UC-UCO-CICESE_UT3-Plenitas Team at HOMO-LAT 2025: Exploring the Challenge of Determining the Opinion of Latin-American Speakers about LGBTQ+ Community

Yoan Martínez-López[1,2,*], Jose Carlos Hernández Nieto[3], Ansel Rodríguez-Gónzalez[4], Julio Madera[3], Ana Orquidea López Correoso[5], Carlos de Castro Lozano[1,2], Jose Miguel Ramírez Uceda[1,2] and José Carlos Arévalo Fernández[2]

[1]*Universidad de Córdoba, Córdoba, Spain*
[2]*Plénitas, C/ Le Corbusier s/n, 14005 Córdoba, Spain*
[3]*Universidad de Camaguey, Circunvalación Norte, Camino Viejo Km 5 y 1/2, Camaguey, Cuba*
[4]*CICESE-UT3, Nayarit, México*
[5]*IPVCE "Máximo Gómez Báez", Circunvalación Norte, Camaguey, Cuba*

Abstract

The HOMO-LAT25 shared task focuses on detecting sentiment polarity in online messages related to the LGBTQ+ community across Latin American Spanish dialects. The UC-UCO-CICESE_UT3-Plenitas team participated in two main subtasks: multi-dialect polarity detection—where the training and evaluation dialects match—and cross-dialect polarity detection—where they differ. These subtasks challenge systems to recognize positive, negative, or neutral sentiment toward specific keywords across varying sociolinguistic contexts. The team employed transformer-based models from the Hugging Face ecosystem, particularly leveraging the Robertuito model from the pysentimiento library, which is pre-trained for Spanish sentiment classification. The overall methodology relied on deep learning techniques to manage unstructured social media text, emphasizing the adaptability of the transformers for multilingual sentiment analysis. Out of 31 teams and 67 total submissions, only 8 teams progressed to the final evaluation phase. The UC-UCO-CICESE_UT3-Plenitas team reached the finals and placed 7th in both subtasks. Their F1-scores were 0.2592 and 0.2588, respectively—demonstrating baseline performance but highlighting a clear gap from top-performing teams. Despite the modest scores, their progression to the final stage reflects the model's initial competence and provides a solid foundation for future enhancements. Key areas for improvement include better handling of dialectal variation, improved feature engineering, and optimization of classification strategies. Overall, the HOMO-LAT25 task represents a critical benchmark for developing inclusive, culturally aware NLP systems, supporting applications such as content moderation, public sentiment tracking, and digital policy efforts aimed at protecting LGBTQ+ communities in Latin America.

Keywords

Sentiment Analysis, Polarity Detection, Latin American Spanish, LGBTQ+, Transformers

1. Introduction

The HOMO-LAT (Human-Centric Polarity Detection in Online Messages Oriented Toward the Latin American LGBTQ+ Population) initiative represents a collaborative effort among researchers across various Latin American countries [1, 2], aiming to investigate societal perceptions of the LGBTQ+ community through computational analysis. The project focuses on the linguistic and social dynamics within LGBTQ+-related discourse on social media platforms, particularly Reddit, analyzing subreddits representative of Spanish-speaking countries throughout Latin America. This approach enables a

IberLEF 2025, septiembre de 2025, Zaragoza, España
*Corresponding author.
yoan.martinez@plenitas.com (Y. Martínez-López); jose.hernandez@reduc.edu.cu (J. C. H. Nieto); ansel@cicese.edu.mx (A. Rodríguez-Gónzalez); julio.madera@reduc.edu.cu (J. Madera); analopezcorreoso@gmail.com (A. O. L. Correoso); carlosdecastrolozano@gmail.com (C. d. C. Lozano); p52raucj@uco.es (J. M. R. Uceda); josecarlos@plenitas.com (J. C. A. Fernández)
0000-0002-1950-567X (Y. Martínez-López); 0000-0002-6709-2443 (J. C. H. Nieto); 0000-0001-9971-0237 (A. Rodríguez-Gónzalez); 0000-0001-5551-690X (J. Madera); 0000-0001-6603-843X (C. d. C. Lozano); 0000-0002-5027-7521 (J. M. R. Uceda); 0009-0004-4155-957X (J. C. A. Fernández)

comprehensive examination of regional and cultural attitudes toward LGBTQ+ issues across diverse sociopolitical contexts.

The HomoLat 2025 shared task centers on polarity detection for a curated set of generic terms commonly used throughout Latin America . It is designed to engage a broad research community, including both academic and industrial stakeholders, by offering a challenging benchmark that combines sentiment analysis with the cultural and linguistic variability inherent in Latin American Spanish dialects [3, 4]. From an academic perspective, this task advances natural language processing (NLP) and computational linguistics by encouraging the development of models capable of nuanced sentiment interpretation across heterogeneous linguistic landscapes.

Beyond academia, the task holds significant practical relevance. LGBTQ+ advocacy organizations can employ polarity detection systems to identify and respond to harmful or exclusionary language in digital discourse [5, 6, 7, 8, 9, 10, 11, 12, 13], thereby fostering safer and more inclusive online environments. Policymakers and regulators may use such tools to monitor hate speech and inform interventions. Additionally, social media platforms can integrate these models into moderation pipelines to detect and mitigate LGBTQ+ phobic content. In the private sector, marketing and public relations professionals can leverage sentiment analysis to design culturally sensitive campaigns and track public opinion, enabling more responsive and ethical engagement with LGBTQ+ communities.

1.1. Related Work

Deep learning and transformer-based architectures have revolutionized natural language processing (NLP). Models such as BERT and its derivatives are capable of modeling linguistic nuance and long-distance dependencies, making them ideal for multilingual tasks. Hugging Face, as an open-source repository and community, facilitates access to these architectures via its Transformers and Datasets libraries. Often referred to as the ´GitHub of Machine Learning´, it offers an extensive repository of pre-trained models, datasets, and interactive applications via its Model Hub. Deep learning is a subfield of machine learning that employs artificial neural networks with multiple layers, hence the term ‘deep’, to automatically learn complex patterns in large datasets. These models excel at processing unstructured data such as text, images, and audio. Rather than relying on explicit rule-based programming, deep learning systems learn to map input to output through iterative training, mimicking some aspects of human cognitive learning [14, 15].Hugging Face is particularly notable for its Transformers library, which provides state-of-the-art implementations of transformer-based architectures for a wide array of tasks, including natural language processing (NLP), computer vision, and speech processing. This platform facilitates rapid experimentation and reproducibility, and most of the models used in this competition were sourced from the Hugging Face Hub [16, 17]. Transformers, introduced by [18] in the seminal paper "Attention is All You Need", represent a significant breakthrough in deep learning. The architecture is centered on a self-attention mechanism that allows models to dynamically weight the importance of each token in an input sequence relative to the others, enabling modeling of long-range dependencies. Unlike recurrent or convolutional architectures, transformers process input tokens in parallel, resulting in superior computational efficiency. Robertuito, developed as part of the pysentimiento library, is a transformer model pre-trained specifically for Spanish sentiment analysis. It has shown robust performance on social media data in Spanish-speaking contexts.

This task ultimately invites the Latin American NLP and AI community to develop and refine applications tailored to the sociolinguistic diversity of the region[19]. In this paper, we describe our participation in the HOMO-LAT25.

2. Methodology

2.1. Dataset

The official HOMO-LAT25 dataset was used. It included annotated social media messages with polarity labels (positive, neutral, negative), alongside dialectal information. The dataset was divided into two

subtasks:

- Subtask 1: Multi-dialect (same dialects in train and test)
- Subtask 2: Cross-dialect (dialects differ in train and test)

Subtask 1: Multi-Dialect Polarity Detection (Multi-Class Classification) The goal of this subtask is to determine the polarity—positive, negative, or neutral—of a social media post with respect to a given keyword. The distinguishing feature of this task is that the dialectal varieties of Spanish used in the training set are the same as those appearing in the evaluation set, allowing models to learn and be evaluated within consistent linguistic boundaries [1, 2].

Subtask 2: Cross-Dialect Polarity Detection (Multi-Label Classification) This subtask also aims to classify the polarity of posts towards a specific keyword but introduces a greater level of difficulty: the Spanish dialects present in the evaluation set differ from those in the training set. This setup tests the cross-linguistic generalizability and robustness of the models, pushing them to adapt to previously unseen dialectal variations in the context of sentiment classification.

Key components of transformers include multi-head self-attention, positional encoding to preserve word order, and feed-forward networks for feature transformation. Layer normalization and residual connections facilitate stable and efficient training. Owing to their scalability and performance, transformers form the foundation of many state-of-the-art models such as BERT, GPT, and T5, and have become the dominant architecture not only in NLP but increasingly in computer vision and speech recognition domains as well.

2.1.1. Robertuito: A Transformer-Based Model for Spanish Sentiment Analysis

Robertuito is a pre-trained natural language processing (NLP) model specifically fine-tuned for sentiment analysis in Spanish. Distributed as part of the pysentimiento library, Robertuito is designed to classify textual data into sentiment categories such as Positive, Negative, Neutral, and in some cases, Mixed or Other. The model is built upon transformer-based architectures, primarily variants of BERT (Bidirectional Encoder Representations from Transformers), and has been adapted to perform effectively on Spanish-language tasks [20]. Robertuito serves as a robust tool for developers and researchers aiming to analyze the emotional tone embedded in Spanish text [21, 22]. Its seamless integration into the pysentimiento framework allows for straightforward deployment in real-world applications, particularly those tailored to Spanish-speaking populations. We used the ‘finiteautomata/beto-sentiment-analysis‘ version of Robertuito from the pysentimiento library. The model was "fine-tuned" using the training split for each subtask. Training was done with the AdamW optimizer and early stopping on macro-F1.

2.2. Preprocessing

Messages were lowercased and cleaned of non-textual noise (e.g., emojis, links). No stemming or lemmatization was applied to preserve dialectal features.

2.3. Evaluation Metrics

To assess the performance of sentiment analysis models submitted by competition participants, the F1-score is employed as the primary evaluation metric [23, 24]. The F1-score captures a balance between precision and recall, providing a holistic measure of model effectiveness. The computation of the F1-score is adapted according to the task setting:

- For multi-class classification, the macro-averaged F1-score is used, ensuring equal weight is given to each class, regardless of its frequency.
- For multi-label classification, the sample-averaged F1-score is applied, evaluating the model’s performance across individual samples with multiple labels.
- For binary classification, the macro-averaged F1-score is again used to balance performance across both classes.

The evaluation used the macro-F1 score for Subtask 1 and the sample-based F1 for Subtask 2. This evaluation framework ensures a fair and comprehensive comparison of models, particularly in settings with class imbalance or multiple sentiment labels.

3. Results

The lowest F1-score was observed for the Positive class, particularly in the cross-dialect scenario. Manual inspection revealed false positives caused by irony or dialect-specific expressions, see table 1.

Subtask	Class	Precision	F1-score
Multi-dialect	Positive	0.28	0.24
	Neutral	0.27	0.25
	Negative	0.29	0.28
Cross-dialect	Positive	0.26	0.22
	Neutral	0.25	0.23
	Negative	0.27	0.28

Table 1
Class-wise performance for each subtask

Example:

> *"Sí, claro que me encanta que nos discriminen "* — incorrectly labeled as Positive.

Competition Results and Performance Analysis: The HOMO-LAT 2025 shared task attracted participation from 31 teams, generating a total of 67 submissions over the course of the competition. From these, only 8 teams advanced to the final evaluation stage, highlighting the selective and competitive nature of the task.

The UC-UCO-CICESE_UT3-Plenitas team (ymlopez) was among the finalists, ultimately ranking 7th out of the finalist teams. The team submitted entries for two subtasks (tracks), achieving the following results:

- Track 1 – Multi-Dialect Polarity Detection (Multi-class) F1 Score: 0.2592, Final Rank: 7th
- Track 2 – Cross-Dialect Polarity Detection (Multi-label) F1 Score: 0.2588, Final Rank: 7th

4. Conclusions

Although the UC-UCO-CICESE_UT3-Plenitas team successfully advanced to the final phase of a highly competitive evaluation involving 31 participating teams, its final placement—7th out of finalist teams—highlights the need for further methodological enhancements. Future research should prioritize the refinement of classification strategies, the development of more robust and representative data features, and the improvement of model generalization. These efforts aim to elevate performance metrics, particularly the F1 score, towards the levels achieved by mid-ranking teams, which consistently reached the 0.40–0.45 range. Such improvements are essential for achieving competitive parity in complex multilingual and sentiment analysis tasks. While the performance of our system ranked among the finalists, our results indicate clear areas for improvement. These include enhancing dialectal generalization, leveraging domain adaptation techniques, and incorporating class-weighted loss functions to handle imbalanced data. Future work will focus on integrating sociolinguistic features and leveraging multilingual embeddings to capture deeper cultural context in sentiment classification.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4 and Grammarly in order to: Grammar and spelling check. After using these services, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] G. Bel-Enguix, H. Gómez-Adorno, G. Sierra, J. Vásquez, S. T. Andersen, S. Ojeda-Trueba, Overview of HOMO-MEX at Iberlef 2023: Hate speech detection in Online Messages directed Towards the MEXican spanish speaking LGBTQ+ population, Procesamiento del lenguaje natural 71 (2023) 361–370.

[2] H. Gómez-Adorno, G. Bel-Enguix, H. Calvo, S. Ojeda-Trueba, J. Andersen, Scott Thomas Vásquez, , S. Ojeda-Trueba, T. Alcántara, M. Soto, M. Cesar, Overview of HOMO-MEX at IberLEF 2024: Hate Speech Detection Towards the Mexican Spanish speaking LGBT+ Population, Procesamiento del lenguaje natural 73 (2024) 393–405.

[3] 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.

[4] G. Bel-Enguix, H. Gómez-Adorno, S. Ojeda-Trueba, G. Sierra, J. Barco, E. Lee, J. Dunstan, R. Manrique, Overview of HOMO-LAT at IberLEF 2025: Human-centric polarity detection in Online Messages Oriented to the Latin American-speaking lgbtq+ populaTion, Procesamiento del lenguaje natural 75 (2025) –.

[5] D. D. Kayande, K. K. Ponnusamy, P. K. Kumaresan, P. Buitelaar, B. R. Chakravarthi, Vel at homo-mex 2024: Detecting lgbt+ phobia in mexican spanish social media (2024).

[6] D. Espinosa, G. Sidorov, E. R. Vázquez, The labocic at homomex 2024: Using bert to classify hatelgtb speech, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024), co-located with the 40th Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), CEURWS. org, 2024.

[7] O. G. Vazquez, M. Cardoso-Moreno, J. A. Torres-León, D. Jiménez, Homocic at homo-mex 2024: Deep learning approaches for classifying homophobic content in tweets and songs: Leveraging llm and nl, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024), co-located with the 40th Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), CEUR-WS. org, 2024.

[8] J. Román Pásaro, Á. Carrillo Casado, J. Mata Vázquez, V. Pachón Álvarez, et al., I2c-uhu at homomex2024: Leveraging large language models and ensembling transformers to identify and classify hate messages towards the lgbtq+ community (2024).

[9] R. Gonzalez-Henao, D. Marrugo-Tobon, J. Martinez-Santos, E. Puertas, Verbanexai lab at homomex 2024: Multiclass and multilabel detection of lgbtq+ phobic content using transformers, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024), co-located with the 40th Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), CEUR-WS. org, 2024.

[10] M. Ramírez-González, D. I. Hernández-Farías, M. Montes-y Gómez, Labtl-inaoe at homo-mex 2024: Distance-based representations for lgbt+ phobia detection (2024).

[11] S. Damián, D. Vázquez, E. Felipe-Riverón, C. Yáñez-Márquez, Dsvs at homo-mex24: Multi-class and multi-label hate speech detection using transformer-based models, children 2 (2024) 4.

[12] L. Ramos, C. Palma-Preciado, O. Kolesnikova, M. Saldana-Perez, G. Sidorov, M. Shahiki-Tash, Intellileksika at homo-mex 2024: Detection of homophobic content in spanish lyrics with machine learning, in: Journal Name or Conference Proceedings Here, 2024.

[13] D. A. Niño, M. M. y Gómez, C. V. Cruz, Colpos at homo-mex 2024: Weighted naive bayes for lgbtq+ phobia detection in spanish text (2024).
[14] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (2015) 436–444.
[15] N. Rusk, Deep learning, Nature Methods 13 (2016) 35–35.
[16] S. M. Jain, Hugging face, in: Introduction to transformers for NLP: With the hugging face library and models to solve problems, Springer, 2022, pp. 51–67.
[17] J. Jones, W. Jiang, N. Synovic, G. Thiruvathukal, J. Davis, What do we know about hugging face? a systematic literature review and quantitative validation of qualitative claims, in: Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, 2024, pp. 13–24.
[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 30, Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
[19] M. Shahiki-Tash, J. Armenta-Segura, Z. Ahani, O. Kolesnikova, G. Sidorov, A. Gelbukh, Lidoma at homomex2023@ iberlef: Hate speech detection towards the mexican spanish-speaking lgbt+ population. the importance of preprocessing before using bert-based models, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), 2023.
[20] J. M. Pérez, D. A. Furman, L. A. Alemany, F. Luque, Robertuito: a pre-trained language model for social media text in spanish, 2022. URL: https://arxiv.org/abs/2111.09453. arXiv:2111.09453.
[21] S. Gonzalez-Lopez, S. Bethard, Transformer-based cynical expression detection in a corpus of Spanish YouTube reviews, in: J. Barnes, O. De Clercq, R. Klinger (Eds.), Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, Association for Computational Linguistics, Toronto, Canada, 2023, pp. 194–201. URL: https://aclanthology.org/2023.wassa-1.18/. doi:10.18653/v1/2023.wassa-1.18.
[22] J. M. Molero, J. Pérez-Martín, A. Rodrigo, A. Peñas, Offensive language detection in spanish social media: Testing from bag-of-words to transformers models, IEEE Access 11 (2023) 95639–95652. doi:10.1109/ACCESS.2023.3310244.
[23] D. Zhang, J. Wang, X. Zhao, Estimating the uncertainty of average f1 scores, in: Proceedings of the 2015 International Conference on The Theory of Information Retrieval, ICTIR '15, Association for Computing Machinery, New York, NY, USA, 2015, p. 317–320. URL: https://doi.org/10.1145/2808194.2809488. doi:10.1145/2808194.2809488.
[24] Z. C. Lipton, C. Elkan, B. Narayanaswamy, Thresholding classifiers to maximize f1 score, 2014. URL: https://arxiv.org/abs/1402.1892. arXiv:1402.1892.

A. Online Resources

The results are available via

- HOMOLAT25 results,
- HOMOLAT Codabench.