en el departamento de Risaralda, Colombia, 2005-2010. En: *Revista de la Facultad de Medicina* 61 (2013), Nr. 1, p. 9–16
- [31] DE LOS RECURSOS DEL SISTEMA GENERAL DE SEGURIDAD SOCIAL EN SALUD, ADRES A.:
<https://www.adres.gov.co/entidades-territoriales/procesos/bdua>
. (2020)
- [32] DE SALUD, Instituto N.: Vigilancia en salud pública de violencia de género e intrafamiliar, Colombia, Periodo V de 2020. (2020)
- [33] DE SALUD MINISTRA DE SALUD, Ministerio: Resolución número 3374 de 2000. (2000)
- [34] SANTOS MOTA, Rosana ; MATÍAS SANTOS, Mariana ; CAMARGO, Climene Laura d. ; PEREIRA GOMES, Nadirlene ; FREIRE DINIZ, Normélia M. ; DINIZ RODRIGUES, Adriana: Sexualidad, embarazo y violencia doméstica: experiencias de adolescentes brasileñas. En: *Index de Enfermería* 23 (2014), Nr. 3, p. 139–143

-
- [35] SECRETARÍA: Secretaría de Bogotá ¿Qué es la violencia intrafamiliar y cómo se puede prevenir? <https://bogota.gov.co/mi-ciudad/integracion-social/que-es-la-violencia-intrafamiliar-y-como-se-puede-prevenir>. (2021)
- [36] SILVA, Thiago C. ; ZHAO, Liang: *Machine learning in complex networks*. Vol. 2016. Springer, 2016
- [37] SUBRAMANI, Sudha ; WANG, Hua ; VU, Huy Q. ; LI, Gang: Domestic violence crisis identification from facebook posts based on deep learning. En: *IEEE access* 6 (2018), p. 54075–54085
- [38] UNICEF [u. a.]. *Ocultos a plena luz: un análisis estadístico de la violencia contra los niños. Resumen*. 2015
- [39] VÁLDEZ-SANTIAGO, Rosario ; SANÍN-AGUIRRE, Luz H.: La violencia doméstica durante el embarazo y su relación con el peso al nacer. En: *Salud Pública de México* 38 (1996), Nr. 5, p. 352–362
- [40] VARGAS, Jean D P. ; ROGETT, Leonardo C.: Apuntes acerca de la violencia, la violencia intrafamiliar y la violencia escolar. En: *Psicogente* 10 (2007), Nr. 17, p. 16–25