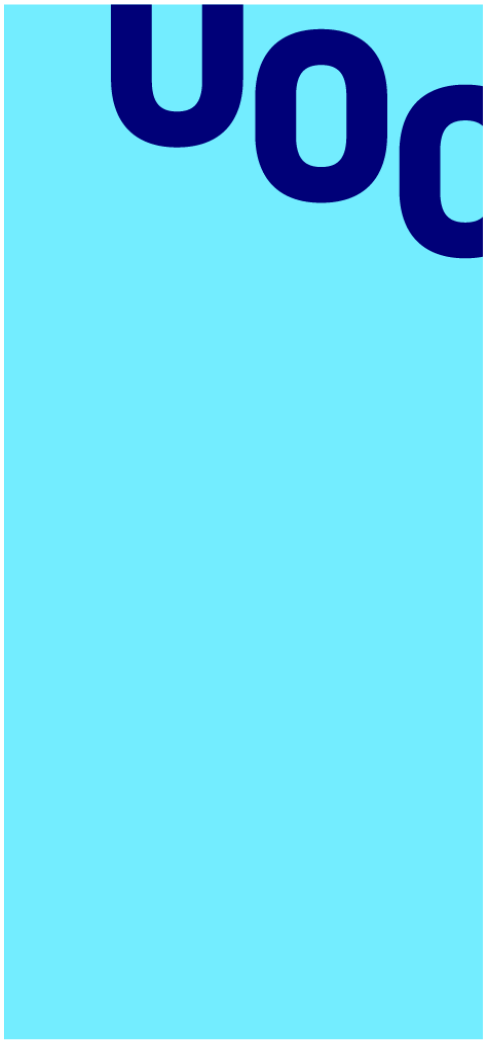
Universitat Oberta de Catalunya

uoc.edu

Arquitectura de un sistema de ayuda a la prevención de casos de violencia de género en España



Universitat Oberta de Catalunya

Javier Plo Moreno

Máster Ciencia de Datos
Area Machine Learning (TFM-ML)

Tutor/a de TF
Laia Subirats Maté
Profesor/a responsable de la asignatura
Ferran Prados Carrasco

Fecha Entrega
01/2023

Copyright

Ficha del Trabajo Final

Título del trabajo:	Arquitectura de un sistema de ayuda a la prevención de casos de violencia de género en España
Nombre del autor/a:	Javier Plo Moreno
Nombre del Tutor/a de TF:	Laia Subirats Maté
Nombre del/de la PRA:	Ferran Prados Carrasco
Fecha de entrega:	01/2023
Titulación o programa:	Máster en Ciencia de Datos
Área del Trabajo Final:	Area 3. Machine Learning (TFM-ML)
Idioma del trabajo:	Castellano
Palabras clave	Violencia de género, Prevención, Machine Learning

Resumen del Trabajo

La violencia de género es uno de los problemas más sangrantes que tenemos actualmente en la sociedad. Según el INE, (Instituto Nacional de Estadística) y otras fuentes gubernamentales, las víctimas de violencia de género en España aumentan cada año, así como la tasa de mujeres mayores de 14 años que son víctimas de dicha violencia.

Este proyecto propone la arquitectura de un sistema que ayude a detectar posibles situaciones de violencia de género construyendo perfiles detallados de los agresores y víctimas en base a casos anteriores, y mediante técnicas de Machine Learning, definir modelos que puedan ser aplicados a la población en general.

Este trabajo pretende sensibilizar sobre la problemática de la violencia de género, aumentar el conocimiento sobre los procesos que llevan a una persona no violenta a serlo en diferentes ámbitos, y a través de los modelos definidos, detectar con antelación posibles casos y relacionar posibles agresores y víctimas.

Este sistema podría permitir a las fuerzas de seguridad del estado y a los jueces determinar zonas geográficas y los momentos óptimos para llevar a cabo acciones concretas a nivel preventivo, así como a otros organismos e instituciones del estado para realizar acciones de formación y sensibilización en materia de violencia de género.

Se trata de una arquitectura verdaderamente ambiciosa, que puede involucrar a muchos interlocutores para obtener la información necesaria: informes policiales, expedientes judiciales, datos económicos y médicos, informes de servicios sociales, registro civil, redes sociales y datos de sistemas existentes como ATENPRO y VIOGEN.

Abstract

Gender violence is one of the bloodiest problems that we currently have in society. According to the INE, (National Institute of Statistics) and other government sources, victims of gender violence in Spain increase each year, as well as the rate of women over 14 years of age who are victims of such violence.

This project proposes the architecture of a system that helps to detect possible situations of gender violence by building detailed profiles of the aggressors and victims based on previous cases, and through Machine Learning techniques, define models that can be applied to the general population.

This work aims to raise awareness about the problem of gender violence, increase knowledge about the processes that lead a non-violent person to be so in different areas, and through the defined models, detect possible cases in advance and relate possible aggressors and victims.

This system could allow the state security forces and judges to determine geographical areas and the most optimal moments in which to carry out concrete actions at a preventive level, as well as other state agencies and institutions to carry out training and awareness actions on gender violence.

It is a truly ambitious architecture, which can involve many interlocutors to obtain the necessary information: police reports, judicial files, economic and medical data, reports from social services, civil registry, social networks, and data from existing systems such as ATENPRO and VIOGEN.

Agradecimientos

En primer lugar, quiero expresar mi agradecimiento a mi tutora Laia Subirats Maté, por el apoyo brindado y los buenos consejos que he recibido a lo largo de todo el proyecto.

Quiero extender mi agradecimiento a Verónica Dahl, profesora emérita en ciencias de la computación de la Simon Fraser University, y reconocida como una de las pioneras de la programación lógica, por haber contribuido a mi conocimiento sobre el papel de la mujer en el mundo digital y sobre la la violencia de género, y haber aportado ideas sobre como enfocar dicho problema desde la perspectiva del agresor.

También a Ramón López de Mántaras, profesor de investigación del IIIA-CSIC de Barcelona, por haber despertado en mí el interés por la inteligencia artificial, al asistir a sus clases de sistemas expertos en la Universidad Autónoma de Barcelona allá por los años 90, y ponerme en contacto con Verónica Dahl.

Por último, pero no menos importante, también quiero agradecer a mi familia el apoyo recibido.

Índice

Lista de Figuras

Lista de Tablas

1. Introducción

1.1. Contexto y justificación del Trabajo

Según el INE, (Instituto Nacional de Estadística – www.ine.es) el número de mujeres víctimas de violencia de género en España aumentó un 3,2% en el año 2021, hasta 30.141. y la tasa de víctimas de violencia de género fue de 1,4 por cada 1.000 mujeres de 14 y más años. Según epdata (www.epdata.es), en base a datos del Ministerio de la Presidencia, Relaciones con las Cortes e Igualdad, hasta septiembre del año pasado, en el último año, fueron asesinadas 34 mujeres en España por violencia de género.

Respecto a las motivaciones personales, no tengo ninguna experiencia laboral ni relación personal con el tema de la violencia de género, pero es obvio que se trata de un problema muy grave, que no solo afecta a víctimas y agresores, sino también a sus familias y entorno, por ejemplo, los hijos en el caso de violencia vicaria.

Sí tengo relación familiar estrecha con temas de bullying, LGTBI y procesos de transición de un sexo a otro, por lo que estoy especialmente sensibilizado con todo lo que tenga que ver con la falta de tolerancia y el maltrato físico y psicológico.

Ojalá no fuera necesario plantearse la idea de construir un sistema de prevención de este tipo. Significaría que la educación de base evita tanto buena parte de la violencia en general, como de la de género en particular.

Se propone la definición de la arquitectura de un sistema que ayude a la prevención de los casos de violencia de género en España mediante la detección previa de los posibles agresores y víctimas que puedan estar involucrados, ya que en este momento no existe ningún sistema similar en funcionamiento.

Par conseguirlo, definiremos perfiles de los agresores y las víctimas. Estos perfiles se basarán en información proveniente de:

- Bases de datos públicas
- Juzgados
- Servicios Sociales
- Policía
- Bomberos
- Ambulancias
- Datos económicos (Hacienda/Entidades Bancarias…)
- Situación Laboral
- Educación y Formación
- Tratamiento de patologías mentales

- Redes Sociales (Análisis de Sentimientos)
- …

En base a trabajos previos y estudios sociológicos sobre qué factores influyen en el hecho de que una persona que no es violenta en el entorno familiar/afectivo, pase a serlo en un momento determinado, se propondrá la ampliación de los datos que son recopilados en la tramitación de los casos de violencia de género, con el objetivo de construir un perfil lo más detallado posible tanto de agresores como de víctimas. A dicha ampliación se contribuirá también con un estudio en Twitter que permita identificar publicaciones relacionadas con la violencia de género.

Dichos perfiles permitirán la definición de modelos que son susceptibles de ser aplicados a grupos de población objetivo, con el fin de poder identificar posibles agresores y víctimas y posibles vínculos entre ellos.

El análisis de los perfiles y sus datos ayudarán tanto a identificar a posibles grupos de población objetivo, como a vincular posteriormente a los posibles agresores y víctimas.

De forma paralela, toda esta información puede permitir, por ejemplo:

- Definir mapas de calor de zonas en las que se pueden dar más casos, con el objetivo de realizar acciones concretas en dichas zonas como:
    - Reforzar vigilancia policial de forma general
    - Agudizar la sensibilización y formación sobre violencia y maltrato en la población
    - Seguimiento de casos puntuales

A continuación, se muestra un esquema de la arquitectura propuesta:

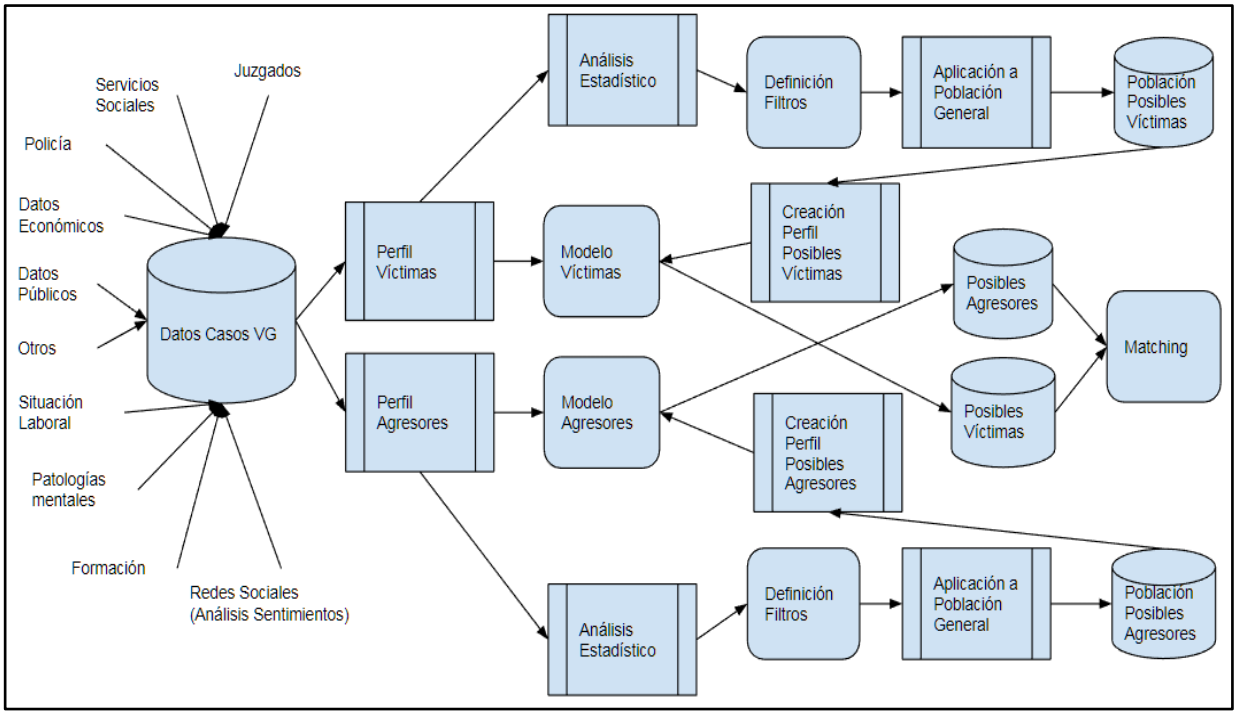


Figura 1: Arquitectura de la solución

Es importante hacer notar, que para llevar a la práctica esta propuesta, se requerirá de un análisis profundo sobre ética en el manejo de datos y leyes de protección de los mismos, y que es muy probable que conlleve la modificación o creación de leyes para permitir a los actores involucrados oportunos acceder a información personal de los ciudadanos: médica, económica, publicaciones públicas/privadas en redes sociales etc.

1.2. Objetivos del Trabajo

El objetivo principal del trabajo es:

- Definir de un sistema que ayude a la detección previa de casos de violencia de género mediante la colaboración de instituciones, organismos, organizaciones, etc. aportando la información necesaria para la creación de perfiles tanto de víctimas como de agresores.

Para conseguir este objetivo principal, abordaremos los siguientes objetivos secundarios

- Sensibilizar sobre la problemática de la violencia de género
- Obtener perfiles lo más detallados posible de agresores y víctimas, identificando los factores que desencadenan la aparición de dicha violencia.

- Realizar un análisis de datos provenientes de redes sociales que permita identificar publicaciones sexistas y topics relacionados con la violencia de género
- Construir modelos para víctimas y agresores, basados en la información proporcionada durante la evolución de los casos de violencia de género ya registrados
- Proporcionar filtros a aplicar sobre datos de población general para obtener poblaciones de posibles víctimas y agresores
- Proporcionar criterios de matching entre posibles víctimas y posibles agresores
- Analizar las implicaciones éticas que tiene el acceso a información personal en relación con la intimidad y a la Ley de protección de datos.

1.3. Impacto en sostenibilidad, ético-social y de diversidad

En la redacción de esta memoria se han considerado todas las aportaciones con independencia del género del autor o de la autora. Durante la fase de diseño se ha considerado si el resultado de dicho diseño es adecuado para todos los usuarios, con independencia de si son hombres o mujeres. Se ha definido un modelo tanto para posibles víctimas como para posibles agresores, sin que se hayan producido sesgos, estereotipos ni preeminencias masculinas ni en las imágenes ni en el discurso.

Entre las razones que motivan la realización de este trabajo no hay ninguna que tenga que ver con la sostenibilidad, sí fundamentalmente con el comportamiento ético, la responsabilidad social, la diversidad y los derechos humanos.

Entiendo que el único impacto negativo en cuanto a aspectos de sostenibilidad ambiental y/o huella ecológica, se centra en la necesidad de diseñar y ejecutar sistemas de recolección y procesamiento de la información, y de sistemas que utilizan dicha información para servir de entrada a algoritmos de Machine Learning, y que dichos sistemas van a utilizar ordenadores para poder conseguir los objetivos definidos, y por lo tanto se van a consumir cantidades indeterminadas de energía, que en este momento no es posible evaluar. En cualquier caso, durante el diseño de dichos sistemas, se ha velado por tratar de reducir la cantidad de información necesaria para llegar a los objetivos, y por tanto el tiempo y la cantidad de energía necesaria.

Respecto a la vertiente del comportamiento ético y de responsabilidad social, se ha de tener en cuenta que los sistemas anteriormente mencionados, van a tener que manejar información confidencial de las personas, por lo que el contenido de dicha información como de los resultados obtenidos, puede dar lugar a comportamientos poco éticos por parte de los propietarios.

También hay que tener en cuenta que el sistema propuesto va a "señalar" a personas como posibles víctimas y agresores, por lo que el tratamiento de esta información ha de ser estrictamente escrupuloso con el derecho a la intimidad de estas personas.

Por estos motivos, en este trabajo se ha incluido una sección específica sobre las implicaciones éticas que tiene el manejo de la información confidencial de las personas en proyectos de Data Science, y de qué formas se pueden intentar mitigar.

Un punto a comentar como posible impacto negativo es que, en la selección de posibles poblaciones objetivo de agresores, va ser muy difícil abstraerse del hecho de que la mayoría de agresores son hombres, principalmente porque los datos así lo indican. Tal como me indicaba una de las personas que aparecen en los agradecimientos, Verónica Dahl, sin ánimo de introducir ningún sesgo previo en cuanto al género de los agresores, y haciendo referencia a la violencia de género: "*La mayoría de los hombres no son violentos, pero es cierto que la mayoría de las personas violentas son hombres*"

Sucede algo parecido en el tema de la edad de dichas poblaciones objetivo, tanto de agresores, como de víctimas. Resulta evidente que podemos establecer una edad, por debajo de la cual, es muy poco probable encontrar un agresor. También es evidente que sí podemos encontrar en la realidad víctimas muy jóvenes, por ejemplo, en el caso de la violencia vicaria.

En cuanto a los impactos positivos en esa vertiente, podemos destacar el hecho de que se trata de un sistema que intenta solventar una situación existente con el objetivo de tratar de reducir los casos de violencia de género y por tanto la desigualdad, contribuyendo al bien común de la sociedad. Todo esto, afecta de forma positiva no solamente a las víctimas, sino también a todo su entorno (familiar, laboral, afectivo) y también a los posibles agresores, sobre los que se pueden aplicar acciones formativas y de sensibilización sobre comportamientos machistas y misóginos, bien sobre grupos de población concretos o sobre poblaciones de personas en zonas geográficas donde se prevea un número elevado de agresores y/o víctimas

Una de las motivaciones principales que han motivado la realización de este trabajo es la de sensibilizar sobre el problema de la violencia de género, por lo que sí existe una preocupación de responsabilidad social que se puede considerar como una de las razones para realizarlo.

Respecto a la dimensión diversidad, género y derechos humanos, y enlazando con la vertiente anterior, comentar que únicamente podemos destacar impactos positivos como: visibilizar el problema, proponer una posible solución preventiva de un problema que afecta fundamentalmente al colectivo de mujeres y niñas mayores de 14 años y a sus descendientes, en el caso de la violencia vicaria, o que afecta al desempeño del trabajo en el caso de la violencia de género que se produce en el ámbito laboral. También el hecho de

que los resultados obtenidos pueden dar pistas sobre en qué comunidades, provincias o zonas, son necesarias acciones de concienciación de base.

Hay que indicar que otra de las motivaciones principales para realizar este trabajo es la preocupación y afectación en el entorno personal propio en relación con los problemas de diversidad/género, y por consiguiente de respeto de los derechos humanos, que cada vez son más graves en nuestra sociedad.

Hay que comentar que todos estos impactos positivos son un efecto inherente de la propia solución y también pueden estar protagonizados por los posibles agresores y víctimas, que sin llegar a saber que lo pueden ser, reciban la información, y la formación suficiente, para no serlo nunca. También pueden estar protagonizados por la población de niños y niñas, sobre los que se debe centrar la formación sobre valores de diversidad, tolerancia y rechazo a la violencia, para tratar de evitar que puedan pasar en un futuro a engrosar las estadísticas de agresores o víctimas.

1.4. Enfoque y método seguido

Se ha aplicado la metodología CRISP-DM, que se considera como una de las metodologías más apropiadas para proyectos dedicados a extraer valor a los datos, y más concretamente para desarrollar un producto nuevo, como es el caso.

Consta de seis fases:

- **Entendimiento del negocio**
- **Comprender los objetivos y requisitos**
- **Entendimiento de los datos**
  - Análisis exploratorio de los datos
- **Preparación de los datos**
  - Construcción de un conjunto de datos definitivo
- **Modelado**
  - Realización de los modelos pertinentes
- **Evaluación**
  - Evaluar la calidad de los resultados y decidir cómo pueden explotarse
- **Despliegue**
  - Puesta en producción

Esta metodología pone al dato en el centro, siendo probable que los propios objetivos no aparezcan al comienzo del proceso, sino que vayan apareciendo en la fase de entendimiento de los datos a través del análisis exploratorio de los mismos.

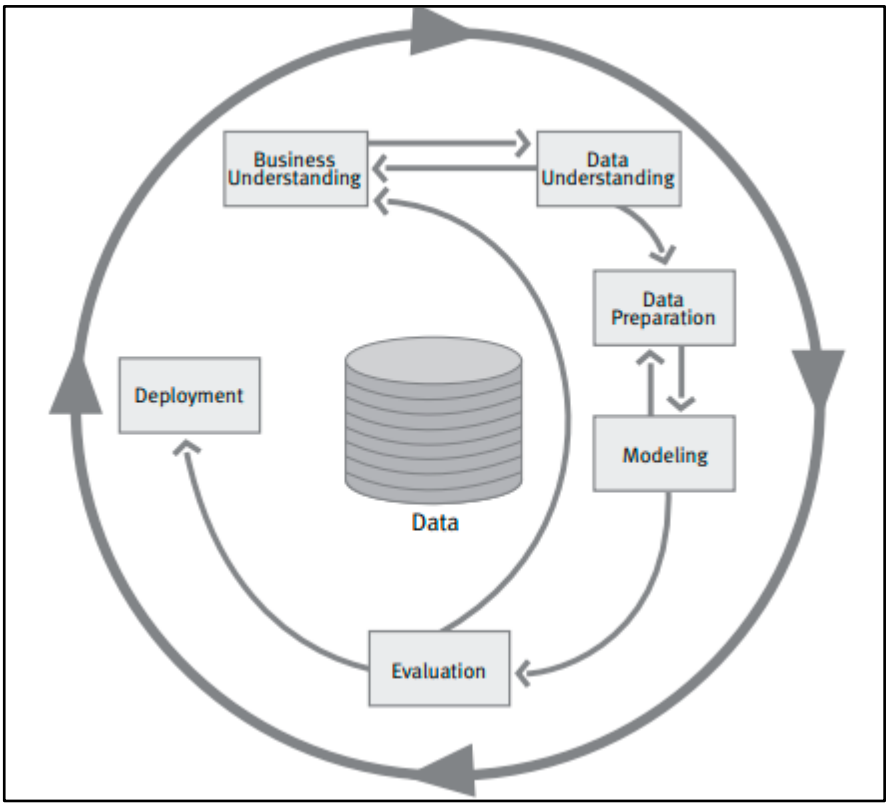


Figura 2: Esquema metodología CRISP-DM.

La fase de Business Understanding se corresponde con la investigación sobre la información disponible de los casos de violencia de género en España y en relación con los marcadores/factores que desencadenan que una persona pueda convertirse en agresor o en víctima. Dicha información se ha utilizado como conjunto de datos base del proyecto

La fase de Data Understanding se ha llevado a cabo mediante un análisis estadístico de los casos, seguida de la fase de Data Preparation, donde se han efectuado acciones de búsqueda de datos nulos, outliers etc.

En la fase de Modelado, en base a los marcadores detectados, se han definido los atributos y variables adicionales a los ya disponibles en el conjunto de casos utilizado como base, con el objetivo de poder definir perfiles/datasets de agresores y víctimas, y se han creado los modelos necesarios.

En la fase de Evaluación, se han ejecutado los modelos definidos sobre los datasets generados y mediante las métricas adecuadas se ha medido su efectividad.

La fase de Deployment queda fuera del alcance de este trabajo. Se trata de una propuesta de arquitectura muy ambiciosa, donde están involucradas muchas instituciones, organizaciones y empresas, y muy probablemente necesitará de un proyecto propio.

1.5. Planificación del trabajo

La planificación del proyecto se basa en siete fases fundamentales:

1. **Definición**
    a. En la que se realizan trabajos de análisis básico de las necesidades del proyecto, interés, relevancia, objetivos, motivación personal, planificación etc.
2. **Estado del arte**
    a. En esta fase se completa la fase anterior profundizando en la situación actual de los trabajos previos, publicaciones y trabajos de investigación para resolver la problemática en cuestión
    b. También se valorará la posible ampliación o modificación de la explicación de la temática escogida
3. **Diseño e implementación**
    a. En esta fase se realizan las tareas definidas en la planificación
4. **Redacción Memoria 1ª Entrega** (Borrador completo de la memoria)
5. **Redacción Memoria 2ª Entrega** (Memoria final)
6. **Preparación Defensa** (Vídeo Presentación)
7. **Defensa Pública**

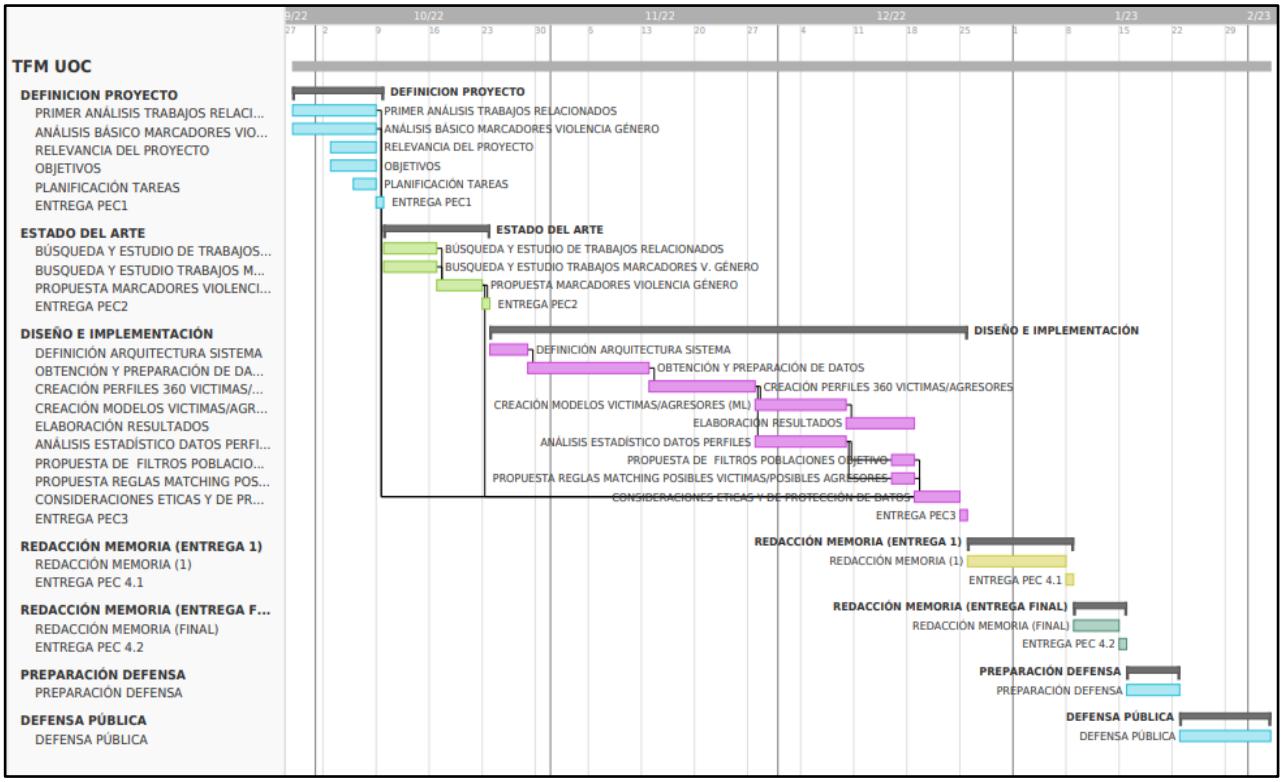


Figura 3: Gantt del proyecto.

1.6. Breve sumario de productos obtenidos

En la realización del presente trabajo se han obtenido los siguientes productos:

- Perfiles de víctimas y agresores y modelos a aplicar sobre poblaciones de posibles víctimas y agresores.
- Modelos para analizar el sexismo en redes sociales.
- Analizador de topics y comparador de documentos similares sobre información en redes sociales

2. Estado del arte

Después de realizar una búsqueda exhaustiva, no se tiene constancia de ningún trabajo que proponga la arquitectura de un sistema similar.

Como se verá seguidamente, la mayoría de las aproximaciones al problema se centran en predicciones a nivel numérico de la evolución de los casos en el tiempo, sobre estudios de términos relacionados con la violencia de género en Redes Sociales o mejoras sobre algún sistema predictivo ya existente en ámbitos muy concretos, pero no se ha encontrado ninguno que plantee la posibilidad de vincular a posibles agresores y víctimas con antelación.

Se va a enfocar este apartado en base a dos tipologías diferentes de estudios previos:

1. Estudios previos donde se aborden los factores que intervienen en la aparición de la violencia de género, tanto genéricos como concretos, para poder definir los atributos necesarios en los modelos tanto de víctimas como de agresores, y poder establecer criterios de filtro de las poblaciones objetivo de ambos tipos y posteriormente criterios de matching entre los posibles agresores y posibles víctimas

2. Estudios y soluciones previas donde se aborde el desarrollo de sistema orientados a la prevención de la violencia de género en los que se utilicen técnicas de Machine Learning. Esto servirá para tener información sobre los objetivos de dichos trabajos y para ayudar a seleccionar las técnicas a aplicar para la creación de los modelos mencionados.

En relación a la primera de las tipologías, tras realizar una búsqueda en PubMed con el término “gender violence”, se observa un incremento sostenido de publicaciones entre los años 2010 y 2021, con un descenso bastante acusado en 2022, tal como se aprecia en la figura 4

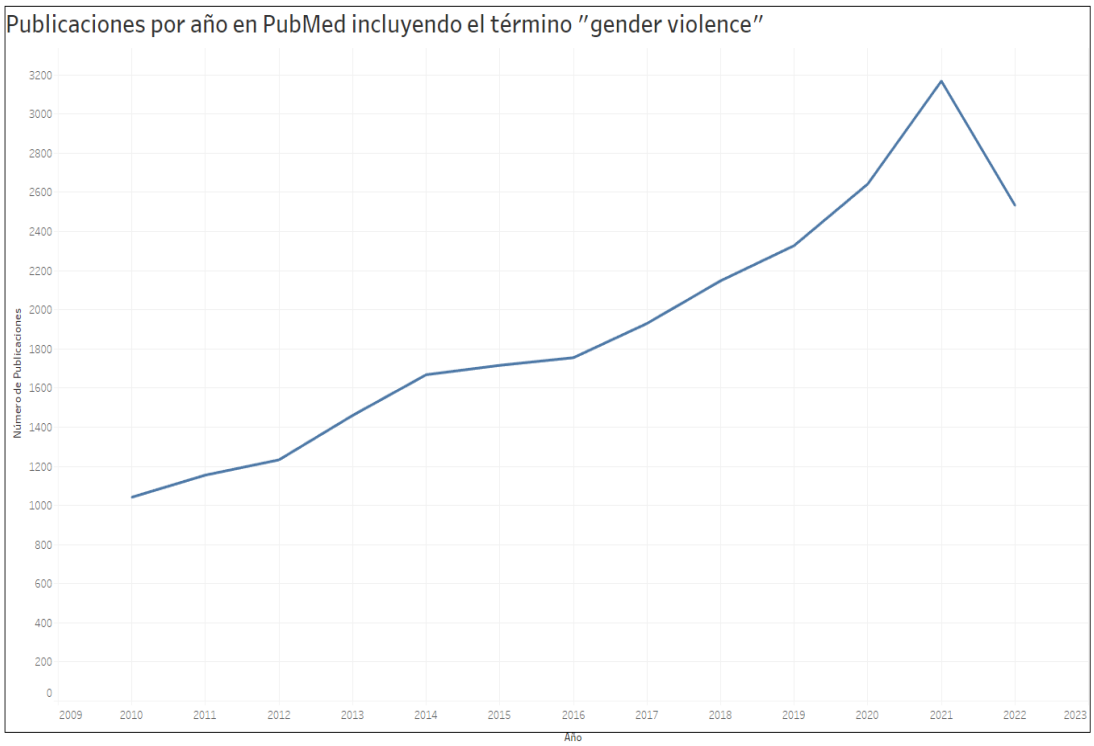


Figura 4. Recuento de publicaciones en PubMed incluyendo el término gender violence

Los artículos encontrados describen los factores bien de forma genérica como en esta publicación de ONU MUJERES [01], donde se categoriza la violencia contra las mujeres y las niñas en cuatro tipologías: **económica, psicológica, emocional, física y sexual** aportando información también sobre los indicios de maltrato en una relación como: acoso, miedo de la pareja, acceso limitado o ninguno a finanzas o la toma de decisiones, cambios en la personalidad y/o conducta.

En esta misma línea encontramos aportaciones como la de Gómez, C en la Revista Iberoamericana de Psicología [02], donde se categorizan estos factores como: biológicos, psicológicos, del contexto social inmediato y factores estructurales

Otras aportaciones como la de la OMS [03] entran más en detalle sobre factores relacionados con la violencia en la pareja y la violencia sexual, y que se producen a nivel individual, familiar, comunitario y social que interactúan entre sí: bajo nivel de formación, exposición al maltrato infantil, haber presenciado escenas de violencia, trastornos de personalidad antisocial, uso nocivo de alcohol, normas que otorgan privilegios o condiciones superiores a los hombres, escaso nivel de acceso a empleo remunerado en el caso de la mujer, antecedentes de violencia, conductas de control.

Algunas aportaciones como la de la Junta de Andalucía [04] se enfocan en la prevención de la violencia de género y protocolos de actuación en el ámbito educativo y en la adolescencia haciendo bastante hincapié en el ciberacoso, el control, el “grooming” o ciber extorsión. En este ámbito los

signos de violencia de género se suelen manifestar en forma de: aislamiento, baja autoestima, bajo rendimiento, entre otras.

Algunas páginas web, como la de la Delegación del Gobierno contra la Violencia de Género [05], identifica como primeros signos del maltrato respecto a la víctima: ignorar o despreciar sus sentimientos, ridiculizar insultar o despreciar a las mujeres en general, humillar, amenazar, agredir físicamente, aislar de la familia y/o amistades, forzar a mantener relaciones sexuales, control monetario y de decisiones, no permitir trabajar, celos, violencia vicaria, etc.

Dentro de esta web también se hace referencia a la violencia ejercida sobre mujeres con discapacidad intelectual, sobre las mujeres en el mundo rural y en el ámbito laboral. En este último se suele dar el acoso sexual y por razón de sexo, tal como se indica en una publicación de Cruz Roja sobre Violencia de género en el ámbito laboral [06], y que se corrobora en una publicación de los servicios de prensa de La Moncloa [07] donde se informa que casi la mitad de las mujeres que sufren acoso sexual en el trabajo, señalan a sus superiores varones como agresores.

De todos estos y otros estudios, se establecen como categorías de violencia para este trabajo las siguientes: **económica, psicológica, emocional, física, sexual, vicaria, ciber violencia** en los ámbitos familiar, afectivo y educativo, y podemos definir también una categoría de **violencia laboral**

En relación con la segunda de las tipologías, tras realizar una búsqueda en PubMed con el término “gender violence machine learning”, se observa que se produce un incremento entre los años 2018 y 2020 descendiendo gradualmente desde entonces, tal como se aprecia en la figura 5

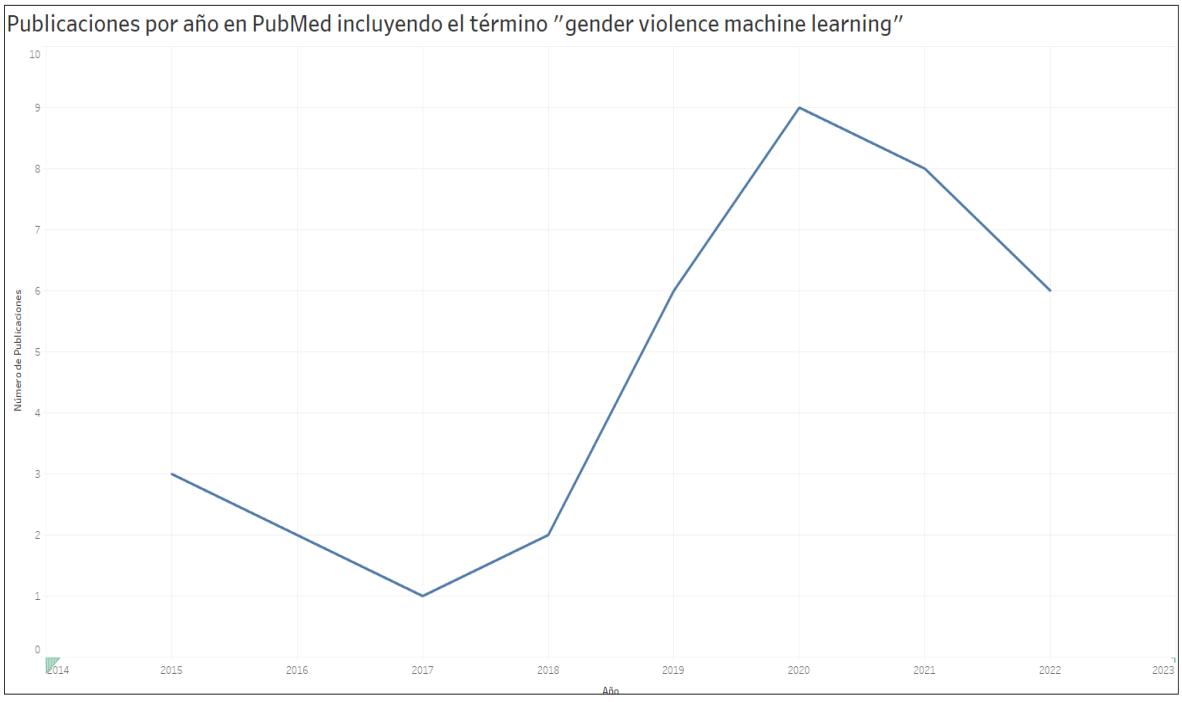


Figura 5. Recuento de publicaciones en PubMed incluyendo el término “gender violence machine learning”

En cualquier caso, vemos que el número de trabajos sobre técnicas de Machine Learning aplicadas al problema de la violencia de género, nunca ha sido elevado en la web utilizada, sin embargo, la misma búsqueda en Google sobre artículos académicos, arroja más de 172000 resultados y 408000 si utilizamos Google Académico.

Algunos de estos trabajos proponen métodos de forma general, y señalan problemas e implicaciones prácticas, como en el estudio de Gemma Ronzano [09], donde se definen dos metodologías de predicción fundamentales: clínicas y actuariales, estas últimas basadas en un número determinado de predictores. Para estas metodologías de apuntan problemas de falsos positivos y negativos, complejidad de predecir el comportamiento humano, aparición de patrones ilusorios que pueden no existir realmente, definir horizontes de predicción correctos. Estas metodologías implican invasión de la privacidad y el hecho de que se dé la posibilidad de señalar a ciertos individuos como posibles agresores, que realmente no tienen intención de cometer ninguna agresión.

En otras publicaciones como la de la doctora Ria Ivandic y colaboradores [10] se reflexiona sobre que técnicas de machine learning para evaluar riesgos en casos de abuso doméstico son más efectivas que los métodos convencionales, haciendo referencia al sistema DASH (Reino Unido) y a un estudio de Jeffrey Roger y colaboradores [11] donde se comparan estos métodos con los convencionales, utilizando los predictores de dicho sistema y estrategias como Random Forest.

En el trabajo de Richard A. Berk y colaboradores [12], se proponen una serie de predictores en un sistema de predicción que sirve como ayuda a los jueces para decidir si en una lectura de cargos, se deja a un delincuente en libertad provisional o no.

Una publicación reciente, y realmente interesante, es la de Colin Lecher en The Markup [13], donde se indica que los cuerpos de policía buscan sistemas de predicción de la violencia doméstica, haciendo referencia de nuevo al sistema DASH y VIOGEN (España). Este último funciona con un algoritmo que se basa en qué factores de un incidente se han relacionado con casos de alto riesgo en el pasado, en base a detalles registrados por la policía.

Se han encontrado también publicaciones muy completas como la de Ignacio Rodríguez-Rodríguez y colaboradores [14], donde se tratan la selección de los atributos mediante diferentes metodologías, los posibles métodos de predicción (Linear Regression / Support Vector Machines / Random Forest / Gaussian Processes), la selección de los datos y la población objetivo, para un sistema de predicción del número de denuncias por violencia de género.

Existe una gran variedad de trabajos específicamente orientados al uso de técnicas de Machine Learning en la detección de la violencia de género en diversos medios. Uno se los centros de referencia en este aspecto es The Data + Feminism Lab, del MIT. En una de sus publicaciones Catherine D'Ignazio y colaboradores [15] esbozan un sistema que automatiza parcialmente la detección de feminicidios para organizaciones de la sociedad civil y activistas que están monitoreando el fenómeno.

Existen publicaciones para identificar y clasificar lenguaje de odio y misógino en Twitter, como la de María Anzovino y colaboradores [16], donde se trabaja con diferentes categorías de misoginia y diferentes técnicas de machine learning (Support Vector Machines / Random Forest / Naïve Bayes / Multilayer Perceptron Neural Network). Relacionada con esta publicación tenemos la de Sarah Hewitt y colaboradores [17] donde se proponen algunos términos a analizar en Twitter directamente relacionados con la misoginia.

Otro trabajo interesante a este respecto es el de Teresa Piñeiro-Otero y Xabier Martínez-Rolán [18], donde se hace un estudio intensivo del discurso de odio contra las mujeres en diferentes medios y en particular sobre insultos sexistas y misóginos

En otras publicaciones como la de Carlos M. Castorena y colaboradores [19], se utiliza Deep Learning para detectar la violencia de género en Twitter, o la de Mohammed Ali Al-Garadi y colaboradores [20], donde se utilizan modelos tradicionales de machine learning, deep learning, y transformer based (BERT y RoBERTa)

Hay que destacar también una interesante publicación de Angel González-Prieto y colaboradores [21], donde se propone una metodología basada en técnicas de ML para ayudar a las autoridades involucradas en la política contra el crimen. En concreto, se aplica al manejo de un conjunto de datos oficial, VIOGEN, dependiente del Gobierno de España, para evaluar el riesgo de revictimización utilizando un modelo híbrido estocástico.

Algunas publicaciones como la de George Karystianis y colaboradores [22] proponen el uso de información policial enlazada con datos del ministerio de salud, aplicando arquitecturas Deep Learning.

Enlazando con sistemas de predicción basados en información policial, tenemos el ya mencionado DASH y un trabajo interesante de Emily Turner y colaboradores [23] donde analizan en profundidad dicho sistema y el cuestionario utilizado por la policía para recabar la información en los casos [24].

En base a la información recopilada en este apartado, podemos considerar dos grupos de factores que influyen en la aparición de la violencia de género, uno relacionado con los agresores y otro relacionado con las víctimas

Factores agresores

- Bajo nivel cultural
- Exposición a maltrato infantil
- Trastornos de personalidad antisocial
- Antecedentes de violencia
- Problemas económicos
- Alcoholemia
- Comportamientos dañinos (múltiples parejas, actitudes de aprobación de la violencia)
- Leyes discriminatorias

Factores víctimas

- Bajo nivel cultural
- Ausencia de empleo remunerado
- Exposición a maltrato infantil
- Tratamientos psiquiátricos (ansiedad/depresión/baja autoestima)
- Antecedentes de violencia
- Alcoholemia
- Leyes discriminatorias

Aparte de estos factores, se deben tener en cuenta otros datos importantes, como los que aparecen en el Boletín estadístico anual (2020) de la Delegación del Gobierno contra la violencia de Género [08] para ambos grupos como: ubicación geográfica, fecha, nacionalidad, edad, discapacidad intelectual, estado civil (Tipo de relación/convivencia), custodia de los hijos, medidas de alejamiento, publicaciones y vínculos en redes sociales donde poder detectar amenazas y términos relativos a la violencia de género, denuncias previas en juzgados, policía -mediante los formularios (VPR) (VPER)-, intervenciones de otros servicios públicos como 016, y sistemas como ATENPRO [25] y VIOGEN [26], ambulancias, bomberos, etc.

Toda esta información se utilizará para definir una serie de atributos que configurarán el perfil de agresores y de víctimas, poder definir filtros a aplicar sobre la población general y obtener poblaciones de posibles agresores y víctimas a los que aplicaremos los modelos obtenidos, y finalmente para definir criterios de matching entre estas poblaciones y tratar de vincular a sus miembros.

Como hemos comentado al inicio de esta sección, no se ha encontrado ninguna propuesta que plantee un sistema similar al que propone este trabajo. Lo cierto es que se trata de una propuesta muy ambiciosa, y que el presente trabajo abarca fundamentalmente una parte de la arquitectura, la relativa a los modelos de agresores y víctimas, y la de análisis en Twitter de términos relacionados con la violencia de género.

Indicar también, que hay muy pocos conjuntos de datos públicos en relación con la violencia de género en España, con volúmenes pequeños de información y con pocos atributos y variables como para poder abordar parte de la arquitectura propuesta.

3. Materiales y métodos

Como se menciona en el apartado 1.4 del presente documento, la metodología usada para el desarrollo de este proyecto ha sido CRISP-DM. Por lo tanto, los apartados que se presentan a continuación para describir el desarrollo del proyecto siguen las fases definidas por esta metodología.

En concreto, en este apartado, se describen los pasos para la obtención y comprensión de los tres conjuntos de datos utilizados:

- Conjunto de datos de casos de violencia de género en España tomado como base
- Conjuntos de datos (perfiles) generados para víctimas y agresores
- Conjunto de tweets proveniente de redes sociales para detección y categorización del sexismo

3.1. Conjunto de datos de casos de violencia de género

El dataset elegido contiene información sobre casos de mujeres asesinadas en España, por sus parejas o exparejas, entre los años 2003 y 2017. Los datos provienen de la web de estadística de la violencia de género en España [27] y se encuentran disponibles en Data World (Domestic Violence in Spain) en un perfil de Pablo Sáenz de Tejada [28]

El conjunto de datos está formado por 10 características y 900 sucesos (registros) se puede ver en la tabla 1

Característica	Tipo
Year	Integer
Month	Text
Autonomous Comunity	Text
Province	Text
Relation	Text
Victim age (interval)	Text
Agressor age (interval)	Text
Previous Abuse Report	Text
Living Together	Text
Victims	Integer

Tabla 1. Características conjunto de datos violencia de género

A partir de este conjunto de datos, se plantea la problemática de determinar qué variables influyen más sobre el hecho de que ya se hubiera producido algún abuso previo, mediante el análisis de las correlaciones entre el hecho de reportes de abusos previos y otras características que definen el suceso, la realización de contrastes de hipótesis que nos proporcionen relaciones interesantes inferidas de los datos de la población, como por ejemplo la cohabitación, las edades, etc. Se estima que el hecho de que ya se hubiera producido algún abuso previo, es un factor determinante en el hecho de que dicho abuso se convierta en agresión y que esta pueda llevar al asesinato.

Tanto el tratamiento de los datos como las actuaciones comentadas se han realizado utilizando el lenguaje R (R-Studio)

3.1.1. Tratamiento y análisis de los datos

Una vez cargado el conjunto de datos se procede a comprobar la existencia de ceros y elementos vacíos y a la identificación de outliers.

De cara a poder realizar un estudio de la homogeneidad y la normalidad de la varianza, se factorizan y convierten a valores numéricos las variables cualitativas a cuantitativas categóricas.

Posteriormente se definen un conjunto de grupos de datos que puedan resultar interesantes de analizar:

- Agrupación por Comunidad Autónoma con temperaturas muy elevadas en verano
  - Andalucía, Aragón, Región de Murcia y Castilla La Mancha
  - Meses: Junio, Julio y agosto
- Agrupación por agresores jóvenes (16-17 años, 18-20 años)
- Agrupación por agresores adultos (21-30 años, 31-40 años, 41-50 años, 51-64 años)
- Agrupación por agresores mayores (65-74 años, 74-85 años, > 85 años)
- Agrupación por Convivientes
- Agrupación por Relation

Seguidamente se comprueba la homogeneidad y normalidad de la varianza utilizando la prueba de normalidad de Shapiro. Se comprueba si el p-valor es superior al nivel de significación prefijado alfa = 0.05. Si esto se cumple, entonces se considera que variable en cuestión sigue una distribución normal

Posteriormente se pasa a estudiar la homogeneidad de varianzas, aplicando el test de Levene entre la variable víctimas y el resto.

3.1.2. Estudio de la correlación

Al tratar variables cuantitativas politómicas y nominales, por lo que la prueba Chi-squared resulta adecuada en algunos casos, y el test exacto de Fisher en otros para valorar la independencia.

La edad de los agresores la consideramos nominal, ya que no las vamos a utilizar estableciendo relaciones de tipo mayor/menor, ni vamos a evaluar distancias entre los diferentes rangos de edades

3.1.3. Modelado y evaluación

Se han planteado modelos de regresión logística multinomial que permitan trabajar con los grupos de datos construidos anteriormente, donde se haya encontrado correlación entre las variables que los componen. Una vez aplicados los modelos se han evaluado los odds ratio.

La regresión multinomial se utiliza para explicar la relación entre una variable nominal dependiente, Previous Abuse Report en este caso, y una o más variables independientes, y es una extensión de la regresión logística binomial. El algoritmo nos permite predecir una variable dependiente categórica que tiene más de dos niveles. Como cualquier otro modelo de regresión, la salida multinomial se puede predecir usando una o más variables independientes. Las variables independientes pueden ser de tipo nominal, ordinal o continuo.

3.2. Conjuntos de datos de víctimas y de agresores

3.2.1. Descripción de los conjuntos de datos

Tal como se ha indicado en el punto 2, no existen conjuntos de datos públicos con un volumen de datos destacable ni con un conjunto de características, que puedan ser utilizados de forma efectiva - salvo para tener una visión estadística del problema- en un sistema de predicción de la violencia de género basado en datos, informaciones y hechos que hacen que una persona pueda llegar a convertirse en agresor o en víctima.

Se ha considerado que para crear los conjuntos de datos de víctimas y agresores no son relevantes los atributos temporales (mes-año), de ubicación (Comunidad autónoma y Provincia), ya que ni el lugar ni el momento en el que suceden los hechos (estrictamente temporalmente hablando) hacen que alguien se pueda convertir en agresor o víctima.

Tampoco se han considerado los atributos relación y convivencia, ya que se han definido conjuntos de datos y modelos separados para víctimas y agresores, donde estos dos campos carecen de sentido al reflejar una situación común a ambos.

De cualquier forma, esta información descartada del conjunto de datos base, será recabada durante la toma de datos en los casos de violencia de género, ya que posteriormente será de utilidad tanto para aplicarla en los criterios de selección de poblaciones objetivo, como en el proceso de vinculación de posibles agresores y víctimas.

En base a la investigación realizada en el punto 2, se han definido los siguientes marcadores (atributos) y variables para poder construir los conjuntos de datos de víctimas y agresores, tomando como base las características del conjunto citado en el punto 3.1, y que se pueden ver en las figuras 2 y 3.

Atributo	Descripción	A	V
Age	Edad	X	X
LCL	Bajo nivel cultural	X	X
Unemployed	Desempleo	X	X
ECHA	Exposición a abusos en la infancia	X	X
PsychoTreatment	Tratamiento psicológico/psiquiátrico	X	X
Addictions	Adicciones	X	X
SharedCustody	Custodia compartida de hijos	X	X
Marital_Status	Estado civil	X	X
Distancing_Measures	Medidas de alejamiento	X	X
PoliceReports	Atestados policiales violencia género	X	X
Previous_abuse_report	Abusos previos	X	X
IntelDissab	Discapacidad intelectual		X
ATENPRO	Aparición en sistema ATENPRO	X	X
VIOGEN	Aparición en sistema VIOGEN	X	X
FireArms	Posesión de armas de fuego	X	
EchoProblems	Problemas económicos	X	
SexismPresence	Sexismo en redes sociales	X	

Tabla 2. Marcadores conjuntos víctimas y agresores

Variable	Descripción	A	V
EchoViolence	Violencia económica	X	X
EmotionalViolence	Violencia emocional	X	X
PhisycalViolence	Violencia física	X	X
SexualViolence	Violencia sexual	X	X
VicariousViolence	Violencia vicaria	X	X
CiberViolence	CiberViolencia	X	X
WorkPlaceViolence	Violencia laboral	X	X

Tabla 3. Variables conjuntos víctimas y agresores

Al no disponer de datos reales que consten de los atributos y variables indicados, se han tomado como base los 900 registros del conjunto base, y se han completado hasta obtener dos conjuntos de 10000 registros, donde se han generado de forma random los datos no disponibles tanto de atributos como variables.

Se ha optado por incrementar el número de registros de forma artificial debido al hecho de que los 900 casos registrados se refieren a asesinatos, no a denuncias o casos que se puedan recopilar en un futuro y que nos permitan disponer de información histórica relativa a los atributos y variables propuestos.

Finalmente, y después de realizar operaciones de "dumming" en los atributos *Previous abuse report y marital status*, este último generado de forma random de entre los siguientes valores: single, married, separated, divorced, widow y factunion, el conjunto de datos de víctimas tiene 29 columnas, 22 atributos y 7 variables, y el conjunto de datos de agresores tiene 31 columnas, 24 atributos y 7 variables. Todos los atributos y variables son binarios, a excepción de Age que es un numérico.

Previamente a la creación de los conjuntos de training y test, se ha procedido a realizar una comprobación de que los datos están balanceados y a un escalado de los mismos (StandarScaler).

En este punto se introduce el concepto de **vector de violencia**, que se puede definir como el conjunto de valores que toman las variables que categorizan los tipos de violencia definidos para cada posible agresor o víctima.

3.2.2. Creación de modelos

Se está claramente ante un problema de clasificación multilabel, donde a diferencia de una clasificación multiclase donde cada registro puede ser etiquetado con una única clase del conjunto de n clases posibles, se puede etiquetar cada registro con 0 o n clases. Es decir, en un problema multiclase, las clases son mutuamente excluyentes, mientras que en un problema multilabel cada etiqueta representa una tarea diferente de clasificación, pero relacionadas entre sí. Es razonable pensar que en un caso real se puedan dar diferentes tipos de violencia simultáneamente y que estén relacionadas entre sí.

Por otra parte, al existir diferentes aproximaciones para abordar un problema de clasificación multilabel, se ha optado por aplicar diferentes métodos y estrategias, utilizando la librería de Python scikit-learn.

Como se puede ver en [29] y en [30] se dispone tanto de algoritmos que soportan la clasificación multilabel de forma nativa, como de otros específicamente desarrollados para este tipo de clasificación (Multilearn Adapt), o bien aplicando aproximaciones que transforman problemas de clasificación multilabel en problemas de clasificación single label, como se puede ver en [31] o utilización del Deep Learning como se puede ver en [32].

Tomando como premisa que cualquier persona puede llegar a ser tanto víctima o agresor, ambos modelos asignarán uno o más tipos de violencia a cualquier miembro de las poblaciones objetivo, lo que no implica, que en el momento de seleccionar dichas poblaciones, no se tengan en cuenta criterios estadísticos o de cualquier otro tipo para establecer filtros en dichas selecciones.

Para la creación de los modelos de víctimas y agresores se han utilizado las estrategias/algoritmos de clasificación de la tabla 4

Tipo/Estrategia	**Algoritmo**
Multilabel nativo	Random Forest
	Multilayer Perceptron
	Decission Tree
	Nearest Neighbours
Multilabel no nativo (Multioutput)	Extended Gradient Boosting
	Logistic Regression
	Gaussian Naive Bayes
	Linear SVC
Problem Transformation	Binary Relevance
	Classifier Chains
	Label Powerset
MultiLearn Adapt	MLkNN
	BRkNNaClassifier
	MLTSVM
Multilabel Deep Learning	Problem Transformation con BinaryRelevance (Keras)

Tabla 4. Estrategias/algoritmos de clasificación víctimas y agresores

En todos los casos se ha utilizado GridSearchCv para realizar búsquedas exhaustivas de parámetros y F1 como métrica, aunque también son comunes métricas como Hamming loss o Jaccard similarity, como se indica en [33] y en [34]

3.3. Conjunto de datos proveniente de redes sociales

3.3.1. Detección y clasificación del sexismo

El Oxford English Dictionary define el sexismo como “prejuicio, estereotipo o discriminación, generalmente contra las mujeres, por motivos de sexo”. La desigualdad y la discriminación contra las mujeres que permanecen arraigadas en la sociedad se replican cada vez más en las redes sociales.

Detectar el sexismo en línea puede ser difícil, ya que puede expresarse de formas muy diferentes. El sexismo puede sonar “amigable”: la afirmación “Las mujeres deben ser amadas y respetadas, trátalas siempre como un vaso frágil” puede parecer positiva, pero en realidad está considerando que las mujeres son más débiles que los hombres.

El sexismo puede sonar “gracioso”, como es el caso de los chistes o el humor sexista (“Hay que amar a las mujeres… solo eso… Nunca las entenderás”).

El sexismo puede sonar "ofensivo" y "odioso", como en "Humíllate, exponte y degrádate como la puta perra que eres si quieres que un hombre de verdad te preste atención".

En este caso, el objetivo es la detección del sexismo en un sentido amplio, desde la misoginia explícita hasta otras expresiones sutiles que implican conductas sexistas implícitas.

Sin embargo, incluso las formas más sutiles de sexismo pueden ser tan perniciosas como las más violentas y afectar a las mujeres en muchas facetas de sus vidas, incluidas las funciones domésticas y de crianza, las oportunidades profesionales, la imagen sexual y las expectativas de vida, por nombrar algunas.

La identificación automática de sexismos en un sentido amplio puede ayudar a crear, diseñar y determinar la evolución de nuevas políticas de igualdad, así como fomentar mejores comportamientos en la sociedad, y en nuestro caso contribuir a la definición del sistema que se aborda en este trabajo.

Para la identificación y categorización del sexismo en redes sociales, se han utilizado los conjuntos de datos públicos de **EXIST: sEXism Identification in Social neTworks** [35], es decir, se ha simulado participar en el reto EXISTS2021 y se han realizado las siguientes tareas que se indican a continuación.

3.3.1.1. Identificación del sexismo

Esta tarea es una clasificación binaria. Los modelos tienen que decidir si un texto dado (tweet o gab) es o no sexista (es decir, es sexista en sí mismo, describe una situación sexista o critica un comportamiento sexista). Los siguientes tuits muestran ejemplos de mensajes sexistas y no sexistas:

Sexista: *"Mujer al volante, tenga cuidado"*

No Sexista: "*Alguien me explica que zorra hace la gente en el cajero que se demora tanto*"

3.3.1.2. Categorización del sexismo

Una vez un mensaje ha sido clasificado como sexista, se ha de categorizar mediante clasificación multiclase de acuerdo con los siguientes tipos:

- **Ideológico y desigualdad**: El texto desacredita el movimiento feminista, rechaza la desigualdad entre hombres y mujeres, o presenta a los hombres como víctimas de la opresión de género. Ejemplo: *"Mi hermana y mi madre se burlan de mí por defender todo el tiempo los derechos de todos y me acaban de decir feminazi, la completaron".*

- **Estereotipos y dominancia**: El texto expresa falsas ideas sobre la mujer que sugieren que son más adecuadas para cumplir determinados roles (madre, esposa, cuidadora de la familia, fiel, tierna, amorosa, sumisa, etc.), o inapropiadas para determinadas tareas (conductora, trabajadora, etc.), o afirma que los hombres son de alguna manera superiores a las mujeres. Ejemplo *"A las mujeres hay que amarlas…solo eso… Nunca las entenderás.".*

- **Objetificación**: El texto presenta a las mujeres como objetos al margen de su dignidad y aspectos personales, o asume o describe ciertas cualidades físicas que las mujeres deben tener para cumplir con los roles tradicionales de género (cumplimiento de los estándares de belleza, hipersexualización de los atributos femeninos, cuerpos de mujeres a disposición de los hombres, etc.). Ejemplo: *"Pareces una puta con ese pantalón" - Mi hermano de 13 cuando me vio con un pantalón de cuero".*

- **Violencia sexual**: Se realizan sugerencias sexuales, solicitudes de favores sexuales o acoso de carácter sexual (violación o agresión sexual). Ejemplo *"#MeToo Estas 4 no han obtenido su objetivo. El juez estima que se abrieron de patas"*

- **Misoginia y violencia no sexual**: El texto expresa odio y violencia hacia la mujer. Ejemplo *"Las mujeres de hoy en día te enseñaran a querer… estar soltero"*

3.3.1.3. Conjunto de datos sexismo

El conjunto de datos EXIST-2021 final consta de 6977 tweets para entrenamiento y 3386 tweets para test en inglés y castellano, donde ambos conjuntos se seleccionaron aleatoriamente de los 9000 y 4000 registros de los conjuntos etiquetados inicialmente, de entrenamiento y test respectivamente, para garantizar el equilibrio de clases de acuerdo con la tarea de identificación.

Además, se añadieron 492 "gabs" en inglés y 490 en español de la red social sin censura Gab.com siguiendo un procedimiento similar al descrito anteriormente. Este conjunto se incluyó en el conjunto de prueba EXIST para medir la diferencia entre las redes sociales con y sin "control de contenido", Twitter y Gab.com respectivamente.

3.3.1.4. Preprocesado y funciones de limpieza y transformación de tweets

Dentro del procesamiento del lenguaje natural existe el concepto de término, que se define como el conjunto de una o más palabras.

Otro concepto importante es el de token, que se define como la unidad textual mínima procesada que no siempre se corresponde con un término.

El tokenizer es una herramienta básica del procesamiento del lenguaje natural que lista los tokens de un texto. El tokenizador puede llegar a distinguir como tokens, símbolos que no son términos. Por este motivo se han definido diferentes funciones de limpieza y transformación de los tweets tales como de conversión a minúsculas, eliminación de: usernames, hashtags, links, números, puntuación y espacios, y palabras de una sola letra.

El concepto de token va ligado directamente con el de n-grama, que se define como una secuencia de tokens consecutivos que tiene un orden de complejidad n, donde n es el número de tokens consecutivos del n-grama

Se han filtrado los n-gramas con stop words, palabras que tienen como función cohesionar un texto, pero que no aportan significado alguno al texto, ni al n-grama: artículos, preposiciones, etc.

De forma general se ha utilizado la librería NLTK, que es una plataforma para crear programas de Python con el fin de trabajar con datos de lenguaje humano. Proporciona

interfaces fáciles de usar para más de 50 corpus y recursos léxicos como WordNet, junto con un conjunto de bibliotecas de procesamiento de texto para clasificación, tokenización, lematización, etiquetado, análisis y razonamiento semántico. Para la tokenización se ha utilizado el método word_tokenize de esta librería

3.3.1.5. Normalización

Una tarea importante del proceso de exploración de los datos es la normalización del texto, que consiste en aglutinar las referencias al mismo concepto en uno solo, por ejemplo, términos como libro y libros son variantes de una forma que aglutina las referencias al concepto de libro.

Una tarea importante dentro del proceso de normalización es la de stemming, que consiste en quitar y remplazar los sufijos de la raíz de la palabra. Otra tarea importante es la de lematización, algo más compleja, e implica hacer un análisis del vocabulario y su morfología para retornar la forma básica de la palabra (sin conjugar, en singular, etc).

Para el stemming se ha utilizado el SnowballStemer de la librería NLTK. Cada idioma tiene sus reglas, en inglés la librería NLTK utiliza el algoritmo de stemming de Porters, en español toma algunas reglas que, en resumen, eliminan diversos sufijos que se atribuyen a acciones (ar, er, ir, ía, en, es, etc), terminaciones plurales, géneros y otros.

3.3.1.6. Vectorización

La vectorización consiste en representar de una forma computacionalmente tratable los términos que son relevantes en un texto. La representación vectorial es la más utilizada.

Una vez se han realizado las operaciones de filtrado, transformación y normalización se ha aplica la vectorización utilizando la función CountVectorizer de la librería scikit-learn de python

3.3.1.7. Creación de modelos y evaluación

Tanto los modelos de clasificación binaria para la identificación del sexismo, como los de clasificación multiclase para la tarea de categorización, han sido evaluados con las siguientes métricas: accuracy, precisión, recall y f1-macro (requerido así en EXIST2021).

Tal como se indica en [33] algunos algoritmos son capaces de manejar múltiples clases como Logistic Regression, Random Forest o Naive Bayes, otros como clasificadores SGD o SVM son estrictamente binarios.

También existen estrategias que permiten realizar clasificaciones multiclase con multiples clasificadores binarios, como OvR (one versus rest) o como OvO (one versus one)

La estrategia OvR consiste en crear un clasificador por clase. Para cada clasificador, la clase se compara con todas las demás clases. Además de su eficiencia computacional (solo se necesitan n_clases clasificadores), tiene como ventaja su interpretabilidad. Dado que cada clase está representada por un único clasificador, es posible obtener conocimiento sobre la clase al inspeccionar su clasificador correspondiente. Esta es la estrategia más utilizada.

La estrategia OvO construye un clasificador por cada par de clases. En el momento de la predicción, se selecciona la clase que recibió más votos. En caso de empate (entre dos clases con el mismo número de votos), selecciona la clase con la mayor confianza de clasificación agregada sumando los niveles de confianza de clasificación por pares calculados por los clasificadores binarios subyacentes. Se trata de un método más lento que el anterior.

Los clasificadores de este punto han sido implementados también utilizando la librería scikit-learn. Dicha librería detecta si se intenta utilizar un clasificador binario para una tarea de clasificación multiclase y automáticamente ejecuta OvR o OvO dependiendo del algoritmo. Se pueden consultar los diferentes algoritmos OvR y OvO en [29]

De forma general se ha utilizado una estragegia de pipeline mediante el método Pipeline de la librería scikit-learn donde primero se aplica la vectorización, utlilizando el método CountVectorizer de la citada librería, y posteriormente el clasificador elegido. El vectorizador base utilizado tiene como parámetros analyzer y tokenizer.

Esta estrategia se ha aplicado a una búsqueda intensiva de parámetros mediante el método GridSearchCv de la misma librería utilizando los parámetros: max_df, min_df, max_features y ngram_range. Los parámetros de cada clasificador dependen del método de clasificación utilizado.

Una vez determinados los mejores valores de los parámetros para cada estimador, se ha procedido a la creación de los modelos correspondientes en base a cada tarea: Identificación (clasificación binaria) o Categorización (clasificación multiclase). La tabla 5 muestra os scorings aplicados a cada clasificador.

Tarea	Clasificador	Scoring
Identificación	SGDClassifier	Accuracy
	LinearSVC	Accuracy
Categorización	LogisticRegression()	F1-macro
	DecisionTreeClassifier()	F1-macro
	SVC()	F1-macro

Tabla 5. Tareas, clasificadores y scorings utilizados

Todos los modelos han sido entrenados y testeados con los tweets en castellano.

3.3.2. Topic Modelling

El topic modeling es una técnica no supervisada de NLP, capaz de detectar y extraer de manera automática relaciones semánticas latentes de grandes volúmenes de información.

Estas relaciones son los llamados topics, que son un conjunto de palabras que suelen aparecer juntas en los mismos contextos y nos permiten observar relaciones que seríamos incapaces de observar a simple vista.

Existen diversas técnicas que pueden ser usadas para obtener estos topics. El principal algoritmo es el modelo *latent dirichlet allocation* (LDA), propuesto por David Blei en 2011 [36], que nos devuelve por un lado los diferentes topics que componen la colección de documentos y por otro lado cuánto de cada topic está presente en cada documento.

Los topics consisten en una distribución de probabilidades de aparición de las distintas palabras del vocabulario.

Se ha utilizado esta técnica para obtener topics a partir de los tweets tanto de training como de prueba indicados en el apartado 3.3.1 y posteriormente se ha aplicado métrica de distancia de Jensen-Shannon a las distribuciones de los topics del modelo LDA obtenido previamente.

Se trata de una medida de distancia estadística entre distribuciones de probabilidad. Esta métrica devuelve un valor comprendido entre 0 y 1, donde cuanto menor sea este valor significa una mayor similitud entre las dos distribuciones.

Para poder aplicar todas estas técnicas se ha utilizado la librería Gensim de python [37]. Se trata de una librería open source que permite la representación de documentos en formato de vectores semánticos y está diseñada para procesar textos digitales sin estructura ("texto sin formato") utilizando algoritmos de aprendizaje automático no supervisados, por lo que solo se necesita un corpus de documentos.

Los algoritmos en Gensim, como Word2Vec, FastText, Latent Semantic Indexing (LSI, LSA, LsiModel), Latent Dirichlet Allocation (LDA, LdaModel), etc., descubren automáticamente la estructura semántica de los documentos examinando patrones estadísticos de coocurrencia dentro de un corpus de documentos de entrenamiento.

Una vez que se encuentran estos patrones estadísticos, cualquier documento de texto sin formato (oración, frase, palabra...) se puede expresar sucintamente en la nueva

representación semántica y se puede consultar la similitud tópica con otros documentos (palabras, frases...).

3.3.2.1. Creación del modelo y evaluación

Para poder aplicar el modelo LDA se debe identificar que elemento simboliza un documento en el contexto del problema a resolver, para comprender que es lo que forma parte del corpus. En este caso cada uno de los tweets representa un documento, por lo que el corpus será la unión de todo el texto de todos los tweets del apartado 3.3.1

Una vez definido el corpus, se ha procedido a la normalización del texto de los tweets, de forma similar a la indicada en el apartado 3.3.1 con procesos de limpieza, tokenización y stemming.

A partir del texto procesado, se ha creado también un diccionario, que asigna un identificador numérico a cada palabra única y que se usa para obtener el identificador a partir de la palabra y viceversa.

Posteriormente se ha inicializado el corpus en base al diccionario creado, y se ha obtenido una bolsa de palabras (Bag of Words) [38] con las frecuencias de aparición de cada palabra.

Tras aplicar esta técnica, cada documento está representado como una lista de tuplas donde el primer elemento es el identificador numérico de la palabra y el segundo es el número de veces que esa palabra aparece en el documento.

Documentos (Tweets)

D1: He is a lazy boy. She is also lazy.

D2: This dog is very lazy.

Transformación vectorial

Bolsa de Palabras

	he	lazy	boy	she	dog	very
D1	1	2	1	1	0	0
D2	0	1	0	0	1	1

Figura 6. Ejemplo de transformación vectorial de varios textos a una Bolsa de Palabras

Tras estas fases de procesamiento de la información se ha construido el modelo LDA con una parametrización base de 50 topics, donde se han indicado también como parámetros el corpus y el diccionario a utilizar.

Se ha utilizado la librería wordcloud para construir nubes de palabras y poder visualizar las palabras más importantes de cada topic

Figura 7. Ejemplo de nube de palabras de uno de los topics

Para evaluar el rendimiento del modelo, se han estimado los topics en dos documentos diferentes. En el primer caso, se ha escogido un documento de entre las noticias utilizadas en el corpus para entrenar el modelo y para la segunda prueba se ha utilizado una nueva noticia.

A partir de aquí es posible obtener la contribución de los topics más importantes

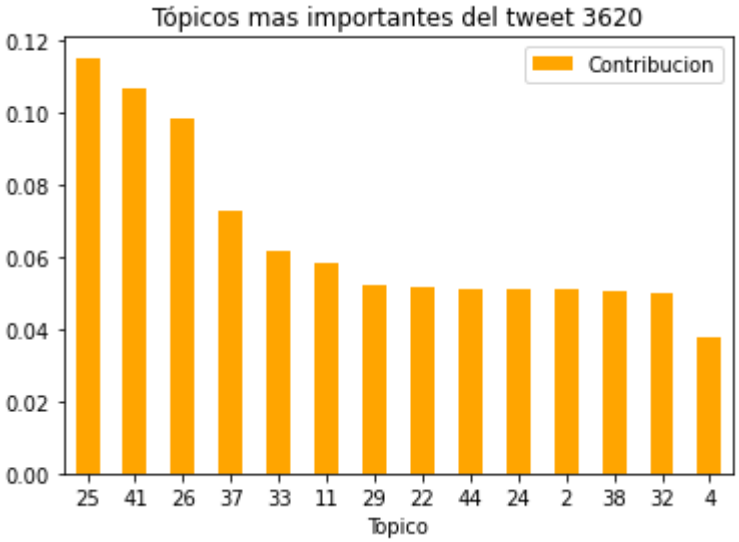


Figura 8. Ejemplo de contribución de los topics más importantes en un tweet

3.3.2.2. Similitud entre textos

Existen un sinfín de técnicas para obtener la similitud entre textos. Las técnicas más básicas para calcular la similitud entre textos solo tienen en cuenta la similitud léxica, esto es la semejanza en las palabras contenidas en los textos comparados
.
Una de estas técnicas consiste en representar los textos como bolsas de palabras (BOW) y aplicar la similitud de coseno para comparar la similitud entre dos documentos. En este caso, la semejanza entre los dos textos dependerá del número de palabras que compartan.

Es importante que los modelos tengan también en cuenta la similitud semántica, es decir, que puedan entender el significado real de las palabras o de la frase en cada contexto. Existen varios métodos para extraer esta similitud semántica de los textos, como los ampliamente utilizados modelos de word embeddings que permiten representar cada palabra como un vector de números reales y capturar de esta forma su información semántica.

Se ha definido una función de búsqueda para mostrar los n tweets más similares a uno dado, y aplicado la métrica de distancia de Jensen-Shannon a las distribuciones de los topics del modelo LDA comentado en el punto 3.3.2 mediante la utilización de la función jensen_shannon de la librería Gensim

4. Selección de poblaciones objetivo y matching

La selección de poblaciones objetivo sobre las que aplicar los modelos de víctimas y agresores, y el proceso de vinculación posterior de ambos conjuntos de posibles agresores y víctimas (matching), son dos procesos muy importantes dentro de la arquitectura del sistema propuesto, ya que en ellos se basa el poder seleccionar de forma eficiente los candidatos a ser valorados por los modelos propuestos y el poder relacionarlos posteriormente de forma adecuada.

Los criterios de selección de poblaciones objetivo pueden ser determinados por:

- Ubicación geográfica de los casos
- Temporalidad de los casos
- Estadísticas ya existentes sobre violencia de género y las que puedan derivarse de la ejecución del sistema propuesto.

Los criterios de vinculación (matching) entre posibles agresores y víctimas pueden venir determinados por:

- Personas con el mismo valor del vector de violencia y /o combinaciones de valores de los atributos y el vector de violencia que se estimen adecuados.
- Datos del registro civil, a efectos de estado civil.
- Datos del padrón, a efectos de comprobación de convivencia.
- Vínculos en redes sociales, fundamentalmente para los casos de ciber violencia.
- Vínculos en el entorno laboral, para los casos de violencia laboral.
- Proximidad geográfica, fundamentalmente para los casos de violencia física
- Trámites de divorcio con custodia compartida, en el caso de violencia vicaria, con el objetivo de detectar posibles víctimas menores de edad.
- Órdenes de alejamiento.

5. Consideraciones éticas y de privacidad

Tal como comenta Rosa Colmenarejo en [39], en la era del big data muchas decisiones que afectan a nuestro día a día son tomadas por modelos matemáticos que se programan y entrenan para tomar decisiones de forma independiente. Dichos algoritmos ocupan la posición de "**sujeto moral**", y por tanto al tratar de identificar al "**sujeto responsable**" nos encontramos con una máquina.

Esto nos lleva a que las preguntas fundamentales de la ética: ¿qué debo hacer?, ¿qué decisión tomar?, ¿qué criterios adoptar?... se reducen a un mero automatismo y al hecho de si hemos de desterrar la ética como disciplina académica o la hemos de aplicar para afrontar una realidad que cambia cada día más rápido.

Existen tres problemas fundamentales que deben ser controlados por la ética aplicada a la gestión de datos:

- **Propiedad de los datos**
- **Privacidad**
- **Identificar al sujeto moral**

La propiedad de los datos se ha transformado en un asunto técnico legal ya que existen sentencias que se posicionan claramente del lado de los usuarios. De cualquier forma la ética sigue siendo necesaria en este caso, ya que es algo que podría no estar nunca regulado en su totalidad debido al volumen y la velocidad de los datos involucrados, que incluso están modificando el concepto identidad como algo que varía en durante el tiempo y que llega a diferenciarse en identidad online e identidad offline y que se transforma y adapta en su interacción con la red.

Tal como indica Agustí Carrillo en [40], se entiende por privacidad el respeto y protección de la información personal. Tiene que ver con el control de sus propios datos, y debe permitir a los individuos mantener el control personal sobre su información con respecto a su recolección, uso y divulgación.

La normativa europea a este respecto se denomina GDPR (General Data Protection Regulation) y refuerzan los derechos de las personas sobre sus datos personales e imponen multas para las organizaciones que no protegen sus datos.

En base a dicha normativa, se definen como datos personales a cualquier información relativa a una persona física identificada o identificable, y se considera como violación de dichos datos cualquier destrucción accidental o ilegal, pérdida, alteración, divulgación no autorizada o acceso a los datos personales transmitidos, almacenados o procesados para otro uso.
La normativa española se puede agrupar en diferentes ámbitos de aplicación:

- **Servicios de la sociedad de la información**
- **Tratamiento de los datos de carácter personal**
- **Propiedad intelectual**
- **Administración electrónica**
- **Firma electrónica**

Debido al auge de las técnicas de Machine Learning, se ha de tener en cuenta que, no únicamente es importante la privacidad de la información, sino también la del algoritmo que procesa dicha información, y aunque no existen todavía normativas en este aspecto, es necesario tener en cuenta principios éticos, o lo que podríamos denominar gobierno de algoritmos:

- **Responsabilidad**: Para gestionar los efectos adversos a la sociedad o a individuos particulares.
- **Explicación**: Se deben poder explicar los efectos de un sistema de algoritmos a las personas afectadas por las decisiones que puedan tomar.
- **Exactitud:** Se deben poder identificar y registrar los errores que puedan cometer dichos algoritmos.
- **Auditoría:** Deben poder ser auditados por terceros.
- **Equidad**: Se debe poder determinar si producen efectos discriminatorios y evitar sesgos de sus programadores o en el propio objetivo de dichos algoritmos.

No se ha introducido ningún tipo de sesgo en los algoritmos utilizados en el presente trabajo, pero, tal como hemos comentado anteriormente, resulta evidente por motivos simplemente estadísticos, que la población mayoritaria de agresores serán fundamentalmente hombres y la de víctimas serán mujeres y menores (violencia vicaria)

Es obvio que todo lo comentado en este punto, afecta de forma importante a las necesidades de información que se plantean en la arquitectura propuesta. Buena parte de los atributos incluidos en los modelos de víctimas y agresores se pueden considerar datos personales, aun cuando únicamente reflejen la presencia o ausencia de una situación personal o patología determinada, y por tanto, solamente disponibles bajo permiso expreso del individuo, y en algunos casos protegidos por el secreto profesional, como en el caso de la información médica o económica.

Para poder implementar el sistema de prevención, sería necesario adaptar la legislación actual y tener acceso a dicha información, total o parcialmente, en la población general, o bien acotar dicha población a aquellas personas que a nivel de información pública, como la proporcionada en las redes sociales, muestren indicios de sexismo y violencia de género, en el caso de los agresores, y en el caso de las víctimas muestren indicios de depresión, ansiedad u otras patologías asociadas a las víctimas de violencia de género.

Para este último caso sería necesario tener acceso a la información personal de los posibles agresores y las posibles víctimas de la que dispongan las redes sociales y contar con su colaboración para analizar esta información.

En cualquier caso, la implementación y gestión de un sistema de prevención con estas necesidades debería estar bajo la supervisión gubernamental a través de los cuerpos y fuerzas de seguridad del estado y la judicatura, que en nuestro caso serían los sujetos morales.

En cuanto a los profesionales involucrados en el desarrollo de un sistema de estas características deberían seguir los siguientes principios en base al objetivo principal del proyecto:

- **Beneficencia**: No causar ningún daño y maximizar los posibles beneficios, disminuyendo los posibles daños
- **Autonomía**:
  - Capacidad racional para elegir de forma crítica.
  - Poseer habilidades mentales y/o físicas para cumplir los objetivos
  - Sentido de la responsabilidad tanto en la deliberación como tras la ejecución de la acción
- **No maleficencia**: Ante todo, no hacer daño. Especialmente útil cuando no se posible la aplicación del resto de principios.
- **Justicia**: Va ligada al principio de No maleficencia. Permite jerarquizar y ordenar las posibilidades de acción de acuerdo con criterios de justicia social

Se ha de recalcar que se debe ser absolutamente escrupuloso en la aplicación de los principios comentados, ya que tal como se ha indicado en el punto 1.3, se trata de un sistema que va a señalar a personas como posibles agresores o víctimas.

El marco legal sobre protección de datos comentado resulta un problema importante para la implementación de la solución propuesta, por lo que sería necesario plantear alguna alternativa para conseguir dicha información, como modificar el marco legal únicamente para permitir el acceso a datos personales relevantes de aquellas personas que muestren trazas significativas de sexismo en redes sociales o bien que tengan antecedentes de violencia.

6. Resultados

6.1. Modelos para el conjunto de casos de violencia de género

Una vez definidos los grupos de datos de trabajo en el punto 3.1.1, se aplica la prueba de Shapiro comprobando que ninguna de las variables tiene una distribución normal. Se puede ver gráficamente en las figuras 9 y 10 con la variable Agressor.Age

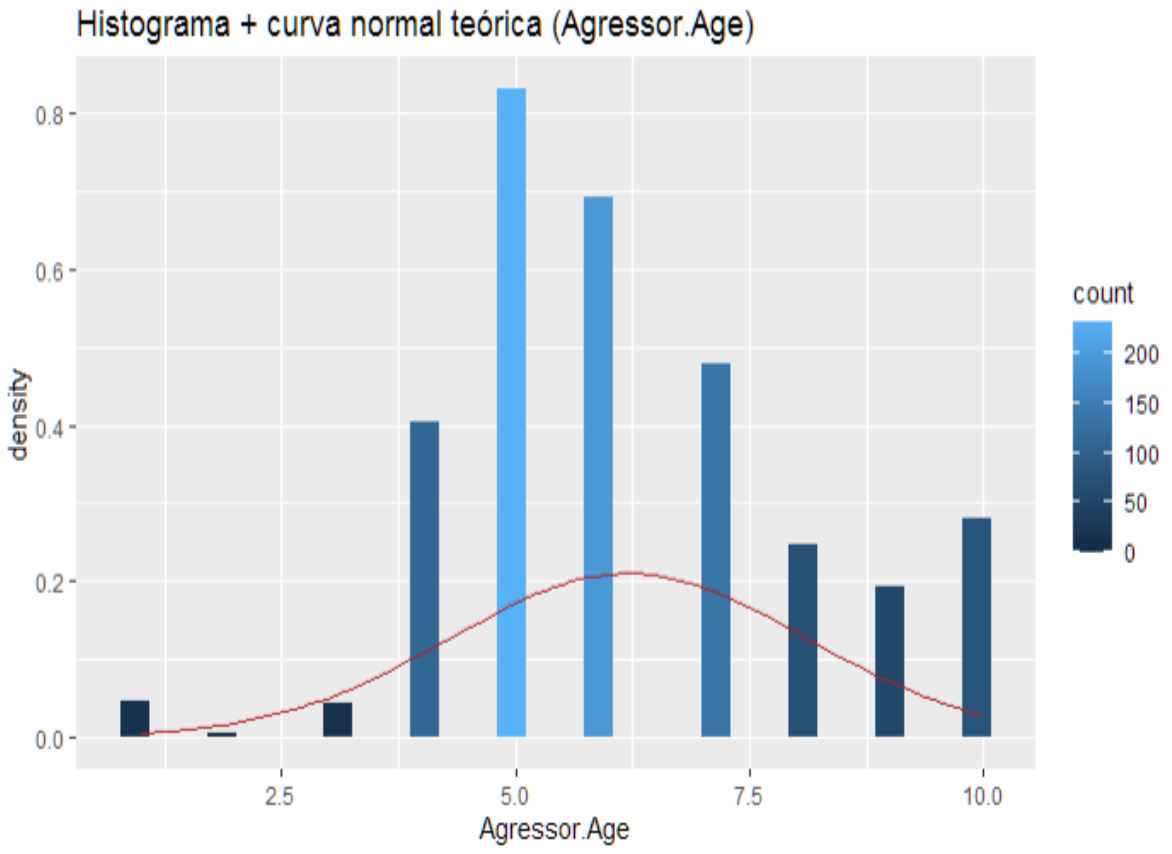


Figura 9. Histograma de la variable Agressor.Age del conjunto de datos

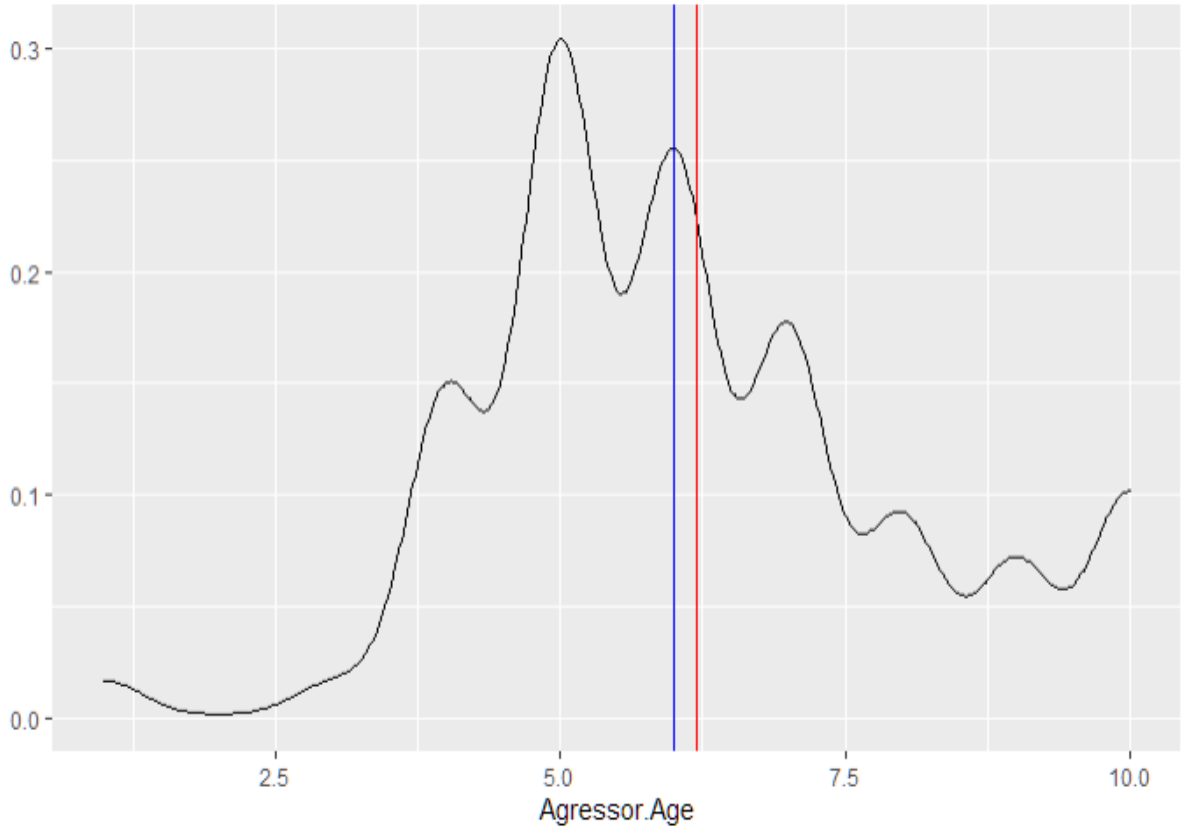


Figura 10. Histograma de la variable Agressor.Age del conjunto de datos con media (azul) y mediana(rojo)

Como ninguna de las variables es normal se ha aplicado el test de Levene entre Victims y el resto de las variables, obteniéndose los valores que se muestran en la tabla 6

Variable	F Value	Pr(>F)
Year	1.0172	0.4332
Month	1.029	0.4181
Autonomous Comunity	0.3649	0.9954
Province	0.3098	1
Relation	0.1716	0.8424
Victim age (interval)	0.2589	0.9894
Agressor age (interval)	0.3173	0.9695
Previous Abuse Report	0.374	0.6881
Living Together	0.2719	0.762

Tabla 6. Test de Levene entre victims y el resto de las variables

Los resultados indican que no hay diferencias significativas entre las varianzas de los grupos, es decir existe homogeneidad de varianza u homocedasticidad. Todas las variables utilizadas son categóricas, por lo que el análisis de sus relaciones se ha obtenido mediante tablas de contingencia y pruebas de Fisher

Las gráficas de las tablas de contingencia se pueden ver en las figuras 11, 12, 13 y 14:

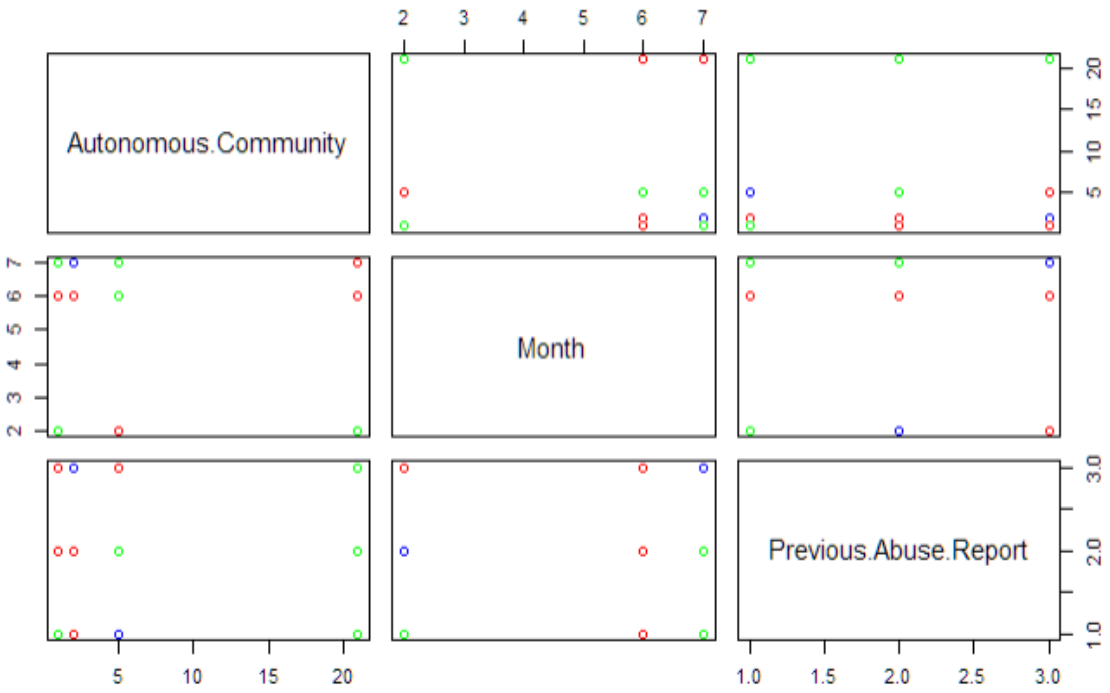


Figura 11. Tabla de contingencia variables Autonomous.Community, Month y Previous.Abuse.Report

Podemos comprobar que donde más asesinatos se producen, es en la Comunidad Andaluza (1) en los meses de Junio (7) y Julio (6), sin que haya constancia de abusos previos, seguida por Castilla La Mancha (5) en el mes de Julio (6), también sin constancia de abusos previos.

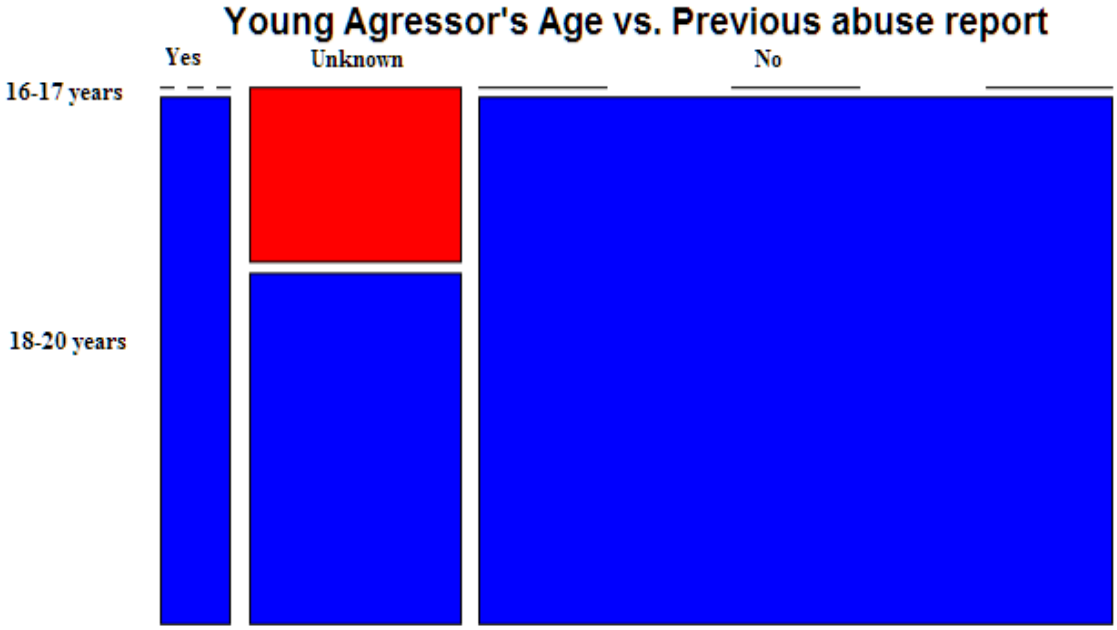


Figura 12. Tabla de contingencia variable Young Agressor's Age y Previous.Abuse.Report

Se comprueba que el mayor número de asesinatos en jóvenes se producen en la franja de 18-20 years y sin constancia de abusos previos

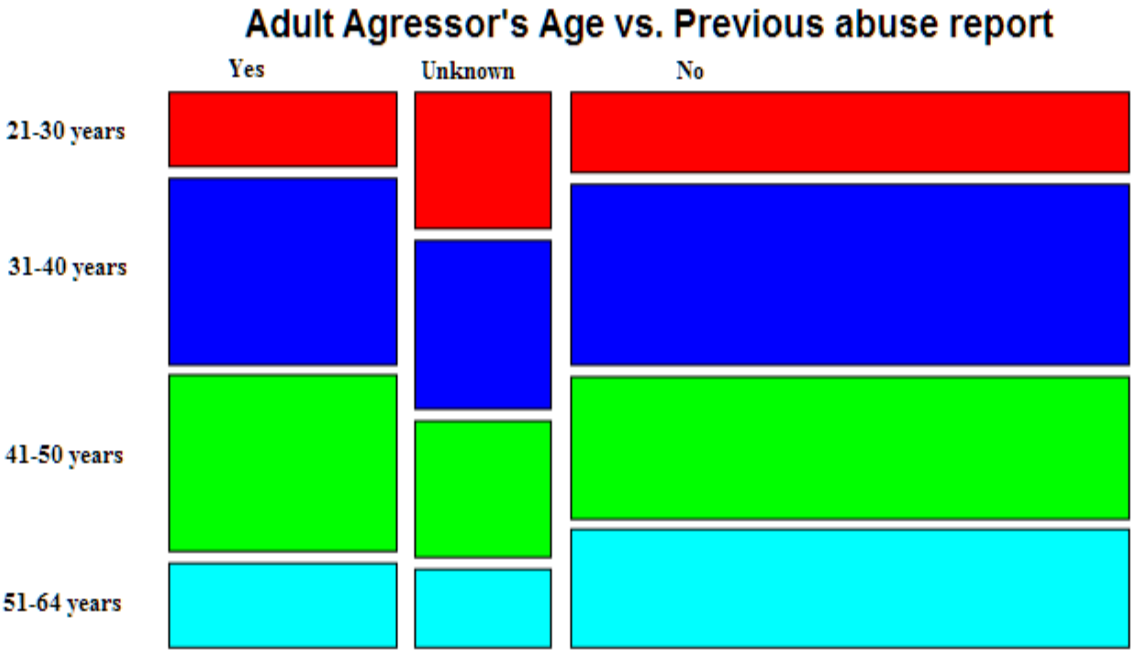


Figura 13. Tabla de contingencia variable Adult Agressor's Age y Previous.Abuse.Report

Se comprueba que el mayor número de asesinatos en adultos se producen en las franjas 31-40 years y 41-50 years

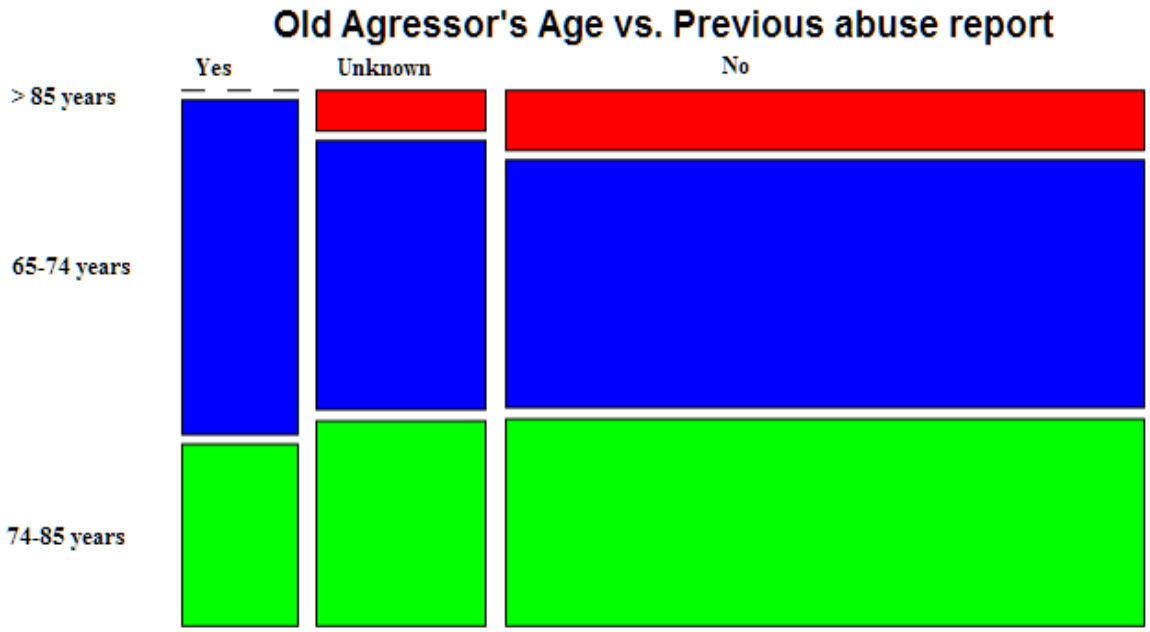


Figura 14. Tabla de contingencia variable Old Agressor's Age y Previous.Abuse.Report

Se comprueba que se produce un número similar de asesinatos, especialmente en las franjas de agresores 65-74 y 74-85, y sin constancia de abusos previos. En la franja de agresores mayores de 85 años, el número se reduce significativamente.

Los resultados de las pruebas de Fisher han arrojado los resultados que se muestran en la tabla 7

Tabla de Contingencia	P Value	Corr.(p_value<=0.05)
Comunidad Autónoma / Mes	0.981	No
Jóvenes / PAR	0.307	No
Adultos / PAR	0.05986	No
Mayores / PAR	0.6504	No
Convivientes / PAR	2.2e-16	Si
Relation / PAR	0.0001419	Si

Tabla 7. Resultados pruebas de Fisher

Los resultados de la aplicación de los modelos de regresión logística planteados para cada grupo de datos han arrojado los datos que se muestran en la tabla 8

model.vgenero.convivientes.yes (Abuso.Previo)	Variables Independientes	B,(EE)	OR	IC95% OR	p-valor
No	Intercept	-2.830,(0.525)	0.058	(-3.858 ;-1.802)	< 0.001
No	Agressor.Age	0.481,(0.077)	1.618	(0.33; 0.632)	< 0.001
Unknown	Intercept	0.806,(0.395)	2.239	(0.031;1.581)	< 0.001
Unknown	Agressor.Age	0.084, (0.064)	1.087	(-0.04; 0.208)	< 0.001
model.vgenero.relation.expartner (Abuso.Previo)	Variables Independientes	B (EE)	OR	IC95% OR	p-valor
No	Intercept	-0.162,(0.958)	0.85	(-0.35;0.22)	< 0.001
No	Relation	-0.162,(0.096)	0.85	(-0.35;0.22)	< 0.001
Unknown	Intercept	0.195,(0.08)	1.22	(0.38;0.352)	< 0.001
Unknown	Relation	0.195, (0.08)	1.22	(0.38;0.352)	< 0.001
model.vgenero.agresores.adultos(Abuso.Previo)	Variables Independientes	B (EE)	OR	IC95% OR	p-valor
No	Intercept	0.685, (0.71)	1.984	(-0.708;2.079)	0.05986
No	Agressor.Age	-0.222, (0.129)	0.801	(-0.477;0.032)	0.05986
Unknown	Intercept	0.603, (0.523)	1.829	(-0.421;1.628)	0.05986
Unknown	Agressor.Age	0.052,(0.093)	1.054	(-0.129;0.235)	0.05986

Tabla 8. Resultados modelos de regresión logística

Se ha considerado también el modelo relativo a agresores adultos, ya que su p-value se encuentra al límite de descartar la hipótesis nula.

Al interpretar las odds ratios de cada variable, se ha de asumir que el resto de las variables independientes se mantienen fijas. Se ha de interpretar cada una de las variables independientes entre los distintos tipos de Abuso.Previo tomando como referencia Abuso.Previo = Yes.

A modo de ejemplo, en la figura 15 se muestra el feedback de creación del modelo multinomial model.vgenero.convivientes.yes (Abuso.Previo)

```
# weights:  9 (4 variable)
initial  value 639.392352
iter  10 value 518.018643
iter  10 value 518.018643
final  value 518.018643
converged
```

Figura 15. Feedback creación modelo multinomial model.vgenero.convivientes.yes (Abuso.Previo)

En la figura 16, se muestra el resumen del modelo:

```
Call:
multinom(formula = Previous.Abuse.Report ~ Agressor.Age, data = ViolenciaGenero.Convivientes.Yes)

Coefficients:
  (Intercept) Agressor.Age
2  -2.8302764    0.4812557
3   0.8063664    0.0840625

Std. Errors:
  (Intercept) Agressor.Age
2   0.5246014   0.07701679
3   0.3954953   0.06371025

Residual Deviance: 1036.037
AIC: 1044.037
```

Figura 16. Feedback resumen creación modelo multinomial model.vgenero.convivientes.yes (Abuso.Previo)

La línea de coeficientes que comienza con 2 hace referencia al modelo comparando la probabilidad de que no sepamos nada sobre el informe previo (Unknown), respecto a que si lo haya (Yes). La línea de coeficientes que comienza con 3 hace referencia al modelo comparando la probabilidad de que no haya informe previo (No), respecto a que si lo haya (Yes)

A partir de los coeficientes del modelo la figura 17 muestra el cálculo de los odds ratio:

```
  (Intercept) Agressor.Age
2  0.05899654     1.618105
3  2.23975489     1.087697
```

Figura 17. Cálculo odds ratio

Y la figura 18 muestra el cálculo de los intervalos de confianza:

```
, , 2

                   2.5 %     97.5 %
(Intercept)  -3.8584763 -1.8020766
Agressor.Age  0.3303056  0.6322058

, , 3

                    2.5 %    97.5 %
(Intercept)    0.03120986 1.5815230
Agressor.Age -0.04080731 0.2089323
```

Figura 18. Cálculo intervalos de confianza

6.2. Modelos para los conjuntos de víctimas y agresores

La aplicación de los modelos de clasificación multilabel definidos para las víctimas arrojan los resultados que se muestran en la tabla 9

Tipo/Estrategia	Algoritmo	F1-macro
Multilabel nativo	Random Forest	0.5054
	Multilayer Perceptron	0.4865
	Decission Tree	0.4903
	Nearest Neighbours	0.4934
Multilabel no nativo (Multioutput)	Extended Gradient Boosting	0.4911
	Logistic Regression	0.4578
	Gaussian Naive Bayes	0.4558
	Linear SVC	0.4578
Problem Transformation	Binary Relevance	0.4522
	Classifier Chains	0.4549
	Label Powerset	0.4849
MultiLearn Adapt	MLkNN	0.4934
	BRkNNaClassifier	0.4958
	MLTSVM	0.6676
Multilabel Deep Learning	Problem Transformation con BinaryRelevance (Keras)	0.4739

Tabla 9. Resultados modelos multilabel para víctimas

En el caso de los modelos definidos para los agresores los resultados se muestran en la tabla 10

Tipo/Estrategia	Algoritmo	F1-macro
Multilabel nativo	Random Forest	0.5121
	Multilayer Perceptron	0.4971
	Decission Tree	0.4971
	Nearest Neighbours	0.5068
Multilabel no nativo (Multioutput)	Extended Gradient Boosting	0.5036
	Logistic Regression	0.5092
	Gaussian Naive Bayes	0.5083
	Linear SVC	0.5093
Problem Transformation	Binary Relevance	0.5078
	Classifier Chains	0.5051
	Label Powerset	0.4959
MultiLearn Adapt	MLkNN	0.5036
	BRkNNaClassifier	0.4981
	MLTSVM	0.6638

Multilabel Deep Learning	Problem Transformation con BinaryRelevance (Keras)	0.4797

Tabla 10. Resultados modelos multilabel para agresores

Se observa que los resultados son ligeramente mejores en los modelos implementados para los agresores cuyo conjunto de datos tiene algunos atributos más. De cualquier forma, se ha de tener en cuenta que, en algunas ocasiones, como es el caso de datos no balanceados, no son utilizables de forma eficiente por los algoritmos estándar de Machine Learning [41]

6.3. Modelos para el conjunto de datos de redes sociales

6.3.1. Detección y clasificación del sexismo

Los resultados de las métricas accuracy y F1-macro obtenidos para los modelos implementados se muestran en la tabla 11

Tarea	Clasificador	Scoring
Identificación	SGDClassifier	Accuracy 0.70
	LinearSVC	Accuracy 0.71
Categorización	LogisticRegression()	F1-macro 0.62
	DecisionTreeClassifier()	F1-macro 0.51
	SVC()	F1-macro 0.60

Tabla 11. Resultados modelos multiclase identificación/categorización sexismo

Los resultados obtenidos son algo inferiores a los más destacados mostrados en el ranking de EXISTS2021 [35] para el caso de la identificación (Accuracy: 0.78) y mejoran sensiblemente con respecto a la categorización (F1-macro: 0.57)

6.3.2. Topic Modelling

Una vez creado el modelo LDA en base al corpus y el diccionario, en la figura 19 se muestran 5 palabras de 20 topics.

```
(47, '0.089*"aunqu" + 0.072*"nenaz" + 0.044*"total" + 0.044*"acept" + 0.035*"particip"')
(30, '0.124*"tip" + 0.063*"manspreading" + 0.061*"vas" + 0.044*"ultim" + 0.037*"drog"')
(3, '0.113*"cort" + 0.070*"fald" + 0.053*"frent" + 0.050*"larg" + 0.031*"ment"')
(40, '0.121*"violacion" + 0.072*"fot" + 0.049*"conden" + 0.040*"segun" + 0.034*"nacional"')
(14, '0.128*"nuev" + 0.120*"llev" + 0.095*"man" + 0.078*"mam" + 0.042*"quien"')
(19, '0.059*"parej" + 0.055*"final" + 0.041*"inclus" + 0.031*"año" + 0.030*"san"')
(34, '0.048*"consider" + 0.044*"vien" + 0.034*"gobiern" + 0.027*"trump" + 0.027*"pedaz"')
(32, '0.079*"vist" + 0.072*"car" + 0.043*"pag" + 0.035*"sirv" + 0.033*"ojal"')
(49, '0.097*"clar" + 0.072*"vec" + 0.036*"violador" + 0.034*"poc" + 0.029*"dond"')
(17, '0.106*"abort" + 0.092*"public" + 0.049*"dar" + 0.040*"mor" + 0.034*"celebr"')
(37, '0.126*"tan" + 0.119*"bien" + 0.066*"mund" + 0.045*"hombr" + 0.040*"conoc"')
(39, '0.117*"dia" + 0.091*"mujer" + 0.080*"hoy" + 0.039*"pod" + 0.035*"ver"')
(22, '0.240*"put" + 0.142*"hij" + 0.114*"parec" + 0.038*"respet" + 0.034*"hermos"')
(29, '0.126*"buen" + 0.081*"va" + 0.058*"tet" + 0.058*"not" + 0.055*"gord"')
(33, '0.170*"chic" + 0.070*"mas" + 0.052*"cos" + 0.041*"hac" + 0.039*"ser"')
(25, '0.252*"muj" + 0.101*"ser" + 0.045*"gan" + 0.026*"quier" + 0.024*"florer"')
(18, '0.068*"nunc" + 0.051*"hag" + 0.047*"pens" + 0.041*"muj" + 0.036*"pon"')
(1, '0.087*"asi" + 0.051*"nadi" + 0.049*"cul" + 0.042*"cambi" + 0.041*"pued"')
(41, '0.090*"dic" + 0.088*"habl" + 0.065*"si" + 0.042*"pues" + 0.034*"embaraz"')
(5, '0.138*"femin" + 0.114*"mujer" + 0.051*"hac" + 0.044*"hombr" + 0.033*"derech"')
```

Figura 19. Ejemplos topics/palabras

En la figura 20 se muestran 4 de dichos topics en formato wordcloud.

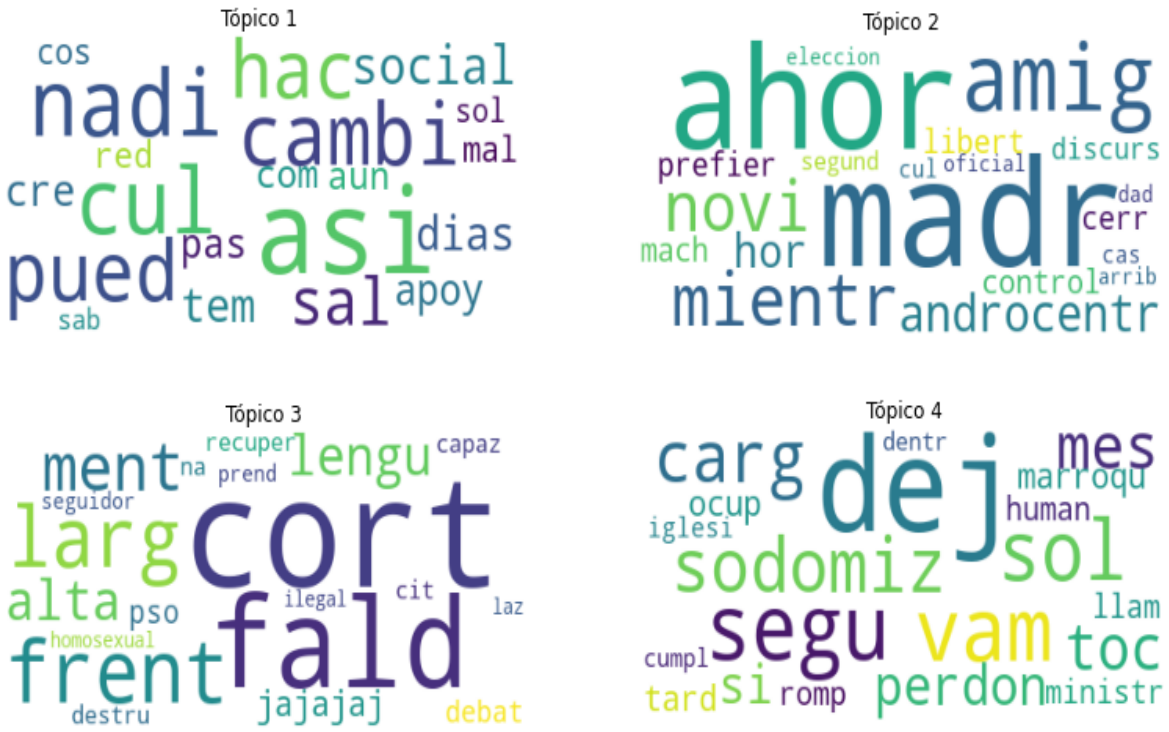


Figura 20. Ejemplos topics en formato wordcloud

A continuación, se ha seleccionado un tweet al azar:

Tweet: *"hasta una feminazi se va a dejar maltratar por un hombre la hora de coger porque cuando una mujer anda con ganas de que le metan la verga se olvida de cualquier ideal principio hasta que se le baje la calentura"*

Su representación Bag Of Words y la distribución aplicando el modelo LDA son las que se muestran a continuación:

BOW: [(1, 1), (2, 1), (8, 1), (12, 1), (165, 1), (201, 1), (233, 1), (356, 1), (603, 1), (708, 1), (818, 1), (879, 1), (1146, 1), (1194, 1), (1296, 1), (1655, 1), (1802, 1), (2710, 1)]

Distribución: [(2, 0.0509694), (4, 0.0379818), (11, 0.05833467), (22, 0.051842775), (24, 0.05105242), (25, 0.11538702), (26, 0.09862312), (29, 0.052431647), (32, 0.050195925), (33, 0.061706264), (37, 0.07299206), (38, 0.050780836), (41, 0.10672502), (44, 0.051318105)]

Posteriormente se han obtenido los índices y la contribución (proporción) de los topics más significativos para el tweet, y que se muestran en la figura 21

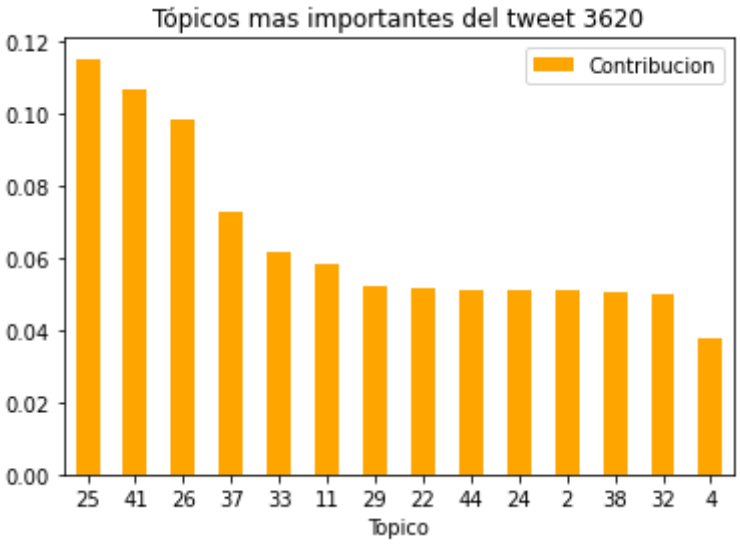


Figura 21. Topics más importantes tweet seleccionado al azar

En la figura 22 se muestran 10 palabras de cada topic

```
*** Tópico: 25 ***
muj, ser, gan, quier, florer, hombr, diner, rebeld, pas, mar

*** Tópico: 41 ***
dic, habl, si, pues, embaraz, ser, hac, pued, feminazi, señorit

*** Tópico: 26 ***
hech, denunci, met, intent, baj, vari, ello, mir, top, articul

*** Tópico: 37 ***
tan, bien, mund, hombr, conoc, sient, peor, moment, sufr, ser

*** Tópico: 33 ***
chic, mas, cos, hac, ser, tont, esper, masculin, rubi, veo

*** Tópico: 11 ***
cuent, mierd, tir, cag, padr, si, vez, relacion, hab, entonc

*** Tópico: 29 ***
buen, va, tet, not, gord, siempr, niñat, perd, encant, llor

*** Tópico: 22 ***
put, hij, parec, respet, hermos, hiz, salg, mil, sep, increibl

*** Tópico: 44 ***
quier, peg, mujer, rob, carcel, respons, pequeñ, recuerd, real, maltrat

*** Tópico: 24 ***
ningun, vide, comentari, izquierd, inform, olvid, deten, via, compart, radical

*** Tópico: 2 ***
madr, ahor, amig, mientr, novi, androcentr, hor, prefier, libert, mach

*** Tópico: 38 ***
tom, comun, sex, famili, sos, si, demas, twitt, jamas, anda

*** Tópico: 32 ***
vist, car, pag, sirv, ojal, golp, pur, disfrut, atencion, complet

*** Tópico: 4 ***
dej, segu, vam, sol, sodomiz, carg, mes, toc, perdon, si
```

Figura 22. Ejemplo palabras de cada topic

Posteriormente se ha evaluado el modelo con un tweet nuevo:

New_tweet: *"@test. Una feminazi maltratada por un hombre olvida ideales si quiere que se la metan"*

Se ha preprocesado el tweet de la misma forma que se hizo con el corpus y obtenido el siguiente conjunto de stems:

['feminazi', 'maltrat', 'hombr', 'olvid', 'ideal', 'si', 'quier', 'met']

De la misma forma que para el tweet obtenido al azar, se han obtenido los índices y la contribución (proporción) de los topics más significativos para el nuevo tweet, y que se muestran en la figura 23

Figura 23. Topics más importantes en tweet nuevo

Y en la figura 24 se muestran las palabras de cada topic:

```
*** Tópico: 41 ***
dic, habl, si, pues, embaraz, ser, hac, pued, feminazi, señorit

*** Tópico: 37 ***
tan, bien, mund, hombr, conoc, sient, peor, moment, sufr, ser

*** Tópico: 44 ***
quier, peg, mujer, rob, carcel, respons, pequeñ, recuerd, real, maltrat

*** Tópico: 24 ***
ningun, vide, comentari, izquierd, inform, olvid, deten, via, compart, radical

*** Tópico: 26 ***
hech, denunci, met, intent, baj, vari, ello, mir, top, articul

*** Tópico: 5 ***
femin, mujer, hac, hombr, derech, gener, odi, dañ, iguald, luch
```

Figura 24. Palabras de los topics del tweet nuevo

Como se puede observar, los topics más relevantes del nuevo tweet son el 41, el 37 y el 44. En ellos se puede apreciar la aparición de los términos feminazi y hombre.

En un caso real, el problema sería más simple, y arrojaría resultados más claros, ya que se trataría de detectar topics relacionados con el sexismo y la violencia de género de entre otros topics muy diferentes que puedan estar contenidos en los tweets que publica una persona, por lo que la técnica de detección de topics puede resultar eficaz a la hora de detectar sexismo y trazas de violencia de género en las redes sociales.

6.3.3. Similitud entre textos

Para validar la utilidad de la similitud entre textos mediante la distancia de Jensen-Shannon, se ha utilizado un tweet nuevo: *"['@test. Todas las feminazis son iguales. Se olvidan pronto de los ideales si conviene']"*

Posteriormente se ha utilizado el método indicado en el punto 3.3.2 para calcular las distancias y obtener los 5 tweets más similares al tweet nuevo

```
1: los que quieran mencionar (0.1583302915096283)
2: imagino que esta feminazi es boba (0.17766991257667542)
3: kokichi se le subió la fama (0.18915718793869019)
4: feminazi tóxica endeluego (0.21620112657546997)
5: la izquierda feminazi no dirá nada de esto (0.2162017822265625)
```

Figura 25. Tweets similares al tweet nuevo

Como se puede ver en la figura 25, los 5 tweets más similares calculados mediante la distancia Jensen-Shannon son, bastante similares al nuevo tweet.

7. Conclusiones y trabajos futuros

7.1. Conclusiones

El objetivo principal del trabajo ha sido definir la arquitectura de un sistema que ayude a la prevención de los casos de violencia de género en España y con ello sensibilizar sobre dicha problemática.

Debido a lo ambicioso de la arquitectura propuesta, se ha tratado de presentar las alternativas que nos ofrecen las técnicas de análisis estadístico y machine learning para la implementación de una solución, más que obtener modelos con eficiencias elevadas. Es decir, se han conseguido todos los objetivos planteados, pero algunos resultados respecto a la eficiencia obtenida en la evaluación de algunos modelos son mejorables.

La metodología utilizada ha sido la adecuada, ya que se adapta a los requerimientos fundamentales del proyecto: tratamiento de la información, modelado y evaluación y ha sido aplicada de la forma prevista.

De forma general, se ha seguido la planificación establecida en relación sobre todo al aspecto temporal, y en relación también a los objetivos y métodos planteados. En relación con el análisis de la información en redes sociales, el planteamiento inicial fue realizar un streaming de publicaciones en Twitter para detectar contenido vinculable a la violencia de género, pero finalmente se ha optado por construir clasificadores en base a datos previamente calificados como sexistas o no sexistas, y categorizados en diferentes tipos de sexismo.

De forma similar, el tratamiento de Topic Modelling y la similitud entre los textos, no se contempló inicialmente dentro del análisis de la información en redes sociales, pero se ha considerado importante incluirlo como herramienta adicional para la detección de posibles perfiles de agresores.

En relación con los criterios de selección de poblaciones objetivo y matching entre posibles agresores y víctimas, se han proporcionado algunos que se han considerado lógicos, pero que pueden ser ampliados por especialistas en campos relacionados más directamente con la estadística y con el comportamiento humano.

Durante el desarrollo del proyecto no han aparecido impactos no previstos en el punto 1.3 (ético-sociales, de sostenibilidad y de diversidad), se han mitigado los relacionados con la diversidad y se ha abordado el impacto negativo relacionado con la privacidad y la información personal en el punto 5 Consideraciones éticas y de privacidad.

7.1.1. Conjunto de datos casos violencia de género

Sobre este conjunto de datos se ha realizado un análisis estadístico básico y posteriormente ha servido como base para definir perfiles (conjuntos de datos) de agresores y víctimas formados por atributos (marcadores) que pueden determinar que una persona sea un posible agresor o víctima.

La definición de dichos marcadores se ha llevado a cabo mediante una investigación sobre publicaciones del ámbito de la sociología, la psiquiatría y la medicina, por lo que sería interesante que, en una posible implementación de la arquitectura, interviniesen profesionales de dichos ámbitos en su definición.

7.1.2. Conjunto de datos de víctimas y agresores

Se han definido modelos de **clasificación multilabel** para víctimas y agresores mediante técnicas de Machine Learning. La evaluación de dichos modelos nos ha llevado a resultados mejorables respecto a los obtenidos de media (agresores F1 macro: 0,5117) y (víctimas f1 macro: 0,4885), teniendo en cuenta que los algoritmos de machine learning convencionales no acostumbran a manejar atributos binarios de forma demasiado eficiente [41].

7.1.3. Conjunto de datos provenientes de redes sociales

Respecto a los modelos de **clasificación binaria y multiclase** definidos para la detección y categorización del **sexismo**, los resultados se han aproximado a los rankings de EXIST2021 [35] en el caso de la detección (Accuracy: 0.71) y se han mejorado en el caso de la categorización (F1 macro: 0.62). Durante la evaluación de los modelos se ha constatado que la forma de sexismo más compleja de categorizar es la relacionada con la **objetización**.

En cuanto a los modelos definidos para el **topic modelling** (LDA) y la **similitud entre textos** se ha comprobado que pueden ser útiles también para la detección de expresiones sexistas y para detectar textos similares a otros con contenidos sexistas.

7.2. Trabajos futuros

Por el hecho de tratarse de un proyecto muy ambicioso, con su realización se abren múltiples de líneas de trabajo futuro:

- Los atributos y variables considerados en los perfiles de agresores y víctimas se han propuesto a partir de la consulta en trabajos publicados sobre violencia de género dentro del ámbito de la medicina, la sociología, la psicología y datos obtenidos de instituciones oficiales, por lo que sería conveniente una revisión de los atributos y variables consideradas en los modelos de víctimas y de agresores dirigida por profesionales de los ámbitos mencionados, para garantizar que se recopila la información adecuada para el objetivo principal del proyecto.

- En la realización del trabajo se han abordado la detección y clasificación del sexismo en redes sociales de cara a identificar posibles agresores, por lo que otra línea de trabajo consistiría en analizar los topics significativos que aparecen en las publicaciones de las víctimas de violencia de género y que pueden estar relacionados con las patologías que acostumbran a afectar a las víctimas: problemas de salud crónicos, lesiones, trastornos mentales como depresión, ansiedad, suicidio, y estrés postraumático, por ejemplo, tal como se indica en la publicación del consejo general de la psicología en España [42].

  Una aproximación realmente interesante en este aspecto serían las técnicas utilizadas por Laura Planas Simón en su trabajo Análisis de la depresión y la ansiedad causadas por un aborto usando datos de Twitter [43] trasladadas al caso de violencia de género.

- Profundizar en los criterios y en la aplicación de sistemas automáticos para poder vincular a posibles agresores y víctimas

- Abordar la obtención de información relevante para los perfiles de agresores y víctimas utilizando la información pública en redes sociales.

- Se ha constatado la falta alarmante de información sobre los casos de violencia de género. El único conjunto de datos público encontrado es realmente pequeño y con unos atributos y variables básicamente temporales, de ubicación, y edad. Otra línea sería definir un sistema ETL que permita recopilar los datos de los casos de violencia de género en diferentes organismos e instituciones, unificando y centralizando la información que se recopile.

- Ampliar la detección y clasificación del sexismo al resto de lenguas oficiales en España y a variantes del castellano habladas en centro américa y Sudamérica.

8. Glosario

ATENPRO acrónimo de sistema de atención y protección para víctimas de violencia de género. Es una modalidad de servicio que, con la tecnología adecuada, ofrece a las víctimas de violencia de género una atención inmediata, ante las eventualidades que les puedan sobrevenir, las 24 horas del día, los 365 días del año

BERT acrónimo de Bidirectional Encoder Representations from Transformers. Es una técnica basada en redes neuronales para el preentrenamiento del procesamiento del lenguaje natural (PLN) desarrollada por Google.

Big Data conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales.

BOW acrónimo de Bag of Words

CRISP-DM acrónimo de Cross Industry Standard Process for Data Mining

DASH modelo de riesgo para los profesionales que trabajan con víctimas de violencia doméstica

data science disciplina científica centrada en el análisis de grandes fuentes de datos para extraer información, comprender la realidad y descubrir patrones con los que tomar decisiones.

deep learning área del machine learning en la que el aprendizaje se realiza en capas sucesivas en las la información va adquiriendo significado progresivamente.

ETL acrónimo de Extracción, Transformación y Carga

Gaussian Processes son un método genérico de aprendizaje supervisado diseñado para resolver problemas de regresión y clasificación probabilística.

GDPR acrónimo de General Protection Data Regulation

GENSIM librería open-source en Python para representar documentos en formato de vectores semánticos con el objetivo de procesar textos digitales desestructurados utilizando algoritmos de machine learning no supervisados

Google Académico motor de búsqueda web de libre acceso que indexa el texto completo o los metadatos de la literatura académica en una variedad de disciplinas y formatos de publicación.

grooming o ciberacoso, del verbo to groom, que alude a conductas de acercamiento o preparación para un determinado fin. Se trata de una serie de conductas y acciones emprendidas por adultos, a través de Internet, con el objetivo deliberado de ganarse la amistad de menores de edad, creando una conexión emocional con los mismos, con el fin de ganarse su confianza y poder abusar sexualmente de ellos

INE: acrónimo de Instituto Nacional de Estadística

LGTBI acrónimo de lesbianas gais transexuales e intersexuales

Linear Regression algoritmo de aprendizaje automático basado en aprendizaje supervisado que realiza una tarea de regresión. La regresión modela un valor de predicción objetivo basado en variables independientes. Se utiliza principalmente para averiguar la relación entre las variables y la previsión.

machine learning subcampo de las ciencias de la computación y una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras aprendan a partir de transformaciones estadísticas lineales.

Multilayer Perceptron Neural Network es una clase totalmente conectada de red neuronal artificial (ANN) feedforward

NLTK acrónimo de Natural Language Toolkit

OMS acrónimo de Organización Mundial de la Salud

ONU acrónimo de Organización de las Naciones Unidas

Random Forest algoritmo de aprendizaje automático de uso común registrado por Leo Breiman y Adele Cutler, que combina la salida de múltiples árboles de decisión para llegar a un solo resultado. Su facilidad de uso y flexibilidad han impulsado su adopción, ya que maneja problemas de clasificación y regresión.

RoBERTa método robustamente optimizado para el entrenamiento previo de sistemas de procesamiento de lenguaje natural (NLP) que mejora las representaciones de codificador bidireccional de Transformers

Support Vector Machines conjunto de métodos de aprendizaje supervisado utilizados para la clasificación, regresión y detección de valores atípicos.

Twitter red social y servicio de microblogging usado para la comunicación en tiempo real utilizado por millones de personas y organizaciones.

VIOGEN sistema policial centralizado en el Ministerio del Interior y destinado al seguimiento y protección de las mujeres víctimas de violencia de género y de sus hijos e hijas en cualquier parte del territorio nacional.

violencia vicaria aquella que tiene como objetivo dañar a la mujer a través de sus seres queridos y especialmente de sus hijas e hijos

violencia de género hace referencia a cualquier acto con el que se busque dañar a una persona por su género. La violencia de género nace de normas perjudiciales, abuso de poder y desigualdades de género.

violencia doméstica es un concepto utilizado para referirse a «la violencia ejercida en el terreno de la convivencia asimilada, por parte de uno de los miembros contra otro, contra algunos de los demás o contra todos ellos

VPR acrónimo de valoración policial de riesgo

VPER valoración policial de evolución del riesgo sin incidente

016 servicio telefónico de información y de asesoramiento jurídico en materia de violencia de género

9. Bibliografía

Se ha utilizado de forma general la bibliografía proporcionada en las asignaturas del máster, tanto los documentos proporcionados por la UOC que se incluyeron en cada asignatura, como de la bibliografía recomendada en cada una de ellas.

01. Mujeres, O. N. U. (2021). Preguntas frecuentes: Tipos de violencia contra las mujeres y las niñas. ONU Mujeres. Recuperado, 4.

02. Acosta, C. A. G. (2014). Factores asociados a la violencia: revisión y posibilidades de abordaje. Revista Iberoamericana de psicología, 7(1), 115-124.

03. Organización Mundial de la Salud: Violencia contra la mujer (12/10/2022). https://www.who.int/es/news-room/fact-sheets/detail/violence-against-women

04. Carolina Alonso Hernández, Rosario Cacho Sáez, Irene González Ramos, Eufemia Herrera Álvarez, Javier Ramírez García: Junta de Andalucía. Consejería de Educación. Dirección General de Participación y Equidad: Guía de buen trato y prevención de la violencia de género. Protocolo de actuación en el ámbito educativo (12/10/2022). https://www.uma.es/media/files/Gu%C3%ADa_buenostratos.pdf

05. Delegación del Gobierno contra la Violencia de Género (12/10/2022): https://violenciagenero.igualdad.gob.es/violenciaEnCifras/home.htm

06. Cruz Roja. 25 NOV Violencia de género en el ámbito laboral (12/10/2022).: https://www2.cruzroja.es/-/violencia-de-g-c3-a9nero-en-el-c3-a1mbito-laboral

07. Servicios de Prensa de la Moncloa. Gobierno de España (14/10/2022).: https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/igualdad/Paginas/2021/290421-acoso_sexual.aspx

08. Delegación del Gobierno contra la violencia de Género. Boletín estadístico anual (2020) (14/10/2022).. https://violenciagenero.igualdad.gob.es/violenciaEnCifras/boletines/boletinAnual/docs/Boletin_estadistico_anual_2020_df.pdf

09. Gemma Ronzano (BSc Psychology). Forecasting Domestic Violence (14/10/2022). Forehttps://psychology.nottingham.ac.uk/staff/ddc/c8cxpa/further/Dissertation_examples/Ronzano_13.pdf

10. Dr. Ria Ivandic The London School of Economics and Political Science. Artificial intelligence could help protect victims of domestic violence (14/10/2022). https://www.lse.ac.uk/News/Latest-news-from-LSE/2020/b-Feb-20/Artificial-intelligence-could-help-protect-victims-of-domestic-violence

11. Grogger, J., Gupta, S., Ivandic, R., & Kirchmaier, T. (2021). Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases. Journal of Empirical Legal Studies, 18(1), 90-130.

12. Berk, R. A., Sorenson, S. B., & Barnes, G. (2016). Forecasting domestic violence: A machine learning approach to help inform arraignment decisions. Journal of empirical legal studies, 13(1), 94-115.

13. Colin Lecher. The Markup (29/02/2022). Police are looking to Algorithms to predict domestic violence (14/10/2022).. https://themarkup.org/the-breakdown/2022/06/29/police-are-looking-to-algorithms-to-predict-domestic-violence

14. Rodríguez-Rodríguez, I., Rodríguez, J. V., Pardo-Quiles, D. J., Heras-González, P., & Chatzigiannakis, I. (2020). Modeling and forecasting gender-based violence through machine learning techniques. Applied Sciences, 10(22), 8244.

15. D'Ignazio, C., Val, H. S., Fumega, S., Suresh, H., & Cruxen, I. (2020). Feminicide & machine learning: detecting gender-based violence to strengthen civil sector activism.

16. Anzovino, M., Fersini, E., & Rosso, P. (2018, June). Automatic identification and classification of misogynistic language on twitter. In International Conference on Applications of Natural Language to Information Systems (pp. 57-64). Springer, Cham.

17. Hewitt, S., Tiropanis, T., & Bokhove, C. (2016, May). The problem of identifying misogynist language on Twitter (and other online social spaces). In Proceedings of the 8th ACM Conference on Web Science (pp. 333-335).

18. Piñeiro-Otero, T., & Martínez-Rolán, X. (2021). Eso no me lo dices en la calle. Análisis del discurso del odio contra las mujeres en Twitter. Profesional de la información (EPI), 30(5).

19. Castorena, C. M., Abundez, I. M., Alejo, R., Granda-Gutiérrez, E. E., Rendón, E., & Villegas, O. (2021). Deep neural network for gender-based violence detection on Twitter messages. Mathematics, 9(8), 807.

20. Al-Garadi, M. A., Kim, S., Guo, Y., Warren, E., Yang, Y. C., Lakamana, S., & Sarker, A. (2022). Natural language model for automatic identification of intimate partner violence reports from twitter. Array, 15, 100217.

21. González-Prieto, Á., Brú, A., Nuño, J. C., & González-Álvarez, J. L. (2021). Machine learning for risk assessment in gender-based crime. arXiv preprint arXiv:2106.11847.

22. Karystianis, G., Cabral, R. C., Han, S. C., Poon, J., & Butler, T. (2021). Utilizing text mining, data linkage and deep learning in police and health records to predict future offenses in family and domestic violence. Frontiers in digital health, 3, 602683.

23. Turner, E., Medina, J., & Brown, G. (2019). Dashing hopes? The predictive accuracy of domestic abuse risk assessment by police. The British Journal of Criminology, 59(5), 1013-1034.

24. DASH questionary (17/10/2022). https://safelives.org.uk/sites/default/files/resources/Dash%20risk%20checklist%20quick%20start%20guidance%20FINAL.pdf

25. ATENPRO (24/10/2022) https://violenciagenero.igualdad.gob.es/informacionUtil/recursos/servicioTecnico/home.htm

26. VIOGEN (24/10/2022) https://www.interior.gob.es/opencms/pdf/archivos-y-documentacion/documentacion-y-publicaciones/publicaciones-descargables/seguridad-ciudadana/La_valoracion_policial_riesgo_violencia_contra_mujer_pareja_126180887.pdf

27. Delegación del gobierno contra la violencia de género. Portal Estadístico. (25/10/2022) http://estadisticasviolenciagenero.igualdad.mpr.gob.es/

28. Pablo Saenz de Tejada. Data World. Domestic Vioence in Spain (25/10/2022). https://data.world/pablosdt

29. Scikit Learn. Multiclass and multioutput algorithms. (27/10/2022) https://scikit-learn.org/stable/modules/multiclass.html

30. Scikit Learn. Multi-Label Classification in Python (27/10/2022) http://scikit.ml/

31. Scikit Multilearn. Api Reference. (27/10/2022) http://scikit.ml/api/skmultilearn.html

32. Scikit Multilearn . Multi-Label Deep Learning with scikit multilearn. (27/10/2022) http://scikit.ml/multilabeldnn.html

33. Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow. O'Reylly 2nd Edition. Pag 100 and 106.

34. Gavin Hackeling. Mastering Machine Learning with scikit-learn. PACKT publishing. Pag 94

35. EXIST: sEXism Identification in Social neTworks. (30/10/2022) http://nlp.uned.es/exist2021/

36. Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3. Enero (2003): 993-1022.

37. GENSIM Topic Modelling for humans. (30/10/2022) https://radimrehurek.com/gensim/

38. Zhang Y, Jin R, Zhou ZH. Understanding bag-of-words model: A statistical framework.International Journal of Machine Learning and Cybernetics. 2010 12;1:43-52.

39. Rosa Colmenajero Fernández. Ética y big data (pid_00243550). Universitat Oberta de Catalunya.

40. Agustí Cerrillo Martinez. Datos Abiertos (pid_00246839). Universitat Oberta de Catalunya

41. Mahmudah, K. R., Indriani, F., Takemori-Sakai, Y., Iwata, Y., Wada, T., & Satou, K. (2021). Classification of Imbalanced Data Represented as Binary Features. *Applied Sciences*, *11*(17), 7825.

42. Infocop Online. Consejo General de Psicología en España. Violencia de género: cómo afecta a la salud de las mujeres. (10/12/2022). https://www.infocop.es/view_article.asp?id=15453

43. Laura Planas Simón. Depression and Anxiety on Miscarriages on Twitter. https://github.com/LauraPlanas/MiscarriageTwitterAnalysis

10. Anexos

El código generado para este proyecto se encuentra alojado en el siguiente repositorio de Github:

https://github.com/javierplo/Gender-Violence-in-Spain