

Minería de datos aplicada a la detección de patrones asociados a la violencia a la mujer.

Agramont Bejarano Ignacio Andre, Saavedra Ayarde Stephanie Belen

December 7, 2022

En el desarrollo de este proyecto de investigación se introduce al concepto y a la implementación del modelo de Minería de Datos "Análisis de correspondencia múltiple", donde se tiene por objetivo identificar patrones asociados a la violencia a la mujer en Bolivia basado en la encuesta de prevalencia de violencia contra la mujer 2016. Se aborda esta temática siguiendo todos los pasos de KDD (Knowledge Discovery in Databases por su siglas en ingles), tomando los datos de manera crítica basada en información debidamente documentada.

Los informes internacionales, a nivel latinoamericano, indican que Bolivia ocupa el primer lugar en violencia física contra la mujer y el segundo lugar en violencia sexual, en las edades de 15 a 49 años. Se hace referencia a las acciones que el Estado boliviano realiza, a través de promulgación de leyes que tienen como meta garantizar una vida libre de violencia a la mujer boliviana, sin embargo, datos actuales muestran que este fenómeno se ha incrementado y recrudecido significativamente; es por esto que a través del modelo elegido se identificará los indicadores relacionados a tal fenómeno. Se precisa cuidado y precisión en cada etapa del proceso de KDD para que de esta forma el análisis y resultados sean correctos y según las expectativas planteadas en el inicio del proceso de investigación.

Palabras Clave: Análisis de Correspondencia Múltiple, KDD, Minería de datos, Preprocesamiento, Violencia a la mujer.

0.1 Introducción

La violencia de género se ha convertido en la primera causa de muerte de las mujeres a nivel mundial, miles de mujeres son asesinadas cada año por el sólo hecho de ser mujeres, tal es así, que en la actualidad se le ha designado el término nuevo de feminicidio para nombrar este fenómeno. Los asesinatos de mujeres vienen a constituir el último eslabón de una cadena de múltiples y variadas formas de violencia, como son las violaciones, abusos, golpes, desprecio, humillación sumados a una serie de malos tratos. La violencia de género se manifiesta en diversas formas y en todas las esferas en las que la mujer se desenvuelve, sea esta laboral, familiar, en la escuela, en la pareja, en ámbitos sociales, comunicacionales, etc. Latinoamérica presenta altos índices de violencia contra la mujer. El informe de 2013 publicado por la OMS/OPS, refieren que la violencia contra la mujer llega a causar daños irreversibles, no sólo para la víctima sino también para su entorno familiar y la sociedad en su conjunto. Ante ello, se propone la implementación de un modelo de minería de datos que pueda identificar los patrones o características asociadas a la violencia a la mujer, es decir tener indicadores que reflejen el por qué del crecimiento de los casos de violencia de género.

La minería de datos o exploración de datos es un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones o reglas significativas en grandes volúmenes de conjuntos de datos, se lo considera como una disciplina dentro del campo de la ciencia de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior. Además de la etapa de análisis en bruto, supone aspectos de gestión de datos y de bases de datos, de procesamiento de datos, del modelo y de las consideraciones de inferencia, de métricas de intereses, de consideraciones de la teoría de la complejidad computacional, de post-procesamiento de las estructuras descubiertas, de la visualización y de la actualización en línea.

0.2 Pregunta de investigación

La pregunta de investigación en la cual se basa el proyecto es: ¿Cuáles son las variables que influyen en que una mujer pueda o no ser víctima de violencia de género? A través de esta pregunta de investigación se analizara los datos para identificar los niveles de influencia de los diferentes elementos a los indicadores de violencia a la mujer en Bolivia.

0.3 Motivacion

La violencia contra las mujeres y niñas es una violación grave de los derechos humanos. Su impacto puede ser inmediato como de largo alcance, e incluye múltiples consecuencias físicas, sexuales, psicológicas, e incluso mortales, para mujeres y niñas. Afecta negativamente el bienestar de las mujeres e impide su plena participación en la sociedad. Además de tener consecuencias negativas para las mujeres, la violencia también impacta su familia, comunidad y el país. Los altos costos asociados, que comprenden desde un aumento en gastos de atención de salud y servicios jurídicos a pérdidas de productividad, impactan en presupuestos públicos nacionales y representan un obstáculo al desarrollo.

Tras varias décadas de movilizaciones promovidas por la sociedad civil y los movimientos de mujeres, se ha conseguido incluir la erradicación de la violencia de género en las agendas nacionales e internacionales. Sin embargo, continúan existiendo desafíos en la aplicación de estas leyes, resultando en una limitada protección y acceso a la justicia por parte de mujeres y niñas. Asimismo, no se hace lo suficiente para prevenir la violencia, y cuando ésta ocurre a menudo queda impune.

Es por eso que debemos implementar nuevas herramientas más allá de las leyes y normativas que una comunidad puede establecer, a herramientas me refiero al uso de la tecnología como vía para reducir los indicadores de violencia, esto a partir de predicciones e identificación de patrones en los datos. Algunos de los estudios, demuestran la utilidad de la minería de datos en el análisis de la violencia, encontrando así información relevante que pueda generar nuevos conocimientos que permitan crear nuevos objetivos y estrategias para la disminución de estos indicadores. La motivación para realizar el presente proyecto es poder aplicar técnicas de minería de datos para identificar indicadores asociados a la violencia de género, que elementos influyen a que una mujer pueda o no ser víctima de violencia.

0.4 Marco Teórico

0.4.1 Violencia de género

Se entiende por violencia de género cualquier acto violento o agresión, basados en una situación de desigualdad en el marco de un sistema de relaciones de dominación de los hombres sobre las mujeres que tenga o pueda tener como consecuencia un daño físico, sexual o psicológico, incluidas las amenazas de tales actos y la coacción o privación arbitraria de la libertad, tanto si ocurren en el ámbito público como en la vida familiar o personal.

La violencia de género constituye una grave violación a los derechos humanos; al mismo tiempo, se trata de un problema de salud y protección que pone en riesgo la integridad, la dignidad, la libertad y hasta la vida de la mujer. La violencia de género ha sido y sigue siendo una de las manifestaciones más claras de la desigualdad, subordinación y de las relaciones de poder de los hombres sobre las mujeres. Este tipo de violencia se basa y se ejerce por la diferencia subjetiva entre los sexos. En pocas palabras, las mujeres sufren violencia por el mero hecho de ser mujeres, y las víctimas son mujeres de cualquier estrato social, nivel educativo, cultural o económico.

La violencia de género es aquella que se ejerce sobre las mujeres por parte de quienes estén o hayan estado ligados a ellas por relaciones de afectividad (parejas o ex-parejas). El objetivo del agresor es producir daño y conseguir el control sobre la mujer, por lo que se produce de manera continuada en el tiempo y sistemática en la forma, como parte de una misma estrategia. Cualquier tipo de violencia se fundamenta en una relación de poder donde alguien trata de dominar a la otra persona por la fuerza, contra su voluntad; trata de obligarla a que haga lo que no desea, a que se ruegue y reconozca su inferioridad y dependencia con respecto a quien ejerce la violencia.

0.4.2 Estado actual de la violencia a la mujer en Bolivia

La violencia de género se ha convertido en la primera causa de muerte de las mujeres a nivel mundial, miles de mujeres son asesinadas cada año por el sólo hecho de ser mujeres, tal es así, que en la actualidad se le ha designado el término nuevo de feminicidio para nombrar este fenómeno macabro y monstruoso.

Los asesinatos de mujeres vienen a constituir el último eslabón de una cadena de múltiples y variadas formas de violencia, como son las violaciones, abusos, golpes, desprecio, humillación sumados a una serie de malos tratos. La violencia de género se manifiesta en diversas formas y en todas las esferas en las que la mujer se desenvuelve, sea esta laboral, familiar, en la escuela, en la pareja, en ámbitos sociales, comunicacionales, etc.

Latinoamérica presenta altos índices de violencia contra la mujer. El informe de 2013 publicado por la OMS/OPS, refieren que la violencia contra la mujer llega a causar daños irreversibles, no sólo para la víctima sino también para su entorno familiar y la sociedad en su conjunto. Ante ello, innumerables instituciones le han dedicado espacio y tiempo a la prevención e intervención a esta problemática. La Organización de Naciones Unidas (ONU), en 1995, incluyó entre sus objetivos estratégicos la lucha contra la violencia hacia las mujeres; la Organización Mundial de la Salud (OMS), por su parte, en 1998, declaró a la violencia de pareja como una prioridad internacional para los servicios de salud y en 2013 reiteró esta prioridad, estimulando la pronta acción a partir de datos encontrados en una pesquisa realizada y publicada, refiere datos como que el 38 por ciento de los asesinatos de mujeres en el mundo son producto de casos de violencia machista por parte de un compañero sentimental, parejas o ex-parejas. En contraste al de hombres muertos a manos de sus parejas, equivalente a un 6 por ciento. (OMS, 2013)

En Octubre de 2016, ONU Mujeres, refiere que alrededor de 60.000 mujeres son asesinadas al año en América Latina, expresó el siguiente comunicado: “La impunidad, que alcanza niveles alarmantes en América Latina y el Caribe, es un elemento central para la perpetuación de la violencia contra las mujeres. Mientras persista la impunidad, las sociedades continuarán aceptando y tolerando actos de violencia contra mujeres y niñas”. América Latina es la región donde se presentan más asesinatos de mujeres por su género: 14 de los 25 países del mundo se encuentran con las tasas elevadas de femicidio y alrededor de 60000 mujeres son asesinadas al año. Según ONU Mujeres, 1 de cada 3 mujeres en el mundo sufre violencia física o sexual a manos de un compañero sentimental.

En Bolivia, las estadísticas y las noticias cotidianas muestran que los índices de denuncias sobre violencia de pareja donde la víctima es la mujer, son altamente significativos, según el informe de OMS/OPS de 2013 sobre violencia contra la mujer en América Latina y El Caribe, Bolivia ocupa el primer lugar en violencia física contra la mujer y el segundo lugar en violencia sexual, en las edades de 15 a 49 años, dicha violencia es infligida por el compañero íntimo, mujeres alguna vez casadas o en unión, el 53,3 por ciento en 2003 y el 27,3 por ciento, datos recabados en 2008.

0.4.3 Minería de datos

El minado de datos es un conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos que expliquen el comportamiento de estos datos. A pesar de que la idea de la Minería de Datos puede parecer una innovación tecnológica muy reciente, en realidad este término apareció en la década de 1970, no obstante, no fue hasta la década de 1980 cuando empezó su consolidación. La minería de datos surgió con la intención o el objetivo de ayudar a comprender una enorme cantidad de datos y que estos pudieran ser utilizados para extraer conclusiones para contribuir en la mejora de los diferentes procesos ya sea en un flujo industrial o en proyectos independientes.

Permite convertir datos sin procesar y estructurarlos en información comprensible sobre diversas áreas de negocios y del mercado. Por tanto, los datos son el medio o la base para llegar a conclusiones y transformar estos datos en información relevante, para que las empresas puedan abarcar mejoras y soluciones que les ayuden a conseguir sus objetivos. Su principal finalidad es explorar, mediante la utilización de distintas técnicas y tecnologías, bases de datos enormes de manera automática. El objetivo es encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos que se han ido recopilando con el tiempo. Estos patrones pueden encontrarse utilizando estadísticas o algoritmos de búsqueda próximos a la Inteligencia Artificial y a las redes neuronales.

Este análisis brinda información importante que ayudará a tu empresa a resolver problemas y a predecir datos que puedan ser útiles para los diferentes procesos que tienes dentro de tu proyecto de investigación.

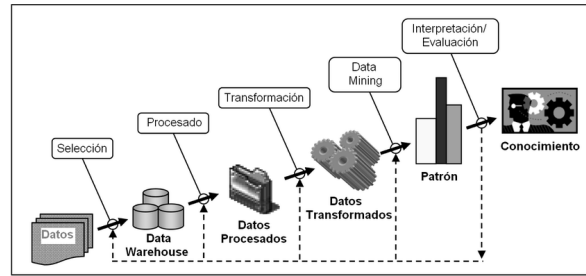


Figure 1: Etapas del proceso de KDD

0.4.4 KDD

El proceso de extraer conocimiento a partir de grandes volúmenes de datos ha sido reconocido por muchos investigadores como un tópico de investigación clave en los sistemas de bases de datos, y por muchas compañías industriales como una importante área y una oportunidad para obtener mayores ganancias. Autores como Fayyad, Piatetsky-Shapiro y Smith lo definen como “El proceso no trivial de identificación de patrones válidos, novedosos, potencialmente útiles y fundamentalmente entendibles al usuario a partir de los datos”.

El Descubrimiento de conocimiento en bases de datos (KDD, del inglés Knowledge Discovery in Databases) es básicamente un proceso automático en el que se combinan descubrimiento y análisis. El proceso consiste en extraer patrones en forma de reglas o funciones, a partir de los datos, para que el usuario los analice. Esta tarea implica generalmente preprocesar los datos, hacer minería de datos y presentar resultados.

El proceso KDD que se muestra en la figura 1 es interactivo e iterativo, involucra numerosos pasos con la intervención del usuario en la toma de muchas decisiones. Se resume en las siguientes etapas:

1. Selección
2. Preprocesamiento - Limpieza
3. Transformación - Reducción
4. Minería de datos
5. Interpretación - Evaluación

0.4.5 Análisis de Correspondencia Múltiple

Es una técnica descriptiva o exploratoria cuyo objetivo es resumir una gran cantidad de datos en un número reducido de dimensiones, con la menor pérdida de información posible. En esta línea, su objetivo es similar al de los métodos factoriales, salvo que en el caso del análisis de correspondencias el método se aplica sobre variables categóricas u ordinales. El análisis de correspondencias simple se utiliza a menudo en la representación de datos que se pueden presentar en forma de tablas de contingencia de dos variables nominales u ordinales. Otras utilizaciones implican el tratamiento de tablas de proximidad o distancia entre elementos, y tablas de preferencias; esto se hace representando los datos como puntos en un espacio euclídeo de baja dimensión.

Si se trata de una tabla de contingencia de dos variables cualitativas, una variable cuyas categorías aparecen en filas y la otra variable cuyas categorías son representadas en columnas, el análisis de correspondencias consiste en resumir la información presente en las filas y columnas de manera que pueda proyectarse sobre un subespacio reducido, y representarse simultáneamente los puntos fila y los puntos columna, pudiéndose obtener conclusiones sobre relaciones entre las dos variables nominales u ordinales de origen. La extensión del análisis de correspondencias simples al caso de varias variables nominales

(tablas de contingencia multidimensionales) se denomina Análisis de Correspondencias Múltiples, y utiliza los mismos principios generales que la técnica anterior. En general se orienta a casos en los cuales una variable representa ítems o individuos y el resto son variables cualitativas u ordinales que representan cualidades.

El Análisis de Correspondencias tiene dos objetivos básicos, la asociación entre categorías de columnas o filas, es decir, medir la asociación de solo una fila o columna, para ver, por ejemplo, si las modalidades de una variable pueden ser combinadas. Y el otro objetivo es la asociación entre categorías de filas y columnas, es decir estudiar si existe relación entre categorías de las filas y columnas. El análisis de correspondencias solo requiere que los datos representen las respuestas a una serie de preguntas y que estén organizadas en categorías. Dependiendo si existen dos o más variables el análisis será simple o múltiple.

0.5 Datos

Para indentificar y analizar los patrones relacionados a la violencia a la mujer se emplean los datos que provienen de la encuesta de prevalencia de violencia contra la mujer en Bolivia del año 2016 que pone a disposición el Instituto Nacional de Estadística (INE) a través de su página web. Dichos datos estan disponibles para su descarga y su respectivo uso en el siguiente link [Dataset — Encuesta de violencia contra la mujer](#).

El proyecto se encuentra en un repositorio público de Github para su revisión y para su respectivo uso como base de otros proyecto o para replicar el mismo, se lo puede encontrar en el siguiente link [Repositorio — Modelo de minería de datos para la identificación de patrones violencia contra la mujer](#)

0.6 Metodología

0.6.1 Selección

En la etapa de selección, se identifica los datos relevantes y prioritarios para cumplir o alcanzar el indicador definido, se selecciona una muestra representativa de este, sobre el cual se realiza el proceso de descubrimiento.

Variables de estudio

Dentro de este proyecto se seleccionaron múltiples variables que nos permitieron realizar el análisis, las variables que intervinieron en el estudio son desarrolladas con mayor detalle a continuación:

- Indicador de humillación: Se construyo un indicador de humillación como tipo de violencia a partir de la sección 1: Ámbitos social, educativo y Laboral, donde se tomo la pregunta 1 en su enunciado 1. Dentro de esta pregunta se tomo el caso más crítico y el caso intermedio, respuesta tal como "Muchas Veces" y "Pocas Veces", donde se evidencia un tipo de violencia.
- Indicador de amenazas: Se construyo un indicador de humillación como tipo de violencia a partir de la sección 1: Ámbitos social, educativo y Laboral, donde se tomo la pregunta 1 en su enunciado 2. Dentro de esta pregunta se tomo el caso más crítico y el caso intermedio, respuesta tal como "Muchas Veces" y "Pocas Veces", donde se evidencia un tipo de violencia.
- Indicador de agresión: Se construyo un indicador de humillación como tipo de violencia a partir de la sección 1: Ámbitos social, educativo y Laboral, donde se tomo la pregunta 3 en su enunciado 1. Dentro de esta pregunta se tomo el caso más crítico y el caso intermedio, respuesta tal como "Muchas Veces" y "Pocas Veces", donde se evidencia un tipo de violencia.
- Indicador de piropos sexuales: Se construyo un indicador de humillación como tipo de violencia a partir de la sección 1: Ámbitos social, educativo y Laboral, donde se tomo la pregunta 4 en su enunciado 1. Dentro de esta pregunta se tomo el caso más crítico y el caso intermedio, respuesta tal como "Muchas Veces" y "Pocas Veces", donde se evidencia un tipo de violencia.

- Indicador de Manoseos: Se construyo un indicador de humillación como tipo de violencia a partir de la sección 1: Ámbitos social, educativo y Laboral, donde se tomo la pregunta 5 en su enunciado 1. Dentro de esta pregunta se tomo el caso más crítico y el caso intermedio, respuesta tal como "Muchas Veces" y "Pocas Veces", donde se evidencia un tipo de violencia.
- Indicador de obligación a tener relaciones sexuales: Se construyo un indicador de humillación como tipo de violencia a partir de la sección 1: Ámbitos social, educativo y Laboral, donde se tomo la pregunta 6 en su enunciado 1. Dentro de esta pregunta se tomo el caso más crítico y el caso intermedio, respuesta tal como "Muchas Veces" y "Pocas Veces", donde se evidencia un tipo de violencia.
- Indicador de antecedentes familiares: Se construyo un indicador de antecedentes de violencia en la familia a partir de la sección 2: Familia de origen y situación patrimonial, donde se tomaron la pregunta 3, 4 y 5. Dentro de estas preguntas se tomo el caso más crítico, respuesta tal como "Muchas Veces", donde se evidencia antecedentes violentos.
- Indicador de violencia en la pareja: Se construyo un indicador de antecedentes de violencia en la familia a partir de la sección 3: Relación Actual, donde se toma la pregunta 10. Dentro de esta pregunta se considera cada subitem con el caso más crítico, respuesta tal como "Muchas Veces", donde se evidencia actos violentos dentro de la pareja.
- Sexo: Se tomo la variable sexo para poder realizar un filtrado en todos los datos y solo analizar data correspondiente a mujeres, esta variable se obtuvo de la pregunta 2 correspondiente a la sección 2: Características sociodemográficas.
- Edad: Se tomo la variable edad para poder observar su influencia con los indicadores de violencia expuestos anteriormente, esta variable se obtuvo de la pregunta 3 correspondiente a la sección 2: Características sociodemográficas.
- Perteneciente a un pueblo indígena: Se tomo la variable pertecene a un pueblo indígena para poder observar su influencia con los indicadores de violencia expuestos anteriormente, esta variable se obtuvo de la pregunta 9 correspondiente a la sección 2: Características sociodemográficas.
- Estado Civil: Se tomo la variable estado civil para poder realizar un filtrado en todos los datos y solo analizar data correspondiente a mujeres casadas, solteras, concubinas, separadas, divorciado segun la base de datos que se esta empleando, esta variable se obtuvo de la pregunta 10 correspondiente a la sección 2: Características sociodemográficas.
- Nivel de educación: Se tomo la variable nivel de educación alcanzado para poder observar su influencia con los indicadores de violencia expuestos anteriormente, esta variable se obtuvo de la pregunta 1 correspondiente a la sección 3: Educación.
- Ingreso: Se tomo la variable ingreso percibido para poder observar su influencia con los indicadores de violencia expuestos anteriormente, esta variable se obtuvo de la pregunta 11a correspondiente a la sección 5: Empleo.

0.6.2 Preprocesamiento - Limpieza

En la etapa de preprocesamiento/limpieza, se analiza la calidad de los datos aplicando operaciones básicas como la remoción de datos ruidosos, se seleccionan estrategias para el manejo de datos desconocidos, datos nulos, datos duplicados y técnicas estadísticas para su reemplazo.

- En el caso de este proyecto se realizo un proceso de unificación de nombres de variables, es decir transformar todos los nombre de variables a minúsculas para que el análisis no se obstaculise por este factor.
- Se omitieron los valores nulos de la data seleccionada y datos extremos, es decir filas con datos atípicos o datos nulos.

0.6.3 Transformación - Reducción

En la etapa de transformación/reducción de datos, se buscan las características útiles para representar los datos bajo el objetivo propuesto.

- En el caso de este proyecto se realizó un proceso de reducción de dimensiones para preparar la data para el modelo, es decir se omiten las columnas de folio dado que no es una variable relevante para el análisis.

$$matrixN < -matrixN[, -c(1)]$$

- Luego de generar la pseudo matriz de contingencia se omitieron las variables iniciales, quedandose solo con las variables cruzadas en formato binario a partir de cada factor y carácter en la data seleccionada.

$$matrizN < -dummy_ols(data, variables)$$

$$matrixN < -matrixN[, -c(1 : 10)]$$

- Una vez se tiene la matriz de contingencia se omiten las variables que no poseen ninguna observación, es decir variables con cero ocurrencias dentro de la matriz, esto con el objetivo de no crear datos extremos que puedan afectar nuestro modelo.

0.6.4 Minería de datos

El objetivo de la etapa minería de datos es la búsqueda y descubrimiento de patrones y características al indicador de la violencia a la mujer. Luego de realizar la selección de variables, su respectiva limpieza y transformación se pasa a la etapa de Minería de datos, tal como se detalla a continuación.

1. A través de la matriz generada se debe obtener el valor de n, el cual representa la cantidad de observaciones que se tiene dentro del modelo.

$$n < -sum(matrixN)$$

2. A partir del valor de n y la matriz de contingencia, se debe calcular la matriz p que representa la división entre ambos valores para sacar las proporciones de participación de cada elemento y relativizar la información.

$$matrixP < -matrixN/n$$

3. Para continuar con los cálculos respectivos se debe convertir el valor de P en una matriz, dado que este es un vector.

$$matrixP < -as.matrix(matrixP)$$

4. Para una tabla de contingencia en forma de matriz, se debe calcular el valor de las marginales de las filas y de las columnas (r = rows, c = cols).

$$r < -margin.table(matrixP, 1)$$

$$c < -margin.table(matrixP, 2)$$

5. Al tener la matriz de contingencia con las variables cruzadas en formato binario a partir de cada factor y carácter en la data seleccionada, se debe calcular la matriz de residuos estandarizados. Este cálculo se lo realiza a través de la fórmula que se muestra a continuación.

$$S < -diag(r^{\wedge} - 0.5)) * (matrixP - r * t(c)) * diag(c^{\wedge} - 0.5))$$

6. Posteriormente se realiza la descomposición SVD, obteniendo el valor de u, v y d para obtener las coordenadas de los ejes posteriormente.

$$u < -svd(S) u$$

$$v < -svd(S) v$$

$$Da < -diag(svd(S) d)$$

7. Se procede a calcular las coordenadas principales de las filas donde se guarda las nuevas coordenadas para generar los gráficos posteriormente y entender su relación una sobre la otra. En FF se guarda la relación entre todas las observaciones en cuanto a las variables seleccionadas inicialmente.

$$FF < -diag(r^{\wedge} - 0.5)) * u * Da$$

8. Se procede a calcular las coordenadas principales de las columnas donde se guarda las nuevas coordenadas para generar los gráficos posteriormente y entender su relación una sobre la otra. En GG se guarda los niveles de todas las columnas dentro de nuestro conjunto de datos.

$$GG < -diag(c^{\wedge} - 0.5)) * v * Da$$

9. Una vez que se tiene las coordenadas principales se calcula el valor de la inercia. Se puede interpretar como la media ponderada de las distancias chi cuadrado entre los perfiles fila y su perfil medio. O, alternativamente, se puede definir del mismo modo para los perfiles columna.

$$sum(((matrixP - r * t(c)) * 2) / (r * t(c)))$$

$$cumsum(svd(S) d / (svd(S) d$$

10. Finalmente se debe graficar los resultados de los datos obtenidos para poder observar las diferentes tendencias que puede existir entre las variables seleccionadas anteriormente.

$$plot(GG[, 1], GG[, 2], type = "n")$$

$$text(GG[, 1], GG[, 2], labels = labels, cex = 0.4)$$

$$points(FF[, 1], FF[, 2], col = "grey")$$

0.7 Resultados y Análisis

En la etapa de interpretación/evaluación, se interpretan los patrones descubiertos y posiblemente se retorna a las anteriores etapas para posteriores iteraciones. En este proyecto de investigación se realizaron diferentes pruebas con diferentes indicadores para saber su influencia en actos de violencia contra la mujer. A continuación se muestran los resultados de todos los casos:

- En este caso se cruzo variables como el nivel de educación, la pertenencia a un grupo indígena, y los indicadores de violencia tales como humillación, amenazas, agresión, piropos sexuales, obligación a tener relaciones sexuales y manoseos. En esta gráfica se evidencia que el nivel de educación si es un factor influyente en los indicadores de violencia contra la mujer, dado que a menor nivel de educación se puede observar tendencia a mayor cantidad de actos de violencia social, educativa y laboral.
- En este caso se cruzo variables como el indicador de antecedentes de violencia familiar y los 6 indicadores de violencia tales como como humillación, amenazas, agresión, piropos sexuales, obligación a tener relaciones sexuales y manoseos. En esta gráfica se evidencia que los antecedentes de violencia familiar si bien pueden influenciar en que una mujer puede o no ser víctima de violencia, este indicador no es determinante por la distancia que toma con los indicadores de violencia. Si bien se encuentra cerca, no en todos los casos aplica que sea condicionante para la presencia de actos de violencia social, educativa y laboral.
- En este caso se cruzo variables como la pertenencia a un grupo indígena y los indicadores de violencia tales como humillación, amenazas, agresión, piropos sexuales, obligación a tener relaciones sexuales y manoseos. En esta gráfica se evidencia que la pertenencia a un grupo indígena no es un factor que influye en que una mujer pueda ser o no víctima de violencia. Los datos se encuentran muy alejados por lo que ser o no parte de un origen indígena no es condicionante para sufrir actos de violencia.

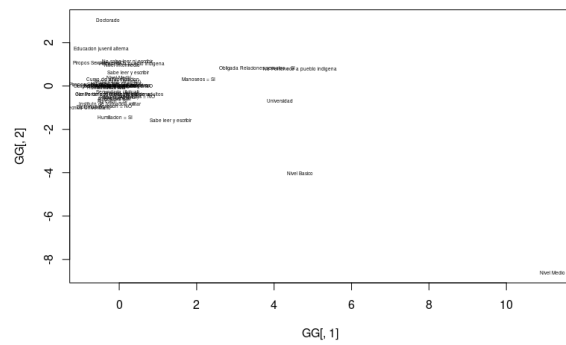


Figure 2: Grafica resultante del primer caso de análisis

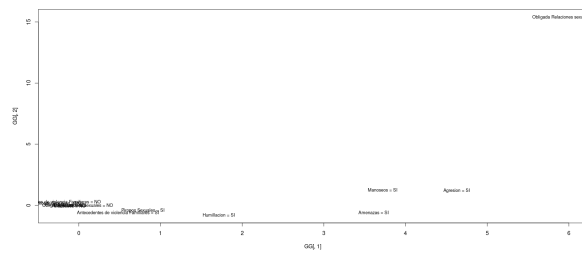


Figure 3: Grafica resultante del segundo caso de análisis

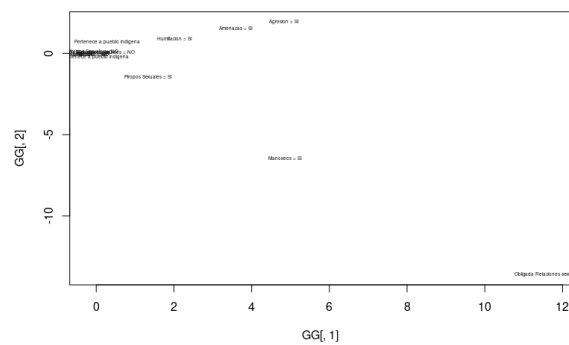


Figure 4: Grafica resultante del tercer caso de análisis

0.8 Conclusiones

Durante el desarrollo de todo el proyecto se pudo observar la importancia de tener bases sólidas en los conceptos y en el uso de modelos de minería de datos para su respectiva aplicación en este tipo de proyectos de investigación, es decir, la selección de los datos, en el tratamiento de la data preseleccionada a través de la limpieza y de la transformación, y así por aplicarlo al modelo de minería de datos establecido. El proyecto aplicó todas las etapas de KDD para poder obtener resultados correctos y una posterior interpretación.

Dentro de este proyecto se presentaron varios desafíos, tal como la correcta selección de las variables a utilizar, la transformación de las mismas y su aplicación en el modelo escogido; esto dado que los indicadores de violencia que se construyeron en base a las preguntas de la encuesta no tenían una relación significativa con las demás variables, es por eso que se realizaron diferentes pruebas, no solo con diferentes variables sino también con diferentes modelos de minería de datos.

Es así que se dan por cumplidos todos los objetivos que fueron planteados al iniciar este proyecto de investigación, concluyendo en la importancia del proceso de creación de proyectos aplicando los diferentes modelos de minería de datos, apoyándonos en la tecnología de la época actual y el hecho de estar actualizándose y optimizando procesos de forma constante.

0.9 Recomendaciones

Luego de finalizar el desarrollo del proyecto de investigación se recomienda los siguientes puntos para el desarrollo de proyectos similares.

- Tener conocimiento de todas las variables que intervienen dentro del conjunto de datos.
- Tener conocimientos sólidos en la identificación de variables y en su respectivo tratamiento.
- Tener un conocimiento sólido en el modelo seleccionado para poder implementarlo en el proyecto.
- Realizar un correcto preprocesamiento para segmentar la data útil para el caso de estudio
- Tener cuidado con no eliminar datos nulos o vacíos, y que estos puedan afectar al análisis del modelo.
- Tener una buena comunicación con todos los miembros del equipo y tener una buena administración de las tareas.
- Para el correcto flujo de trabajo, respetar deadlines de las diferentes actividades.

0.10 Referencias

Agrawal, R. y Srikant, R. (1994). Fast Algorithms for Mining Association Rules. vldb Conference, Santiago de Chile.

Agrawal, R., Ghosh S., Imielinski, T., Iyer, B. y Swami, A. (1992). An Interval Classifier for Database Mining Applications. Proceedings VDLB Conference, Vancouver.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. y Zanasi A. (1998). Discovering Data Mining from Concept to Implementation, Prentice Hall. Recuperado de <http://dl.acm.org/citation.cfm?id=270298>

Chapman, P., Clinton, J., Randy, K., Khabaza, T., Reinartz, T., Shearer, C. y Wirth, R. (2000). Step-by-Step Data Mining Guide. Recuperado de <http://www.crisp-dm.org/CRISPWP-0800.pdf>

Han, J. y Kamber, M. (2001). Data Mining Concepts and Techniques. San Francisco: Morgan Kaufmann Publishers.

Hernández, J., Ramírez, M. y Ferri, C. (2005). Introducción a la Minería de Datos. Madrid: Editorial Pearson Educación SA.

0.11 Anexos

Repositorio de Github: <https://github.com/StephanieSaavedraAyarde/Modelo-mineria-de-datos-violencia-a-la-mujer>

Base de datos: <https://www.ine.gob.bo/index.php/censos-y-banco-de-datos/censos/bases-de-datos-encuestas-sociales/>