



TRABALLO FIN DE GRAO
GRAO EN ENXEÑARÍA INFORMÁTICA
MENCIÓN EN COMPUTACIÓN



**Amosando o Lado Escuro das redes sociais:
desenvolvemento do primeiro corpus en
lingua galega para a detección de misoxinia
en Twitter e Mastodon**

Estudante: Lucía María Álvarez Crespo

Dirección: Laura Milagros Castro Souto

A Coruña, xuño de 2023.

Á miña irmá. A todas as irmás.

Agradecementos

Agradecemento en especial á Cátedra CICAS pola bolsa de axudas económicas recibida para a realización deste traballo e a Rafael Alcalde, administrador de *mastodon.gal* e antigo compañeiro da FIC, pola súa axuda desinteresada. Tampouco vos esquezo a vós: Eva Blanco, Iria Calvelo, Icia Carro, Azahara de Miguel, Jairo Díaz, Jorge Díaz, Isabel Díaz, Marta Ferre, David García, David González, Rubén Guede, Daniel Feito, Diego Leis, Íñigo López-Riobóo, David Martos, Elena Martínez, Victoria Martínez, Beatriz Molina, David Novoa, Francisco Pascual, Paloma Piot, Delfina Ramos, Laura Revenga, Alex Rodríguez, Samuel Suárez, Aida Vidal e Juanma Zapata; que aínda que non o saibades sodes cruciais na parte emocional deste proxecto. Grazas por axudar xa sexa tomándonos un café, saíndo de festa, debatendo nas xuntanzas de *Hello Sisters!*, xogando ao *CS:GO* ou simplemente queréndome dende a distancia. Pero especialmente a ti, Alex, por ser o meu compañeiro dende aqueles chistes malos cos *algoritmos voraces* de segundo de carreira e aínda así decidir quedarte. Non quero rematar isto sen facer unha mención especial a Laura Castro, a miña titora, que non dubidou nunca de min e das miñas capacidades. Grazas, Laura, por facer de novo posíbel o imposíbel.

Resumo

O presente Traballo Fin de Grao constitúe un proxecto de investigación que ten como obxectivo o desenvolvemento do primeiro corpus en lingua galega para a detección de misoxinia en Twitter e Mastodon. Buscamos recoller e analizar datos lingüísticos en galego presentes nestas redes sociais co fin de comprender e identificar a manifestación da misoxinia na comunicación dixital. O proceso de desenvolvemento do corpus consta de varias etapas. Comeza coa recollida dos datos en galego a través das plataformas de Twitter e Mastodon, seguido pola selección e normalización dos textos. A continuación, realizamos unha limpeza exhaustiva dos datos para preparalos para a análise e estudo posterior. Un aspecto destacado deste proxecto é a aplicación de técnicas de aprendizaxe automática para adestrar modelos capaces de identificar e clasificar de forma precisa a presenza de misoxinia nos textos analizados. Estes modelos contribuirán á detección temperá da misoxinia, permitindo intervir e prever comportamentos prexudiciais. O corpus resultante será estruturado e organizado de forma eficiente, facilitando a consulta e análise por parte de equipos de investigación con interese en abordar diversos aspectos da lingua galega relacionados coa misoxinia. A dispoñibilidade deste corpus proporcionará unha valiosa fonte de datos para aprofundar na comprensión deste fenómeno na comunidade galegofalante. En resumo, o desenvolvemento deste corpus en lingua galega para a detección de misoxinia en Twitter e Mastodon representa un avance significativo na investigación científica, permitindo a análise sistemática e exhaustiva dos datos lingüísticos en galego relacionados con este fenómeno. Este proxecto contribuirá á creación dunha sociedade máis inclusiva, fomentando o respecto e a igualdade na comunicación dixital en lingua galega.

Abstract

The present Bachelor's Thesis constitutes a research project aimed at developing the first corpus in the Galician language for the detection of misogyny on Twitter and Mastodon. We aim to collect and analyze linguistic data written in Galician on these social media platforms in order to understand and identify the manifestation of misogyny in digital communication. The corpus development process consists of several stages. It begins with the collection of data in Galician from Twitter and Mastodon platforms, followed by the selection and normalization of the Galician texts. Subsequently, a thorough data cleaning is performed to prepare them for further analysis and study. An important aspect of this project is the application of Machine Learning techniques to train models capable of accurately identifying and classifying the presence of misogyny in the analyzed texts. These models will contribute to the

early detection of misogyny, enabling intervention and prevention of harmful behaviors. The resulting corpus will be structured and organized in an efficient manner, facilitating its consultation and analysis by researchers interested in addressing various aspects of the Galician language related to misogyny. The availability of this corpus will provide a valuable source of data to deepen the understanding of this phenomenon in the Galician-speaking community. In summary, the development of this Galician language corpus for the detection of misogyny on Twitter and Mastodon represents a significant advancement in scientific research, allowing for systematic and comprehensive analysis of linguistic data related to this phenomenon in Galician. This project will contribute to the creation of a more inclusive society, promoting respect and equality in digital communication in the Galician language.

Palabras chave:

- Misoxinia
- Corpus
- Aprendizaxe automática
- Procesamento de linguaxe natural
- Galego
- Twitter
- Mastodon

Keywords:

- Misogyny
- Corpus
- Machine Learning
- Natural Language Processing
- Galician
- Twitter
- Mastodon

Índice Xeral

1	Introdución	1
1.1	Motivación e xustificación	1
1.2	Obxectivos do traballo	3
1.2.1	Fases da investigación	4
1.2.2	Obxectivos concretos	5
1.3	Planificación	5
1.3.1	Metodoloxía	6
2	Descrición do problema	9
2.1	Estado da arte	9
2.2	Desafíos asociados á lingua galega	11
3	Desenvolvemento do corpus	14
3.1	Fases necesarias	14
3.2	Colleita de datos	15
3.2.1	Colleita de mostras non misóxinas	16
3.2.2	Colleita de mostras misóxinas	22
3.3	Descrición do conxunto de datos proposto	32
3.3.1	Estrutura	32
3.3.2	Tamaño	33
3.3.3	Publicación	34
4	Validación do corpus	37
4.1	Preprocesamento do conxunto de datos seleccionado	37
4.1.1	Limpeza das mostras procedentes de Mastodon	38
4.1.2	Limpeza das mostras procedentes de Twitter	40
4.1.3	Preprocesamento común das mostras	41
4.1.4	Creación dos embeddings	43

4.2	Experimentos realizados	44
4.2.1	Iteración 1: RF, SVM e LSVM	44
4.2.2	Iteración 2: undersampling con Random Undersampling	47
4.3	Resultados e comparativas	49
4.3.1	Resultados da iteración 1	50
4.3.2	Resultados da iteración 2	51
5	Seguimento	54
5.1	Control de tarefas	54
5.2	Custos do proxecto	57
6	Conclusións	60
6.1	Obxectivos alcanzados	60
6.2	Aplicabilidade	61
6.3	Leccións aprendidas	62
6.4	Traballo futuro	63
	Relación de Acrónimos	65
	Glosario	67
	Bibliografía	69

Índice de Figuras

1.1	Enquisa “The state of online harassment” (fonte: [1])	2
1.2	Diagrama de Gantt proposto para o proxecto	7
2.1	Exemplo de chío misóxico pasivo non incluído no corpus (fonte: [2])	10
3.1	Fase 1 do fluxo de traballo	14
3.2	De esquerda a dereita: centralizada, federada e distribuída (fonte: [3])	28
3.3	Exemplo de conversa na que se gardaba o primeiro chío e non o segundo, que é o relevante	28
3.4	Fase 2 do fluxo de traballo	32
3.5	Mostras misóxicas vs. non misóxicas do dataset proposto	34
3.6	O chío coa palabra <i>puta</i> está clasificado no MisoCorpus como misóxico, mais no contexto vemos que en realidade ten un ton satírico	35
4.1	Fase 3 do fluxo de traballo	37
4.2	Fase 4 do fluxo de traballo	44
4.3	Fase 5 do fluxo de traballo	49
5.1	Diagrama de Gantt do proxecto	56
6.1	Exemplo de chío enviado polo ParityBot	62

Índice de Táboas

4.1	Resumo dos hiperparámetros utilizados	45
4.2	Métricas da iteración 1	51
4.3	Métricas iteración 2	53
5.1	Custos totais estimados	57
5.2	Custos de persoal asociados ao proxecto	59
5.3	Especificacións técnicas das máquinas	59

Introdución

Ao longo deste primeiro capítulo abordarase o marco teórico e contextual necesario para comprender profundamente a problemática da **misoxinia** nas redes sociais, cun especial enfoque en **Twitter** e **Mastodon** en lingua galega. Examinaranse conceptos chave relacionados co **Procesamento de Linguaxe Natural (NLP)**, a detección de contido ofensivo e a **discriminación** por razón de xénero, proporcionando así os fundamentos esenciais para a comprensión e análise da detección de **misoxinia** nestas contornas virtuais. Alén diso, salientarase a importancia crítica de contar cun **corpus** específico en lingua galega e describirase a metodoloxía empregada para a súa construción, sentando así os alicerces para o desenvolvemento do sistema de detección automática que será abordado nos seguintes capítulos.

1.1 Motivación e xustificación

Nos últimos anos, as redes sociais veñen experimentado un auxe exponencial no seu uso, mudando en espazos virtuais onde millóns de persoas comparten ideas, opinións e experiencias. Porén, esta ampla interacción social tamén expón a existencia de comportamentos negativos e prexudiciais, como o acoso e a **misoxinia**. A **misoxinia**, particularmente, é un fenómeno preocupante que perpetúa a desigualdade de xénero e socava a dignidade e os dereitos das mulleres.

As redes sociais son as plataformas en liña máis mencionadas tanto por homes como por mulleres como escenario de acosos sufridos. Segundo un estudo realizado polo Pew Research Center¹ (centro de investigación que realiza estudos e análises sobre tendencias sociais, demografía, opinión pública e outros eidos a nivel global), aproximadamente catro de cada dez estadounidenses teñen experimentado acoso en liña [1]. Este estudo revelou resultados significativos en relación ao xénero, mostrando que as mulleres son máis propensas que os homes a denunciar casos de acoso, tanto sexual (16% fronte a 5%) como xeral (13% fronte 9%). Entre as

¹<https://www.pewresearch.org/>

mulleres afectadas, salienta que o 33% das mulleres menores de 35 anos sufriron acoso sexual en liña, en comparación co 11% dos homes da mesma idade. Entre os adultos acosados en liña, case a metade das mulleres (47%) acreditan que foron vítimas de acoso en liña debido á súa condición de muller, en comparación co 18% dos homes acosados que opinan o mesmo (ver figura 1.1). Estes resultados subliñan a necesidade de abordar e comprender a problemática do acoso en liña, especialmente no que respecta ás mulleres, co fin de promover unha maior seguridade e benestar nas plataformas dixitais.

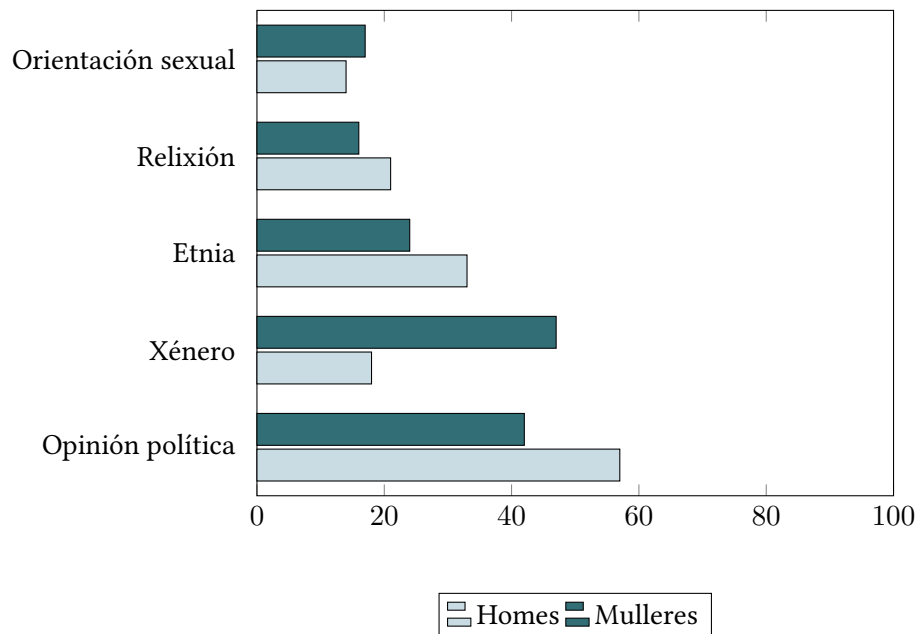


Figura 1.1: Enquisa “The state of online harassment” (fonte: [1])

A **misoxinia**, definida como o odio ou o prexuízo contra as mulleres, pode manifestarse de diversas maneiras a nivel lingüístico, como a exclusión social, a **discriminación**, a hostilidade, as ameazas de violencia e a cosificación sexual [4]. A **misoxinia** en liña foi comparada coa caza de bruxas, xa que amosa unha función semellante: coaccionar ás mulleres para evitar que progresen [5]. Este tipo de violencia, tamén coñecida como “Violencia cara as mulleres na política” (**Violence Against Women In Politics (VAWIP)** polo seu acrónimo en inglés) [6, 7], afecta ás mulleres de éxito, xa que sofren ataques sexistas motivados tanto polo seu xénero como pola súa visibilidade pública. Neste contexto, o **Procesamento de Linguaxe Natural** emerge como unha ferramenta prometedora para analizar e comprender o discurso das redes sociais. O **NLP** mestura técnicas de lingüística computacional, **aprendizaxe automática** e procesamento de datos para extraer información valiosa de grandes volumes de texto [8]. A aplicación destas técnicas ao estudo da **misoxinia** nas redes sociais permite identificar **padróns**, tenden-

cias e manifestacións específicas deste fenómeno, o que á súa vez pode contribuír á adopción de medidas preventivas e á sensibilización social.

Por unha banda, as tecnoloxías de **análise de sentimentos** teñen un gran potencial para extraer información crítica sobre as opinións compartidas nas redes sociais que poden axudar a identificar o discurso de odio e a **discriminación** [9]. Estas tecnoloxías foron aplicadas en múltiples tarefas de clasificación de textos, como a ironía [10] ou a detección do discurso de odio [11]. Se consideramos que a **misoxinia** é unha forma de discurso de odio, entón os detectores de discurso de odio deberían funcionar perfectamente ben ao analizar texto que contén trazas misóxinas. Porén, nalgúns casos, a **misoxinia** é moi sutil e profunda, polo que pode non ser doada de identificar. Ademais, as diferenzas culturais e de contexto poden complicar ese labor de identificación. A pesar destas dificultades, a identificación automática da **misoxinia** está gañando relevancia como unha tarefa desafiante e, en consecuencia, foi proposta en conferencias relacionadas co **NLP** debido á súa relevancia social [12, 13].

Para rematar, é importante destacar que a maior parte da investigación en **NLP**, **análise de sentimentos** e discurso de odio centrouse en idiomas como o inglés e, en menor medida, o español. Isto implica que as linguas minorizadas, como o galego neste caso, quedan fóra do alcance destas investigacións, o que pode dificultar a extrapolación dos resultados e técnicas desenvolvidas para estes idiomas específicos. É fundamental abordar a investigación e desenvolvemento de ferramentas de detección de **misoxinia** en linguas minorizadas, como o galego, para poder comprender e abordar axeitadamente este fenómeno en contextos lingüísticos específicos e promover unha maior inclusión na **análise de sentimentos** e detección de discurso de odio. Isto require adaptar e mellorar as técnicas existentes e desenvolver recursos específicos para estas linguas minorizadas.

1.2 Obxectivos do traballo

Neste **Traballo de Fin de Grao (TFG)**, abordarase a tarefa de detección da **misoxinia** nunha das redes sociais máis populares, **Twitter** [14], así como na súa alternativa libre, **Mastodon**. Estas dúas plataformas ofrecen un amplo espazo para a interacción social e a expresión de opinións, xogando co anonimato como unha característica destacada [15], razón pola que constitúen fontes valiosas de datos para a análise. Ademais, dado o enfoque específico deste estudo na lingua galega, prestarase especial atención á detección da **misoxinia** nas mensaxes escritas nesta lingua.

Poñendo de manifesto a necesidade de dispor de recursos lingüísticos específicos para a detección da **misoxinia** na lingua galega, a achega máis relevante deste **TFG** é a **creación do primeiro corpus en lingua galega para a detección da misoxinia**. Este **corpus**, integrado por un conxunto de textos recompilados nas redes sociais **Twitter** e **Mastodon**, constituirá

unha base de datos fundamental para o adestramento e avaliación dos modelos de [aprendizaxe automática](#) desenvolvidos neste traballo. Resulta especialmente salientable que, no momento da realización deste estudo, non se atopou ningún [corpus](#) dispoñíbel en lingua galega para a detección da [misoxinia](#), o que confire unha importancia pioneira a este traballo.

O obxectivo secundario deste TFG é desenvolver un sistema de detección automática de [misoxinia](#) nas mensaxes de [Twitter](#) e [Mastodon](#) en lingua galega, facendo uso do [corpus](#) creado como un recurso fundamental. Para acadar este obxectivo, aplicaranse técnicas de [NLP](#). Asemade, exploraranse e avaliaranse diferentes enfoques e algoritmos de [aprendizaxe automática](#) co fin de acadar un elevado rendemento e precisión na detección da [misoxinia](#).

A través deste traballo, o obxectivo último é contribuír ao coñecemento sobre a presenza e manifestación da [misoxinia](#) nas redes sociais en lingua galega, así como proporcionar unha ferramenta útil para a identificación temperá da [misoxinia](#). Este traballo procura sentar as bases para a implementación de estratexias de intervención e concienciación que promovan unha cultura dixital máis inclusiva e respectuosa.

1.2.1 Fases da investigación

O presente [Traballo de Fin de Grao](#) consta de dúas etapas fundamentais. Na primeira etapa, desenvólvese unha recompilación automatizada de textos en lingua galega extraídos de redes sociais. Na segunda etapa, adóptanse enfoques baseados en técnicas de [Machine Learning \(ML\)](#) clásicas para demostrar a utilidade do conxunto de datos obtido.

Os exemplos reais serán obtidos de plataformas como [Twitter](#) e [Mastodon](#). A elección destas dúas redes sociais, [Twitter](#) e [Mastodon](#), reviste unha importancia crucial para obter datos representativos e significativos. En comparación con [Twitter](#), [Mastodon](#) presenta unha particularidade única e relevante para o noso estudo. Existe unha [instancia](#) específica en lingua galega chamada [mastodon.gal](#)², na que o galego é a lingua principal empregada polas persoas usuarias. Isto supón unha valiosa fonte de mostras en galego que non se atopa noutras redes sociais [16].

Ademais, nesta [instancia](#) de [Mastodon](#) os comentarios misóxinos están explicitamente prohibidos [17], o que ofrece un entorno máis seguro para a interacción entre usuarios e crea unha oportunidade valiosa para investigar a [misoxinia](#) en lingua galega. Por iso, a existencia desta [instancia](#) de [Mastodon](#) é especialmente relevante, xa que nos permite obter un conxunto de datos que representa a clase non-misóxina. Estes datos son fundamentais para o desenvolvemento dun modelo de clasificación binaria que sexa capaz de distinguir entre mensaxes misóxinas e non misóxinas. Utilizando este conxunto de datos como referencia, poderemos adestrar o modelo para que aprenda a identificar e clasificar con precisión as mensaxes que conteñen contido misóxino. Así, esta [instancia](#) de [Mastodon](#) proporciona unha valiosa contri-

²[mastodon.gal](#)

bución á investigación, permitiéndonos abordar a problemática da [misoxinia](#) na lingua galega e axudando a crear unha plataforma máis segura e respectuosa para as mulleres.

Pola súa banda, en [Twitter](#), a diferenza da [instancia](#) galega de [Mastodon](#), os comentarios misóxinos non son [moderados](#), polo que a presenza de comentarios misóxinos nesta rede é máis frecuente [18] e constituirá a fonte fundamental de mostras para a clase misóxina. En resumo, faremos uso tanto da [instancia](#) galega de [Mastodon](#) como de [Twitter](#) para recompilar un [corpus](#) diverso e representativo que nos permita realizar unha análise exhaustiva da [misoxinia](#) en lingua galega. Esta aproximación combinada permitiranos obter un conxunto de mostras rico e variado, que servirá como base para desenvolver técnicas avanzadas de [Procesamento de Linguaxe Natural](#) e contribuír ao avance do coñecemento no campo da lingüística computacional e da [análise de sentimentos](#).

Así, o [corpus](#) conterá miles de mostras concretas de comentarios sexistas e non sexistas realizados en redes sociais en lingua galega, o que permitirá o desenvolvemento de técnicas eficaces baseadas en [NLP](#) para combater a [misoxinia](#) en liña. Para demostrar a utilidade deste [corpus](#), adestraranse modelos clásicos de [Machine Learning](#), como os [Random Forest \(RF\)](#), que permitirán identificar se un texto é ou non misóxino.

1.2.2 Obxectivos concretos

En resumo, os obxectivos da nosa investigación son:

- Desenvolver un [corpus](#) en lingua galega para a detección de [misoxinia](#) en redes sociais, utilizando exemplos reais obtidos de [Twitter](#) e [Mastodon](#).
- Realizar o preprocesamento do conxunto de datos e adestrar modelos de [Machine Learning](#).
- Avaliar a eficacia dos modelos de [ML](#) adestrados co [corpus](#) desenvolvido na identificación de textos misóxinos.

1.3 Planificación

Nesta sección abordarase a planificación do presente [Traballo de Fin de Grao](#) para garantir a consecución dos obxectivos propostos. Para isto, presentaremos unha visión xeral do enfoque metodolóxico empregado, xunto coa descrición das etapas, actividades e recursos necesarios para a realización deste proxecto.

A continuación, exploraremos as tarefas que se anticipaban para a realización deste [TFG](#) nun período de entre 300 e 360 horas, correspondente ao esforzo asociado aos 12 créditos da materia:

- Planificación do proxecto
- Recompilación de datos
- Preprocesamento do corpus
- Adestramento de modelos
- Análise de resultados
- Redacción da memoria

O diagrama de Gantt proposto pode observarse na imaxe 1.2. Para o adecuado seguemento, recomendamos consultar a sección correspondente no capítulo correspondente (capítulo 5).

1.3.1 Metodoloxía

Para manter un control das tarefas realizadas durante o transcurso dos meses, óptase por seguir un ciclo de desenvolvemento de software de estilo Scrum [19] cunha única desenvolvedora, que inclúe a creación de prototipos [20]. A razón para isto foi definir entregas durante cada unha das avaliacións semanais coa titora do proxecto.

Como xa indicamos, o obxectivo principal deste traballo é elaborar un primeiro **corpus**, ou dataset, para a detección de **misoxinia** en lingua galega e, como obxectivo secundario, o adestramento de varios modelos de **aprendizaxe automática** que demostran a utilidade do dataset. Ademais dos incrementos adicionais tanto para o estudo inicial dos algoritmos de **ML** como para a súa análise e refinamento posteriores, a fronteira entre as iteracións corresponde a cada aumento e mellora do proxecto definido nas reunións semanais. Estas iteracións son chamadas sprints. Cada sprint ten un límite de tempo establecido e busca estender a funcionalidade ou o valor do produto software ao seu remate.

A través da implementación de prácticas áxiles inspiradas en Scrum xunto coa construción de prototipos, o equipo de desenvolvemento pode obter características e funcionalidades de forma iterativa e incremental. A cada reunión semanal, establecece un obxectivo novo que permite explorar diferentes enfoques e mellorar o rendemento do proxecto. Esta abordaxe promove a adaptabilidade, a colaboración e a entrega de valor de forma continua. A combinación destas dúas metodoloxías proporciona un marco sólido para un desenvolvemento eficiente e axustado ás necesidades do proxecto, permitindo así a identificación temperá de problemas e a realización de melloras constantes. En resumo, a adopción de elementos da metodoloxía Scrum baseada na construción de prototipos é unha escolla estratéxica que facilita un proceso de desenvolvemento efectivo e axustado ás necesidades deste proxecto [21].

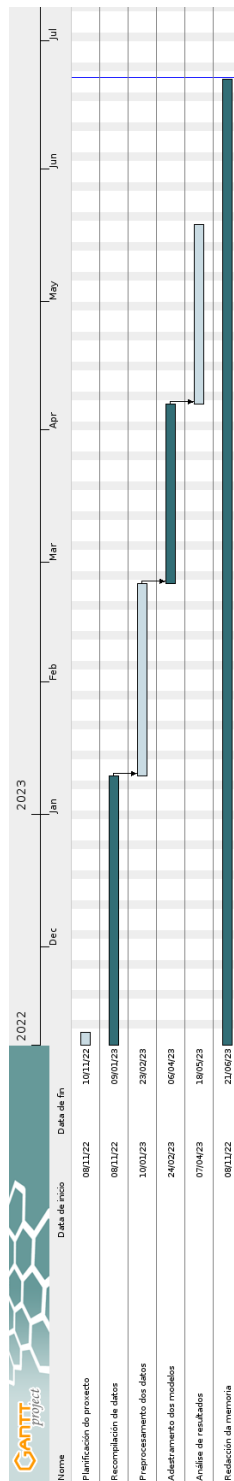


Figura 1.2: Diagrama de Gantt propuesto para el proyecto

A este ciclo de desenvolvemento engádesse o control do proxecto coas correccións que considere oportunas a titora, ademais dun control de versións en GitHub. O obxectivo principal dun repositorio de GitHub é proporcionar un lugar centralizado para almacenar, organizar e colaborar no desenvolvemento de proxectos de software utilizando o sistema de control de versións Git, así como facilitar un seguimento eficaz dos cambios diarios no código e posibles cambios de requisitos na explicación a medida que se avanza nas iteracións.

Descrición do problema

No contexto actual, onde as redes sociais se converteron nun espazo fundamental para a expresión e interacción social, a detección da [misoxinia](#) en liña adquire unha relevancia crucial. A lingua galega non escapa desta realidade e é fundamental abordar este fenómeno co rigor científico que require. É por isto que neste capítulo mergullamos nunha exhaustiva análise do estado da arte da detección de [misoxinia](#) en lingua galega en redes sociais, especialmente en [Twitter](#) e [Mastodon](#). Este capítulo serve como punto de partida para a nosa investigación, na que nos propoñemos desenvolver non só unha primeira proposta de [corpus](#) en lingua galega para este propósito, senón unha ferramenta de detección de [misoxinia](#) en lingua galega que sexa precisa, robusta e adaptábel.

2.1 Estado da arte

A detección de discursos misóxinos e comportamentos ofensivos en redes sociais presenta un desafío complexo e multidimensional. A través dunha revisión exhaustiva da literatura científica e das investigacións existentes, exploramos as abordaxes máis innovadoras e eficaces na detección da [misoxinia](#).

Nos últimos anos, moitos equipos investigadores veñen traballando na [análise de sentimentos](#) nas redes sociais, especialmente no contexto do [Twitter](#). Mediante a análise da literatura, atopamos contribucións chave que apoian a identificación de comportamentos e situacións das usuarias e usuarios en diversos contextos ao redor do mundo. Estas tecnoloxías proporcionan a capacidade de comprender as actitudes e sentimentos da cidadanía, o que incentivou enfoques en diferentes dominios. Por exemplo, Manguri *et al* [22] utilizaron a [análise de sentimentos](#) para o estudo das opinións e novidades sobre a COVID-19 no [Twitter](#).

No contexto do Xapón, realizouse un estudo exploratorio sobre as instancias de discurso de odio misóxino ou sexista e linguaxe abusiva dirixida ás mulleres políticas no [Twitter](#) ([VAWIP](#)) utilizando ferramentas e métodos computacionais de análise lingüística de [corpus](#) e

análise de sentimentos, complementados cunha análise cualitativa en profundidade do abuso verbal sufrido por catro destacadas mulleres políticas: Renhō, Tsujimoto Kiyomi, Yamao Shiori e Koike Yuriko [23]. Deste xeito, combinaron enfoques cuantitativos e estatísticos con enfoques cualitativos e hermenéuticos, obtendo unha perspectiva completa e enriquecedora do fenómeno.

A prevalencia dos abusos misóxinos nas redes sociais, como o *Twitter*, motiva considerables esforzos para abordar este problema. O alto volume e a persistencia dos *chíos* abusivos presenta desafíos tanto para as usuarias e usuarios como para os provedores destas plataformas. A detección automatizada pode axudar a identificar e combater o contido abusivo rapidamente. Neste contexto, o traballo realizado por Hewitt *et al* [24] explora a investigación previa sobre a *misoxinia* en liña, presenta un experimento que subliña os desafíos que supón a *análise de sentimentos* para a detección de *misoxinia* utilizando datos do *Twitter* e propón traballos futuros para combater eficazmente este problema.



Figura 2.1: Exemplo de *chío* misóxino pasivo non incluído no *corpus* (fonte: [2])

Unha das dificultades na detección de *misoxinia* nas redes sociais radica no feito de que estas mensaxes poden ser mensaxes *activas* ou *pasivas* dependendo de se van dirixidas ou non a algunha muller en particular. Así, o IberEval¹ organizou unha tarefa de *Automatic Misogyny Identification (AMI)* con dous *corpus* dispoñíbeis, un en castelán e outro en inglés, nos que os *chíos* de ambos *corpus* estaban etiquetados ou ben como *activos* ou *pasivos* [4]. Esta abordaxe da problemática da detección de *misoxinia* como un problema de clasificación binaria tamén foi observada no EvalIta², que tamén organizou unha tarefa de *AMI*, mais neste caso os *corpus* estaban en inglés e italiano [25].

Outra das dificultades que inflúen na detección de *misoxinia* nas redes sociais é o uso frecuente dunha linguaxe informal non sempre correctamente rexistrada nos *corpus*, o que

¹<https://sites.google.com/view/ibereval-2018>

²<https://www.evalita.it/>

subliña a necesidade dun **corpus** verdadeiramente representativo. Neste sentido, Lynn *et al* [26] empregaron o Urban Dictionary³ para recoller *slangs* relacionados coa **misoxinia** e así comparar o seu rendemento.

Na plataforma **Mastodon**, a literatura científica dispoñíbel no ámbito da **análise de sentimentos** é escasa e limitada. A pesar dos avances na **análise de sentimentos** noutras redes sociais, como o **Twitter**, existe unha notable falta de investigacións centradas especificamente na **análise de sentimentos** na rede descentralizada de **Mastodon**. Recentemente, comezan a emerxer os primeiros estudos relacionados coa **análise de sentimentos** en **Mastodon** [27, 28]. Porén, esta falta de literatura académica dedicada a este tema representa un desafío para comprender as dinámicas misóxinas que poden ocorrer entre as comunidades de **Mastodon**.

Ademais das tarefas mencionadas anteriormente, a literatura relevante tamén inclúe máis investigacións sobre a identificación da **misoxinia**, aínda que a meirande parte delas se realizou en lingua inglesa. Hai algunhas primeiras aproximacións na literatura en lingua castelá [29], mais no transcurso da realización deste traballo non se atopou ningunha en lingua galega. A falta de literatura dispoñíbel aplicada á lingua galega representa o principal reto deste traballo, que detallaremos na seguinte sección.

Como mencionamos anteriormente, un dos obxectivos deste proxecto é optimizar os resultados dun sistema de detección, concretamente un sistema de detección de **misoxinia** na lingua galega en **Twitter** e **Mastodon**. Para iso, crearemos diferentes sistemas de clasificación binaria para identificar mensaxes que conteñan contido misóxino. A idea é desenvolver un modelo de **aprendizaxe automática** capaz de identificar e clasificar con precisión estas mensaxes. É importante sinalar que, ata o momento, hai pouca literatura dispoñíbel sobre este tema en particular na lingua galega e nas redes sociais de **Twitter** e **Mastodon**. Polo tanto, este proxecto representa unha contribución significativa á investigación neste campo e á comprensión dos fenómenos de **misoxinia** nestas plataformas en lingua galega.

2.2 Desafíos asociados á lingua galega

No contexto da nosa investigación sobre a detección de **misoxinia** en lingua galega en **Twitter** e **Mastodon**, atopámonos con varias dificultades relacionadas coa lingua galega. Estas dificultades engloban aspectos como a escaseza de recursos e literatura científica dispoñíbel en galego sobre esta temática específica, así como a existencia de características propias da lingua que requiren unha atención especial no proceso de análise e clasificación. Ademais, a falta de **corpus** representativos e anotados en galego complica a tarefa de adestrar os modelos de **aprendizaxe automática** en detección de **misoxinia**. Estas dificultades sinalan a necesidade de abordar o estudo da **misoxinia** na lingua galega desde un enfoque multidisciplinar e cola-

³ <https://www.urbandictionary.com/>

borativo, buscando solucións innovadoras e adaptadas ás particularidades da nosa lingua. A pesar destes desafíos, o noso proxecto representa unha oportunidade para avanzar nesta área de investigación e contribuír ao coñecemento sobre a problemática da [misoxinia](#) na lingua galega nas redes sociais.

Existe algunha literatura que se enfoca na [análise de sentimentos](#) de [chíos](#) en galego. Con todo, é importante ter en conta que moitas veces estes estudos traducen os [chíos](#) do galego ao inglés para poder aplicar as técnicas de [análise de sentimentos](#) existentes [30]. Esta tradución ao inglés pode xerar problemas, xa que a nosa lingua ten características propias que se perden nesta tradución. De xeito máis destacado, resulta problemática a perda do xénero das palabras, que é un aspecto moi relevante na detección e interpretación da [misoxinia](#).

A creación de [corpus](#) en lingua galega para investigacións científicas presenta dificultades debido á escaseza de recursos dispoñíbeis. A falta de [corpus](#) adecuados en galego dificulta a realización de estudos e a aplicación de técnicas específicas para a [análise de sentimentos](#) e detección de [misoxinia](#). En Ortega *et al* [31] explorouse a tradución automática entre diferentes linguas, incluíndo o galego. O equipo investigador propuxo un enfoque que aproveita a proximidade entre o portugués e o galego para desenvolver un sistema de tradución automática. Esta técnica implicou a transliteración, que é a acción de transcribir palabra a palabra os termos escritos dunha lingua na outra, neste caso do portugués ao galego, e o uso de dúas arquitecturas neuronais diferentes para crear sistemas de tradución automática para os pares de linguas castelán-galego e inglés-galego. A pesar de que este estudo se enfocou na tradución de textos, a utilización da transliteración entre linguas relacionadas pode ofrecer unha estratexia valiosa tamén para a nosa investigación sobre a detección de [misoxinia](#). Ao abordar a [análise de sentimentos](#) directamente na lingua galega sen necesidade de traducións, pero si de transliteracións de linguas semellantes, podemos obter resultados máis precisos e evitar a perda de información importante, como o xénero das palabras, que como indicamos é fundamental na identificación da [misoxinia](#).

Debemos mencionar tamén a Fernández *et al* [32], que propoñen unha metodoloxía semiautomática para xerar recursos para unha aplicación de [análise de sentimentos](#) en galego, aproveitando recursos en español e utilizando o portugués como lingua intermediaria debido á súa proximidade co galego. Destacan a escaseza de recursos e aplicacións para linguas diferentes do inglés no campo do [Procesamento de Linguaxe Natural](#). Os investigadores utilizaron un corpus de críticas de películas en español e un vocabulario controlado como punto de partida, traducíndoos ao galego e ao portugués. Aplicaron o algoritmo de [Support Vector Machine \(SVM\)](#) para adestrar un módulo de [análise de sentimentos](#) e avaliaron os resultados. Con todo, non había un padrón de referencia dispoñíbel para comparar o rendemento do sistema de clasificación de sentimentos en galego e español. Por iso, construíron manualmente un [corpus](#) de proba pequeno, extraendo textos clasificados como críticas positivas ou

negativas de sitios como Google Maps⁴, Booking⁵ e a App Store de Apple⁶. Os resultados do estudo mostraron un rendemento prometedor do sistema proposto, pero necesítanse máis investigacións para unha validación completa.

A pesar dos esforzos para abordar a falta de recursos e **corpus** en lingua galega, aínda persiste o problema da falta de ferramentas e técnicas específicas para o **Procesamento de Linguaxe Natural** en galego. As tecnoloxías e métodos desenvolvidos para outras linguas, como o inglés ou o castelán, non sempre se aplican directamente ao galego debido ás súas particularidades lingüísticas. Agerri *et al* [33] describen o desenvolvemento de recursos e ferramentas de procesamento de **NLP** para o galego, incluíndo **corpus** anotados manualmente e módulos de **NLP** específicos. Aínda así, estas ferramentas e información non son todo o útiles que poderían para o noso proxecto, xa que utilizan como fontes de datos páxinas web como Wikipedia⁷ ou a Xunta de Galicia⁸. O problema é que a linguaxe utilizada nestes lugares en galego é máis formal e non reflicte necesariamente a linguaxe informal que se utiliza nas redes sociais ou noutros contextos en liña. Isto pode levar a discrepancias entre os datos recollidos da Wikipedia e a linguaxe real utilizada polos usuarios nas plataformas de redes sociais. É importante ter en conta esta diferenza e considerar a necesidade de obter datos que reflectan máis de preto a linguaxe informal que se pretende analizar neste proxecto.

En conclusión, dadas as dificultades e limitacións existentes en canto a recursos e ferramentas para a detección de **misoxinia** na lingua galega, así como as posibles perdas de información ao traducir textos ao inglés, a perspectiva de utilizar a transliteración do portugués ao galego preséntase como unha opción interesante. Ao crear un **corpus** en galego desde cero, poderemos capturar de maneira máis precisa a variante da lingua utilizada nas redes sociais e, ao mesmo tempo, preservar características lingüísticas relevantes, como o xénero das palabras. Este enfoque multidisciplinar permitiranos avanzar na investigación sobre a detección de **misoxinia** na lingua galega e contribuír ao coñecemento neste ámbito. A pesar dos desafíos que comporta, a creación dun **corpus** en galego preséntase como unha oportunidade única para comprender e abordar a problemática da **misoxinia** na nosa sociedade.

⁴<https://www.google.es/maps/preview>

⁵<https://www.booking.com/>

⁶<https://www.apple.com/es/app-store/>

⁷<https://gl.wikipedia.org/wiki/Portada>

⁸<https://www.xunta.gal/portada>

Desenvolvemento do corpus

PARA abordar o desafío da detección de **misoxinia** na lingua galega en redes sociais, é fundamental contar cun **corpus**, ou conxunto de datos, adecuado e representativo. Neste capítulo, describimos o proceso de desenvolvemento do noso conxunto de datos, que servirá como base para a nosa investigación e para adestrar os modelos de **aprendizaxe automática** na detección de contido misóxico.

3.1 Fases necesarias

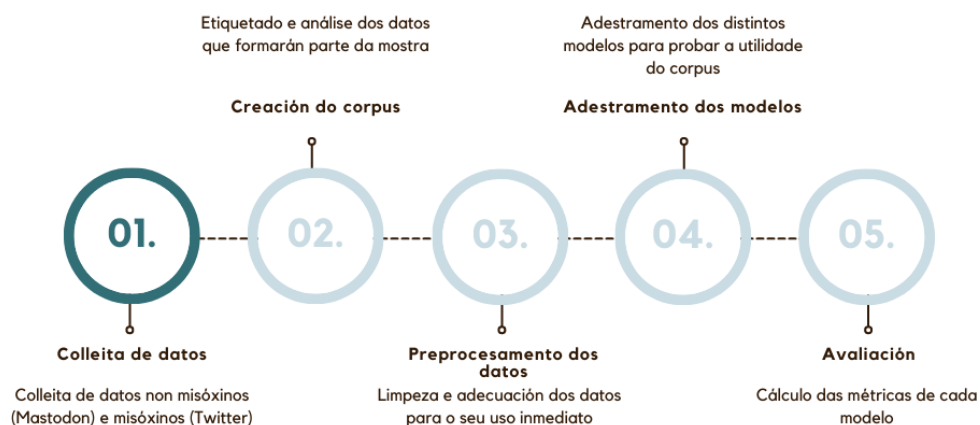


Figura 3.1: Fase 1 do fluxo de traballo

No proceso de detección de **misoxinia** nas redes sociais, séguense varias etapas para obter e preparar os datos necesarios, como se pode ver na figura 3.1. En primeiro lugar, realízase a recolleita de datos relevantes das redes sociais, tanto non misóxicos como misóxicos, utilizando **Application Programming Interface (API)** das plataformas ou técnicas de **web scraping**.

Na etapa de selección e etiquetaxe de datos, realízase unha selección dos datos relevantes para cada mostra obtida. Ademais, etiquetaranse os datos para identificar as mostras que

conteñen [misoxinia](#) e as que non. Esta etiquetaxe pode realizarse de maneira automatizada debido a que a colleita de ámbalas clases farase en momentos distintos.

Unha vez recopilados os datos, lévase a cabo a limpeza para eliminar información non desexada ou ruído que poida afectar á análise posterior. Isto implica filtrar caracteres especiais, eliminar enlaces ou etiquetas [HyperText Markup Language \(HTML\)](#), normalizar o texto e realizar outros procesos de limpeza e preprocesamento.

Unha vez preparados os datos, procede-se á análise e modelado dos mesmos. Isto implica o uso de técnicas de [Procesamento de Linguaxe Natural](#) e [Machine Learning](#) para extraer características relevantes dos textos e construír modelos predictivos de detección de [misoxinia](#). En casos nos que exista un desequilibrio no número de mostras misóxinas e non misóxinas, levarase a cabo un equilibrado de mostras.

Finalmente, avalíanse e válidanse os modelos construídos utilizando métricas axeitadas, como a puntuación *F1-score*.

3.2 Colleita de datos

A etapa de colleita de datos desempeña un papel crítico no desenvolvemento de calquer [corpus](#), pero especialmente do noso conxunto de datos, buscando obter unha mostra abundante e diversa de textos en lingua galega que reflecta a linguaxe e os temas comúns presentes nas redes sociais, incluíndo a presenza de contido misóxico. Para alcanzar este obxectivo, empregamos técnicas de colleita de datos automatizadas. A idea é empregar un enfoque de clasificación binaria, tal como se observou na literatura [4, 25, 29] para a detección de [misoxinia](#). Para iso, é fundamental contar cun dataset que conteña mostras tanto de [Twitter](#) como de [Mastodon](#) en lingua galega, diferenciando entre textos misóxicos e non misóxicos.

Iniciamos o proceso obtendo unha clase non misóxina no noso conxunto de datos mediante a recollida de [toots](#) da [instancia](#) galega de [Mastodon](#). Para isto, empregamos diversas ferramentas de rastreo e captura de datos, configurándoas para buscar e descargar [toots](#) publicados en lingua galega dende as contas da [instancia](#) galega. Un aspecto salientable é a confianza que temos en non atopar contidos misóxicos ao recoller [toots](#) desta [instancia](#), grazas ás estritas directrices de moderación que promoven unha comunicación respectuosa e inclusiva [17]. Esta vantaxe da [instancia](#) galega de [Mastodon](#) permite seleccionar [toots](#) para a clase non misóxina do noso conxunto de datos sen a necesidade de realizar unha revisión exhaustiva dos contidos descargados.

A continuación, avanzamos na obtención do conxunto de datos misóxicos en [Twitter](#), tendo en conta a importancia de considerar que a moderación dos contidos nesta plataforma é moito menos restritiva que na [instancia](#) galega de [Mastodon](#), isto implica unha maior probabilidade de atopar [chíos](#) con contidos misóxicos. Así, debemos proceder con cautela ao tomar

decisións relacionadas coa obtención de datos. É esencial ter en conta que, no contexto de [Twitter](#), non todos os [chíos](#) identificados como misóxinos poderían cumprir coas definicións establecidas no noso proxecto.

Ao combinar os datos recollidos da [instancia](#) galega de [Mastodon](#), que representan a clase non misóxina, cos [chíos](#) misóxinos obtidos en [Twitter](#), crearemos un conxunto de datos completo e equilibrado, que nos permitirá adestrar e avaliar con precisión o noso modelo de clasificación binaria. Esta abordaxe dual garantirá que o noso sistema sexa capaz de distinguir eficazmente entre contidos misóxinos e non misóxinos en lingua galega nas redes sociais. Ao final deste proceso, estaremos en posición de abordar a problemática da detección de [misoxinia](#) nas plataformas dixitais.

3.2.1 Colleita de mostras non misóxinas

Nesta subsección, abordaremos o proceso de obtención de [toots](#) representativos da clase non misóxina no noso conxunto de datos, cun enfoque especial na [instancia](#) galega do [Mastodon](#). Para obter esas mostras, decidimos utilizar a [API](#) do [Mastodon](#), aproveitando as súas funcionalidades para acceder aos [toots](#) publicados nesta [instancia](#). Esta abordaxe foi elixida en base á comprensión de que a [instancia](#) galega do [Mastodon](#) segue directrices de moderación rigorosas, que promoven unha comunicación respectuosa e inclusiva. Ao utilizar a [API](#) para capturar todos os [toots](#) publicados nesta [instancia](#), podemos garantir que estamos recollendo unha mostra ampla e diversificada de textos en lingua galega, con alta probabilidade de non seren misóxinos sen a necesidade de revisión individual.

Para coller os [toots](#) da [instancia](#) galega do [Mastodon](#), decidimos aproveitar a [API](#) do [Mastodon](#). A [API](#) permite acceder aos datos públicos e recuperar [toots](#) a través de peticións [HTTP](#). Esta abordaxe automatizada simplifica o proceso de colleita, permitindo obter os [toots](#) de forma eficiente e sistemática.

Durante a colleita de [toots](#) no [Mastodon](#), atopamos algúns problemas que esixiron atención e solución. Un dos principais problemas foi a limitación da [API](#) en canto á cantidade de datos recuperados por petición. A [API](#) do [Mastodon](#) establece un límite no número de [toots](#) que se poden descargar dunha soa vez, o que dificultou a obtención da mostra completa. Para superar esta limitación, implementamos unha abordaxe de paxinación, na que iteramos polas páxinas de [toots](#) usando un parámetro de ID máximo (*max_id*, de acordo coa [API](#)) para obter de xeito acumulado un conxunto de datos máis grande.

Implementouse un bucle que realizaba peticións sucesivas ata que se esgotaran os [toots](#) dispoñíbeis, utilizando o ID do último [toots](#) obtido como referencia para a seguinte petición. Almacenamos todos os [toots](#) nunha colección e, posteriormente, obtivemos os nomes dos campos (*fieldnames*, de acordo coa [API](#)) [3] dos [toots](#) para escribilos nun ficheiro de tipo [Comma-Separated Values](#) (CSV).

Problemática da variabilidade das instancias no Fediverso

Con todo, durante a execución do código, atopamos un problema no que o ficheiro `CSV` resultante estaba baleiro. Identificamos algunhas posibles causas deste problema, como respostas `HTTP` mal sucedidas, respostas `JavaScript Object Notation (JSON)` sen datos ou discrepancias nos nomes dos campos dos `toots`. Para solucionar isto, decidimos comprobar se a resposta `HTTP` era satisfactoria, para garantir que a resposta `JSON` conteña os datos esperados.

O problema principal que estaba a causar o ficheiro `CSV` baleiro estaba nos nomes dos campos dos `toots`. Ao analizar a estrutura dos datos devoltos pola `API` do `Mastodon`, percibimos que a orde dos nomes dos campos poden variar entre diferentes instancias do `Mastodon`. Esta discrepancia na orde dos nomes dos campos impedía a escrita correcta dos datos no ficheiro `CSV`, xa que algunhas entradas non correspondían cos nomes dos campos especificados na documentación da `API` do `Mastodon`. Para solucionar este problema, adaptamos o código para obter dinamicamente os nomes dos campos dos `toots` a partir da resposta da `API`, en lugar de usar nomes fixos especificados manualmente. Desta forma, garantimos que os datos se escriban correctamente, independentemente dos posibles cambios nos nomes dos campos.

Existe unha clara diferenza entre `Twitter` e `Mastodon` en canto á súa natureza como redes sociais. `Twitter` é unha plataforma centralizada e propietaria que está baixo o control exclusivo da empresa `Twitter, Inc.` As usuarias e os usuarios acceden ao servizo despois de rexistrarse e a arquitectura interna do software, así como a maioría dos seus aspectos de funcionamento, non son transparentes para o público. Por outro lado, `Mastodon` é unha rede social de código aberto que se basea en software libre. Calquera persoa ten a posibilidade de despreparar unha instancia de `Mastodon` nun servidor controlado por ela mesma, onde as usuarias e os usuarios poden rexistrarse e utilizar o servizo. Un exemplo disto é o administrador `nekerafa`, que puxo en funcionamento a instancia `mastodon.gal`. As distintas instancias de `Mastodon` federan entre si (ver 3.2), formando colectivamente o coñecido como Fediverso, grazas ao protocolo `ActivityPub` [34]. Este protocolo aberto permite descubrir outras instancias e intercambiar información, como `toots`, respostas a `toots`, acceso a perfís de usuarios, mensaxes directas e marcadores de favoritos, independentemente da instancia na que cada persoa usuaria estea rexistrada.

Desta forma, a diferenza fundamental entre `Twitter` e `Mastodon` radica na súa natureza centralizada e privativa fronte á natureza federada e de código aberto de `Mastodon`, que promove a descentralización e a transparencia do sistema.

A `API` de `Mastodon` está deseñada para ser independente das instancias individuais de `Mastodon`. Con todo, cada instancia de `Mastodon` pode ter a súa propia implementación e configuración personalizada, o que significa que a funcionalidade e os datos dispoñíbeis a través da `API` poden variar entre as diferentes instancias.

Cada instancia de Mastodon ten a súa propia Uniform Resource Locator (URL) da API á que é posible acceder para interactuar coa plataforma. A URL da API de Mastodon varía en función da instancia que se estea a utilizar. Para obter información específica sobre os endpoints dispoñibles e os nomes dos campos asociados nunha instancia concreta, é necesario consultar a documentación da API desa instancia en particular. Ademais da documentación da API, algunhas instancias de Mastodon poden proporcionar a súa propia documentación adicional no seu sitio web. Esta documentación pode ofrecer detalles específicos sobre os endpoints, os parámetros aceptados e os nomes dos campos dispoñíbeis na instancia específica. Debido a que para cada instancia de Mastodon os nomes dos campos específicos e a súa estrutura poden variar, é esencial consultar a documentación específica da instancia utilizada, como *mastodon.gal*, para obter información precisa sobre os nomes dos campos dispoñíbeis e a forma de acceder a eles.

Ao traballar coa instancia de *mastodon.gal*, atopamos que as particularidades técnicas desta instancia non estaban documentadas adecuadamente. A variación nos nomes dos campos dificultou o proceso de colleita e procesamento dos datos. Tivemos que adaptar constantemente o noso código para manexar os diferentes nomes dos campos en cada toots, o que requiriu un esforzo adicional.

Sen a información adecuada sobre os nomes dos campos específicos de *mastodon.gal*, tivemos que adoptar un enfoque máis complexo para xestionar a variabilidade dos nomes dos campos. Durante a iteración dos toots recollidos, tivemos que realizar comprobacións adicionais para identificar e manexar os diferentes nomes dos campos en cada toots. Iso implicou unha adaptación constante do noso código e unha maior atención para asegurarnos de capturar todos os datos relevantes, independentemente dos nomes dos campos específicos utilizados en cada caso.

Para resolver esta situación, implementamos unha solución na que obtemos dinamicamente os nomes dos campos dos toots descargados. Percorremos cada toots e actualizamos un conxunto de nomes dos campos cos keys de cada toots. Deste xeito, obtemos un conxunto completo de nomes dos campos que se aplicará ao escribir os datos no arquivo CSV.

Nesta problemática concreta de traballar coa instancia de *mastodon.gal*, tamén nos atopamos con situacións onde algúns toots tiñan un campo adicional chamado *application*, que en ocasións aparecía e noutras non. Debido á súa presenza inconsistente e á falta de información documentada sobre este campo na instancia de *mastodon.gal*, foi necesario realizar unha investigación adicional para determinar a súa relevancia e a súa inclusión nos datos recollidos.

Foi necesario establecer comunicación co administrador da instancia para obter aclaracións sobre este campo específico e determinar se tiña unha importancia significativa para os nosos propósitos de investigación. A través desta comunicación conxunta co administrador de *mastodon.gal*, puideron ofrecerse detalles adicionais sobre o campo *application* e as súas

implicacións. O administrador explicou que a presenza inconsistente deste campo estaba relacionada coas opcións de privacidade dispoñíbeis en *Mastodon*. Estas opcións permiten aos usuarios establecer diferentes niveis de privacidade para os seus *toots*, o que podería explicar a variación na aparición do campo *application* nos datos recollidos. Esta información foi valiosa para a nosa investigación, xa que nos permitiu entender mellor a orixe da variación nos nomes dos campos e descartar a inclusión do campo *application* nos datos recollidos, xa que non era relevante para os nosos obxectivos de estudo.

A falta de documentación adecuada e a variación dos nomes dos campos entre os *toots* de *mastodon.gal* presentaron un desafío significativo neste proxecto. Tivemos que dedicar tempo e recursos adicionais para comprender e adaptarnos ás particularidades desta *instancia* específica. A pesar destas dificultades, conseguimos implementar as solucións necesarias para obter os datos desexados, aínda que requiriu unha maior dedicación e esforzo pola nosa parte. Este caso destaca a importancia de contar cunha documentación completa e precisa das particularidades técnicas de cada *instancia* de *Mastodon*, así como unha comprensión clara dos nomes dos campos e a súa variabilidade. Unha documentación exhaustiva e actualizada axudaría a simplificar o proceso de colleita e procesamento de datos e a minimizar os problemas asociados á variación dos nomes dos campos.

Problemática dos tokens de acceso

O seguinte problema a afrontar foi o feito de que, aínda que se estableceu unha paxinación para obter os *toots* iterativamente, a resposta só devolvía 5 *toots*, e ademais escritos en inglés. Esta discrepancia entre o límite especificado, a lingua e o número real de *toots* devoltos xerou confusión e a necesidade de investigar a causa do problema. A pesar de ter implementada correctamente a lóxica do bucle para obter *toots* adicionais mediante a actualización do parámetro *max_id*, a API de *Mastodon* respondía con só 5 *toots* en cada solicitude. Este comportamento inesperado levantou interrogantes sobre posibles limitacións na API ou cambios no seu funcionamento que non foron tidos en conta inicialmente.

A análise do código revelou que hai unha limitación na obtención dos *toots* desexados que no momento de realizar este proxecto non se atopaba especificada na documentación oficial. O problema débese a que a API pública de *Mastodon* só inclúe as peticións, e polo tanto os *toots*, á *instancia* de *mastodon.social*¹, que é considerada a *instancia* de referencia ou principal dentro do Fediverso. Isto é, aínda que as peticións tipo GET se fixesen a través da API de *Mastodon* e especificando como URL a web *mastodon.gal*, automaticamente eran realizadas en *mastodon.social*.

Para acceder aos *toots* da *instancia Mastodon mastodon.gal* e obter máis de 5, chegouse á conclusión de que se precisa unha chave de acceso (*access key*). Esta conclusión foi alcanzada

¹mastodon.social

despois dunha investigación e análise exhaustiva da documentación da [API de Mastodon](#). A [API de Mastodon](#) require unha autenticación para realizar certas accións, como obter [toots](#) dunha liña de tempo específica. Para realizar a autenticación, é necesario obter unha chave de acceso válida. Sen esta chave, o acceso á [API](#) está limitado e só se poden recuperar un número limitado de [toots](#).

A obtención dunha chave de acceso require seguir un proceso específico. Neste caso, requiriu a creación dunha conta na [instancia](#) de [mastodon.gal](#), o rexistro como desenvolvedora de aplicacións na [instancia](#), e o seguimento das súas directrices para obter as credenciais necesarias. Isto incluíu a creación dunha nova aplicación na [instancia](#) e a xeración dun par de chaves: unha chave de cliente e unha chave secreta do cliente. Para acceder aos recursos públicos da [API de Mastodon](#), só se require a chave do cliente, que se obtén ao crear unha aplicación na [instancia de Mastodon](#). A chave secreta do cliente utilízase para acceder a recursos protexidos, como a liña de tempo de inicio do usuario ou as mensaxes directas.

Unha vez que se teñen as credenciais necesarias, pódese realizar unha solicitude á [API de Mastodon](#) incluíndo as credenciais no encabezado da solicitude. A [API](#) verificará as credenciais e, se son válidas, fornecerá unha chave de acceso que permitirá obter máis de 5 [toots](#) da liña de tempo desexada.

Ademais, é importante mencionar que o proceso de obtención dunha chave de acceso para acceder aos [toots](#) de [mastodon.gal](#) foi demorado pola necesidade de que o administrador da [instancia de Mastodon](#) proporcione manualmente o acceso á creación de novas contas. Neste caso, o administrador da [instancia mastodon.gal](#) debe realizar unha revisión manual das solicitudes de novas usuarias ou usuarios e aprobalas individualmente. Isto implica que, despois de rexistrarse como nova usuaria e así obter as credenciais necesarias, o proceso de obtención dunha chave de acceso pode experimentar atrasos debido á dependencia da intervención manual do administrador. A necesidade de que o administrador revise manualmente as novas solicitudes é unha medida de seguridade implementada para evitar abusos ou *spam* na [instancia Mastodon](#). Aínda que esta medida é importante para manter a integridade e calidade da comunidade [mastodon.gal](#), tamén pode implicar menor eficiencia no proceso de obtención de chaves de acceso.

Asemade, durante o período de espera para obter a chave de acceso, aproveitouse o tempo para familiarizarse coa librería de [Mastodon](#) para Python. Esta librería facilita enormemente o proceso de interacción coa [API de Mastodon](#) e o uso das chaves de acceso unha vez obtidas. A librería de [Mastodon](#) para Python proporciona un conxunto de funcións e métodos predefinidos que simplifican a autenticación, o envío de solicitudes e a xestión das respostas da [API](#). Púidose aproveitar esta ferramenta para acelerar o proceso de desenvolvemento, xa que non se tivo que implementar cada unha das chamadas e funcionalidades manualmente. Ao usar a librería de [Mastodon](#) para Python, puidémonos centrar máis no proceso de filtrado e análise

dos `toots` obxecto de estudo, en lugar de investir tempo en cuestións técnicas relacionadas coa interacción coa `API`. Isto axudou a simplificar e acelerar o proceso de uso da `API` e das claves de acceso unha vez que foron obtidas.

Problemática da eficiencia na utilización de `batches`

Co obxectivo de evitar esgotar a memoria `Random Access Memory (RAM)` dispoñíbel durante o proceso de recuperación de datos, tomouse a decisión de substituír a paxinación convencional e implementar unha estratexia de recuperación por `batches`. En lugar de obter todos os datos dunha vez, o código foi modificado para recuperar os datos en grupos de 1.000 en 1.000. Este enfoque por `batches` ofrece vantaxes significativas. Ao procesar os datos en pequenos grupos, a cantidade de memoria necesaria para gardalos e manipularlos é moito menor. Deste xeito, prevéñense problemas relacionados co esgotamento de memoria `RAM`, especialmente en conxuntos de datos moi grandes. Ademais, a implementación por `batches` tamén permite unha maior eficiencia no procesamento dos datos. Ao recoller os datos en grupos, é posible realizar operacións e análises parciais durante o proceso de recuperación, sen ter que agardar á finalización completa da recuperación de todos os datos. Isto acelera o proceso de análise e facilita unha maior reactividade na obtención de resultados intermedios.

Porén, a utilización de `batches` tivo un efecto non previsto que nos fixo pensar na posibilidade de que o algoritmo estivese a entrar nun ciclo infinito durante o proceso de recuperación de `toots` por `batches`, sen chegar a completar ningunha acción. Esta disfunción pensouse estar relacionada co obxecto `Mastodon`, a conexión coa `instancia` de `Mastodon` ou coa forma na que se estaba a empregar o obxecto `Mastodon` no código. O obxecto `Mastodon` créase utilizando a librería de Python para `Mastodon`, que proporciona as funcións e métodos necesarios para interactuar coa plataforma de `Mastodon`. O obxecto `Mastodon` é unha parte fundamental da interacción coa plataforma de `Mastodon`. Este obxecto é unha `instancia` deseñada para facilitar o acceso á `API` de `Mastodon` e interactuar cos seus recursos. A través do obxecto `Mastodon`, é posíbel realizar chamadas á `API` para recuperar `toots` por `batches`, publicar novos `toots`, obter información de usuarios e realizar outras accións necesarias na análise dos datos. Este obxecto é a ponte entre o código Python desenvolvido e a plataforma de `Mastodon`, permitindo así a manipulación dos datos e a realización das tarefas desexadas na análise dos `toots`. É grazas a este obxecto que se establece a conexión coa `instancia` de `Mastodon` e se pode acceder ás funcionalidades ofrecidas por esta rede social federada. Tamén era posíbel que existisen problemas coa configuración da rede, como unha devisa que restrinxa a conexión. Finalmente, determinouse que o aparente bloqueo observado era resultado dos longos tempos de resposta do `Mastodon` ao recuperar `batches` do tamaño indicado, e non por un erro de sintaxe ou lóxica no código desenvolvido.

Rango temporal empregado

Unha vez rematado o proceso de desenvolvemento do código para a colleita de mostras non misóxinas, procedeuse á discusión sobre o período de tempo axeitado para recoller a mostra de datos. Inicialmente, considerouse que sería apropiado centrarse no período de outubro-novembro de 2022. Neste período, Elon Musk adquiriu [Twitter](#) [35] e moitos usuarios mostraron o seu descontento migrando a [Mastodon](#) [36]. No entanto, despois dunha análise exhaustiva, chegouse á conclusión de que sería máis beneficioso recoller todos os datos dispoñíbeis dende a creación da [instancia](#) en maio de 2022 ata febreiro de 2023, pois a mostra dun mes non era significativa abondo e ademais non había datos públicos dispoñíbeis no momento de realizar este traballo que nos indicasen que período de tempo fora o de máis actividade na rede social. Esta elección permitiríanos obter unha perspectiva máis completa dos [padróns](#) e comportamentos dos usuarios ao longo dun período de tempo máis estendido.

Conclusións

A pesar das dificultades atopadas, conseguimos resolvelas mediante adaptación e investigación. A documentación completa, a obtención de claves de acceso, a implementación por [batches](#) e o tratamento dos caracteres indesexados foron aspectos fundamentais para obter os datos desexados. A experiencia resalta a importancia de ter un enfoque flexíbel e áxil ao traballar con plataformas como [Mastodon](#), onde as particularidades técnicas e as variacións na [API](#) poden ser comúns. A capacidade de adaptarse rapidamente e realizar investigación adicional para solucionar problemas é crucial para obter os datos desexados con éxito. Tamén subliña a necesidade de documentación precisa e completa por parte das instancias de [Mastodon](#) para axudar aos desenvolvedores a navegar e utilizar a súa [API](#) de forma eficiente. Coñecer as ferramentas dispoñíbeis, como a obtención de claves de acceso e a implementación por [batches](#), pode mellorar o rendemento e a eficiencia do proceso de recollida de datos. En resumo, a experiencia demostrou que a adaptabilidade, a investigación, a documentación e o coñecemento das ferramentas son elementos fundamentais para superar as dificultades e obter éxito na obtención de datos en plataformas como [Mastodon](#).

3.2.2 Colleita de mostras misóxinas

Nesta sección, exploraremos o proceso de colleita de mostras misóxinas, abordando os métodos e as ferramentas utilizadas para recompilar datos relevantes. Tamén analizaremos os desafíos que puideron xurdir durante este proceso e as estratexias para superalos.

Aínda que inicialmente consideramos a opción de empregar [mastodon.gal](#) debido á súa popularidade e participación activa da comunidade galegofalante, chegamos á conclusión de que esta contorna non nos proporcionaría as mostras necesarias para o noso estudo. A [instancia](#)

aplica unha política de tolerancia cero en relación aos contidos e comportamentos misóxinos, entre outras toxicidades, o que constitúe unha medida moi positiva para garantir un ambiente seguro e respectuoso para os seus usuarios. Non obstante, dado que o noso obxectivo radica na análise e comprensión dos comportamentos misóxinos presentes en liña, tivemos que explorar outras opcións que nos permitisen recoller datos relevantes e representativos.

Tamén se considerou a posibilidade de obter mostras de comportamento misóxino da *instancia masto.pt*², a plataforma portuguesa da rede Mastodon. Esta decisión foi motivada pola similitude existente entre o galego e o portugués, así como pola existencia de ferramentas de transliteración que nos permitirían converter os textos en portugués a galego [31]. No entanto, despois dunha análise exhaustiva do código de conduta de masto.pt, chegamos á conclusión de que esta *instancia* tamén prohibe explicitamente os comportamentos misóxinos e, polo tanto, os *toots* son *moderados* [37].

Tras considerar detidamente as opcións dispoñíbeis, chegamos á conclusión de que a mellor alternativa para obter mostras de comportamento misóxino en galego é facer uso do conxunto de datos Spanish MisoCorpus-2020 [29]. Este *corpus*, enfocado especificamente na *misoxinia* en español, demostrou ser unha fonte valiosa para a nosa investigación. O Spanish MisoCorpus-2020³ é un *corpus* equilibrado que contén exemplos representativos de distintos tipos de comportamento misóxino extraídos directamente da rede social Twitter. A través da súa análise, atópase clasificado o *corpus* en tres subconxuntos interrelacionados. O primeiro subconxunto aborda a violencia contra mulleres relevantes, proporcionando mostras específicas deses comportamentos. O segundo subconxunto refírese a mensaxes que acosan a mulleres en español de España e español de América Latina, ofrecendo unha visión ampla deste problema en diferentes contextos lingüísticos. O terceiro subconxunto engloba trazos xerais relacionados coa *misoxinia*, permitíndonos estudar a súa manifestación en diversas formas. A pesar de que o *corpus* está en castelán, temos á nosa disposición ferramentas de transliteración do castelán ao galego [31]. Isto permitiunos converter os textos do *corpus* ao galego de maneira automática, garantindo así a súa adaptación á nosa lingua de estudo.

Problemática da obtención das mostras orixinais

Cabe destacar que a utilización do Spanish MisoCorpus-2020 non estivo exenta de desafíos. Este *corpus* presentou unha dificultade particular: en lugar de contar con texto completo, estaba composto por IDs de *chíos* clasificados por expertos como misóxinos. Este enfoque tiña a finalidade de preservar a privacidade dos usuarios implicados, así que para obter os textos orixinais precísase facer uso da API de Twitter e así recuperar os textos a partir dos IDs.

²masto.pt

³<https://pln.inf.um.es/corpora/misogyny/misocorpus-spanish-2020.rar>

No momento de realizar o noso proxecto, atopámonos cun desafío inesperado. A [API de Twitter](#) deixou de estar dispoñíbel de forma gratuíta debido ás novas políticas implementadas despois da adquisición da plataforma por parte de Elon Musk [38]. Esta situación presentou un obstáculo na nosa investigación, así como nas investigacións en curso de equipos en todo o mundo [39], xa que non contabamos co acceso directo aos datos desexados. Ante esta circunstancia, vímonos na obriga de explorar e implementar técnicas de web scraping para obter os textos dos [chíos](#) relacionados coa [misoxinia](#).

O web scraping é unha técnica que nos permite extraer información de páxinas web de forma automatizada. Utilizando esta tecnoloxía, desenvolvemos ferramentas personalizadas para rastrexar as contas e as conversacións pertinentes nas redes sociais, recompilando así os datos necesarios para o noso estudo. A pesar de que o web scraping supón unha solución alternativa, tamén presentou desafíos propios. Ademais, o proceso de web scraping é máis complexo e lento en comparación co acceso directo á [API de Twitter](#).

Para realizar o web scraping decidimos utilizar a ferramenta Selenium [40]. Selenium é unha ferramenta de automatización de probas en navegadores web que nos permite interactuar coas páxinas web de forma programática. Esta elección baséase nas vantaxes que ofrece Selenium en termos de flexibilidade e control sobre a navegación e a extracción de datos.

O proceso metodolóxico adoptado consiste na aplicación da biblioteca Selenium en conxunto co navegador Chrome para a obtención de mostras de [chíos](#) misóxinos. A utilización do Selenium posibilita a interacción automatizada co navegador, permitíndonos extraer a información desexada para análises posteriores. O primeiro paso será ler os IDs dos [chíos](#) procedentes do conxunto de datos. Iníciase o navegador web Chrome, constrúese a [URL](#) do chío e captúrase o texto que logo será volcado nun [CSV](#) para o seu posterior procesamento. Esta abordaxe permite a sistematización do fluxo de traballo e a eficiente recollida das mostras de interese.

```
1 # Iterate over the tweet IDs
2 for tweet_id in tweet_ids:
3     url = f'https://twitter.com/anyone/status/{tweet_id}'
4     driver.get(url)
5     time.sleep(5) # Wait 5 seconds to load the page
6
7     # Download the HTML of the page
8     text = driver.find_elements(By.CSS_SELECTOR,
9     '[data-testid="tweetText"]')
10
11     # If the tweet does not exist, the page will be empty
12     if len(text) != 0:
13
14         # Write the tweet to a CSV file
```

```

14     with open(os.path.dirname(__file__) +
15               "../data/tweets.csv", mode='a', newline='', encoding='utf-8')
16     as file:
17         writer = csv.writer(file)
18         tweet_text = []
19         for t in text:
20             tweet_text.append(t.text)
21             writer.writerow([tweet_id, " [response]
22                             ".join(tweet_text)])
23 # Close the browser
24 driver.quit()

```

A construción das URLs de [Twitter](#) sen coñecer o nome de usuario, pero tendo o ID do chío, segue unha convención estándar na plataforma. A URL dun chío ten a seguinte estrutura:

```
1 https://twitter.com/{nome_usuario}/status/{ID_do_chío}
```

No noso caso, non dispoñemos do nome de usuario asociado a cada ID do chío. Para superar esta limitación e acceder á información do chío, construímos a URL utilizando o ID do chío coñecido. O ID do chío é único e non cambia independentemente do nome de usuario. Así, substituímos *nome_usuario* pola cadea de texto *anyone* na URL de [Twitter](#) para que teña o seguinte aspecto:

```
1 https://twitter.com/anyone/status/{ID_do_chío}
```

Desta forma, a URL construída permite acceder á páxina do chío directamente, sen necesidade de coñecer o nome de usuario asociado a ese ID específico. É importante destacar que a utilización desta convención estándar asume que o chío está dispoñíbel publicamente e non está suxeito a restricións de privacidade ou bloqueo por parte do usuario, así como que non foi borrado desde a súa inclusión no [corpus](#).

Problemática do acceso ás publicacións nas redes sociais ao longo do tempo

Precisamente, a creación do conxunto de datos presentou dificultades significativas debido á falta de dispoñibilidade dalgúns [chíos](#). Estas dificultades xurdiron cando algúns [chíos](#) non puideron ser atopados a través do seu identificador (ID) debido a que as contas asociadas estaban configuradas como privadas ou foran sancionadas por [Twitter](#). Esta situación afectou de maneira substancial a recompilación de datos, xa que impediu o acceso a un número considerable de [chíos](#) relevantes para o conxunto de datos. Aínda que se fixeron esforzos para obter a maior cantidade posible de datos, a limitación imposta polas contas privadas ou sancionadas resultou na exclusión destes [chíos](#) na creación do conxunto de datos. É importante ter en

conta estas limitacións ao interpretar e analizar os resultados do conxunto de datos, xa que a falta de acceso a certos **chíos** pode introducir nesgos ou limitacións na representatividade da mostra.

Ademais, a creación do conxunto de datos tamén se viu afectada pola eliminación de certos **chíos** por parte dos seus autores. Ademais dos desafíos mencionados anteriormente, onde algúns **chíos** non estaban dispoñibles debido a contas privadas ou sancionadas, tamén atopamos casos nos que os propios autores decidiron borrarlos. Esta situación engadiu un nivel adicional de complexidade, xa que resultou na perda de información relevante para o conxunto de datos. A pesar dos nosos esforzos por recompilar unha mostra representativa de **chíos**, a eliminación de **chíos** por parte dos usuarios afectou á dispoñibilidade de datos e, polo tanto, á integridade do conxunto de datos final. É importante ter en conta as limitacións impostas por contas privadas, sancionadas e os **chíos** eliminados ao analizar os resultados do conxunto de datos, xa que estas circunstancias poden afectar á representatividade e á validez das conclusións derivadas dos datos recompilados.

Crear un modelo máis específico que teña en conta o contexto vólvese máis desafiante debido á dificultade de acceder aos **chíos** orixinais que foron eliminados polos seus autores. No contexto de comentarios misóxicos en resposta a estes **chíos**, a falta de dispoñibilidade dos **chíos** orixinais limita a nosa capacidade de comprender plenamente o contexto no que se produciron os comentarios ofensivos. Cando os **chíos** orixinais son eliminados, perdemos información valiosa sobre o contido e a intención detrás dos comentarios misóxicos. Sen o contexto proporcionado polos **chíos** orixinais, é difícil determinar con precisión o propósito e o ton dos comentarios ofensivos e avaliar correctamente o seu impacto.

Problemática do contido non textual

Ademais dos desafíos mencionados anteriormente, outro obstáculo importante ao realizar o web scraping de **Twitter** para obter datos é a perda de información relacionada cos **emojis**. Os **emojis** son elementos visuais amplamente utilizados nas interaccións en liña e poden desempeñar un papel importante na comunicación emocional e na interpretación do ton e a intención dunha mensaxe. Non obstante, durante o proceso de web scraping, é común perder a información dos **emojis** presentes nos **chíos**. Isto débese a que os **emojis** a miúdo se representan como caracteres especiais que poden ser difíciles de capturar e almacenar correctamente utilizando técnicas de extracción de texto en bruto. A perda de información dos **emojis** pode ser problemática, xa que estes elementos poden proporcionar contexto adicional e axudar a comprender mellor o contido dun chío. No caso específico de abordar comentarios misóxicos, os **emojis** poden transmitir emocións, actitudes ou matices que son relevantes para identificar e clasificar o contido ofensivo de maneira máis precisa. Por tanto, a falta de captura dos **emojis** durante o proceso de web scraping limita a riqueza e a completitude dos

datos recompilados, o que pode ter un impacto na capacidade do modelo para capturar adecuadamente o contexto emocional e o ton dos **chíos**, especialmente no caso de comentarios misóxicos. É importante ter en conta esta limitación ao desenvolver un modelo que busque comprender e abordar o contido ofensivo nas interaccións en liña.

Ademais dos desafíos mencionados anteriormente, ao realizar o web scraping de **Twitter**, outro aspecto que contribúe á dificultade de capturar un contexto máis completo é a incapacidade de recuperar imaxes e GIFs asociados aos **chíos**. **Twitter** é unha plataforma que permite aos usuarios engadir imaxes e GIFs ás súas publicacións, o que pode ser especialmente relevante no caso de comentarios ofensivos e misóxicos, xa que a miúdo se utilizan imaxes gráficas ou *memes* para transmitir mensaxes ofensivas de forma visual. Non obstante, durante o proceso de web scraping, a extracción de imaxes e GIFs vólvese complexa, xa que estes elementos non se almacenan directamente en forma de texto e requiren técnicas adicionais para a súa recuperación. Isto implica que a información visual e contextual asociada aos **chíos**, como imaxes ou GIFs que poidan estar directamente relacionados cos comentarios misóxicos, non se recolle no conxunto de datos. A falta de acceso a estas imaxes e GIFs pode limitar a comprensión completa do contexto no que se producen os comentarios ofensivos, xa que a información visual pode desempeñar un papel crucial na interpretación e análise das mensaxes. A omisión desta información contextual importante pode afectar a capacidade do modelo para capturar a totalidade do contido e comprender plenamente o alcance dos comentarios misóxicos en liña.

Problemática da información contextual

Ao revisar os **chíos** descargados, detectamos un problema adicional. Descubrimos que había un *bug* polo que só se gardaba o primeiro chío de todo o fío de conversa, o que non necesariamente era o que precisábamos para a nosa análise (ver figura 3.3). Para solucionar este problema, implementamos unha nova estratexia: descargamos a conversa enteira, separando as mensaxes individuais por *[response]*. A pesar de que a nova abordaxe resolveu o problema de gardar só o primeiro chío do fío de conversa, tamén trouxo consigo un desafío adicional. Ao descargar a conversa completa e separar as mensaxes por *[response]*, o proceso de obtención dos datos ralentizouse considerablemente.

Criterio de parada

O criterio de parada para a recollida de mostras misóxicas non está baseado no tempo, como no caso das mostras non misóxicas, senón no tamaño do MisoCorpus orixinal. Neste proxecto, non establecemos un límite temporal específico para a recollida de mostras misóxicas, pois dependemos dun **corpus** previo. En vez diso, colleitamos todas e cada unha das

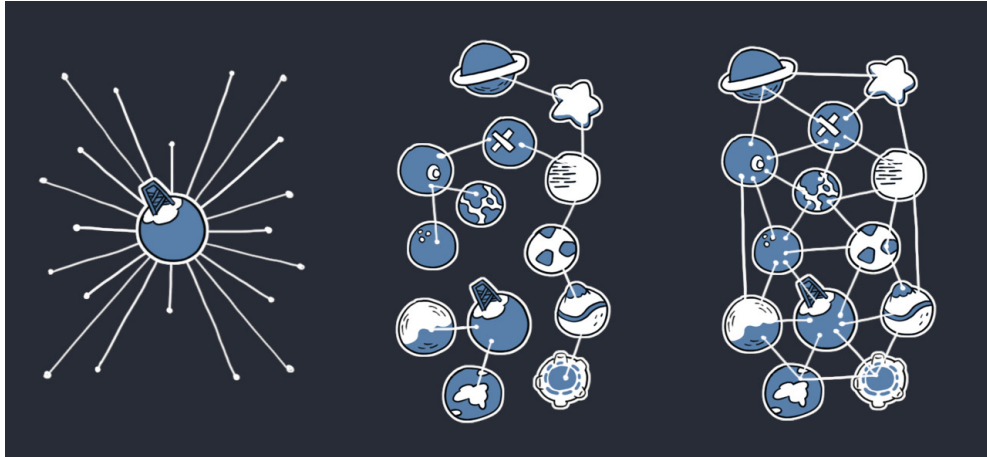


Figura 3.2: De esquerda a dereita: centralizada, federada e distribuída (fonte: [3])



Figura 3.3: Exemplo de conversa na que se gardaba o primeiro chío e non o segundo, que é o relevante

mostras dispoñíbeis no MisCorpus. Máis adiante neste traballo, abordaremos a cuestión do tamaño das mostras tanto para os **chíos** misóxinos como para os non misóxinos.

Limpeza das mostras

Despois de descargar os datos, tivemos que realizar unha depuración manual do dataset. O obxectivo era eliminar os **chíos** que non se correspondían co ID marcado previamente polos expertos do MisoCorpus. Esta tarefa de depuración requiriu unha atención detallada e un traballo meticuloso para asegurarnos de que só conservabamos os **chíos** relevantes para a nosa investigación. A depuración manual do dataset implicou revisar cada chío descargado, identificar os **chíos** do fío da conversa e eliminar aqueles que non coincidían co ID desexado. Aínda que este proceso foi laborioso e requiriu un esforzo considerable, era necesario para garantir a calidade e a precisión dos datos que íamos analizar. A pesar de que a depuración manual engadiu unha etapa adicional ao noso proceso, tamén nos asegurou de que o noso dataset estaba libre de **chíos** non desexados e estaba en consonancia coas pautas establecidas polos expertos do MisoCorpus. Esta depuración axudou a obter un conxunto de datos máis coherente e fiable para a nosa análise.

Transliteración

Despois de completar a depuración manual do dataset, o seguinte paso consistiu en traducir ou transliterar automaticamente os **chíos** do castelán ao galego. Para esta tarefa, optamos por utilizar a ferramenta proporcionada polo Proxecto Nós [41]. O Proxecto Nós é unha iniciativa da Xunta de Galicia e da Universidade de Santiago de Compostela que ten como obxectivo promover a lingua galega na era dixital, ofrecendo, entre outros, ferramentas de tradución e transliteración para facilitar a adaptación de contidos de castelán a galego. A elección do Nós Tradutor [31] do Proxecto Nós como ferramenta de tradución e transliteración foi motivada polo seu compromiso coa promoción e o uso do galego, así como pola súa calidade e precisión nos resultados.

Inicialmente, só tiñamos unha web dispoñíbel para acceder á ferramenta de tradución ou transliteración do Proxecto Nós. No entanto, ao intentar acceder á URL que se mencionaba no artigo, descubrimos que estaba caída. Para solucionar este problema, puxémonos en contacto coa autora principal do artigo para informarlle da situación. A autora explicou que a demostración que se montou para a conferencia na que se presentara a ferramenta non estaba xa dispoñíbel naqueles momentos, debido a que estaban a actualizar os sistemas coa última versión dos modelos de tradución neuronal do Proxecto Nós. Asegurounos que estes novos modelos estarían dispoñíbeis nas seguintes semanas, tanto a través dunha interface web como do propio modelo e datos de adestramento en HuggingFace [42].

Para acelerar o proceso, contactouse co autor doutro artigo do Proxecto Nós para informar da caída da páxina web mencionada no artigo *A neural machine translation system for Galician from transliterated Portuguese text* [31]. Solicitouse acceso á ferramenta ou a algún repositorio que contivera os modelos de tradución debido á súa utilidade para o traballo de investigación. O autor respondeu pedindo desculpas polo problema e asegurando que estaban a traballar xunto co coautor do artigo para solucionalo o máis axiña posible. Tamén se ofreceu a realizar tarefas de tradución mentres se solucionaba a situación.

Finalmente, conseguimos acceso tanto aos modelos xa adestrados como á web do tradutor. Con todo, debido á falta dunha API na páxina web mencionada, non puidemos automatizar o proceso de tradución, polo que decidimos intentar usar os modelos directamente. Desafortunadamente, atopámonos con dificultades debido á falta de documentación no repositorio. Ademais do erro mencionado anteriormente, tamén nos atopamos con outro desafío. O noso sistema resultou ser incompatible coa ferramenta OpenNMT [43], que era necesaria para facer funcionar os modelos de tradución. En concreto, a versión de Torch [44] que era posible instalar no noso sistema non era compatible coa versión necesaria de OpenNMT. Isto significaba que non podíamos facer uso da funcionalidade completa de OpenNMT debido á incompatibilidade entre as versións. Torch é un framework de Deep Learning (DL) de código aberto que proporciona un amplo conxunto de ferramentas e bibliotecas para desenvolver e adestrar modelos de redes neuronais. Foi deseñado para ser flexíbel e eficiente, permitindo a experimentación rápida e a implementación eficiente de algoritmos de Deep Learning. Torch é amplamente utilizado na comunidade de investigación e desenvolvemento de aprendizaxe automática debido á súa potencia e facilidade de uso. Pola súa banda, OpenNMT é unha biblioteca de código aberto desenvolvida en Torch que se enfoca especificamente na tarefa de tradución automática neuronal. Proporciona unha infraestrutura completa para construír e adestrar modelos de tradución automática neuronal, que son modelos baseados en redes neuronais artificiais. OpenNMT permite experimentar con diferentes arquitecturas de modelos, funcións de perda e técnicas de adestramento para mellorar a calidade da tradución automática. A través da combinación de Torch e OpenNMT, os investigadores e desenvolvedores teñen acceso a ferramentas potentes para abordar os desafíos da tradución automática. A falta de compatibilidade entre as ferramentas foi unha limitación significativa para a nosa capacidade de implementar e executar os modelos de tradución desexados. Tivemos que buscar alternativas ou solucións para superar esta incompatibilidade e avanzar co proceso de tradución.

Na busca de alternativas para a tradución ao galego, atopamos un tradutor dispoñíbel no CIXUG [45]. Esta ferramenta permitía a tradución de arquivos de texto (.txt), o que axudou a cubrir as nosas necesidades de tradución. Utilizando esta ferramenta, fomos capaces de procesar e traducir os nosos datos ao galego de forma efectiva. Non obstante, atopamos unha

limitación co tradutor do CIXUG, xa que non era capaz de traducir correctamente os **chíos** en español latino. Isto foi unha desvantaxe significativa, xa que gran parte do noso **corpus** de datos estaba en español latino e precisabamos traducilos ao galego. Como solución alternativa, decidimos limitar o noso estudo e utilizar só aqueles **chíos** que estivesen xeolocalizados en España. Ao facer iso, reducimos a necesidade de tradución de español latino ao galego, xa que a meirande parte dos **chíos** nesta rexión están escritos na variante do español (castelán) de máis uso en España. Como resultado desta restrición xeográfica, a nosa mostra de **chíos** reduciuse de forma significativa. Ao limitar a análise aos **chíos** xeolocalizados en España, excluimos unha gran cantidade de **chíos** escritos noutras rexións ou países de fala hispana. Isto pode ter afectado a representatividade da nosa mostra e a diversidade dos datos analizados. É importante ter en conta esta limitación ao interpretar os resultados e xeneralizar as conclusións da nosa investigación. A pesar desta redución na mostra, consideramos que esta decisión foi necesaria para asegurar a precisión e a coherencia das nosas análises dentro do noso contexto específico.

Conclusións

En resumo, a utilización da técnica de web scraping en conxunto coa ferramenta Selenium para obter os **chíos** xa clasificados do MisoCorpus proporcionou unha maneira eficiente e controlada de obter mostras de comportamento misóxino en liña. Aínda que houbo desafíos relacionados coa perda de información sobre **emojis** e a imposibilidade de recuperar imaxes e GIFs, así como dificultades coa tradución ao galego, os datos recollidos a través deste proceso permitíronnos avanzar na nosa investigación e comprender mellor este fenómeno. A través do web scraping, puidemos acadar unha cantidade significativa de información valiosa que nos axudará a analizar e abordar o comportamento misóxino en liña nun contexto máis amplo.

3.3 Descrición do conxunto de datos proposto

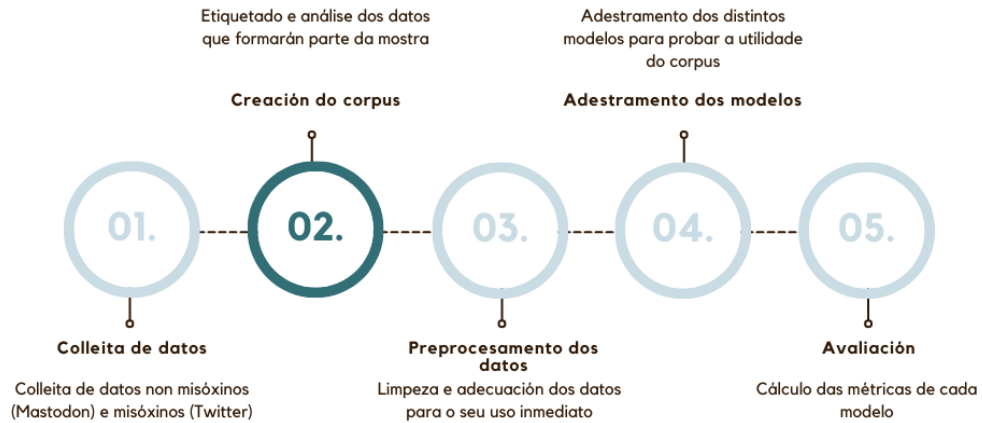


Figura 3.4: Fase 2 do fluxo de traballo

Nesta sección, presentamos a descrición do conxunto de datos proposto para a nosa investigación sobre a detección de **misoxinia** na lingua galega en redes sociais, especialmente en **Twitter** e **Mastodon**. O desenvolvemento dun **corpus** adecuado é fundamental para adestrar e avaliar os modelos de **aprendizaxe automática** utilizados na detección de contido misóxico. Dado que a falta de recursos anotados en galego representa un desafío significativo, propoñemos a creación dun novo conxunto de datos que sexa representativo e adecuado para a nosa investigación.

3.3.1 Estrutura

O conxunto de datos proposto para a nosa investigación é unha colección de mensaxes en galego recollidas de **Twitter** e **Mastodon**. Este conxunto de datos consta de dous ficheiros **CSV**. O primeiro ficheiro, chamado `toots.csv`, contén unha mostra de mensaxes non misóxicas obtidas de **Mastodon**. Estas mensaxes foron seleccionadas para representar a linguaxe galega utilizada de forma xeral na plataforma, sen contido misóxico. O segundo ficheiro, chamado `tweets.csv`, contén unha mostra de mensaxes misóxicas obtidas de **Twitter**. Estas mensaxes foron recollidas utilizando criterios específicos do MisoCorpus que logo foron traducidos a lingua galega.

Ámbolos ficheiros, `toots.csv` e `tweets.csv`, teñen unha estrutura semellante e conteñen as seguintes columnas:

- `id`: esta columna contén un identificador único para cada mensaxe no conxunto de datos que é o mesmo que o da rede social de orixe. O uso dun identificador único axuda a identificar e relacionar cada mensaxe de forma unívoca.

- `language`: nesta columna, rexístrase o idioma no que está escrita cada mensaxe.
- `content`: esta columna contén o texto ou contido das mensaxes. Aquí é onde se atopan os datos reais que serán analizados para detectar a presenza ou ausencia de [misoxinia](#).

3.3.2 Tamaño

O ficheiro `toots.csv` contén un amplo conxunto de 19.387 mostras, as cales non están restrinxidas á lingua galega. En realidade, permitiuse o uso de varias linguas na [instancia](#) galega de [Mastodon](#), incluíndo o galego (gl), castelán (es), inglés (en), asturiano (ast), catalán (ca), italiano (it) e portugués (pt). Durante a análise exhaustiva deste conxunto de datos, identificamos unha situación interesante en relación coa atribución do idioma. Observamos que un número substancial de mostras etiquetadas como portugués eran, na realidade, escritas en galego na súa variante lusista ou reintegracionista [46]. Isto débese ao feito de que os propios usuarios teñen a facultade de elixir o idioma no que desexan no que se vexa identificado o seu `toot` antes de publicalo.

A posibilidade de que os usuarios escollan manualmente a etiqueta do idioma nos seus contidos pode levar a discordancias entre o idioma real do texto e o valor asignado no campo `language`. Neste conxunto de datos, observamos que algúns usuarios poden cometer erros ao identificar de forma precisa o idioma no que están a escribir. Iso pode resultar nunha discrepancia entre o idioma real dunha mostra e o seu correspondente valor asignado no campo `language`. Estas discrepancias deben ser tidas en conta durante o proceso de análise e interpretación dos datos para obter resultados máis precisos e contextualmente relevantes.

O arquivo `tweets.csv` contén unha mostra considerablemente máis pequena, cun total de 1.307 mostras. Esta diminución no tamaño do conxunto de datos débese á eliminación das mostras que non tiñan información de xeolocalización en España. Non obstante, a pesar das traducións realizadas ao galego, é importante sinalar que non todas as mostras contidas no arquivo `tweets.csv` están escritas nesta lingua (gl). De feito, tamén se recolleron mostras en catalán (ca), as cales xa estaban identificadas como tal no MisoCorpus orixinal.

Este [corpus](#) de datos proposto non está balanceado en termos de mostras misóxinas e non misóxinas. O arquivo `toots.csv` contén un total de 19.387 mostras non misóxinas recollidas da [instancia](#) galega de [Mastodon](#), mentres que o arquivo `tweets.csv` contén unha mostra máis pequena de 1.307 mostras etiquetadas como misóxinas, o que indica que, por conseguinte, a porcentaxe de mostras misóxinas é aproximadamente do 6.74% en relación ás mostras non misóxinas neste [corpus](#) de datos (ver 3.5). Esta discrepancia no número de mostras entre os dous grupos pode influenciar a análise e interpretación dos resultados. A pesar da falta de balanceamento no [corpus](#) de datos proposto, este proxecto representa un

importante comezo para a detección de *misoxinia* na lingua galega. Ata o momento, non hai ningún recurso ou *corpus* dispoñíbel para abordar este problema específico nesta lingua. Este *corpus* de datos proporciona un punto de partida valioso para futuras investigacións e desenvolvemento de modelos de detección de *misoxinia* en galego.

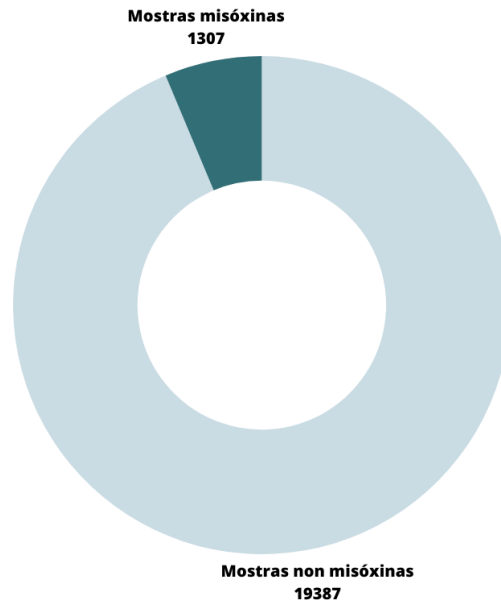


Figura 3.5: Mostras misóxicas vs. non misóxicas do dataset proposto

Ademais, é importante mencionar que non estamos completamente de acordo coa etiqüete dos *chíos* como misóxicos no MisoCorpus. Cremos que algunhas mostras transmiten un ton satírico ou irónico que pode dificultar unha interpretación clara da súa intención misóxina. Neste sentido, sería necesario exercer unha análise máis profunda e contextualizada para determinar se os contidos realmente reflicten actitudes ou discursos misóxicos ou se se trata dunha expresión satírica ou humorística que pode ser malinterpretada sen a debida consideración do seu contexto, como podemos ver na imaxe 3.6.

3.3.3 Publicación

O conxunto de datos proposto foi liberado publicamente⁴ baixo unha licenza Mozilla [47], co obxectivo de fomentar a súa utilización e facilitar a investigación futura. Ao proporcionar acceso aberto a este dataset, preténdese contribuír ao avance do coñecemento na área de análise de discurso misóxico en redes sociais. A dispoñibilidade deste conxunto de datos permitirá que máis grupos de investigación poidan analizar, validar e ampliar os resultados obtidos, axudando así a acadar unha comprensión máis precisa e exhaustiva deste fenómeno.

⁴<https://github.com/luciamariaalvarezcrespo/GalMisoCorpus2023/>



Figura 3.6: O chío coa palabra *puta* está clasificado no MisoCorpus como misóxino, mais no contexto vemos que en realidade ten un ton satírico

no. Ao fomentar a colaboración e a reutilización dos datos, espérase que dean lugar a novas perspectivas e descubertas que poidan contribuir á loita contra a *misoxinia* en liña.

A licenza Mozilla foi escollida debido á súa natureza aberta e permisiva, ademais de por ser requisito para a obtención das axudas da Cátedra CICAS [48], permitindo a libre distribución e uso do conxunto de datos por parte da comunidade científica e garantindo a súa reproducibilidade e transparencia.

No caso da distribución pública do conxunto de datos cunha licenza Mozilla, é importante ter en conta as políticas de uso de *Twitter*⁵. Concretamente, o campo `content` do arquivo `tweets.csv`, é necesario que estea baleiro para cumprir con estas políticas. Isto ocorre debido ás restricións impostas por *Twitter* en relación á difusión de contidos da súa plataforma. Porén, esta restrición non se aplica ao conxunto de datos de *Mastodon*, xa que as súas políticas permiten a distribución dos contidos completos dos *toots*. É esencial respectar estas políticas e asegurarse de que o conxunto de datos cumpre cos requisitos de cada plataforma antes de distribuílo publicamente.

Ademais, para protexer a identidade dos usuarios, asegurouse de que os datos fornecidos nos ficheiros `toots.csv` e `tweets.csv` non conteñan información persoal directamente identificable, como nomes de usuarios. Ao adoptar medidas adecuadas para garantir o anonimato e privacidade, podemos utilizar estes datos para obter información valiosa sobre a *misoxinia* na lingua galega sen comprometer a seguridade ou privacidade dos individuos involucrados.

⁵<https://developer.twitter.com/en/developer-terms/policy>

Validación do corpus

COMO seguinte fase do traballo e máis desta memoria, exploraremos a continuación a validación do corpus a través do adestramento de modelos de *Machine Learning*, co obxectivo de realizar a súa avaliación. Presentaremos en detalle os procedementos de adestramento utilizados para obter modelos que proporcionen resultados precisos e consistentes. Ademais, debateremos a selección de métricas apropiadas para a avaliación do rendemento dos modelos aplicados ao corpus. Daremos énfase á utilización de técnicas de validación cruzada, que permiten obter estimacións robustas do rendemento do modelo. Previamente a isto, farase un preprocesado dos datos.

4.1 Preprocesamento do conxunto de datos seleccionado



Figura 4.1: Fase 3 do fluxo de traballo

Nesta sección falarase das etapas do tratamento dos datos antes de seren utilizados no proceso de análise. Realizaremos unha serie de transformacións e limpeza nos datos seleccionados co obxectivo de preparalos para a seguinte fase de análise.

A etapa de preprocesamento dos datos comprende varias tarefas importantes que contribuirán á calidade e fiabilidade dos resultados finais. Estes procesos poden ser: eliminación de caracteres ou símbolos innecesarios, eliminación de etiquetas [HTML](#) e eliminación doutros caracteres que non achegan información relevante para a análise textual. Ademais, aplicaremos técnicas de normalización, como a conversión a minúsculas, para garantir a consistencia na comparación de palabras. Alén diso, abordaremos a tarefa de eliminación de [emojis](#), símbolos ou caracteres especiais que poidan estar presentes nos datos. Aínda que algúns [emojis](#) poidan conter información relevante, a súa interpretación e análise requiren ferramentas específicas e, por tanto e neste contexto, optaremos por eliminarlos para simplificar a análise textual.

4.1.1 **Limpeza das mostras procedentes de Mastodon**

Nesta subsección abórdase o proceso de limpeza e preparación das mostras recollidas da plataforma [Mastodon](#). Nesta etapa, aplicaranse técnicas e métodos para garantir a calidade e fiabilidade dos datos utilizados no estudo. Describíranse as etapas de preprocesamento, incluíndo a eliminación de información non desexada, normalización de texto, tratamento de ruídos e eliminación de datos irrelevantes. A limpeza das mostras procedentes de [Mastodon](#) é fundamental para obter un conxunto de datos coeso e consistente, que poida ser adecuadamente utilizado nas etapas posteriores da investigación.

Unha vez completada a recuperación de todos os [toots](#), puido observarse que o texto destes contiña certos caracteres estraños que resultaban problemáticos para o tratamento posterior dos datos. Estes caracteres incluían [emojis](#), etiquetas [HTML](#) e [emojis](#) específicos e exclusivos da [instancia mastodon.gal](#). Para resolver esta cuestión, desenvolveuse un *script* que empregaba expresións regulares para eliminar todos estes elementos indesexados do texto dos [toots](#). A través dun proceso de filtrado, puido eliminarse de forma eficiente calquera característica non desexada e manter só o texto procesábel dos [toots](#).

Especificamente en relación cos [emojis](#), ao realizar a limpeza destes símbolos do texto, xeralmente búscase simplificar a análise e o procesamento dos datos, tornándoos máis axeitados para técnicas e algoritmos de procesamento de linguaxe natural. En primeiro lugar, non todos os algoritmos e modelos de análise de texto son capaces de lidar directamente cos [emojis](#), pois son caracteres especiais e poden introducir ruído indesexado nas análises. Ademais, os [emojis](#) poden ter diferentes representacións en distintas plataformas e sistemas operativos, o que pode dificultar a interpretación consistente e precisa. Tamén, a escaseza de ferramentas e recursos de [NLP](#) en galego pode dificultar a capacidade de procesar e interpretar adecuadamente os [emojis](#) presentes nas mensaxes. A falta de modelos preadestrados ou [lexicóns](#) específicos para o galego pode levar á necesidade de adaptar e desenvolver algoritmos propios ou realizar tarefas de tradución ou substitución dos [emojis](#) por palabras descritivas, o

que complica o proceso de análise. Tamén se decidiu eliminar os **emojis** no proceso de preprocesado debido á falta de captura destes caracteres no conxunto de datos misóxinicos. Para manter a concordancia e comparabilidade entre ambos os conxuntos (misóxico e non misóxico), optouse por realizar a eliminación dos **emojis** en ambos os casos. Deste xeito, aseguramos unha comparación equitativa e coherente entre os textos procesados, sen a presenza destes elementos visuais que poderían introducir diferenzas significativas na análise dos datos.

No caso dos **emojis** propios da instancia de Mastodon galega, foi necesario obter unha relación actualizada destes **emojis**, coñecida como *EMOJI_DATA*, para poder incorporalos ao proceso de filtrado mediante expresións regulares. Para facelo, utilizouse a función *mastodon.custom_emojis()*, que devolve unha listaxe de dicionarios de **emojis**. Con esta información, realizouse unha extensión na lista de **emojis** existente, na que se engadiron os **emojis** específicos da instancia de Mastodon empregada neste traballo. Isto realizouse a través dunha comprensión de lista, na que se percorrían os **emojis** personalizados e se extraía o seu *shortcode*. Posteriormente, estes **emojis** foron codificados e decodificados en UTF-8 para garantir a súa correcta inclusión na expresión regular. Ao obter os **emojis** propios da instancia de *mastodon.gal* utilizando a función *mastodon.custom_emojis()*, estes **emojis** son devoltos como cadeas de caracteres codificadas. Para poder incorporalos correctamente á expresión regular, é necesario asegurarse de que esas cadeas están na codificación adecuada. Neste caso, a utilización do método *encode('utf-8')* convérteas cadeas de caracteres nos seus equivalentes codificados en UTF-8. Isto asegura que os caracteres especiais ou **emojis** propios da instancia sexan tratados correctamente. Logo, co método *decode('utf-8')*, as cadeas de caracteres codificadas en UTF-8 son decodificadas de volta á súa forma textual, permitindo a súa inclusión adecuada na expresión regular.

```

1 def get_emoji_regexp():
2     # Sort emoji by length to make sure multi-character emojis are
3     # matched first
4     emojis = sorted(emoji.EMOJI_DATA, key=len, reverse=True)
5     # Add custom emojis
6     custom_emojis = mastodon.custom_emojis()
7     # Add custom emojis to the list of emojis
8
9     emojis.extend([emoji['shortcode'].encode('utf-8').decode('utf-8')
10                  for emoji in custom_emojis])
11    # Create regular expression pattern
12    pattern = u'(' + u'|'.join(re.escape(u) for u in emojis) + u)'
```

Deste xeito, grazas á eliminación de caracteres estraños e á incorporación dos **emojis** propios da instancia de Mastodon na expresión regular, puido obterse un texto limpo e coherente, listo para continuar co proceso de análise e tratamento dos datos. A través desta solución, pui-

do garantirse unha adecuada manipulación dos **toots** sen perder a información relevante que contiñan.

Un problema adicional que se detectou foi que os **emojis** especiais, que se atopaban entre dous “:” (dous puntos), non estaban a ser tratados correctamente ao eliminar o emoji do medio. Isto ocorreu porque, inicialmente, a expresión regular utilizada non tiña en conta estes “:” que rodeaban os **emojis**. Máis tarde, durante o preprocesamento de texto previo ao adestramento de modelos, elimináronse.

É importante mencionar que o conxunto de datos resultante despois do preprocesado descrito anteriormente tamén se fará publicamente dispoñíbel no repositorio. A existencia dun dataset xa procesado e listo para o uso presenta varias vantaxes significativas. En primeiro lugar, ao realizar o preprocesado previo dos datos, eliminamos a necesidade de que os usuarios interesados realicen eses pasos por si mesmos. Isto aforra tempo e esforzo, permitindo que se enfoquen directamente na análise e interpretación dos datos. En segundo lugar, ao contar cun dataset preprocesado, aseguramos unha maior consistencia e uniformidade nos datos. Igualmente, publicar o **corpus** sen procesar, especialmente tendo en conta a información proporcionada polos **emojis**, é de gran interese polas seguintes razóns. En primeiro lugar, ao dispoñer dun **corpus** sen procesar, os grupos de investigación interesados poderán acceder a unha versión orixinal dos datos, sen cambios ou alteracións. Isto permite unha maior transparencia e trazabilidade dos datos, permitindo que outros investigadores repliquen os experimentos e realicen análises adicionais. En segundo lugar, a información contida nos **emojis** pode ser moi relevante para a detección e análise de mensaxes misóxinas. Ao tamén publicar o **corpus** sen procesar, os usuarios poderán acceder á información dos **emojis** tal e como foi recollida, sen perder a súa riqueza semántica e interpretativa. Iso permitirá explorar o seu uso en conxunto co texto para obter unha comprensión máis completa e precisa das mensaxes misóxinas.

4.1.2 Limpeza das mostras procedentes de Twitter

A diferenza das mostras obtidas de **Mastodon**, as mostras procedentes de **Twitter** non requiren unha limpeza específica; xa que, ao seren recollidas por técnicas de web scrapping, foron capturadas sen etiquetas **HTML** e sen **emojis**.

Unha vez considerado isto, remitimos á sección seguinte. Nesa sección, abordaremos o preprocesamento que se aplica tanto ás mostras procedentes de **Twitter** como ás mostras de **Mastodon** despois de seren sometidas á limpeza. Detallaremos as tarefas comúns de preprocesamento, como a eliminación de **hashtags** e a normalización de texto, entre outras. Estas etapas aseguran que as mostras de ambos orixes sigan o mesmo proceso de preprocesamento e se integren nun único conxunto de datos coherente e preparado para a análise posterior.

4.1.3 Preprocesamento común das mostras

Aquí abordaremos o proceso de preprocesamento que se aplica de forma semellante a ambos os conxuntos de datos, tanto aos *toots* non misóxinos como aos *chíos* misóxinos. Nesta sección, describiranse as etapas e técnicas específicas utilizadas para limpar e preparar as mostras antes de sometelas ao adestramento dos modelos. O obxectivo deste preprocesamento común é garantir a consistencia e comparabilidade dos datos en ambos os conxuntos, facilitando así a extracción de características relevantes e a avaliación precisa do desempeño dos modelos aplicados. A través deste proceso, aseguramos unha base sólida e uniforme para a análise posterior, permitindo unha interpretación coherente e unha comparación adecuada entre os resultados obtidos nos dous conxuntos de datos.

Ao realizar o preprocesamento común das mostras, seguimos exactamente o mesmo procedemento utilizado no MisoCorpus [29], xa que as mostras misóxinas son extraídas deste *corpus*. Isto asegura a consistencia e coherencia nos métodos de preprocesamento aplicados en ambas as partes do noso estudo. Ao seguir este enfoque, garantimos que as mostras misóxinas beneficien das mesmas etapas de limpeza e preparación que as non misóxinas, o que nos permite obter resultados comparables e asegurar a coherencia nas nosas análises. Ademais, ao utilizar o mesmo preprocesamento que no MisoCorpus, tamén facilitamos a integración dos nosos resultados coñecidos nesta área de investigación e posibilitamos a comparación directa coas achegas previas realizadas nese *corpus*.

O *pipeline* proposto polos autores [29] consta dos seguintes pasos:

1. Converter cada *chío* en minúsculas
2. Eliminar liñas en branco e etiquetas *HTML*
3. Eliminar mencións e *hashtags*
4. Corrixir erros ortográficos
5. Eliminar símbolos repetidos continuamente

No caso das *URL*, *Twitter* utiliza un mecanismo de acurtamento que xera versións reducidas de *URL* que conteñen o prefixo `http://t.co` seguido dun identificador. Con todo, como a mesma *URL* pode xerar dous identificadores diferentes, isto dificulta a busca de mensaxes duplicadas e, por iso, decidiron eliminar o identificador e representar as *URL* co *token* `http://t.co`. Por último, no caso do símbolo *hashtag*, elimínano na versión normalizada do *chío*, pero mantendo a palabra, o que significa que o *hashtag* `#feminista` convértese en *feminista*.

Nós propoñemos engadir unha etapa previa na que identificamos se o campo `content` está ou non baleiro e eliminar eses textos do conxunto de adestramento. Isto é especialmente

relevante tendo en conta que, durante o proceso de eliminación dos **emojis** dos **toots**, podería ocorrer que quedase un texto baleiro (por exemplo, no caso de que a mensaxe contivera só ese tipo de caracteres). Deste xeito, asegurámonos de eliminar eses casos e garantir que só se manteñen as mostras con contido textual relevante para o noso proceso de preprocesamento e análise.

Como segundo paso do preprocesamento, todas as mostras de texto foron convertidas a minúsculas. Isto foi feito co obxectivo de unificar a forma en que as palabras están escritas. Ao pasar todo o texto a minúsculas, asegurámonos de que non hai diferenzas na representación das palabras por mor das maiúsculas ou minúsculas.

O terceiro paso, entón, foi a eliminación das liñas en branco, que non conteñen ningún contido textual. Estas liñas baleiras non contribúen á análise e poden ser omitidas, o que resulta nun texto máis coherente e compacto. O cuarto paso consistiu en eliminar os identificadores das **URLs**, polo motivo explicado anteriormente. O quinto paso consistiu en eliminar os símbolos de **hashtags**, tal como se indica no artigo de referencia [29]. Desta forma, obtemos un texto máis limpo e coherente para a análise.

Durante o sexto, eliminamos as mencións a outras contas/usuarios (@) como se indica no paper orixinal. As mencións son eliminadas co obxectivo de eliminar, como xa se comentou, as referencias directas a usuarios específicos e considerar unicamente o texto sen esas mencións. Isto axuda a manter a privacidade dos usuarios e a asegurar a confidencialidade dos datos. Alén diso, tamén facilita a análise posterior.

É importante mencionar que non se realizou un paso específico para corrixir as faltas ortográficas presentes nas mostras en lingua galega. Este enfoque ten en conta a realidade dos recursos limitados dispoñíbeis para o procesamento de textos en lingua galega. A corrección de faltas ortográficas é unha tarefa complexa que require ferramentas concretas. Debido á falta de ferramentas específicas e **corpus** de referencia para a lingua galega, non se puido realizar unha corrección ortográfica nas mostras do **corpus**. Así, as faltas ortográficas presentes nos datos non foron modificadas durante o preprocesamento. É importante ter en conta esta limitación ao analizar e interpretar os resultados derivados deste **corpus** preprocesado.

Adicionalmente, como último paso do preprocesamento, procedemos á eliminación de símbolos que se repiten de forma continua en cada mostra de texto. Esta acción ten en conta que, en moitos casos, a repetición excesiva de certos símbolos non achega información relevante para a análise textual e pode afectar negativamente ás etapas posteriores do proceso. Así, ao eliminar símbolos continuamente repetidos, buscamos reducir o ruído e garantir unha representación máis limpa e concisa do contido textual das mostras.

4.1.4 Creación dos embeddings

A continuación abordaremos o proceso de xeración de *sentence embeddings* a partir dos textos preprocesados. Estas representacións, coñecidas como *sentence embeddings*, capturan a información semántica e contextual dos textos, permitindo a súa comparación e análise posterior. Nesta sección, presentaremos o método utilizado para xerar os *embeddings* e a fonte dos modelos predestrados empregados neste proceso. O obxectivo é obter representacións eficientes e informativas dos textos que servirán como base para as tarefas de clasificación e *análise de sentimentos*.

Utilizamos a versión preprocesada do texto para xerar as súas *sentence embeddings*. As *sentence embeddings* están compostas por *word embeddings*, que son representacións densas de palabras dentro dun espazo de alta dimensionalidade, creando agrupacións de palabras que son semanticamente semellantes. As *sentence embeddings* poden ser representadas como un *Average of Word Embeddings (AWE)* no texto. As *sentence embeddings* compórtanse de forma semellante ás *word embeddings*, xa que comparten as mesmas propiedades principais [49]. Neste traballo, aplicamos o modelo Galician FastText [50], que contén *word embeddings* predestradas a partir do Common Crawl¹ e da Wikipedia.

No entanto, é importante salientar que, ao contrario do que foi feito no estudo orixinal [29], neste proxecto non foi posíbel realizar a etapa de extracción das *Linguistic Features (LF)* utilizando a ferramenta UMUTextStats [51]. Esta ferramenta é específica para a lingua castelá e permite obter información lingüística detallada sobre os textos, como o reconto de palabras, a frecuencia de letras, entre outros. Debido á falta de ferramentas equivalentes en galego, non puideron ser extraídas as *Linguistic Features* neste traballo. Porén, os demais pasos de preprocesamento e xeración de *embeddings* foron realizados para obter unha representación rica e informativa dos *toots* en lingua galega.

Ao non ter as *LF* en galego, perdemos unha fonte valiosa de información sobre os aspectos específicos da lingua que poderían influír na detección de mensaxes misóxinas. As *LF* inclúen elementos como a estrutura gramatical, o uso de certas palabras ou expresións, e as características propias do idioma. Estes aspectos son importantes para comprender completamente o contido dos textos e detectar sutilezas ou matices que poidan indicar contido misóxino. Sen esta información, poderíamos perder oportunidades de identificar mensaxes misóxinas que se expresan de forma particular na lingua galega.

¹commoncrawl.org

4.2 Experimentos realizados

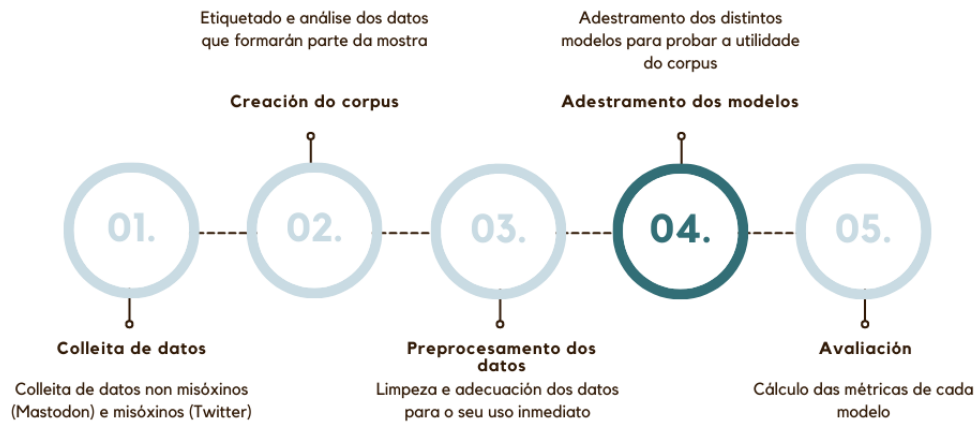


Figura 4.2: Fase 4 do fluxo de traballo

Nesta sección abordarase a fase de adestramento e avaliación dos modelos de [aprendizaxe automática](#) no noso proxecto. A continuación, exploramos a aplicación de algoritmos populares propostos no paper de referencia [29], como [Random Forest \(RF\)](#) [52], [Support Vector Machine \(SVM\)](#) [53] e [Linear Support Vector Machine \(LSVM\)](#) [54], utilizando a biblioteca [Scikit-Learn](#) [55]. Durante os experimentos, realizamos varias iteracións dependendo das técnicas de balanceamento de datos aplicadas. O obxectivo principal destes experimentos é analizar o rendemento dos modelos en termos de precisión, cobertura e outras métricas relevantes, probando así a utilidade do [corpus](#) para a detección de contido misóxico en [Twitter](#) e [Mastodon](#) en lingua galega.

4.2.1 Iteración 1: RF, SVM e LSVM

Na primeira iteración dos experimentos, adestramos os modelos [Random Forest \(RF\)](#), [Support Vector Machine \(SVM\)](#) e [Linear Support Vector Machine \(LSVM\)](#) sen realizar ningunha modificación no desequilibrio dos datos. O obxectivo desta etapa foi avaliar o rendemento destes algoritmos utilizando o conxunto de datos orixinal, sen aplicar técnicas de balanceamento. Iso permitiunos ter unha liña de base para comparar cos resultados das iteracións seguintes, onde se aplicaron enfoques para abordar o desequilibrio das clases.

Cada un dos algoritmos utilizados nos experimentos ten características e métodos de [aprendizaxe](#) diferentes:

- **Random Forest (RF):** é un algoritmo de [aprendizaxe supervisada](#) que utiliza un conxunto de árbores de decisión para realizar predicións. Cada árbore axústase de forma independente e o resultado final obtense combinando as predicións de todas as árbo-

res. Os Random Forests son coñecidos pola súa capacidade para manexar conxuntos de datos grandes e variables, e a súa habilidade para capturar relacións non lineais e complexas entre as variables de entrada. En canto aos seus hiperparámetros, deixáronse os por defecto do Scikit-Learn, igual que no traballo orixinal [29].

- **Support Vector Machines (SVM):** é un algoritmo de aprendizaxe supervisada utilizado tanto para clasificación como para regresión. O SVM busca atopar o hiperplano óptimo que mellor separa as clases no espazo de características. Pode manexar tanto datos linearmente separables como non linearmente separables utilizando diferentes funcións de kernel que mapean os datos a un espazo de maior dimensionalidade. O SVM destaca pola súa capacidade para manexar eficientemente conxuntos de datos de alta dimensionalidade. Os hiperparámetros escollidos manualmente son kernel polinomial e $C=1$, tal e como se indica no artigo orixinal [29].
- **Linear Support Vector Machines (LSVM):** é unha variante lineal do algoritmo SVM. A diferenza do SVM, o LSVM utiliza un hiperplano lineal para separar as clases no espazo de características. Aínda que o LSVM é máis simple en comparación co SVM, segue sendo eficaz en problemas de clasificación linearmente separables. Os hiperparámetros escollidos son penalización L1 e perda de tipo *squared hinge*, xa que son os escollidos no paper orixinal [29].

Random Forest	Support Vector Machine	Linear Support Vector Machine
n_estimators = 100	kernel = polinomial C = 1	penalty = L1 Loss = Squared Hinge

Táboa 4.1: Resumo dos hiperparámetros utilizados

Os autores do paper de referencia [29] non mencionan especificamente o tamaño dos conxuntos de adestramento e proba. A división 70-30 que decidimos facer, na que utilizamos o 70% dos datos para o adestramento e o 30% para a proba, é unha elección común na experimentación de algoritmos de aprendizaxe automática [56]. Esta división permite avaliar o rendemento dos modelos en datos non vistos previamente e axuda a estimar a súa capacidade de xeneralización.

Ao igual que nos experimentos mencionados en [29], tamén se empregou a técnica de *cross validation* de *10-fold* nos nosos experimentos. A *cross validation* é unha metodoloxía comunmente utilizada para avaliar o rendemento dun modelo de aprendizaxe automática. No caso da validación cruzada de *10-fold*, o conxunto de datos completo divídese en 10 partes, ou *folds*, de igual tamaño. Despois, realízanse 10 iteracións, nas cales se selecciona un *fold*

diferente como conxunto de proba en cada unha, e os restantes *folds* como conxunto de adestramento. Nesta etapa, o modelo é adestrado utilizando os datos do conxunto de adestramento e avaliado utilizando os datos do conxunto de proba. No traballo que nos serve de referencia tómasse a *accuracy* como métrica comparativa, mais nós modificaremos o *cross validation* para que tome o *F1-score* .

A escolla do *F1-score* fronte á *accuracy* vén dada pola natureza do noso proxecto. O *F1-score* é unha métrica que combina precisión (*precision*) e sensibilidade (*recall*). Esta combinación é especialmente útil cando temos desbalanceamento nas clases, como no noso caso (ver Fig. 3.5), o cal tamén é común en moitos proxectos de clasificación. O *F1-score* proporciona así unha medida máis completa do rendemento do modelo, tendo en conta tanto os falsos positivos como os falsos negativos. Debido a que o desbalanceamento nas clases é significativo, o *F1-score* semella ser unha opción máis axeitada para comparar o rendemento entre modelos.

Avaliamos a nosa proposta aplicando un modelo de representación de texto baseado en *Bag of Words (BoW)* . O *Bag of Words* é unha representación simplificada de documentos de texto utilizada na área de *NLP* . Neste enfoque, cada documento é tratado como un conxunto desordenado de palabras sen ter en conta a súa orde ou estrutura gramatical. O *BoW* foi seleccionado pola súa facilidade de implementación e porque produce bos resultados en tarefas de *NLP* [57]. Con todo, o *BoW* tende a crear un alto número de características que consomen moitos recursos en termos de tempo e memoria. As características de cada documento con *BoW* foron realizadas utilizando a versión normalizada do texto. Para calcular a porcentaxe de unigramas (palabras individuais) nos documentos, obtívose o *Term Frequency – Inverse Document Frequency (TF-IDF)* para medir a relevancia de cada característica dentro do *corpus* (ver ec. 4.1), utilizando a frecuencia do termo normalizada (ver ec. 4.2) para evitar sesgos con unigramas comúns.

$$tf - idf(t, d) = tf(t, d) * idf(t) \quad (4.1)$$

$$tf = \frac{\text{number of occurrences}}{\text{number of grams}} \quad (4.2)$$

$$idf(t) = \log \left(\frac{1 + n}{1 + df(t)} \right) + 1 \quad (4.3)$$

Na investigación que nos serve de referencia [29] non se especifica que algoritmo de selección de características para filtrar os unigramas máis discriminatorios utilizan, polo que utilizamos o método de chi-cadrado como unha opción comunmente empregada na selección de características para tarefas de clasificación de textos [58]. No noso proxecto, escollemos o método de chi-cadrado para filtrar os unigramas que son máis importantes para distinguir entre os textos misóxinos e non misóxinos. Este método baséase nunha proba estatística cha-

mada chi-cadrado, que nos axuda a medir a relación entre variables categóricas. No noso caso, consideramos cada unigrama como unha variable categórica e queremos determinar cales son os unigramas máis relevantes para a clasificación entre textos misóxinos e non misóxinos. Ao aplicar o método de chi-cadrado, podemos calcular unha puntuación de importancia para cada unigrama en relación coa variable obxectivo, que é a clasificación como misóxino ou non misóxino. Os unigramas que teñen unha puntuación de chi-cadrado máis alta son considerados como máis relevantes e teñen unha maior influencia na clasificación entre os dous tipos de textos. Estes unigramas poden axudarnos a distinguir entre os textos misóxinos e non misóxinos no noso conxunto de datos.

En resumo, o noso [pipeline](#) está composto por estes pasos:

1. Converter o texto nunha representación de [Bag of Words \(BoW\)](#).
2. Calcular a importancia de cada unigrama nos documentos aplicando a fórmula [Term Frequency – Inverse Document Frequency \(TF-IDF\)](#) utilizando a frecuencia do termo normalizada, que mide a relevancia dun unigrama en relación co conxunto de documentos.
3. Utilizar a función de puntuación chi-cadrado (χ^2) para realizar unha selección de atributos. Este paso filtra os unigramas máis discriminantes e selecciona os k mellores.
4. Empregar o clasificador utilizado para a clasificación dos textos. Neste caso, utilizamos cada un dos clasificadores propostos anteriormente ([Random Forest](#), [Support Vector Machine](#) e [Linear Support Vector Machine](#)) cos seus respectivos hiperparámetros seleccionados (ver táboa 4.1).

4.2.2 Iteración 2: undersampling con Random Undersampling

Na segunda iteración do noso proxecto, exploramos a técnica da submostraxe aleatoria, coñecida como [Random Undersampling \(RUS\)](#) [59], para tratar o desbalanceo dos nosos datos. O desbalanceo ocorre cando temos unha distribución desproporcionada das clases no noso conxunto de datos, o que pode afectar o rendemento dos nosos modelos de [aprendizaxe automática](#). Nesta subsección, describiremos en detalle o proceso de submostraxe aleatoria utilizado, así como os seus efectos na nosa avaliación dos modelos. O obxectivo é obter un conxunto de datos equilibrado, onde as clases minoritarias estean representadas de forma máis equitativa coas clases maioritarias, permitindo así unha mellor capacidade de xeneralización dos nosos modelos de clasificación.

Na segunda iteración do noso proxecto, continuamos a adestrar os mesmos modelos descritos na iteración anterior e utilizamos o mesmo [pipeline](#) para o adestramento dos modelos.

No entanto, unha diferenza chave nesta iteración é a aplicación da técnica de **RUS** aos datos da clase maioritaria, que neste caso son as mostras non misóxinas. O obxectivo é reducir a cantidade de mostras da clase maioritaria para equilibrar a distribución das clases no conxunto de datos. Isto axuda a mitigar o impacto do desbalanceo de clases nos nosos modelos de **aprendizaxe automática**. Ao aplicar o **RUS**, eliminamos aleatoriamente mostras da clase maioritaria ata que a proporción entre as clases minoritarias e a clase maioritaria sexa máis equitativa ou, no noso caso, exactamente igual. Deste xeito, aseguramos que o noso conxunto de datos teña unha representación máis balanceada das diferentes clases, o que podería mellorar o rendemento dos nosos modelos de clasificación.

Debido ás características especiais do noso conxunto de datos e ás súas necesidades de balanceo, descubrimos que a implementación estándar de **Random Undersampling** fornecida pola biblioteca `imbalanced-learn` (`imblearn`) [60] non era apropiada para o noso proxecto, xa que os nosos datos son de texto plano e non numéricos. Consecuentemente, tivemos que desenvolver unha implementación personalizada do **RUS** desde cero, específica para as necesidades deste proxecto. Esta implementación personalizada tivo en conta os requisitos particulares do noso conxunto de datos e as súas características únicas, co obxectivo de obter un balanceamento óptimo e obter mellores resultados nos nosos experimentos. Ao crear esta implementación personalizada do **RUS**, puidemos adaptar as estratexias e os métodos de submostraxe de maneira precisa á nosa tarefa de clasificación, o que nos permitiu controlar e axustar o balanceamento dos datos de forma máis efectiva.

Con esta implementación personalizada de **RUS**, superamos os desafíos específicos do noso proxecto en relación ao desbalanceo de clases. Ao aplicar esta técnica de submostraxe aleatoria aos nosos datos, conseguimos obter un conxunto de datos máis equilibrado e representativo de todas as clases. Isto permitiunos mitigar o posíbel impacto negativo do desbalanceo nas nosas avaliacións dos modelos e mellorar a súa capacidade de xeneralización. O noso obxectivo final é desenvolver modelos de clasificación robustos e precisos, e a utilización do **RUS** como parte do noso enfoque de preprocesamento de datos axudounos a avanzar neste sentido. Na seguinte sección, presentaremos os resultados dos nosos experimentos e compararemos o rendemento dos modelos adestrados na segunda iteración coa iteración anterior.

4.3 Resultados e comparativas

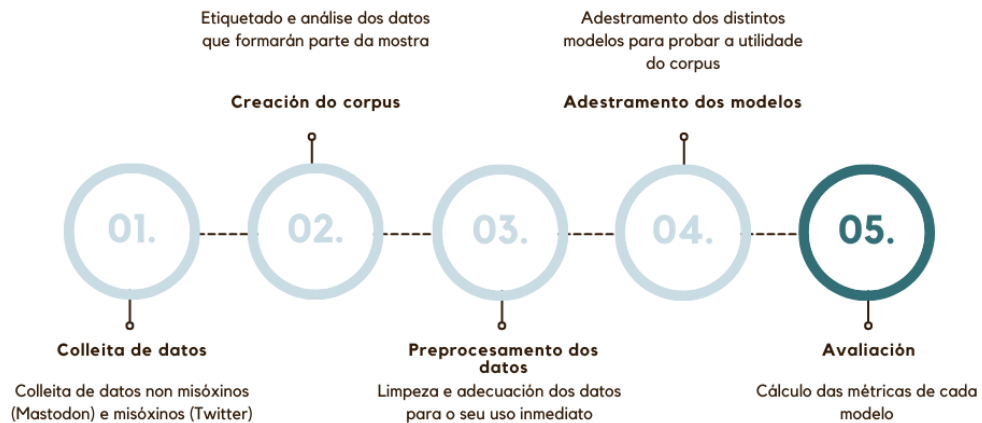


Figura 4.3: Fase 5 do fluxo de traballo

Esta última sección presenta os principais resultados obtidos ao longo do noso proxecto de clasificación de textos. Nela, analizamos o desempeño dos modelos de [aprendizaxe automática](#) adestrados nas diferentes iteracións, tendo en conta as estratexias de balanceamento empregadas. Avaliamos o rendemento dos modelos utilizando métricas relevantes como *recall*, *precision*, *accuracy* e *F1-score*. Tamén, realizamos comparativas entre os diferentes modelos e iteracións para determinar a súa eficacia relativa e identificar aqueles que ofrecen os mellores resultados.

Para calcular as métricas, primeiro debemos calcular o seguinte:

- **TP** (verdadeiros positivos): número de verdadeiros positivos, isto é, de predicións correctas para a clase positiva.
- **FN** (falsos negativos): número de falsos negativos, isto é, a predición é negativa cando realmente o valor debería ser positivo.
- **FP** (falsos positivos): número de falsos positivos, isto é, a predición é positiva cando realmente o valor debería ser negativo.
- **TN** (verdadeiros negativos): número de verdadeiros negativos, isto é, de predicións correctas para a clase negativa.

Feito isto, procederemos a calcular as métricas. O *recall* é a proporción $tp/(tp + fn)$. Esta métrica é intuitivamente a capacidade do clasificador para atopar todas as mostras positivas. A *precision* é a ratio $tp/(tp + fp)$ e é intuitivamente a capacidade do clasificador de non etiquetar como positiva unha mostra que é negativa. No caso do *accuracy*, este sería a taxa

de mostramos correctamente clasificadas fronte ao total de exemplos e se calcula como $(tp + tn)/(tp + fp + tn + fn)$. Finalmente, o *F1-score* pode ser interpretada como unha media harmónica da *precision* e o *recall*, onde un *F1-score* acada o seu valor óptimo en 1 e a súa puntuación peor en 0. A contribución relativa da *precision* e do *recall* ao *F1-score* é igual. A fórmula para calculalo é $2 * (precision * recall)/(precision + recall)$.

Para facilitar a comprensión desta importante parte desta memoria, presentamos gráficos e táboas para visualizar e comparar de forma clara os resultados obtidos. Analizamos os puntos fortes e debilidades de cada modelo, destacando aqueles que mostran un rendemento salientable e ofrecen unha boa capacidade de xeneralización. Ademais, proporcionamos unha análise detallada das diferenzas observadas entre as iteracións e as técnicas de balanceamento empregadas, identificando as súas implicacións no rendemento dos modelos.

4.3.1 Resultados da iteración 1

Como podemos observar polos datos obtidos na táboa 4.2, en xeral, os resultados das métricas para os tres modelos (RF, SVM, LSVM) son bastante semellantes. O *F1-score*, que é unha medida que combina a *precision* e o *recall*, é alto para os tres modelos, con valores próximos a 0.90, o que indica un bo equilibrio entre precisión e recuperación dos resultados. A precisión tamén é alta para os tres modelos, con valores superiores a 0.86. Isto indica que os modelos teñen unha alta capacidade para clasificar correctamente os casos positivos e minimizar os falsos positivos. O *recall*, que representa a capacidade do modelo para atopar todos os casos positivos, tamén é alto, con valores próximos a 0.93. Isto significa que os modelos teñen unha boa capacidade para identificar os casos positivos no conxunto de datos. A precisión e o *recall* son consistentes coa medida de *accuracy*, que é a proporción de predicións correctas en xeral. Os tres modelos teñen unha alta precisión e *recall*, o que se reflicte en valores de *accuracy* similares para os tres modelos, arredor de 0.93.

Neste contexto, o modelo de *Support Vector Machine (SVM)* é a mellor escolla entre os tres modelos avaliados, pois presenta o mellor *F1-score*, cun valor de aproximadamente 0.9101. O *F1-score* é unha medida que combina a precisión e o *recall*, e canto máis alto sexa, mellor será o equilibrio entre estas dúas métricas. O SVM tamén ten unha alta precisión, cun valor de aproximadamente 0.9428. Isto significa que o modelo ten unha excelente capacidade para clasificar correctamente os casos positivos e minimizar os falsos positivos. A alta precisión do SVM indica que está a realizar unha clasificación precisa e fiábel. Ademais, o *recall* do SVM é de aproximadamente 0.9391, o que indica a súa capacidade para identificar a maioría dos casos positivos no conxunto de datos. O alto *recall* implica que o modelo está a atopar a maior parte dos casos de interese, sen deixar moitos sen detectar. Tendo en conta estes factores, o SVM destaca como a mellor opción debido á súa combinación de alto *F1-score*, alta precisión e alto *recall*. Estes resultados indican que o modelo SVM ten un excelente rendemento na tarefa de

detección de *misoxinia* no *corpus* en lingua galega.

	RF	SVM	LSVM
F1-score	0.9038	0.9101	0.8975
Precision	0.9390	0.9428	0.8664
Recall	0.9348	0.9391	0.9308
Accuracy	0.9348	0.9391	0.9308

Táboa 4.2: Métricas da iteración 1

En resumo, os resultados das métricas son prometedores e suxiren que os modelos teñen un bo rendemento na tarefa de clasificación.

4.3.2 Resultados da iteración 2

Na segunda iteración do estudo, introduciuse unha técnica chamada *Random Undersampling* que non se empregou na primeira iteración. Esta técnica consiste en reducir aleatoriamente a cantidade de exemplos da clase maioritaria para equilibrar a proporción entre as clases do conxunto de datos. Esta técnica aplicouse co obxectivo de mitigar o problema de desequilibrio de clases e así tratar de mellorar o rendemento da clasificación na presenza de datos misóxinos e non misóxinos.

Ao observar os resultados das métricas na táboa 4.3, pódese apreciar que os valores para o *F1-score*, *precision*, *recall* e *accuracy* son bastante baixos en xeral. Iso indica que os modelos (RF, SVM, LSVM) non están a obter un bo rendemento na tarefa de detección. O *F1-score*, que combina a precisión e o recall, é especialmente baixo para os tres modelos, con valores que van desde o 0.3226 ata o 0.5118. Isto indica que os modelos están a ter dificultades para acadar un equilibrio entre a precisión e a capacidade de atopar os casos positivos nos datos. A precisión tamén é baixa, con valores que oscilan entre o 0.2404 e o 0.5425. Isto significa que os modelos están a cometer moitos falsos positivos ao clasificar os casos. O *recall*, que representa a capacidade de atopar os casos positivos, tamén é baixo, con valores que van desde o 0.4736 ata o 0.5375. Isto implica que os modelos están a deixar pasar moitos casos positivos sen detectar. O *accuracy* tamén é baixo, con valores que van desde o 0.4736 ata o 0.5375. Isto indica que os modelos non están a realizar predicións correctas en xeral.

Ao analizar os resultados das métricas na táboa, é evidente que os valores do *F1-score*, *precision*, *recall* e *accuracy* son bastante baixos en xeral. Isto pode ser atribuído ao uso da técnica de *Random Undersampling* nos datos. A submostraxe aleatorio é un enfoque comunemente utilizado para tratar con conxuntos de datos desequilibrados, onde a clase minoritaria é reducida para equilibrar a proporción coa clase maioritaria. Con todo, esta técnica pode

levar á perda de información e á redución do rendemento do modelo. No caso destes resultados, é posible que a aplicación do RUS teña levado a unha perda significativa de información relacionada coa clase misóxina, o que resultou en valores baixos de *F1-score*, *precision*, *recall* e *accuracy*. A submostraxe pode ter afectado a capacidade dos modelos de identificar correctamente os casos de *misoxinia*, o que resulta nun rendemento global insatisfactorio. Neste sentido, é importante considerar outras abordaxes para tratar os datos desequilibrados, como a sobre-mostraxe da clase minoritaria ou o uso de algoritmos de aprendizaxe de máquina deseñados para tratar directamente o desequilibrio de clases. Ademais, é fundamental analizar en detalle o impacto do RUS nos resultados e explorar outras estratexias para mellorar o rendemento do modelo na detección de *misoxinia*, como a optimización de hiperparámetros, a selección de características relevantes ou o uso de técnicas avanzadas de *Procesamento de Linguaxe Natural* específicas para a tarefa.

En resumo, os resultados das métricas indican que os modelos *RF*, *SVM*, *LSVM* non están a obter un bo rendemento na tarefa de detección de *misoxinia*. Sería necesario explorar posibles melloras nos modelos ou no conxunto de datos.

	RF	SVM	LSVM
F1-score	0.5118	0.4484	0.3226
Precision	0.5425	0.4766	0.2404
Recall	0.5375	0.4736	0.4903
Accuracy	0.5375	0.4736	0.4903

Táboa 4.3: Métricas iteración 2

Seguimento

O seguimento do proxecto é esencial para asegurar a súa eficaz execución e obter resultados significativos. Neste capítulo, abordaremos de forma sistemática e detallada a evolución do proxecto, desde control das tarefas ata os custos. Realizaremos unha análise exhaustiva das etapas e actividades realizadas, así como tamén avaliaremos o progreso xeral do traballo.

5.1 Control de tarefas

Para levar un control das tarefas realizadas durante o transcurso dos meses, optouse por manter un caderno de notas de traballo no que se tiña un rexistro das actividades a realizar día por día que se obtiñan durante as reunións semanais coa directora do proxecto, estimando a grandes trazos as horas de esforzo para cada unha das tarefas. A utilización deses documentos a modo de *backlog* serviu como guía para axustar a planificación e os prazos.

Inicialmente, levouse a cabo unha fase de recompilación de información e estudo teórico, na que se investigaron as principais técnicas de NLP utilizadas na extracción de comentarios misóxinicos en redes sociais, así como as investigacións máis recentes relacionadas co NLP en lingua galega. A continuación, procedeuse á creación, análise e preprocesamento do conxunto inicial de datos para obter un primeiro conxunto de adestramento. Neste punto, foi necesario aplicar os coñecementos adquiridos na fase anterior e empregar distintas estratexias debido á dificultade de recompilación dos nosos datos de entrada, tal e como se explicou en detalle no capítulo 3. Despois disto, realizouse unha análise exhaustiva dos algoritmos e métodos utilizados na literatura para a detección de comentarios misóxinicos en redes sociais. Tendo en conta os obxectivos e as características do proxecto, seleccionáronse os modelos máis adecuados, que foron posteriormente adestrados utilizando o conxunto de adestramento previamente preparado. Durante o adestramento, foron aplicadas técnicas de axuste de balanceo no conxunto de datos para tratar de mellorar o rendemento dos modelos, como po-

demos observar no capítulo 4. Unha vez adestrados os modelos, procedeuse á súa avaliación utilizando diversas métricas de avaliación, como *precision*, *recall*, *accuracy* e *F1-score*. Ademais, realizáronse comparativas entre os distintos modelos para identificar o máis eficaz na detección de comentarios misóxinos en lingua galega. Finalmente, avaliáronse os resultados e comentáronse as conclusións desta análise, proporcionando unha visión completa do proceso de elección e adestramento dos modelos utilizados neste estudo.

En resumo, as tarefas realizadas foron:

- Investigación en tecnoloxías
 - Recompilación do estado da arte
 - Apredizaxe de tecnoloxías
- Creación dun conxunto de datos
 - Clase non misóxina (Mastodon)
 - Clase misóxina (Twitter)
- Iteración 1: RF, SVM e LSVM
 - Deseño e implementación
 - Adestramento e probas
 - Análise de resultados
- Iteración 2: undersampling con RUS
 - Deseño e implementación
 - Adestramento e probas
 - Análise de resultados
- Documentación

A dimensión temporal de cada unha destas tarefas reflectiuse no Diagrama de Gantt da figura 5.1. O detalle do esforzo de cada tarefa coméntase en máis detalle na sección seguinte.

No desenvolvemento deste proxecto, un dos aspectos máis salientables foi a recompilación dos datos, que requiriu moito máis tempo do inicialmente planificado. A obtención dun conxunto de datos representativo para a detección de comentarios misóxinos en lingua galega foi un proceso complexo e desafiante. A pesar dos esforzos e das estratexias empregadas, a recompilación dos datos mostrouse máis laboriosa e demorada do que se agardaba inicialmente. Este aspecto destacado tivo un impacto significativo na planificación e na execución

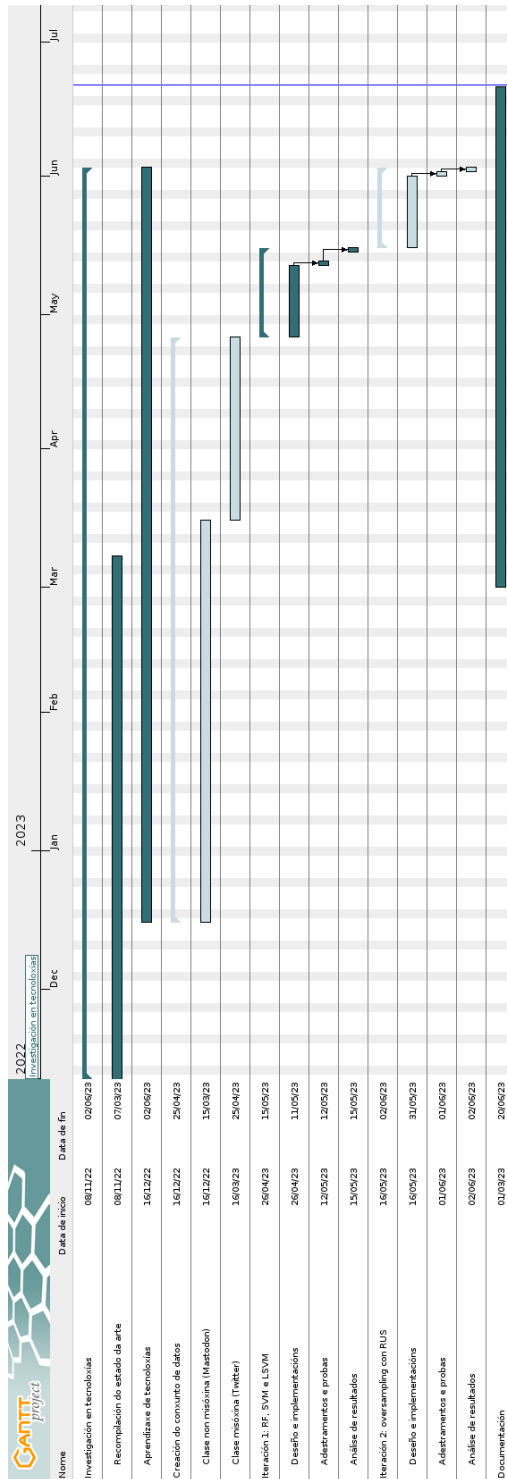


Figura 5.1: Diagrama de Gantt do proxecto

do proxecto, requirindo adaptacións e reaxustes para cumprir cos obxectivos establecidos. A pesar dos obstáculos e da maior demanda de tempo, a importancia deste proceso radica na fiabilidade e na calidade do conxunto de datos resultante, que constitúe a base para o desenvolvemento das técnicas de detección de *misoxinia* neste proxecto.

5.2 Custos do proxecto

Para poñer en valor o traballo realizado, preséntase unha estimación aproximada dos custos baseada no esforzo en horas-persoas e tomando como referencia unha tarifa por hora de traballo. Ademais, tense en conta a necesidade de asignar recursos para a adquisición de tecnoloxía e hardware especializado para o correcto desenvolvemento deste proxecto.

Os recursos de persoal necesarios para a realización do proxecto consistiron nunha desenvolvedora junior e unha directora de proxecto. Para determinar os custos de persoal, tomouse en consideración o soldo bruto anual establecido no convenio colectivo [61] para a categoría de desenvolvedor junior, que é de 15.860,56€ brutos/anuais, mesmo sendo conscientes de que o mercado actual móvese xeralmente por riba desta cifra. Con base neste convenio, o custo por hora de traballo foi calculado en arredor de 8,81€ brutos/hora. As horas de traballo no caso da desenvolvedora junior calculamos que son arredor de 1.800 horas e no caso da directora do proxecto o número de horas anuais ten un máximo de 1.642 [62]. Se consideramos o soldo base bruto dun titular de universidade sen complementos en 37.567,18€ [63], estimamos o custo por hora de traballo da directora do proxecto nuns 22,88€ brutos/hora. Pódense observar todos estes datos reflectidos na táboa 5.1.

Rol	Custo por hora	Estimación total de horas	Custo total
Desenvolvedora junior	8,81€	360	3.171,6€
Directora do proxecto	22,88€	32	732,16€

Táboa 5.1: Custos totais estimados

Na táboa 5.2 presentamos o detalle dos custos asociados a cada unha das tarefas planificadas, tendo en conta as horas estimadas de traballo bruto. Deste xeito, o custo total do proxecto ascende a 3.908,73€ brutos. Esta análise permite unha avaliación precisa dos custos do proxecto e axuda a garantir unha xestión financeira responsable e eficiente, pois a través dunha planificación e cálculo precisos, asegúrase a viabilidade económica e a correcta asignación dos recursos.

En canto aos recursos técnicos utilizados, optouse por empregar un ordenador portátil persoal e un ordenador de sobremesa. Estes equipos, adquiridos previamente, ofrecían as características e prestacións necesarias para o desenvolvemento do proxecto sen implicar custos

adicionais. Os detalles técnicos destas máquinas pódense consultar na táboa 5.3.

Dado que non se realizou unha asignación orzamentaria específica para a execución deste *Traballo de Fin de Grao*, senón que os equipos utilizados estaban previamente dispoñíbeis, excluimos da análise económica o custo destes recursos técnicos, de modo que o custo total do proxecto está constituído exclusivamente polo custo do persoal. Ademais, non houbo necesidade de incorrer en gastos relacionados coa adquisición de licenzas de software nin co acceso á *API* de Twitter, debido a que se situaba nuns 100\$ ao mes e por iso optamos por realizar web scrapping, eliminando así posíbeis custos asociados a estas áreas.

Tarefa	Duración	Custo
Investigación en tecnoloxías	226 h	1.991,06 €
Creación do conxunto de datos	140 h	1.233,40 €
Iteración 1: RF, SVM e LSVM	18 h	158,58 €
Iteración 2: undersampling con RUS	4 h	35,24 €
Documentación	148 h	1.303,88 €
Esforzo desenvolvedora junior	360 h	4.722,16 €
Esforzo titora	32 h	732,16 €
Total proxecto	392 h	3.908,73 €

Táboa 5.2: Custos de persoal asociados ao proxecto

Características	Ordenador portátil	Ordenador de sobremesa
Sistema operativo	Ubuntu 22.04.2 LTS	Microsoft Windows 10 Pro
CPU	Intel®Core™i7-4510U CPU @ 2.00Ghz	Intel®Core™i7-10700KF CPU @ 3.80Ghz 3.79 Ghz
RAM	8 GB	32 GB
Placa base	Dell Inc. XPS13 9333	ASUS GAMING Z490-PLUS
GPU	<i>Integrada</i>	NVIDIA GeForce RTX 3080

Táboa 5.3: Especificacións técnicas das máquinas

Conclusións

DURANTE este capítulo, presentamos a conclusión final do noso proxecto, onde faremos un resumo da situación final e destacaremos as leccións aprendidas ao longo do proceso. Tamén exploraremos as posibles liñas futuras de traballo que poderían ampliar e mellorar os resultados obtidos. O noso obxectivo é ofrecer unha visión global do proxecto, reflectir sobre os seus logros e dificultades e proporcionar recomendacións e suxestións para futuras accións e investigación. Neste peche final, poñeremos en perspectiva o impacto e as aplicacións prácticas do noso traballo.

6.1 Obxectivos alcanzados

Como comentamos no primeiro capítulo, a pesar da grande popularidade que está a ter a [análise de sentimentos](#), non existen moitas solucións aplicadas á detección de [misoxinia](#), e especialmente en linguas minorizadas como o galego. Se pensamos no alto fluxo de persoas que usan [Twitter](#) nun mes, podemos facernos unha idea do impacto que un sistema de detección de [misoxinia](#) podería ter en termos de identificar e eliminar contido misóxico, promover unha comunidade en liña máis respectuosa e fomentar un ambiente seguro e inclusivo para as usuarias. Ademais, a detección e a prevención da [misoxinia](#) en liña son cada vez máis importantes para combater a discriminación, promover a igualdade de xénero e protexer os dereitos humanos nas plataformas dixitais.

O obxectivo principal deste traballo foi desenvolver un primeiro [corpus](#) en lingua galega para a detección de mensaxes misóxicas en [Twitter](#) e [Mastodon](#). Ademais diso, tamén construímos un sistema de clasificación baseado en algoritmos de [aprendizaxe automática](#) para identificar e etiquetar automaticamente as mensaxes misóxicas no [corpus](#). Isto deu lugar a sistemas de clasificación resultantes capaces de analizar novas mensaxes en lingua galega en tempo real e determinar se conteñen contido misóxico ou non.

Como consecuencia do cumprimento do primeiro obxectivo, o **corpus** está dispoñíbel tanto na súa forma orixinal como despois de ser sometido a un proceso de limpeza. Os datos están accesíbeis a través da seguinte dirección web: [GalMisoCorpus 2023](#). Adicionalmente, respecto ao segundo obxectivo, desenvolveuse e configurouse un sistema de clasificación que pasou por varias iteracións individuais, sendo avaliadas e comparadas mediante distintas métricas.

En xeral, os resultados obtidos mostraron un rendemento prometedor na detección de mensaxes misóxinas en lingua galega en [Twitter](#) e [Mastodon](#). Os modelos da primeira iteración acadaron valores altos de *precision*, *recall* e *F1-score*, o que indica unha capacidade adecuada para identificar e clasificar correctamente as mensaxes misóxinas.

É importante destacar que este traballo foi pioneiro na creación do primeiro **corpus** en lingua galega para este fin, o cal representa unha contribución significativa á investigación en lingua galega e ás ferramentas dispoñíbeis para a detección de mensaxes misóxinas nesta lingua. O **corpus** creado proporciona unha base de datos valiosa que permitirá a realización de estudos adicionais, o desenvolvemento de novos algoritmos e a avaliación continua do desempeño dos sistemas de detección.

6.2 Aplicabilidade

En canto ás aplicacións inmediatas que pode ter este **corpus**, o ParityBot [64] é un exemplo dunha aplicación real que está a ser utilizada para detectar e combater a **misoxinia** nas redes sociais. Desenvolto pola empresa Areto Labs¹ de Canadá, o ParityBot utiliza algoritmos de **aprendizaxe automática** para identificar comentarios misóxinos en [Twitter](#) en tempo real. Este bot monitoriza continuamente os **chíos** que se publican cara mulleres políticas en Canadá e Nova Zelandia e analiza o seu texto en busca de contido misóxino. Cando se detecta un comentario misóxino, o ParityBot responde automaticamente ao chío cunha mensaxe positiva para tratar de contrarrestalo. O obxectivo do ParityBot é destacar a presenza e a frecuencia da **misoxinia** nas redes sociais e promover unha cultura de debate máis respectuosa e inclusiva. Ademais de chamar a atención sobre o problema, tamén ofrece ás mulleres políticas unha ferramenta para informarse sobre os comentarios misóxinos que reciben e para acadar unha maior visibilidade dos mesmos. O ParityBot é un exemplo de como as aplicacións de detección automatizada de **misoxinia** están a comezar a ser implementadas no mundo real. Estas tecnoloxías teñen o potencial de facer fronte á **misoxinia** en tempo real e de fomentar unha cultura en liña máis respectuosa e igualitaria. Porén, é importante sinalar que o ParityBot actualmente só é aplicábel ao idioma inglés. Por conseguinte, co uso deste **corpus**, ábrese a posibilidade de iniciar a investigación e desenvolvemento dunha versión en lingua galega. Co tempo, é posíbel que vexamos máis iniciativas semellantes que utilicen algoritmos

¹<https://www.aretolabs.com/>

de [aprendizaxe automática](#) para abordar o problema da [misoxinia](#) nas redes sociais e noutras plataformas en liña.

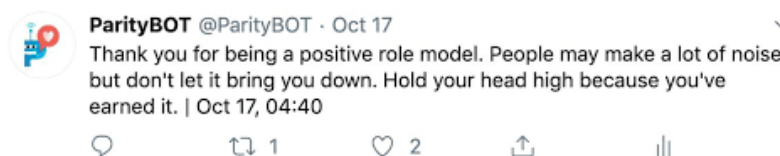


Figura 6.1: Exemplo de chío enviado polo ParityBot

En resumo, este traballo demostrou a viabilidade de desenvolver un [corpus](#) exclusivo para a detección de mensaxes misóxinas en lingua galega en [Twitter](#) e [Mastodon](#). Os resultados obtidos proporcionan unha base sólida para futuras investigacións neste campo e para o desenvolvemento de sistemas de detección de contido ofensivo e discriminatorio en contornas de redes sociais en galego.

6.3 Leccións aprendidas

Durante o desenvolvemento do [corpus](#) de detección de [misoxinia](#) en lingua galega, xurdiron diversas leccións aprendidas que contribuíron á mellora continua do proxecto e ao desenvolvemento persoal no campo do [Machine Learning](#) e do [Procesamento de Linguaxe Natural](#). Ao longo deste percorrido, identificamos a necesidade de ter en conta as particularidades lingüísticas dunha lingua para unha análise precisa da [misoxinia](#). Ademais, mantivemos un enfoque ético e respectuoso ao garantir a privacidade e anonimato dos usuarios implicados. Estas leccións aprendidas resultaron fundamentais para acadar un [corpus](#) actualizado, de calidade e adaptado á evolución da linguaxe e das formas de expresión misóxina na comunidade galega.

No transcurso do proceso de desenvolvemento do [corpus](#), unha das leccións de maior relevancia que se destacou foi a importancia crucial de dispor dunha documentación completa e detallada. A documentación cumpre un papel fundamental ao proporcionar unha guía clara e precisa sobre os distintos usos e particularidades dun software específico. A existencia dunha documentación adecuada tería axudado a garantir a coherencia e a consistencia durante as tarefas de recompilación de textos, así como a prevención de numerosos problemas que se manifestaron ao longo do proxecto. Este achado subliña a necesidade esencial de contar cunha documentación exhaustiva, a cal non só facilitaría unha orientación clara para os usuarios, senón que tamén tería atenuado os contrastempos e as dificultades que tiveron xurdido ao longo do proceso de desenvolvemento do [corpus](#).

Outra lección significativa que se obtivo durante o desenvolvemento do [corpus](#) foi a constatación da falta de ferramentas de [Procesamento de Linguaxe Natural](#) dispoñíbeis para a lin-

gua galega. Este descubrimento puxo de manifesto unha necesidade urxente de investigación e desenvolvemento de recursos e técnicas específicas para o NLP en galego. A escaseza de ferramentas e recursos para a lingua galega representa unha barreira significativa para a análise e comprensión de textos nesta lingua, limitando así as posibilidades de aplicación do NLP en ámbitos como a clasificación de textos, a *análise de sentimentos* e outras tarefas relacionadas. Esta lección resalta a importancia de promover a investigación e o desenvolvemento de ferramentas e recursos de NLP para as linguas minorizadas, co obxectivo de contribuír á defensa dos dereitos lingüísticos das comunidades correspondentes, a través da súa maior igualdade e representatividade lingüística no ámbito do *Procesamento de Linguaxe Natural*.

É importante salientar a importancia de garantir a reproducibilidade dos resultados na divulgación de resultados científicos. Ao longo do desenvolvemento do proxecto, evidenciouse que a reproducibilidade é fundamental para validar e verificar as descubertas obtidas, xa que no noso proxecto non foi posíbel replicar completamente o traballo orixinal por falta de transparencia. Ao compartir os datos, o código e os procedementos utilizados, bríndase aos outros equipos investigadores a oportunidade de replicar os experimentos e validar os resultados de forma independente. Isto non só fortalece a confianza nos resultados presentados, senón que tamén fomenta a colaboración e o avance da comunidade científica en xeral. É por iso que, neste sentido, subliñouse a importancia de compartir os conxuntos de datos utilizados, respectando as consideracións de privacidade e confidencialidade, para que outros grupos de investigación poidan avaliar e comparar os resultados de forma rigorosa e transparente.

Para rematar, tamén é relevante mencionar as materias do grao que resultaron máis beneficiosas para a realización deste proxecto. Entre elas destacan: Recuperación da Información, na que se abordaron conceptos como o TF-IDF que resultaron de grande utilidade para a extracción e análise dos datos; Aprendizaxe Automática, onde se afondou na clasificación binaria e se exploraron algoritmos clásicos de ML, como as SVM, que axudaron na implementación do sistema de clasificación; e Deseño de Sistemas Intelixentes, onde se adquiriron coñecementos en planificación e deseño de proxectos de ML, os cales foron aplicados na organización e desenvolvemento deste traballo. A través destas materias, foi posíbel adquirir as habilidades e os coñecementos necesarios para abordar con éxito os retos e os obxectivos deste proxecto.

6.4 Traballo futuro

Finalmente, propoñemos diversas liñas de investigación e desenvolvemento que poderían ser exploradas para mellorar e ampliar as capacidades do noso sistema de detección de mensaxes misóxinas en lingua galega. Estas propostas teñen como obxectivo seguir avanzando no campo da detección de contido misóxino nas redes sociais.

Como traballo futuro, suxírese ampliar o conxunto de datos utilizado para o adestramento,

xa que a dispoñibilidade dunha maior cantidade e variedade de mensaxes podería mellorar aínda máis o desempeño do sistema. A pesar de que desenvolvemos un primeiro *corpus* para este traballo, é necesario seguir recompilando e etiquetando máis datos en lingua galega para enriquecer e diversificar o conxunto de adestramento dos nosos modelos. Isto axudaría a aumentar a súa precisión e capacidade para identificar novos *padróns* de *misoxinia*.

Unha das ideas de traballo futuro que emerxe deste traballo é a posibilidade de solicitar acceso aos *toots moderados* na *instancia mastodon.gal*. A colaboración co administrador desta instancia podería permitir obter datos en lingua galega, o que reduce a necesidade de tradución e ofrece unha valiosa fonte de información. A análise dos *toots moderados* podería ofrecer datos relevantes para a detección de comentarios misóxinos en lingua galega, sempre e cando estes *toots* estean dispoñíbeis e se conserven. A inclusión destes datos no *corpus* permitiría ampliar a cobertura do proxecto e afondar na análise da presenza de *misoxinia* nas redes sociais. Ademais, a colaboración e o intercambio de coñecementos con comunidades e administradores de instancias semellantes poderían enriquecer futuras investigacións neste e noutros campos relacionados coa lingua galega e as redes sociais.

Tamén, outra liña de traballo futuro importante é o desenvolvemento de *lexicóns* ou modelos que dean soporte aos *emojis* na detección de mensaxes misóxinas. Os *emojis* son elementos moi utilizados nas redes sociais e poden transmitir emocións, actitudes ou sentimentos concretos. No contexto da detección de *misoxinia*, os *emojis* poderían ser indicadores importantes para identificar contido misóxino ou ofensivo. Ademais, é necesario explorar a incorporación de *Linguistic Features* en galego no sistema de detección de mensaxes misóxinas. Ao facelo, poderíamos obter unha comprensión máis profunda das particularidades da lingua galega e capturar os matices específicos que poderían indicar contido misóxino.

Para rematar, sería interesante considerar como traballo futuro probar o noso *corpus* con técnicas de *oversampling*. O *oversampling* é unha técnica utilizada para abordar o desequilibrio de clases nos datos de adestramento, que é moi común en problemas de detección de contido misóxino. Neste traballo, só abordamos directamente o desequilibrio de clases a través do *undersampling*, pero a aplicación de técnicas de *oversampling* podería axudar a mellorar a precisión da nosa detección.

Relación de Acrónimos

- AMI** Automatic Misogyny Identification. 10
- API** Application Programming Interface. 14, 16–24, 30, 58
- AWE** Average of Word Embeddings. 43
- BoW** Bag of Words. 46, 47
- CSV** Comma-Separated Values. 16–18, 24, 32
- DL** Deep Learning. 30
- HTML** HyperText Markup Language. 15, 38, 40, 41
- HTTP** Hypertext Transfer Protocol. 16, 17
- JSON** JavaScript Object Notation. 17
- LF** Linguistic Features. 43, 64
- LSVM** Linear Support Vector Machine. 44, 45, 47, 50–52
- ML** Machine Learning. 4–6, 15, 37, 62, 63
- NLP** Procesamento de Linguaxe Natural. 1–5, 12, 13, 15, 38, 46, 52, 54, 62, 63
- RAM** Random Access Memory. 21
- RF** Random Forest. 5, 44, 47, 50–52
- RUS** Random Undersampling. 47, 48, 51, 52

SVM Support Vector Machine. 12, 44, 45, 47, 50–52, 63

TF-IDF Term Frequency – Inverse Document Frequency. 46, 47

TFG Trabajo de Fin de Grao. 3–5, 58

URL Uniform Resource Locator. 18, 19, 24, 25, 29, 41, 42

VAWIP Violence Against Women In Politics. 2, 9

Glosario

- análise de sentimentos** Proceso de análise de texto dixital para determinar se o ton emocional da mensaxe é positivo, negativo ou neutro. 3, 5, 9–12, 43, 60, 63
- aprendizaxe automática** Eido da Intelixencia Artificial que trata o desenvolvemento de algoritmos e modelos estatísticos que utilizan os sistemas de computación co fin de levar a cabo tarefas sen instrucións explícitas, no canto de basearse en padróns e inferencias. 2, 4, 6, 11, 14, 30, 32, 44, 45, 47–49, 60–62
- batches** conxunto de tarefas ou procesos que son executados de xeito secuencial e sen necesidade de intervención manual durante a execución. 21, 22
- chíos** Mensaxes ou publicacións na rede social Twitter. 10, 12, 15, 16, 23–27, 29, 31, 34, 41, 61
- corpus** Conxunto de textos ou datos lingüísticos recompilados e organizados para o seu estudo e análise. iii, 1, 3–6, 9–15, 23, 25, 27, 31–34, 37, 40–42, 44, 46, 51, 60–62, 64
- discriminación** Trato desigual ou inxusto cara unha persoa ou grupo por razóns de sexo, raza, relixión.... 1–3
- embeddings** representación numérica de palabras ou documentos que captura información semántica e estrutural das súas características. 43
- emojis** Xeroglífico ou debuxo insertado en forma de texto. 26, 31, 38–40, 42, 64
- hashtags** palabras chave que se pode premer e que van precedidas por un cancelo (#). 40–42
- instancia** Cada versión independente da rede social Mastodon, con comunidades de usuarios e normas propias. 4, 5, 15–23, 33, 38, 39, 64
- lexicóns** conxunto de palabras e as súas correspondentes informacións léxicas, como a súa forma, significado e atributos sintácticos. 38, 64

- Mastodon** Plataforma de rede social descentralizada alternativa a Twiter. 1, 3–5, 9, 11, 15–23, 32, 33, 36, 38–40, 44, 60–62
- misoxinia** Odio ou prexuízo contra as mulleres. 1–6, 9–16, 23, 24, 32–34, 36, 51, 52, 57, 60–62, 64
- moderados** Sometidos a algún tipo de control ou supervisión para garantir a súa adeuación ás normas establecidas. 5, 23, 64
- padróns** Regularidades ou estruturas recorrentes. 2, 22, 64
- pipeline** cadea de procesos conectados de xeito que a saída de cada elemento é a entrada do seguinte. 41, 47
- slangs** Xerga ou linguaxe coloquial e informal propia do inglés. 11
- toots** Mensaxes ou publicacións na rede social Mastodon. 15–21, 23, 36, 38, 40–43
- Twitter** Rede social amplamente utilizada para compartir mensaxes curtas. 1, 3–5, 9–11, 15–17, 22–27, 32, 36, 40, 41, 44, 60–62

Bibliografía

- [1] E. A. Vogels, “The state of online harassment,” 2021, consultado o 22 de xuño de 2023. [En liña]. Dispoñible en: <https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/>
- [2] L. LC, “Misogyny manifestation across all social media platforms,” 2020, consultado o 22 de xuño de 2023. [En liña]. Dispoñible en: <https://ohpi.org.au/misogyny-manifestation-across-all-social-media-platforms/>
- [3] “Apis - mastodon documentation,” consultado o 22 de xuño de 2023. [En liña]. Dispoñible en: <https://docs.joinmastodon.org/api/>
- [4] E. Fersini, P. Rosso, and M. Anzovino, “Overview of the task on automatic misogyny identification at ibereval 2018.” *Ibereval@ sepln*, vol. 2150, pp. 214–228, 2018.
- [5] E. Siapera, *Online Misogyny as Witch Hunt: Primitive Accumulation in the Age of Techno-capitalism*. Cham: Springer International Publishing, 2019, pp. 21–43. [En liña]. Dispoñible en: https://doi.org/10.1007/978-3-319-96226-9_2
- [6] I.-P. Union, “Sexism, harassment and violence against women in parliaments in europe,” Inter-Parliamentary Union, Tech. Rep., 2018.
- [7] M. L. Krook and J. Restrepo Sanín, “The cost of doing politics? analyzing violence and harassment against female politicians,” *Perspectives on Politics*, vol. 18, no. 3, p. 740–755, 2020.
- [8] M. Z. Kurdi, *Natural language processing and computational linguistics 2: semantics, discourse and applications*. John Wiley & Sons, 2017, vol. 2.
- [9] M. D. Lytras, A. Visvizi, and J. Jussila, “Social media mining for smart cities and smart villages research,” *Soft Computing*, vol. 24, no. 15, pp. 10 983–10 987, 2020.

- [10] S. Zhang, X. Zhang, J. Chan, and P. Rosso, “Irony detection via sentiment-based transfer learning,” *Information Processing & Management*, vol. 56, no. 5, pp. 1633–1644, 2019. [En liña]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0306457318307428>
- [11] M. Corazza, S. Menini, E. Cabrio, S. Tonelli, and S. Villata, “A multilingual evaluation for online hate speech detection,” *ACM Trans. Internet Technol.*, vol. 20, no. 2, mar 2020. [En liña]. Disponible en: <https://doi.org/10.1145/3377323>
- [12] W. Yin, V. Agarwal, A. Jiang, A. Zubiaga, and N. Sastry, “Annobert: Effectively representing multiple annotators’ label choices to improve hate speech detection,” *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 17, no. 1, pp. 902–913, Jun. 2023. [En liña]. Disponible en: <https://ojs.aaai.org/index.php/ICWSM/article/view/22198>
- [13] R. Priyadharshini, B. R. Chakravarthi, S. Cn, T. Durairaj, M. Subramanian, K. Shanmugavadivel, S. U Hegde, and P. Kumaresan, “Overview of abusive comment detection in Tamil-ACL 2022,” in *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 292–298. [En liña]. Disponible en: <https://aclanthology.org/2022.dravidianlangtech-1.44>
- [14] M. A. Alqahtani, “Factors affecting cybersecurity awareness among university students,” *Applied Sciences*, vol. 12, no. 5, p. 2589, Mar. 2022. [En liña]. Disponible en: <https://doi.org/10.3390/app12052589>
- [15] O. Parlangeli, E. Marchigiani, M. Bracci, A. M. Duguid, P. Palmitesta, and P. Marti, “Offensive acts and helping behavior on the internet: An analysis of the relationships between moral disengagement, empathy and use of social media in a sample of italian students,” *Work*, vol. 63, no. 3, pp. 469–477, 2019.
- [16] IGE, “Enquisa estrutural a fogares. coñecemento e uso do galego.” 2019, consultado o 22 de xuño de 2023. [En liña]. Disponible en: https://www.ige.gal/estatico/estatRM.jsp?c=0206004&ruta=html/gl/OperacionsEstruturais/Resumo_resultados_EEF_Galego.html
- [17] R. Alcalde-Azpiazu, “Acerca de - mastodon.gal,” 2023, consultado o 22 de xuño de 2023. [En liña]. Disponible en: <https://mastodon.gal/about>
- [18] A. International, “Toxic twitter - a toxic place for women,” 2018, consultado o 22 de xuño de 2023. [En liña]. Disponible en: <https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1-1/>

- [19] J. Sutherland and K. Schwaber, “The scrum papers,” *Nuts, Bolts and Origins of an Agile Process*, 2007.
- [20] S. Tanvir, M. Safdar, H. Tufail, and U. Qamar, “Merging prototyping with agile software development methodology,” in *International Conference on Engineering, Computing & Information Technology (ICECIT 2017)*, 2017, pp. 50–54.
- [21] C. Hansen, R. Arlitt, T. Eifler, and M. Deininger, “Design by prototyping: Increasing agility in mechatronic product design through prototyping sprints,” *Proceedings of the Design Society*, vol. 2, pp. 201–210, may 2022.
- [22] K. H. Manguri, R. N. Ramadhan, and P. R. M. Amin, “Twitter sentiment analysis on worldwide covid-19 outbreaks,” *Kurdistan Journal of Applied Research*, pp. 54–65, 2020.
- [23] T. Fuchs and F. Schäfer, “Normalizing misogyny: hate speech and verbal abuse of female politicians on japanese twitter,” in *Japan forum*, vol. 33, no. 4. Taylor & Francis, 2021, pp. 553–579.
- [24] S. Hewitt, T. Tiropanis, and C. Bokhove, “The problem of identifying misogynist language on twitter (and other online social spaces),” in *Proceedings of the 8th ACM Conference on Web Science*, ser. WebSci ’16. New York, NY, USA: Association for Computing Machinery, 2016, p. 333–335. [En línea]. Disponible en: <https://doi.org/10.1145/2908131.2908183>
- [25] E. Fersini, D. Nozza, and P. Rosso, “Ami @ evalita2020: Automatic misogyny identification,” in *Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian*, 2020, pp. 21–28.
- [26] T. Lynn, P. T. Endo, P. Rosati, I. Silva, G. L. Santos, and D. Ging, “A comparison of machine learning approaches for detecting misogynistic speech in urban dictionary,” in *2019 International Conference on Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA)*, 2019, pp. 1–8.
- [27] C. Cerisara, S. Jafaritazehjani, A. Oluokun, and H. Le, “Multi-task dialog act and sentiment recognition on mastodon,” *arXiv preprint arXiv:1807.05013*, 2018.
- [28] P. Monachelis, P. Kasnesis, L. Toumanidis, C. Patrikakis, and P. Papadopoulos, “Evaluation and visualization of trustworthiness in social media – eunomia’s approach,” in *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2022, pp. 217–222.

- [29] J. A. García-Díaz, M. Cánovas-García, R. Colomo-Palacios, and R. Valencia-García, “Detecting misogyny in spanish tweets. an approach based on linguistics features and word embeddings,” *Future Generation Computer Systems*, vol. 114, pp. 506–518, 2021. [En liña]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0167739X20301928>
- [30] M. L. Loureiro, M. Alló, and P. Coello, “Hot in twitter: Assessing the emotional impacts of wildfires with sentiment analysis,” *Ecological Economics*, vol. 200, p. 107502, 2022. [En liña]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0921800922001641>
- [31] J. Ortega, I. De-Dios-Flores, P. Gamallo, and J. Campos, “A neural machine translation system for spanish to galician through portuguese transliteration,” in *Annual Conference of the Spanish Society for Natural Language Processing*, 2022.
- [32] P. M. Fernández and J. R. P. Campos, “Generación semiautomática de recursos de opinion mining para el gallego a partir del portugués y el español,” in *Workshop on Iberian Cross-Language Natural Language Processing Tasks*, 2011.
- [33] R. Agerri, X. G. Guinovart, G. Rigau, and M. A. S. Portela, “Developing new linguistic resources and tools for the galician language,” in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.
- [34] “Activitypub,” consultado o 22 de xuño de 2023. [En liña]. Disponible en: <https://www.w3.org/TR/activitypub/>
- [35] SEC, “United states. securities and exchange commission.” 2022, consultado o 22 de xuño de 2023. [En liña]. Disponible en: <https://www.sec.gov/ix?doc=/Archives/edgar/data/1418091/000119312522272772/d411753d8k.htm>
- [36] H. B. Zia, J. He, A. Raman, I. Castro, N. Sastry, and G. Tyson, “Flocking to mastodon: Tracking the great twitter migration,” *arXiv preprint arXiv:2302.14294*, 2023.
- [37] H. Gameiro, “Sobre - mastodon (pt),” 2023, consultado o 22 de xuño de 2023. [En liña]. Disponible en: <https://masto.pt/about>
- [38] B. Fung, “Academic researchers blast twitter’s data paywall as ‘outrageously expensive’,” 2023. [En liña]. Disponible en: <https://edition.cnn.com/2023/04/05/tech/academic-researchers-blast-twitter-paywall/index.html>
- [39] J. Barnes, “Twitter ends its free api: Here’s who will be affected,” consultado o 22 de xuño de 2023. [En liña]. Disponible en: <https://www.forbes.com/sites/jenaebarnes/2023/02/03/twitter-ends-its-free-api-heres-who-will-be-affected/>

- [40] M. Sharma, “Selenium tool: A web based automation testing framework,” 2014.
- [41] A. I. Vladu, I. de Dios-Flores, C. Magariños, J. E. Ortega, J. Ramom, G. González, S. Barro, and X. L. Regueira, “Proxecto nós: Artificial intelligence at the service of the galician,” in *Annual Conference of the Spanish Society for Natural Language Processing*, 2022.
- [42] P. Gamallo, D. Bardanca, J. R. Pichel, M. García, S. Rodríguez-Rey, and I. de Dios-Flores, “Nos-*mt-opennmt-es-gl*,” 2023, consultado o 22 de xuño de 2023. [En liña]. Dispoñible en: <https://huggingface.co/proxectonos/NOS-MT-OpenNMT-es-gl>
- [43] G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. Rush, “OpenNMT: Open-source toolkit for neural machine translation,” in *Proceedings of ACL 2017, System Demonstrations*. Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 67–72. [En liña]. Dispoñible en: <https://www.aclweb.org/anthology/P17-4012>
- [44] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, “Automatic differentiation in pytorch,” in *NIPS-W*, 2017.
- [45] “Cixug - tradutor,” consultado o 22 de xuño de 2023. [En liña]. Dispoñible en: <https://tradutor.cixug.gal/index.spa.html#docTranslation?dir=spa-glg>
- [46] S. D. Collazo, “O estándar galego: reintegracionismo vs. autonomismo,” *Romanica Olo-mucensia*, no. 1, pp. 1–13, 2014.
- [47] “Mozilla public license,” consultado o 22 de xuño de 2023. [En liña]. Dispoñible en: <https://www.mozilla.org/en-US/MPL/2.0/>
- [48] “Convocatoria de axudas para a realización de traballos de fin de grao,” consultado o 22 de xuño de 2023. [En liña]. Dispoñible en: <https://www.fundacion.udc.es/curso.asp?id=1584&idioma=GA>
- [49] S. Arora, Y. Liang, and T. Ma, “A simple but tough-to-beat baseline for sentence embeddings,” in *International conference on learning representations*, 2017.
- [50] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of tricks for efficient text classification,” *arXiv preprint arXiv:1607.01759*, 2016.
- [51] J. A. García-Díaz, P. J. Vivancos-Vicente, A. Almela, and R. Valencia-García, “Umutextstats: A linguistic feature extraction tool for spanish,” in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 2022, pp. 6035–6044.
- [52] L. Breiman, “Random forests,” *Machine learning*, vol. 45, pp. 5–32, 2001.

- [53] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 1999.
- [54] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.
- [55] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [56] B. Vrigazova, "The proportion for splitting data into training and test set for the bootstrap in classification problems," *Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy*, vol. 12, no. 1, pp. 228–242, 2021.
- [57] E. M. Cámara, M. T. M. Valdivia, J. M. P. Ortega, and L. A. U. López, "Técnicas de clasificación de opiniones aplicadas a un corpus en español," *Procesamiento del Lenguaje Natural*, vol. 47, pp. 163–170, 2011.
- [58] A. Moh'd A Mesleh, "Chi square feature extraction based svms arabic language text categorization system," *Journal of Computer Science*, vol. 3, no. 6, pp. 430–435, 2007.
- [59] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intelligent data analysis*, vol. 6, no. 5, pp. 429–449, 2002.
- [60] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, no. 17, pp. 1–5, 2017. [En línea]. Disponible en: <http://jmlr.org/papers/v18/16-365.html>
- [61] "Resolución de 22 de febrero de 2018, de la dirección general de empleo, por la que se registra y publica el xvii convenio colectivo estatal de empresas de consultoría y estudios de mercado y de la opinión pública." 2018, consultado o 22 de xuño de 2023. [En línea]. Disponible en: [https://www.boe.es/eli/es/res/2018/02/22/\(3\)](https://www.boe.es/eli/es/res/2018/02/22/(3))
- [62] "Resolución de 28 de febrero de 2019, de la secretaria de estado de función pública, por la que se dictan instrucciones sobre jornada y horarios de trabajo del personal al servicio de la administración general del estado y sus organismos públicos." 2019, consultado o 22 de xuño de 2023. [En línea]. Disponible en: [https://www.boe.es/eli/es/res/2019/02/28/\(1\)](https://www.boe.es/eli/es/res/2019/02/28/(1))
- [63] "Táboas retributivas." 2023, consultado o 22 de xuño de 2023. [En línea]. Disponible en: https://www.udc.es/es/gobierno/equipo_reitoral/xerencia/servizos/retribucions_seguridade_social_e_accion_social/taboa_retributivas/

- [64] L. Cuthbertson, A. Kearney, R. Dawson, A. Zawaduk, E. Cuthbertson, A. Gordon-Tighe, and K. Mathewson, “Women, politics and twitter: Using machine learning to change the discourse,” 11 2019.