

# DSVS at MiSonGyny 2025: Multiple Instance Learning for Misogyny Speech Detection in Song Lyrics

Sergio Damián-Sandoval<sup>1,\*</sup>, David Vázquez-Santana<sup>1</sup>

<sup>1</sup>Centro de Investigación en Computación (CIC), Instituto Politécnico Nacional (IPN), México

## Abstract

Detecting misogynistic content in song lyrics is a challenging task due to the informal, metaphorical, and often ambiguous nature of language used in music. In this work, we explore the use of Multiple Instance Learning (MIL) for the automatic classification of Spanish song lyrics as either misogynistic or non-misogynistic. Each song is represented as a bag of sentences to mitigate the limitations of sequence length in transformer-based models and to enable the model to focus on the most informative parts of the lyrics. We employ an attention-based pooling mechanism that learns to weight sentences according to their relevance to the task. Additionally, we address class imbalance by applying class weighting for binary classification and focal loss for the multiclass setting. We experiment with three different dataset versions, varying the granularity and overlap of instances within each bag. Our results show that the MIL framework, combined with a transformer model pre-trained on social media data, achieves robust performance on both binary and multi-class misogyny detection tasks, demonstrating its effectiveness in handling weak supervision scenarios and noisy textual data.

## Keywords

Multiple Instance Learning, Misogyny Detection, Attention Pooling, Spanish Lyrics, Transformer Models, Focal Loss

## 1. Introduction

Misogyny is a term used to describe a group of beliefs, actions, or expressions that show disrespect, animosity, or prejudice against women [1]. It can manifest explicitly, such as through verbal or physical abuse, or covertly, such as through stereotypes or other content that perpetuates women's objectification or subordination. This is a systemic problem with roots in several institutional and cultural frameworks that continue to have some influence on women's lives today.

Misogyny has a prominent and alarming social impact. Women, who are the most impacted group, may experience low self-esteem and the normalization of violence and discrimination as a result of being exposed to misogynistic attitudes and discourse through interpersonal relationships or cultural products like fashion, music, or movies [2]. Women have historically faced many obstacles to their professional and personal growth [3, 4], including disproportionately bearing the burden of unpaid and frequently unappreciated domestic work, limited access to leadership roles, and exclusion from educational institutions. Their mental and physical health have surely been directly impacted by these disparities, which still exist today in more subdued but no less detrimental forms, like the wage gap and workplace harassment.

Misogyny also has a negative impact on men and society as a whole by maintaining toxic forms of masculinity, limiting male emotional expression, and reinforcing strict gender roles. Some sectors have normalized the notion that men are innately oppressive [5], while others continue to hold the view that feminism hates men [6]. This polarization of views makes mutual understanding harder and creates mistrust, which gets in the way of building healthy and positive relationships between men and women.

This way, misogyny not only keeps gender inequality going, but also twists human relationships, stopping us from building connections based on respect, fairness, and empathy.

In this context, music holds a central role as one of the most widely used forms of cultural expression. Song lyrics not only convey emotional states or personal narratives, but also serve as mirrors of

*IberLEF 2025, September 2025, Zaragoza, Spain*

✉ [sdamians2019@cic.ipn.mx](mailto:sdamians2019@cic.ipn.mx) (S. Damián-Sandoval); [dvazquez2019@cic.ipn.mx](mailto:dvazquez2019@cic.ipn.mx) (D. Vázquez-Santana)

🌐 <https://github.com/sdamians> (S. Damián-Sandoval); <https://github.com/Hiram02> (D. Vázquez-Santana)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

broader social conditions and, in many cases, actively contribute to shaping them. Various studies have documented how certain music genres function as vehicles for expressing structural issues. For example, *corridos* and *narcocorridos* have been extensively analyzed for their role in normalizing narco culture [7, 8], reflecting environments marked by violence, organized crime, and the tensions of illicit economic power. Rap and hip-hop, which emerged from marginalized urban contexts, have served as powerful tools of protest against poverty, racism, police brutality, and drug addictions [9], although they have also been criticized for reproducing misogynistic patterns [10]. In the case of reggaetón—a genre with global reach—numerous analyses [11, 12] have highlighted its frequent use of sexualized language, stereotypical representations of women, and the glorification of male dominance. Even genres such as rock or Latin American regional music are not exempt, often portraying female figures in subordinate or idealized roles rooted in patriarchal perspectives.

## 1.1. Motivation

Analyzing song lyrics to detect misogynistic content is an important task for several reasons. Music reaches a wide audience and can strongly influence how people think and feel. When listeners are repeatedly exposed to sexist messages or lyrics that promote the idea of female inferiority, it can make these ideas seem normal, reinforce harmful beliefs, and contribute to gender inequality. This is especially worrying in the case of young people, whose ideas about gender roles and relationships are still being shaped.

In addition, previous studies have shown that sexist content in media, including music [13, 14, 15], can increase tolerance for violence against women and reduce empathy for victims. For this reason, identifying and helping to reduce these messages is an important social responsibility.

The MiSonGyny shared task [16], organized as part of IberLEF 2025 [17], directly addresses this challenge by promoting the development of automatic systems for detecting misogynistic speech in song lyrics. This initiative provides a valuable framework for advancing research on harmful discourse in culturally influential domains, such as music.

## 1.2. Relevance and challenges of the tasks

This study addresses the need for tools capable of automatically detecting and characterizing misogynistic discourse in songs through the use of natural language processing (NLP) techniques. The proposed tasks allow for evaluating the effectiveness of models not only in identifying the presence of misogynistic content, but also in assessing their ability to distinguish between different types of misogyny.

These tasks present several significant challenges. One of the most prominent is the ambiguity inherent in artistic language. Song lyrics often rely on figurative expressions such as metaphors, symbolism, and poetic structures, which complicate the direct interpretation of the message. The frequent use of words or phrases deeply rooted in specific cultural contexts further increases the complexity of automatic detection and analysis.

Another major challenge is the subjectivity involved in interpreting misogynistic content. The same phrase may or may not be considered misogynistic by different groups, depending on cultural context or the listener’s prior knowledge of the artist or song. This subjectivity poses a challenge both for the annotations of the corpus and for the evaluation of detection models.

Despite these difficulties, this work represents an initial step toward the development of automated tools capable of accurately identifying and finely classifying various types of misogynistic discourse in music, which could be highly valuable for monitoring new content. The source code for this work can be found at [https://github.com/sdamians/DSVS\\_MiSonGyny2025](https://github.com/sdamians/DSVS_MiSonGyny2025).

## 2. Related Work

Research on misogyny detection has gained momentum in recent years, particularly in the context of social media platforms, where users often express gender-based hostility either explicitly or implicitly. Many studies have focused on developing computational approaches to detect sexist and misogynistic content on X, Reddit, or YouTube comments [18, 19, 20]. These approaches typically leverage supervised learning models trained on annotated datasets, and often utilize linguistic features, syntactic patterns, and more recently, contextual embeddings from pretrained language models such as BERT or RoBERTa [21, 22].

A notable portion of this research has centered on the automatic classification of hate speech and offensive language [23]. While misogyny is often a subset of these broader categories, dedicated misogyny detection remains a relatively underexplored area, particularly when considering fine-grained categorization of types of misogynistic discourse, such as sexualization, verbal aggression, or dehumanization.

In contrast to the significant attention given to misogyny in social media, relatively little research has addressed the presence of gendered hate speech in musical lyrics. Some exceptions include qualitative analyses in the humanities and cultural studies, which have examined the prevalence of sexist or violent messages in genres such as rap and reggaetón [10, 11, 12]. These works provide critical insight into how misogyny is perpetuated through artistic expression.

More recently, there has been a growing interest in using NLP techniques to study song lyrics. Efforts include sentiment analysis of lyrics across genres [24], as well as the use of machine learning for genre classification [25]. However, the application of NLP for misogyny detection in music remains a nascent field. Existing datasets for misogyny detection rarely include song lyrics.

To address the challenges of our corpus, mainly long lyrics where misogynistic content appears in only a few lines, we explore the use of Multiple Instance Learning (MIL). This approach allows the model to detect harmful expressions within a larger text, even when only song-level labels are available. MIL has been successfully used in similar weakly supervised tasks like sentiment analysis in long documents [26], making it suitable for identifying localized misogynistic content without needing detailed annotations.

Furthermore, due to the imbalanced and multi-class nature of the corpus in the second task, we use Focal Loss [27] to improve learning. Originally developed for object detection, Focal Loss helps the model focus on underrepresented classes and rare instances, such as the sparse presence of misogynistic expressions in lyrics, by reducing the influence of well-classified examples.

## 3. Dataset and Preprocessing

We used the datasets provided by the shared task composed of Spanish song lyrics (with some segments occasionally written in English), curated for the task of misogyny detection. Due to the length of most songs exceeding the input limits of standard Transformer-based models, each song was segmented into smaller textual units and treated as a *bag of sentences*, aligning with the Multiple Instance Learning (MIL) framework.

For the first subtask, which involves binary classification (Misogynistic – M / Non-Misogynistic – NM), the training dataset comprises a total of 2,104 songs, of which 1,462 are labeled as NM and 642 as M. For the second subtask, which addresses multi-class classification, the training set consists of 1,168 entries distributed across four categories: NR (Not Related) with 526 samples, S (Sexualization) with 435, V (Violent) with 129, and H (Hate) with 78. The corresponding test sets contain 527 and 293 samples respectively. Due to the class distribution in both tasks, specific techniques were applied to address the class imbalance.

### 3.1. Preprocessing

Several preprocessing steps were applied to ensure high-quality input data for MIL-based learning:

- **Character encoding correction:** Accent marks and special characters were misread due to encoding inconsistencies; these were systematically corrected.
- **Structural tag removal:** Sentences containing metatextual cues used to structure lyrics, such as *Coro*, *Verso*, *Letra de*, and copyright notices (e.g., *todos los derechos reservados*) were eliminated.
- **Normalization of contractions:** A custom dictionary was built to standardize common informal contractions in lyrics, such as *pa'*  $\rightarrow$  *para* or *to's*  $\rightarrow$  *todos*.
- **Handling bilingual content:** Some lyrics contain English sentences. These were retained and normalized using the same contraction rules when applicable.
- **Sentence capitalization:** All sentences were capitalized to maintain consistency across instances.
- **Removal of low-information tokens:** Interjections and non-contextual elements such as *Ey*, *Yeah*, *Ah*, *Oh*, and *jaja* were removed, as they do not contribute meaningful content to the task.

To explore different levels of granularity in the MIL setup, we created three dataset versions for each subtask:

- **v1:** Each instance corresponds to a single sentence from the song.
- **v2:** Each instance is a fixed group of four consecutive sentences without overlap.
- **v3:** Each instance is a group of four sentences with overlap, meaning that consecutive instances share some sentences.

These versions allow us to investigate the impact of instance length and contextual overlap on the performance of MIL-based misogyny detection.

## 4. Methodology

In this work, we address misogyny detection in Spanish song lyrics using a Multiple Instance Learning (MIL) framework. Each song is represented as a *bag* of textual instances (sentences or grouped sentences), allowing the model to make predictions at the bag level while learning from instance-level evidence.

### 4.1. Model Architecture

We adopted a small Transformer-based model (SLM) as the sentence encoder, specifically the `pysentimiento/robertuito-base-cased` pretrained model [28]. This language model is tailored for informal Spanish text, trained on corpora from social media platforms such as X, making it particularly well-suited to the slang, profanity, and informal expressions found in song lyrics. Due to the model's token limit of 128, we verified that none of the sentence instances exceeded 64 tokens, which simplified the training process and avoided the need for aggressive truncation.

### 4.2. MIL and Attention Pooling

To aggregate information from the bag of instances, we implemented an attention-based pooling mechanism [29]. The model learns attention weights over the instances within each bag, emphasizing those that contribute more significantly to the presence of misogynistic content. These weights are learned during training and allow the model to focus on contextually relevant sentences.

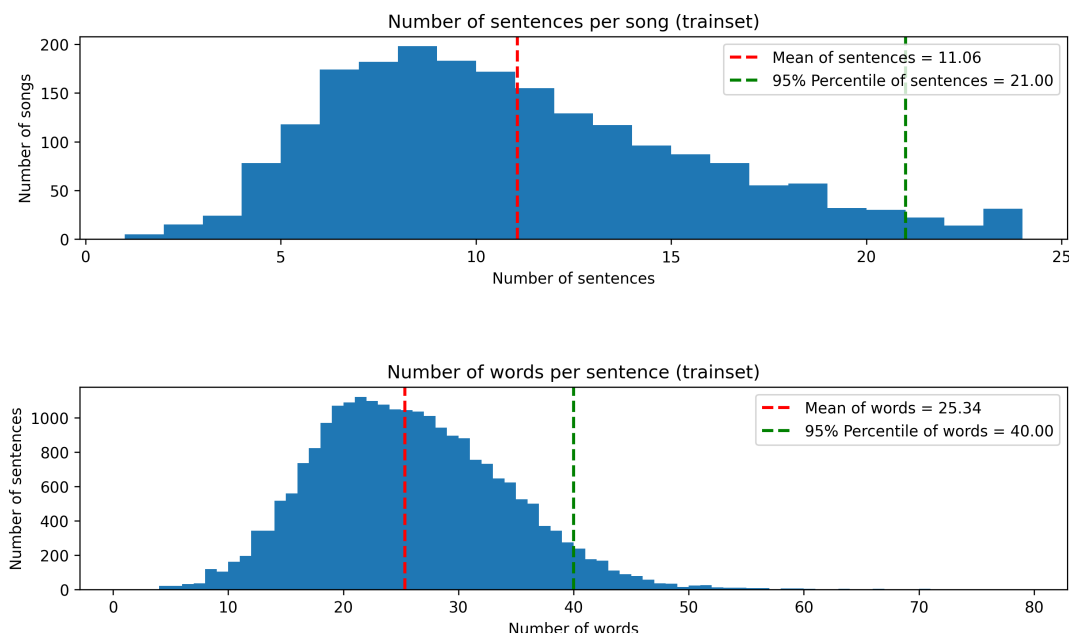
### 4.3. Loss Functions and Class Imbalance Handling

Different strategies were used to address class imbalance in the two subtasks:

- For the **binary classification** task (Subtask 1), we computed *class weights* based on the class distribution in the training set and incorporated them into the loss function to reduce bias toward the majority class.
- For the **multi-class classification** task (Subtask 2), which presents a more severe class imbalance, we employed *Focal Loss* to dynamically scale the loss, focusing learning on harder-to-classify minority classes.

### 4.4. Dataset Variants and Instance Configuration

We evaluated three versions of the dataset: Version 1 serves as the baseline; Version 2 applies Multiple Instance Learning (MIL) with one bag of instances per song lyric; and Version 3 introduces overlapping instances to create some data redundancy, aiming to preserve the sequential order of the song. **Figure 1** and **Figure 2** illustrate this configuration.

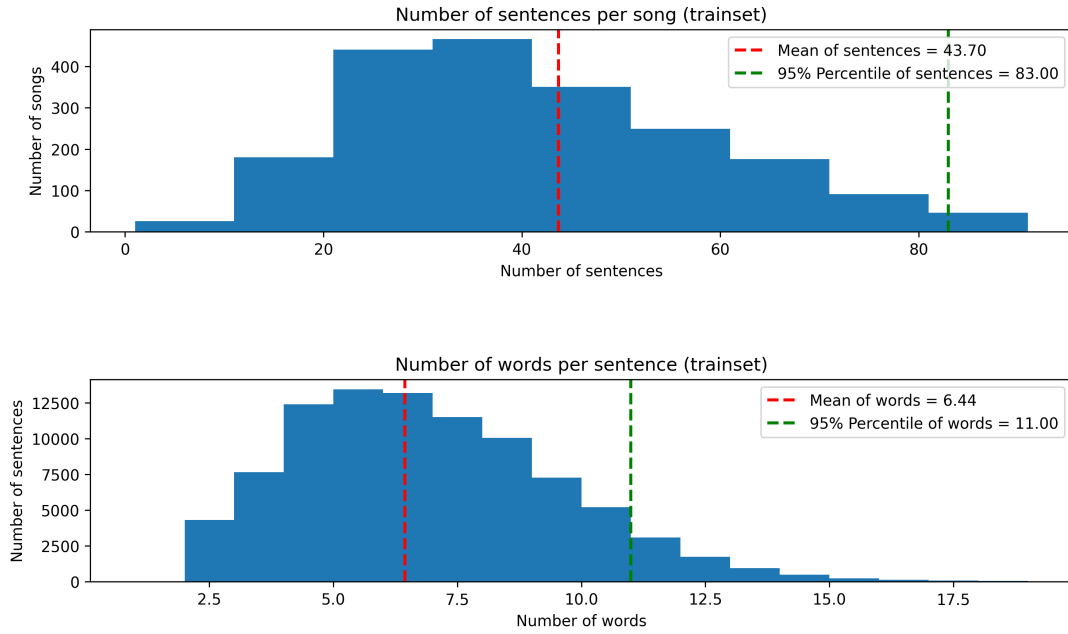


**Figure 1:** Analysis of number of sentences per song and number of words per sentences for subtask 1 (Useful to determine the `batch_size` and the `max_length` hyperparameters).

### 4.5. Training and Evaluation

We employed cross-validation to ensure robustness in model performance. Experiments were conducted using 3-fold, 5-fold, and 10-fold cross-validation schemes. Given the presence of class imbalance in both subtasks, we focused our final evaluation on the F1-score, which provides a balanced measure of precision and recall. To better understand the variability and robustness of the models across different folds, we report the following:

- **F1-score lower bound:** the minimum F1-score observed across the  $k$  folds, indicating the worst-case performance.
- **F1-score upper bound:** the maximum F1-score observed, showing the best-case scenario.
- **F1-score std dev:** the standard deviation of the F1-scores, reflecting the model’s consistency across the folds.



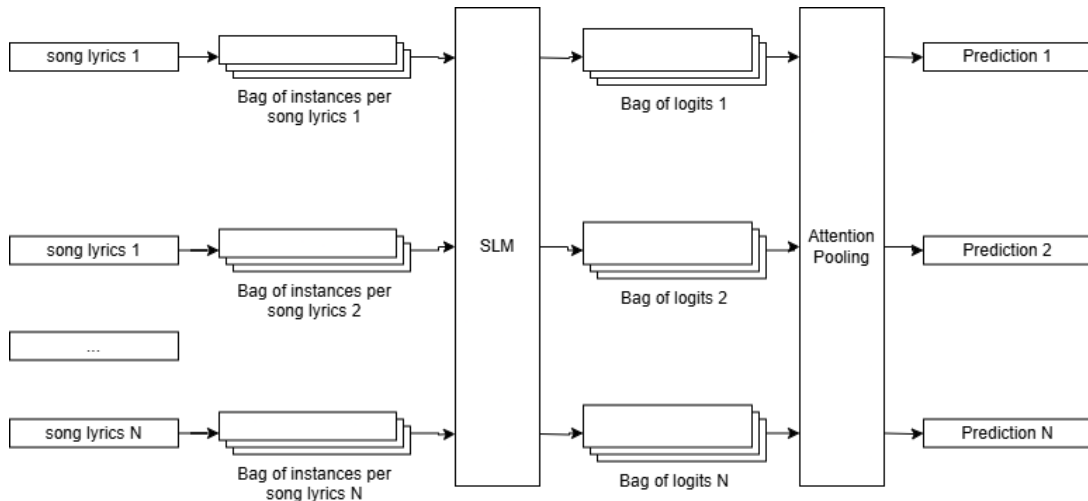
**Figure 2:** Analysis of number of sentences per song and number of words per sentences for subtask 2 (Useful to determine the batch\_size and the max\_length hyperparameters).

- **F1-score mean:** the average F1-score across all folds, providing an overall estimate of model performance.

Model performance was compared across the three dataset versions (v1, v2, v3), for both subtasks.

#### 4.6. System Overview

**Figure 3** presents a comprehensive overview of the entire methodology, illustrating how the raw song lyrics are first segmented and transformed into instances, how these instances are organized into bags, and finally how the MIL model processes and aggregates these bags to perform the classification task.



**Figure 3:** Overall pipeline for misogyny detection using Multiple Instance Learning.

## 5. Results

This section presents the results obtained during the model validation phase and describes the final submissions made to the test set. The performance of each dataset version (v1, v2, v3) was evaluated using 5-fold cross-validation and several classification metrics.

### 5.1. Cross-Validation Performance

Table 1 and Table 2 present the results of the experiment conducted using 5-fold cross-validation. Similar trends were observed with 3-fold and 10-fold configurations. For each dataset version, the reported metrics correspond to the observed F1-score, including the lower bound, upper bound, mean, and standard deviation across the folds.

Dataset Version	lower bound	upper bound	std dev	mean
v1 (single sentence - baseline)	0.6321	0.6417	0.0033	0.6369
v2 (4-sentence blocks)	0.7475	0.7531	0.0022	0.7503
<b>v3 (overlapping blocks)</b>	<b>0.7731</b>	<b>0.7878</b>	<b>0.0058</b>	<b>0.7805</b>

**Table 1**

Average performance across 5-fold cross-validation for each dataset version for subtask 1.

Dataset Version	lower bound	upper bound	std dev	mean
v1 (single sentence - baseline)	0.4256	0.4326	0.0022	0.4291
<b>v2 (4-sentence blocks)</b>	<b>0.4150</b>	<b>0.5645</b>	<b>0.0553</b>	<b>0.4768</b>
v3 (overlapping blocks)	0.4476	0.4584	0.0034	0.4530

**Table 2**

Average performance across 5-fold cross-validation for each dataset version for subtask 2.

### 5.2. Observations

The results indicate that the use of overlapping windows in grouped sentences (v3) yielded the best performance for Subtask 1, suggesting that overlapping context can enhance the model’s ability to detect subtle misogynistic patterns distributed across song lyrics. In contrast, for Subtask 2, the model performed best with version 2, which uses a single bag of non-overlapping instances per song, indicating that overlapping introduced noise in certain classification tasks.

The attention-based pooling mechanism enabled the model to focus on semantically rich instances, which proved especially effective when contextual redundancy was beneficial, as in version v3. However, these findings also suggest that overlapping context is not universally advantageous and its impact may vary depending on the specific task.

Moreover, the use of class weighting in Subtask 1 and Focal Loss in Subtask 2 helped mitigate class imbalance, particularly in the multi-class setting where certain classes had significantly fewer training examples.

### 5.3. Test Set Submissions

Three separate submissions were made to the test set, one for each dataset version. Table 3 and Table 4 summarize the results obtained.

Version v3 achieved the highest F1-score in subtask 1, while version v2 achieved the highest in subtask 2, confirming the findings from the validation phase. These results demonstrate the effectiveness of incorporating instance overlap and leveraging sentence-level attention in a Multiple Instance Learning setup.



Submit	Dataset Version	Precision	Recall	F1-score
1	v2	0.7299	0.7708	0.7211
2	v3	0.7599	0.7849	0.7685
3	v3	<b>0.8036</b>	<b>0.8015</b>	<b>0.8025</b>

**Table 3**

Official results for the three test set submissions for subtask 1.

Submit	Dataset Version	Precision	Recall	F1-score
1	v1	0.4467	0.4205	0.4007
2	v3	0.4757	0.4767	0.4667
3	v2	<b>0.4861</b>	<b>0.4835</b>	<b>0.4840</b>
4	v3	0.4875	0.4766	0.4784

**Table 4**

Official results for the three test set submissions for subtask 2.

#### 5.4. Preliminary Error Analysis (optional)

Although detailed error analysis will be presented in future work, initial observations show that:

- Songs mixing English and Spanish can confuse the model, especially when misogynistic expressions appear in English.
- Informal contractions and slang may still pose challenges despite the use of a domain-specific language model.
- Some false positives involve songs that use misogynistic language ironically or metaphorically, or the intention is to generate general hate speech which the model fails to disambiguate.

## 6. Conclusions

In this work, we explored the use of Multiple Instance Learning (MIL) for the detection of misogynistic content in song lyrics written in Spanish. Due to the length and informal nature of the texts, traditional document-level classification approaches are not well-suited. Instead, we modeled each song as a bag of sentences, allowing the model to focus on the most relevant parts of the text through an attention-based pooling mechanism.

Our experiments with three different dataset versions demonstrated that using grouped and overlapping sentence blocks (v3) provided the best performance, suggesting that additional context and redundancy are beneficial for detecting subtle or context-dependent expressions of misogyny.

The incorporation of techniques to address class imbalance—such as class weighting and Focal Loss—contributed to improving the learning process, particularly in the more challenging multi-class setting. Furthermore, the use of a domain-specific transformer model pre-trained on social media texts (robertuito-base-cased) allowed for better handling of informal language, slang, and contractions frequently present in lyrics.

Our findings show that MIL is a promising framework for tasks involving weak supervision and noisy data, where the label of a full document does not necessarily apply uniformly to all its parts.

### Future Work

Future directions for this work include:

- A deeper error analysis to better understand model behavior in edge cases, such as irony and code-switching.
- Exploring hierarchical MIL models or hybrid architectures with token- and sentence-level supervision.



- Expanding the dataset with more annotated samples to improve generalization, particularly for the underrepresented classes in Subtask 2.
- Incorporating linguistic features or external knowledge bases to support better interpretability of the model decisions.

## Acknowledgments

This work was done with partial support from the Mexican Government through Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) and Instituto Politécnico Nacional (IPN).

## Declaration on Generative AI

During the preparation of this work, ChatGPT-4.0, ChatGPT-3.5 and Writefull for Overleaf were utilized as assistants for the following tasks: Grammar and spelling check, Paraphrase and reword and Text Translation (Spanish to English). After using these tools, the authors reviewed, verified and edited the content as required. The authors take full responsibility for the content of this work.

## References

- [1] M. P. Johnson, K. J. Ferraro, Research on domestic violence in the 1990s: Making distinctions, *Journal of Marriage and Family* 62 (2000) 948–963. doi:<https://doi.org/10.1111/j.1741-3737.2000.00948.x>.
- [2] P. Homan, Health consequences of structural sexism: Conceptual foundations, empirical evidence and priorities for future research, *Social Science & Medicine* 351 (2024) 116379. doi:<https://doi.org/10.1016/j.socscimed.2023.116379>, gender, power, and health: Modifiable factors and opportunities for intervention.
- [3] K. Miner-Rubino, L. M. Cortina, Beyond targets: Consequences of vicarious exposure to misogyny at work., *Journal of Applied psychology* 92 (2007) 1254.
- [4] E. M. Archer, Gender-based inequality in the modern american society, *Exploring Gender at Work: Multiple Perspectives* (2021) 45–63.
- [5] I. Jablonka, *A history of masculinity: From patriarchy to gender justice*, Penguin UK, 2022.
- [6] A. Hopkins-Doyle, A. L. Petterson, S. Leach, H. Zibell, P. Chobthamkit, S. Binti Abdul Rahim, J. Blake, C. Bosco, K. Cherrie-Rees, A. Beadle, et al., The misandry myth: An inaccurate stereotype about feminists’ attitudes toward men, *Psychology of women quarterly* 48 (2024) 8–37.
- [7] M. Mulligan, *Corridos en La guerra contra el narco: Estéticas necropolíticas en México*, Ph.D. thesis, University of Missouri–Columbia, 2021.
- [8] S. Sorcia Reyes, *The music ecology of narcocorridos in mexico’s everyday life* (2024).
- [9] H. Malcomson, Making subjects grievable: narco rap, moral ambivalence and ethical sense making, in: *Ethnomusicology Forum*, volume 30, Taylor & Francis, 2021, pp. 205–225.
- [10] R. Weitzer, C. E. Kubrin, Misogyny in rap music: A content analysis of prevalence and meanings, *Men and masculinities* 12 (2009) 3–29.
- [11] S. S. Campo, A. Faure-Carvallo, A. M. V. Carrasco, Reguetón y representaciones de la mujer: un estudio en educación secundaria, *Revista Internacional de Educación Musical* 10 (2022) 25–32.
- [12] E.-J. Díez-Gutiérrez, E. Palomo-Cermeño, B. Mallo-Rodríguez, (in)equality and the influence of reggaeton music as a socialisation factor: A critical analysis, *Gender Studies* 21 (2022) 66–85. URL: <https://doi.org/10.2478/genst-2023-0005>. doi:10.2478/genst-2023-0005.
- [13] S. Galdi, F. Guizzo, Media-induced sexual harassment: The routes from sexually objectifying media to sexual harassment, *Sex Roles* 84 (2021) 645–669.
- [14] J. Burnay, S. Kepes, B. J. Bushman, Effects of violent and nonviolent sexualized media on aggression-related thoughts, feelings, attitudes, and behaviors: A meta-analytic review, *Aggressive behavior* 48 (2022) 111–136.

- [15] D. Yadav, D. Kalyani, Role of media in increasing violence against women and girls in the society, *International Journal of Scientific Research in Modern Science and Technology* 3 (2024) 21–26. doi:10.59828/ijrmst.v3i2.183.
- [16] T. Alcántara, M. Soto, C. Macias, O. Garcia-Vazquez, A. Espinosa-Juarez, H. Calvo, J. E. Valdez-Rodríguez, E. Felipe-Riveron, Overview of MiSonGyny at IberLEF 2025: Misogyny Speech Detection in Spanish Language Song Lyrics, *Procesamiento del Lenguaje Natural* 75 (2025).
- [17] J. Á. González-Barba, L. Chiruzzo, S. M. Jiménez-Zafra, Overview of IberLEF 2025: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025)*, co-located with the 41st Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS. org, 2025.
- [18] Z. Talat, D. Hovy, Hateful symbols or hateful people? predictive features for hate speech detection on twitter, in: *Proceedings of the NAACL student research workshop*, 2016, pp. 88–93.
- [19] A. Founta, C. Djouvas, D. Chatzakou, I. Leontiadis, J. Blackburn, G. Stringhini, A. Vakali, M. Sirivianos, N. Kourtellis, Large scale crowdsourcing and characterization of twitter abusive behavior, in: *Proceedings of the international AAAI conference on web and social media*, volume 12, 2018.
- [20] G. K. Shahi, A. K. Jaiswal, D. Nandini, L.-D. Ibáñez, T. Mandl, H. Liu, Report on the 1st workshop on diffusion of harmful content on online web (dhow) at websci 2024, in: *Companion Publication of the 16th ACM Web Science Conference*, 2024, pp. 60–64.
- [21] T. Caselli, V. Basile, J. Mitrović, M. Granitzer, Hatebert: Retraining bert for abusive language detection in english, *arXiv preprint arXiv:2010.12472* (2020).
- [22] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, R. Kumar, Predicting the type and target of offensive posts in social media, *arXiv preprint arXiv:1902.09666* (2019).
- [23] M. Mozafari, R. Farahbakhsh, N. Crespi, Hate speech detection and racial bias mitigation in social media based on bert model, *PloS one* 15 (2020) e0237861.
- [24] D. Yang, W.-S. Lee, Music emotion identification from lyrics, in: *2009 11th IEEE International Symposium on Multimedia*, IEEE, 2009, pp. 624–629.
- [25] Q. H. Nguyen, T. T. Do, T. B. Chu, L. V. Trinh, D. H. Nguyen, C. V. Phan, T. A. Phan, D. V. Doan, H. N. Pham, B. P. Nguyen, et al., Music genre classification using residual attention network, in: *2019 International conference on system science and engineering (ICSSE)*, IEEE, 2019, pp. 115–119.
- [26] S. Angelidis, M. Lapata, Multiple instance learning networks for fine-grained sentiment analysis, *Transactions of the Association for Computational Linguistics* 6 (2018) 17–31.
- [27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [28] J. M. Pérez, D. A. Furman, L. Alonso Alemany, F. M. Luque, RoBERTuito: a pre-trained language model for social media text in Spanish, in: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, 2022, pp. 7235–7243. URL: <https://aclanthology.org/2022.lrec-1.785>.
- [29] M. Ilse, J. Tomczak, M. Welling, Attention-based deep multiple instance learning, in: *International conference on machine learning*, PMLR, 2018, pp. 2127–2136.