

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

Autor: Javier Plo Moreno
Directora: Laia Subirats Maté
Profesor: Ferran Prados Carrasco
01/2023

Índice

1. Introducción y estado del arte

- a. Contexto
- b. Arquitectura
- c. Objetivos
- d. Impacto en sostenibilidad, ético-social y diversidad
- e. Metodología

2. Desarrollo del proyecto

- a. Conjunto de datos casos de violencia de género
- b. Conjuntos de datos de víctimas y agresores
- c. Conjunto de datos de redes sociales
- d. Selección de poblaciones objetivo y matching

3. Experimentos y resultados

- a. Modelos conjunto de casos violencia de género
- b. Modelos conjuntos de víctimas y agresores
- c. Modelos conjunto de datos redes sociales
- d. Topic modelling
- e. Similitudes entre textos

4. Conclusiones y trabajo futuro

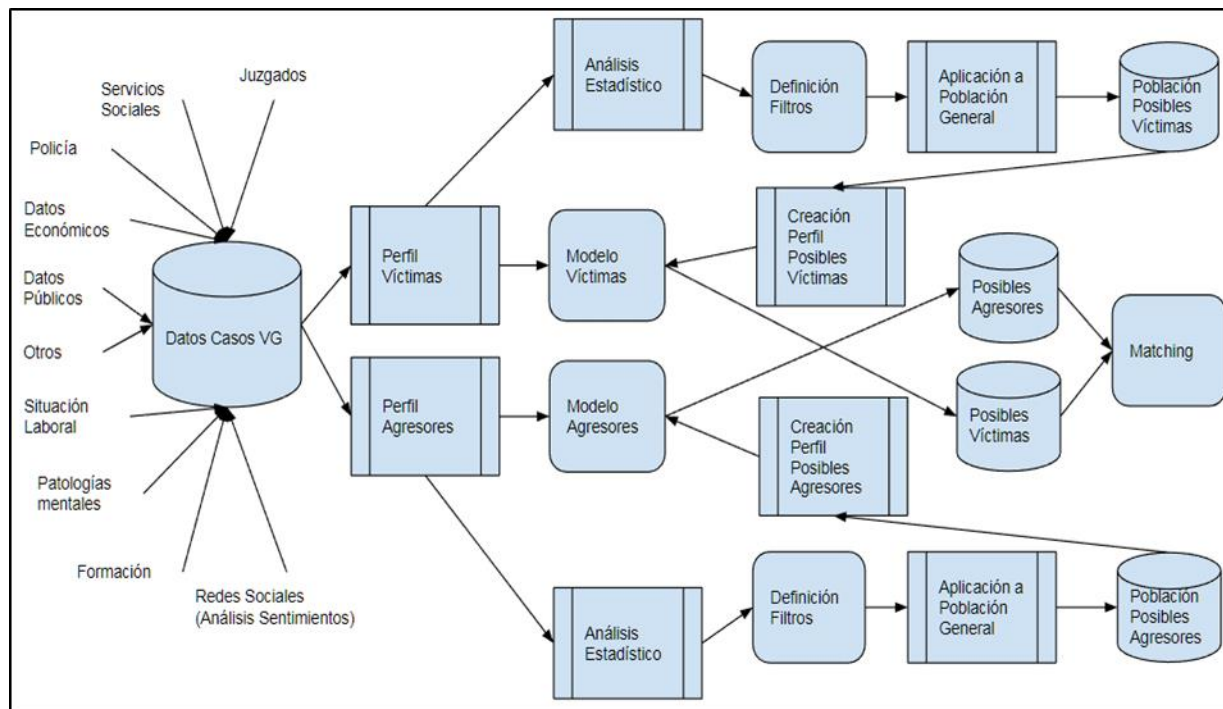
- a. Conclusiones
- b. Trabajo futuro

Introducción y estado del arte

Contexto

- La violencia de género es uno de los problemas más sangrantes de la sociedad actual. La tasa de víctimas de violencia de género en el año 2021 fue de 1,4 por cada 1000 mujeres mayores de 14 años, y durante el último año fueron asesinadas 34 mujeres en España
- Existen sistemas de ayuda a las víctimas en España como VIOGEN y ATENPRO, pero ninguno que plantee el hecho de vincular posibles agresores y víctimas con antelación.
- Los marcadores que intervienen en la aparición de la violencia de género, tanto a nivel de agresor como de víctima conforman ambos perfiles.
- Buena parte de dichos marcadores se basan en información personal, y por tanto privada, por lo que las implicaciones éticas y de protección de datos son un punto esencial a tener en cuenta en la solución.

Arquitectura del sistema



Objetivos

- Definir la arquitectura de un sistema que ayude a la detección previa de casos de violencia de género.
- **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.
- Construir **modelos** para **víctimas y agresores**, basados en la información proporcionada durante la evolución de los casos de violencia de género ya registrados
- Realizar un **análisis de datos** provenientes de redes sociales, que permita identificar publicaciones **sexistas** y topics relacionados con la violencia de género

Impacto en sostenibilidad, ético-social y diversidad

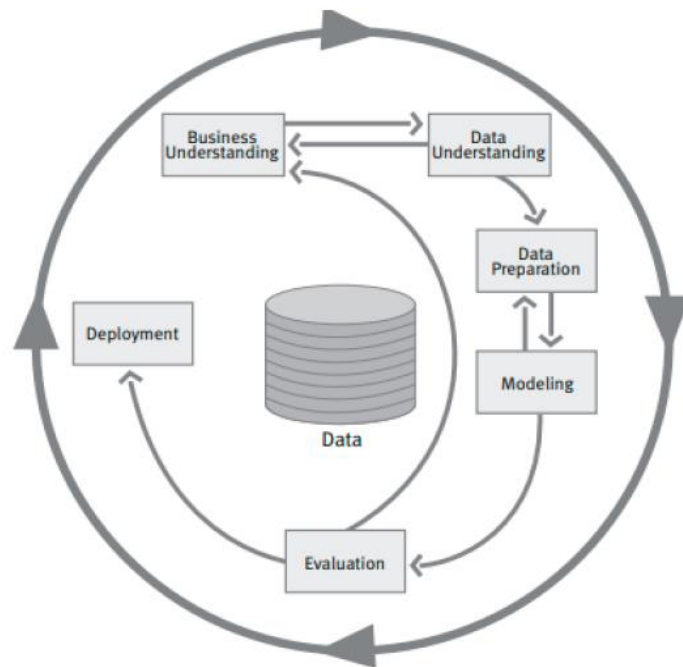
- Impacto Negativo sostenibilidad: Consumo energético equipos informáticos necesarios
- Impacto Positivo diversidad: Definición de modelos tanto para posibles víctimas como para posibles agresores, sin sesgos, estereotipos ni preeminencias masculinas ni en las imágenes ni en el discurso.
- Impacto Negativo comportamiento ético y de responsabilidad social : El manejo de información confidencial y el hecho de “**señalar**” a posibles víctimas y agresores
- Impacto Negativo ético social. Selección de posibles agresores (sesgo): *“La mayoría de los hombres no son violentos, pero es cierto que la mayoría de las personas violentas son hombres”*

Metodología

CRISP-DM

Cross Industry Standard Process for Data Mining

1. Comprensión del negocio
2. Comprensión de los datos
3. Preparación de los datos
4. Modelado
5. Evaluación del modelo
6. Despliegue



Fases de CRISP-DM

Estado del arte

- Aproximaciones al problema centradas en predicciones de evolución de casos o sistemas de ayuda a la decisión con ámbitos de aplicación reducidos, no en la detección previa.
- Numerosas publicaciones sobre violencia de género a nivel médico, sociológico, psicológico etc.
- Estudios previos relacionados con los factores que intervienen en la aparición de la violencia de género
- Estudios y soluciones donde se utilizan técnicas de Machine Learning para tratar de detectar el sexismo.

Desarrollo del proyecto

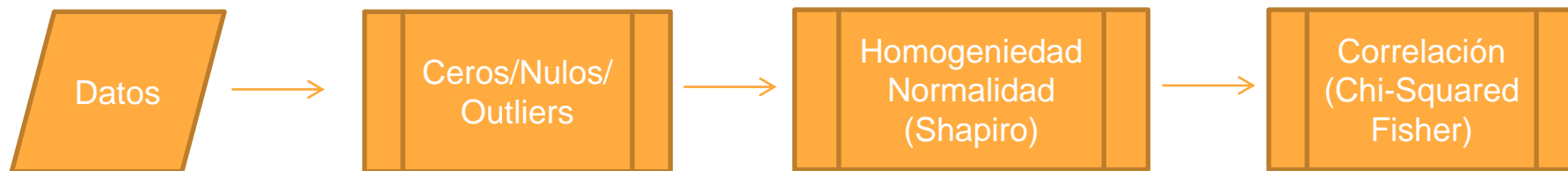
Conjunto de datos casos de violencia de género

- 900 casos con 10 características obtenidos de <https://data.world/pablosdt> y que provienen de la web de estadística de violencia de género en España
- Determinar variables que influyen más sobre Abuso Previo
- Agrupaciones de datos:
 - Comunidad Autónoma/Meses calurosos
 - Agresores jóvenes
 - Agresores adultos
 - Agresores mayores
 - Convivientes
 - Relation

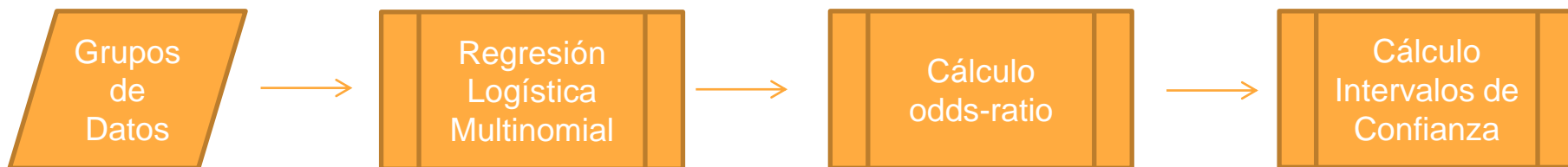
Característica	Tipo
Year	Integer
Month	Text
Autonomous Comunity	Text
Province	Text
Relation	Text
Victim age (interval)	Text
Agressor age (interval)	Text
Previous Abuse Report	Text
Living Together	Text
Victims	Integer

Conjunto de datos casos de violencia de género

Tratamiento de los datos y correlación



Modelado y evaluación



Conjuntos de datos de víctimas y agresores (marcadores)

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	

Conjuntos de datos de víctimas y agresores (variables)

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

- **Vector de violencia:** Conjunto de valores que toman las variables que categorizan los tipos de violencia definidos para cada agresor o víctima.

Conjuntos de datos de víctimas y agresores. Creación de modelos

Tipo/Estrategia	Algoritmo
Multilabel nativo	Random Forest
	Multilayer Perceptron
	Decision 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)

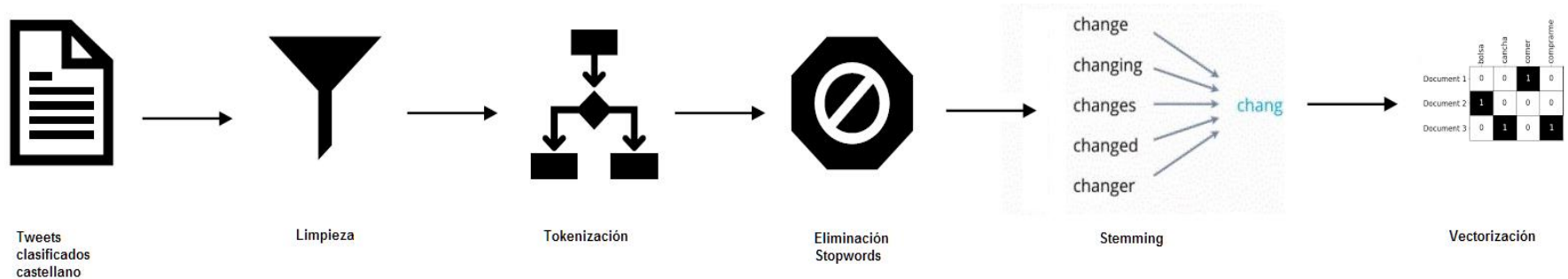
- Clasificadores en función de su capacidad de tratamiento multilabel nativa o no y otras estrategias.
- Datos: Conjunto de datos violencia de género + datos aleatorios (10000 registros)

Conjunto de datos de redes sociales

- **EXIST: sEXism Identification in Social neTworks** : Detección y clasificación del sexismo a partir de información publicada en Twitter y Gab
 - **Identificación**
 - **Sexista:** *“Mujer al volante, tenga cuidado”*
 - **No Sexista:** *“Alguien me explica que zorra hace la gente en el cajero que se demora tanto”*
 - **Categorización**
 - **Ideológico y desigualdad:** *“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:** *“A las mujeres hay que amarlas...solo eso... Nunca las entenderás.”.*
 - **Objetificación:** *“Pareces una puta con ese pantalón” - Mi hermano de 13 cuando me vio con un pantalón de cuero”.*
 - **Violencia sexual:** *“#MeToo Estas 4 no han obtenido su objetivo. El juez estima que se abrieron de patas”*

Conjunto de datos de redes sociales. Transformación

- EXIST2021:** 6977 tweets para entrenamiento y 3386 tweets para test en inglés y castellano, identificados en sexistas/no sexistas y categorizados (sexistas).



Conjunto de datos de redes sociales : Creación de modelos y evaluación

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

- Modelos de clasificación **binaria** y **multiclase** para la **identificación** y **categorización** del sexismo con las estrategias de scoring correspondientes (scoring propuesto por EXIST2021)

Conjunto de datos de redes sociales : Topic modelling

- Obtención de topics: **LDA** (Latent Dirichlet Allocation)
- Similitud entre textos: **Jensen-Shannon** aplicada a las distribuciones de los topics del modelo LDA

Conjunto de datos de redes sociales. Topic Modelling.

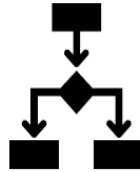
Transformación de datos



Tweets training-test
clasificados castellano



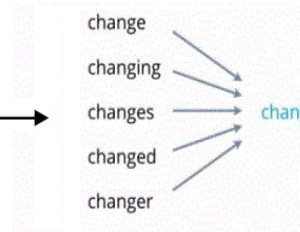
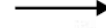
Limpieza



Tokenización



Eliminación
Stopwords



Stemming



I love this movie! It's sweet,
but with a little humor. The
dialogue is great and the
advertisements are fun...
It manages to be whimsical
and romantic while laughing
at the conventions of the
fairytale genre. I would
recommend it to just about
anyone. I've seen it several
times, and I'm always happy
to see it again whenever I
have a friend who hasn't
seen it yet!

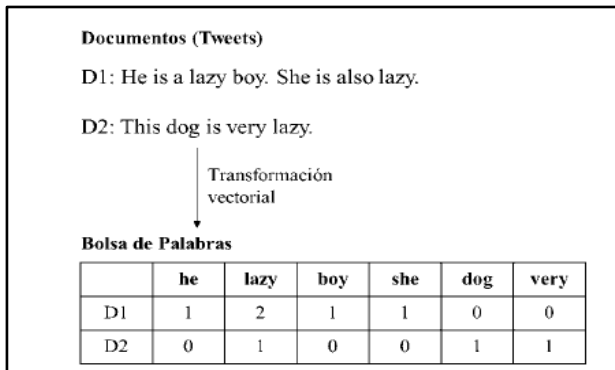


Diccionario y BOW

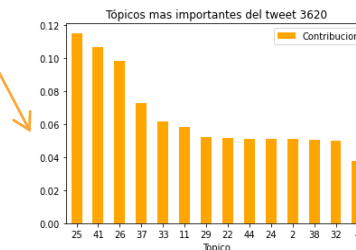
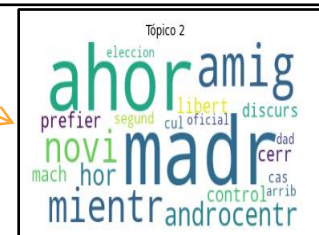
```
I 1
love 1
this 1
movie! 1
It's 1
sweet, 1
but 1
with 1
a 1
little 1
humor. 1
The 1
dialogue 1
is 1
great 1
and 1
the 1
advertisements 1
are 1
fun... 1
It 1
manages 1
to 1
be 1
whimsical 1
and 1
romantic 1
while 1
laughing 1
at 1
the 1
conventions 1
of 1
the 1
fairytale 1
genre. 1
I 1
would 1
recommend 1
it 1
to 1
just 1
about 1
anyone. 1
I've 1
seen 1
it 1
several 1
times, 1
and 1
I'm 1
always 1
happy 1
to 1
see 1
it 1
again 1
whenever 1
I 1
have 1
a 1
friend 1
who 1
hasn't 1
seen 1
it 1
yet!
```

Conjunto de datos de redes sociales. Topic Modelling.

Creación del modelo



```
(67, '0.889**aunq' + 0.872**'menaz' + 0.844**'total' + 0.844**'accept' + 0.835**'particip')
(58, '0.114**'tip' + 0.863**'mensreading' + 0.861**'cas' + 0.844**'licita' + 0.837**'dreg')
(5, '0.113**'cort' + 0.878**'fald' + 0.853**'frec' + 0.858**'larg' + 0.831**'reest')
(49, '0.121**'violacion' + 0.872**'fot' + 0.849**'conden' + 0.848**'segun' + 0.834**'nacional')
(14, '0.128**'nuev' + 0.128**'llev' + 0.895**'man' + 0.878**'man' + 0.842**'quien')
(19, '0.859**'parej' + 0.895**'final' + 0.841**'includ' + 0.831**'alo' + 0.830**'san')
(54, '0.848**'consider' + 0.844**'ider' + 0.834**'gobien' + 0.827**'trung' + 0.827**'pedes')
(31, '0.879**'vist' + 0.872**'car' + 0.843**'pag' + 0.835**'sirv' + 0.833**'ofal')
(49, '0.897**'clar' + 0.872**'vec' + 0.836**'violador' + 0.834**'poc' + 0.829**'dond')
(17, '0.186**'abort' + 0.892**'public' + 0.845**'dar' + 0.840**'mor' + 0.834**'celebr')
(37, '0.125**'tan' + 0.119**'bien' + 0.866**'muo' + 0.845**'homb' + 0.840**'conoc')
(59, '0.113**'sida' + 0.893**'mujer' + 0.888**'hoy' + 0.839**'pod' + 0.835**'ve')
(21, '0.249**'put' + 0.142**'hij' + 0.114**'parec' + 0.838**'respet' + 0.834**'hermos')
(29, '0.126**'buen' + 0.881**'va' + 0.858**'tet' + 0.858**'not' + 0.855**'gord')
(33, '0.178**'chic' + 0.878**'mas' + 0.851**'cos' + 0.841**'hac' + 0.839**'ser')
(25, '0.252**'muj' + 0.181**'ser' + 0.849**'gan' + 0.839**'quien' + 0.824**'flore')
(18, '0.888**'muro' + 0.851**'hag' + 0.843**'pues' + 0.841**'mej' + 0.838**'pa')
(1, '0.897**'asi' + 0.851**'nadi' + 0.849**'cul' + 0.842**'combi' + 0.841**'pued')
(41, '0.898**'dic' + 0.888**'habl' + 0.865**'si' + 0.842**'pues' + 0.834**'embaraz')
(5, '0.138**'femin' + 0.114**'mujer' + 0.851**'hac' + 0.844**'homb' + 0.833**'derech')
```



Selección de poblaciones objetivo y matching

- **Selección de poblaciones objetivo**
 - Ubicación geográfica de los casos.
 - Temporalidad de los casos.
 - Estadísticas ya existentes.
- **Matching entre posibles agresores y posibles víctimas**
 - Vector de violencia similar.
 - Datos registro civil, padrón, certificados de convivencia etc.
 - Vínculos en redes sociales y laborales.
 - Proximidad geográfica.
 - Tramites de divorcio con custodia compartida.
 - Órdenes de alejamiento.

Consideraciones éticas y de privacidad

- “Sujeto moral” y “Sujeto responsable”
- **Ética**
 - Propiedad de los datos.
 - Privacidad.
 - Identificar al sujeto moral.
- **Algoritmo:** Señalará posibles agresores y posibles víctimas.
 - Responsabilidad.
 - Explicación.
 - Exactitud.
 - Auditoría.
 - Equidad.
- **Principios:** Beneficencia, autonomía, no maleficencia, justicia .
- **Legislar** para el acceso a información personal.

Experimentos y resultados

Modelos conjunto de casos de violencia de género

Variable	F Value	Pr(>F)
Year	1.0172	0.4332
Month	1.029	0.4181
Autonomous Community	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

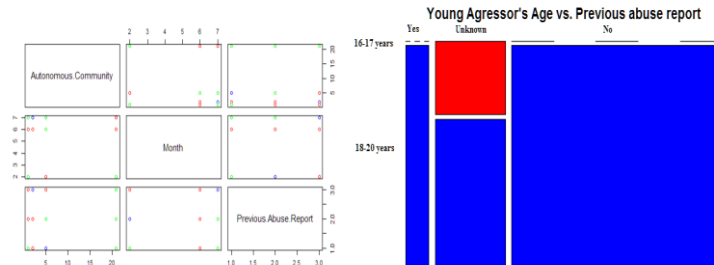
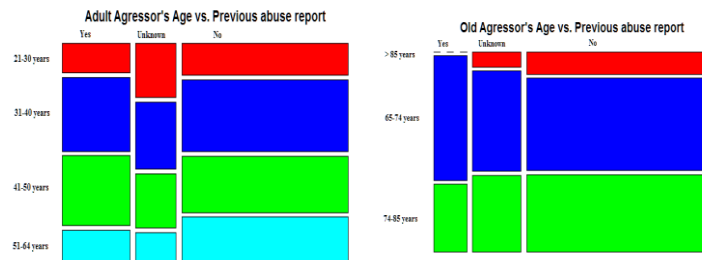


Tabla de Contingencia	P Value	Corr.(p_value<=0.05)
Comunidad Autónoma / Mes	0.981	No
Jóvenes / PAR	0.307	No
Adultos / PAR	0.05988	No
Mayores / PAR	0.6504	No
Convivientes / PAR	2.2e-16	Si
Relation / PAR	0.0001419	Si



Resultados pruebas de Fisher

Tablas de contingencia

Test de Levene entre Victims y el resto de variables

Modelos conjunto de casos de violencia de género. Regresión logística multinomial

model.vgenero.convivientes.yes (Abuso.Previo)	Variables Independientes	B _i (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

- **Odds ratio:** Se interpreta cada una de las variables independientes entre los distintos tipos de Abuso.Previo tomando como referencia Abuso.Previo = Yes.

Modelos conjuntos de víctimas y agresores

Víctimas

Tipo/Estrategia	Algoritmo	F1-macro
Multilabel nativo	Random Forest	0.5054
	Multilayer Perceptron	0.4865
	Decision 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

Agresores

Tipo/Estrategia	Algoritmo	F1-macro
Multilabel nativo	Random Forest	0.5121
	Multilayer Perceptron	0.4971
	Decision Tree	0.4971
	Nearest Neighbours	0.5088
	Extended Gradient Boosting	0.5036
Multilabel no nativo (Multioutput)	Logistic Regression	0.5092
	Gaussian Naive Bayes	0.5083
	Linear SVC	0.5093
	Binary Relevance	0.5078
Problem Transformation	Classifier Chains	0.5051
	Label Powerset	0.4959
	MLkNN	0.5036
MultiLearn Adapt	BRkNNaClassifier	0.4981
	MLTSVM	0.6638
Multilabel Deep Learning	Problem Transformation con BinaryRelevance (Keras)	0.4797

Se observa que los resultados son ligeramente mejores en los modelos implementados para los **agresores** cuyo conjunto de datos tiene algunos atributos más.

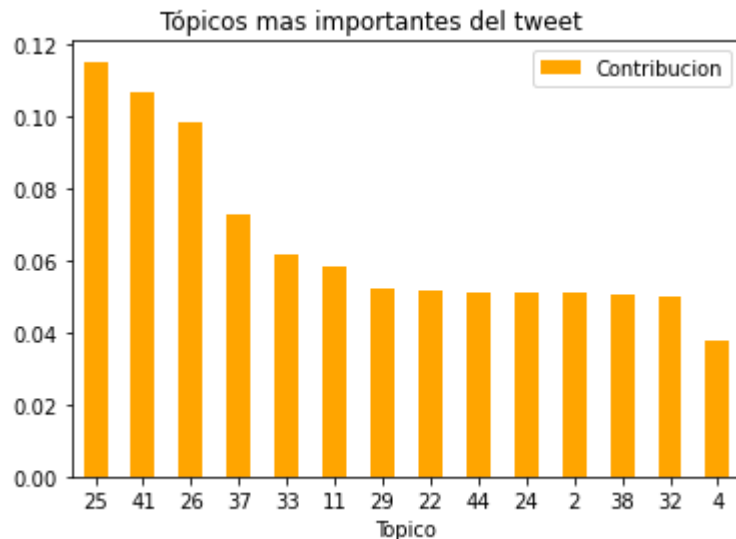
Modelos conjunto de datos en redes sociales. Identificación y clasificación del sexismo

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

Los resultados obtenidos son algo inferiores a los más destacados mostrados en el ranking de EXIST2021 para el caso de la identificación (Accuracy: 0.78) y mejoran sensiblemente con respecto a la categorización (F1-macro: 0.57)

Topic modelling

Tweet (al azar) : *“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”*



```

*** 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, cog, 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, prefer, libert, mach

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

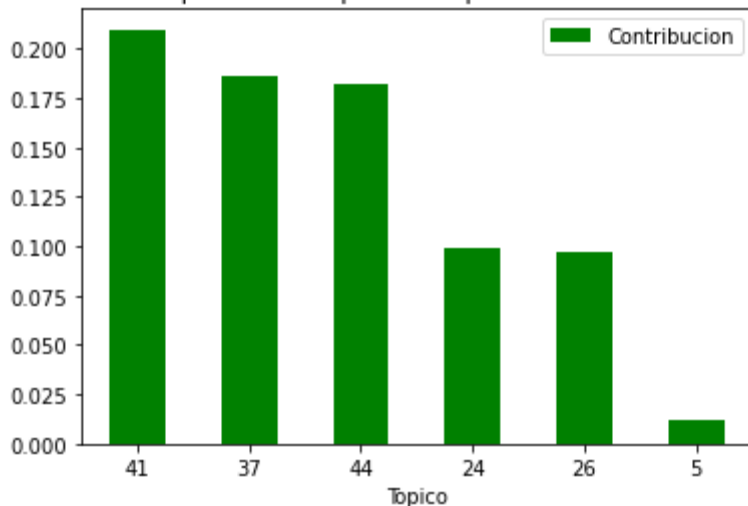
*** Tópico: 32 ***
vist, car, pag, sirv, ojal, golp, pur, disfrut, atencion, complet

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

Topic modelling. Topics relevantes

Tweet nuevo : “@test. Una feminazi maltratada por un hombre olvida ideales si quiere que se la metan”

Tópicos más importantes para tweet nuevo



```
*** Tópico: 41 ***
dic, habl, sí, 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
```

Topics más relevantes del tweet nuevo: 41, 37 y 44. En ellos se puede apreciar la aparición de los términos **feminazi** y **hombre**

Topic modelling. Similitud entre textos

Tweet nuevo : “@test. Una feminazi maltratada por un hombre olvida ideales si quiere que se la metan”

```
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)
```

- Cinco tweets más similares al **tweet nuevo**.
- Similitud calculada utilizando la distancia de **Jensen-Shannon**

Conclusiones y trabajo futuro

Conclusiones

- **Definición de la arquitectura y sensibilización**
- **Conjunto de datos de casos de violencia de género**
- **Conjunto de datos víctimas y agresores**
 - Definición de perfiles.
 - Definición de criterios de selección de poblaciones objetivo y matching.
 - Creación de modelos multilabel .
 - Agresores f1-macro: 0,5117
 - Víctimas f1-macro: 0,4885
- **Conjunto de datos identificación y categorización del sexismo**
 - Creación de modelos de clasificación binaria y multiclase.
 - Identificación Accuracy: 0,71. EXIST2021 (0.78)
 - Categorización f1-macro: 0,62. EXIST2021 (0.57)
- **Topic modelling y similitud entre textos**
 - Creación de modelo LDA y prueba de similitud.
- **Análisis de implicaciones éticas y de privacidad**

Trabajo futuro

- **Victimas y agresores**
 - Revisión marcadores y variables definidas.
 - Profundizar en los criterios para selección de poblaciones objetivo y vinculación posterior (matching).
- **Detección** en redes sociales de las patologías que sufren las víctimas.
- **Definición** de un sistema **ETL** a nivel estatal para recopilar la información de los casos.
- **Ampliar** la detección y clasificación del sexismo a otras lenguas del estado y variantes del castellano (Sudamérica y Centroamérica).
- Estudio en profundidad de las **implicaciones legales** a nivel de privacidad y protección de datos.

FORMAR TRANS- FORMAR

UOC

¡Muchas gracias!

 UOC.universitat

 @UOCuniversidad

 UOCuniversitat



Universitat Oberta
de Catalunya

25 años