Dataset for Identification of Homophobia and Transphobia for Telugu, Kannada, and Gujarati

Prasanna Kumar Kumaresan[1], **Rahul Ponnusamy**[1], **Dhruv Sharma**[2], **Paul Buitelaar**[1], **Bharathi Raja Chakravarthi**[1]
[1]Data Science Institute, University of Galway, Ireland,
[2]Indian Institute of Technology (BHU), Varanasi
{P.Kumaresan1, R.Ponnusamy1, paul.buitelaar, bharathi.raja}@universityofgalway.ie
dhruv.sharma.ece21@iitbhu.ac.in

Abstract

Users of social media platforms are negatively affected by the proliferation of hate or abusive content. There has been a rise in homophobic and transphobic content in recent years targeting LGBT+ individuals. The increasing levels of homophobia and transphobia online can make online platforms harmful and threatening for LGBT+ persons, potentially inhibiting equality, diversity, and inclusion. We are introducing a new dataset for three languages, namely Telugu, Kannada, and Gujarati. Additionally, we have created an expert-labeled dataset to automatically identify homophobic and transphobic content within comments collected from YouTube. We provided comprehensive annotation rules to educate annotators in this process. We collected approximately 10,000 comments from YouTube for all three languages. Marking the first dataset of these languages for this task, we also developed a baseline model with pre-trained transformers.

Keywords: Homophobia, Transphobia, Hate speech, Dravidian Languages, Dataset creation, Low-Resourced Languages

1. Introduction

In recent years, social media has become an integral part of our daily lives, facilitating communication, information sharing, and networking among individuals across the globe. While these platforms offer unprecedented opportunities for connecting and engaging with diverse communities, they also provide a space where hate speech and discrimination can thrive. Homophobia and transphobia (Chakravarthi, 2023), (Chakravarthi et al., 2022a), in particular, are pervasive issues that persistently manifest themselves in the form of harmful comments and expressions on social media platforms. In Diefendorf and Bridges (2020) authors give a brief explanation which is, that homophobia refers to the irrational fear, hatred, or discrimination against individuals who identify as homosexual or are perceived as such, while transphobia encompasses similar sentiments directed at transgender individuals (Thavareesan and Mahesan, 2020), (Thavareesan and Mahesan, 2019). The harmful impact of homophobia and transphobia on the well-being and mental health of LGBTQ+ individuals is well-documented (Meyer, 2003). These prejudices not only contribute to social exclusion but also perpetuate a hostile online environment that can deter vulnerable individuals from fully participating in the digital discourse (Chakravarthi, 2023). The rise of hate speech on social media (Andrew, 2023),(Kumaresan et al., 2022) is concerning and poses a significant challenge to platforms and society at large. Consequently, there is a growing need for effective tools and methodologies to detect, monitor, and mitigate instances of homophobia and transphobia in social media comments (Chakravarthi, 2020). Such tools not only have the potential to protect vulnerable individuals but also contribute to fostering more inclusive and respectful online communities.

Detecting homophobia and transphobia in social media comments is a multifaceted task, as it requires the analysis of various languages, contexts, and intent (Shanmugavadivel et al., 2022), (Wong et al., 2023). Researchers have made significant strides in developing computational models that can identify and categorize hate speech, but the challenges persist due to the constantly evolving nature of online discourse (Davidson et al., 2017). Therefore, this paper will explore various strategies employed to address these challenges, including data analysis, and pre-trained model training. This research paper explores the development and implementation of machine learning algorithms and natural language processing (NLP) techniques for the automated detection of homophobia and transphobia in social media comments. By examining the current state of research, we aim to contribute to a deeper understanding of the complexities surrounding hate speech detection (Mandl et al., 2021), with a specific focus on homophobia and transphobia.

LREC-COLING 2024, pages 4404–4411

Labels	Telugu	Kannada	Gujarati
Homophobia	4,119	3,949	3,275
Transphobia	3,823	4,058	2,894
None of the categories	4,987	6,369	5,430

Table 1: Number of comments in each label for Telugu, Kannada, and Gujarati

Furthermore, we provide a study of the state of the art in detecting homophobia and transphobia in social media comments specifically for Telugu, Kannada, and Gujarati. Telugu, spoken primarily in the Indian states of Andhra Pradesh and Telangana, boasts over 80 million native speakers, one of the Dravidian languages. Gujarati, predominantly found in the western Indian state of Gujarat, has approximately 55 million speakers, while Kannada, the official language of Karnataka in southern India, is spoken by over 40 million people. These languages reflect the linguistic diversity and rich cultural heritage of India. By investigating the challenges, methodologies, and results in the below sections, we hope to contribute to the ongoing efforts to create safer and more inclusive online spaces for all individuals, regardless of their sexual orientation or gender identity.

2. Related Work

YouTube is a popular social networking site that allows its users to create personal profiles and upload videos as well as share views with other users. Techniques like 'likes' and 'shares' are used to attract a large audience on the platform's behalf whereby hundreds of followers view video clips or join the discussion. The contemporary manifestations of social media are frequently subject to exploitation, with an emphasis on the propagation of violent messages, hate speech, and offensive remarks (Waseem and Hovy, 2016). A variety of studies have examined the interactions among individuals regarding these issues, encompassing posts, videos, and comments. Their goal is to assess whether elements of aggression (Aroyehun and Gelbukh, 2018), misogyny, racism, harassment, and violence are prevalent within the realm of social media (Glasgow and Schouten, 2014).

In their 2017 study, Wu and Hsieh (2017) delved into the linguistic patterns present in Chinese texts created by the LGBT+ community. Their research revealed that conventional systems trained to identify gender from text struggled to handle the complexities of these texts effectively. In 2020, Ljubešić et al. (2020) developed emotion lexicons for Croatian, Dutch, and Slovene languages. They employed these lexicons to identify texts that either contained or did not contain socially unacceptable discourse related to the subjects of migrants and the LGBT+ community. As this field is still in its early stages, it grapples with various shortcomings linked to the specific objectives and variation of offensive language regarding homophobia and transphobia, as well as the general challenges associated with classification tasks (Ninalga, 2023), (Sureshnathan et al., 2023). These factors hinder systems from achieving optimal results in this area (Pannerselvam et al., 2024).

In a study by Wenpeng Yin and Roth (2019), they conducted an in-depth examination of the limitations found in prior research on zero-shot text classification. They brought attention to the challenges in accurately defining the problem, the significance of labels, and the overall confusion related to datasets and evaluation setups (Subramanian et al., 2022), (Lande et al., 2023). To address these issues, they took the initiative to create a benchmark for zero-shot text classification by standardizing datasets and evaluation protocols.

Chakravarthi (2023), in their 2021 study, introduced a novel hierarchical taxonomy designed for categorizing instances of homophobia and transphobia on the internet. They also provided a meticulously classified dataset, allowing for the automated detection of homophobic and transphobic content. To address the sensitivity of this issue, detailed annotation guidelines were provided to the annotators. The dataset consists of 15,141 comments written in English, Tamil, and a combination of Tamil-English, each of which has undergone thorough annotation. Building on this dataset, (Chakravarthi et al., 2022b), organized a shared task intending to advance research in identifying homophobia and transphobia. This initiative garnered participation from 10 systems for the Tamil language, 13 systems for English, and 11 systems for the Tamil-English language combination. Additionally, they carried out more work in 2023 (Chakravarthi et al., 2023) as well as in 2024 for English, Tamil, Hindi, Malayalam, and Tamil-English.

Furthermore, inspired by these research works about homophobia and transphobia on YouTube comments, we developed datasets for three languages such as Telugu, Kannada, and Gujarati.

3. Dataset Description

The datasets employed in this research were precisely compiled, drawing from a combination of manual collection methods and the utilization of

Sets	Telugu	Kannada	Gujarati
Train	9,050	10,063	8,119
Test	1,939	2,156	1,740
Development	1,940	2,157	1,740
Total	12,929	14,376	11,599

Table 2: Dataset Statics for training, development, and test sets for Telugu, Kannada, and Gujarati

a specialized YouTube comment scraper[1]. These datasets encompass three distinct languages: Telugu, Kannada, and Gujarati. The paramount concern was to ensure the quality and precision of the datasets. To achieve this, a rigorous annotation process was executed, meticulously adhering to established annotation guidelines detailed in the paper by Chakravarthi (2023).

This annotation process played a pivotal role in the systematic categorization of comments within each dataset. Our team consists of diverse annotators who have played a crucial role in the dataset creation process, each bringing a range of qualifications and language expertise. We have three annotators, two male and one female, who possess postgraduate degrees and have a strong command of the language. It enabled us to partition the data into three distinct sets, namely the training set, the testing set, and the development set.

Each of these sets underwent a comprehensive and systematic analysis, during which comments were meticulously categorized into various labels. These labels encompassed categorizations such as homophobia, and transphobia, and comments that did not fit into either of these categories come under none of the categories. The results of our initial analysis are elaborated upon and presented in detail in Table 2, offering insights into the distribution of comments across the training, testing, and development sets for the Telugu, Kannada, and Gujarati languages. This detailed process was crucial in ensuring the accuracy and comprehensiveness of our dataset for subsequent research and analysis. Examples of the dataset for each dataset are shown in Figure 1. These datasets were used in our CodaLab competition shared task organized by LT-EDI-2024@EACL[2] (Chakravarthi et al., 2024).

The labels for the comments have been quantified for each language, as illustrated in Table 1. The datasets for Telugu, Kannada, and Gujarati languages, used to detect homophobia and transphobia, are annotated and organized into training, test, and development sets. The volume of comments and distribution of labels within each set provides a resource for study, contributing to the dataset's utility for this research.

[1] https://github.com/philbot9/youtube-comment-scraper

[2] https://codalab.lisn.upsaclay.fr/competitions/16056

4. Methodology

Our aim of the homophobia and transphobia approach is to analyze the three labels. To create baselines for the datasets, we have used the Transformer Deep Learning architecture, particularly based on the BERT architecture (Devlin et al., 2018). A transformer is a Deep Learning Architecture, which relies on the Multi-Head Attention Mechanism, proposed by Google (Vaswani et al., 2017). The BERT transformer has recently gained widespread attention due to its exceptional performance in multiple benchmarks.

The primary reasons are the ability to capture contextual meaning by a bi-directional approach, to better understand the variations of language, unlike RNN and LSTM-based models, it captures long-range dependencies due to its transformer architecture, and focuses on relevant information, pre-training on large amounts of unlabelled data, thus being able to easily understand general language representations while fine-tuning. The variants used by us in this study are XLM-RoBERTa (Conneau et al., 2020), IndicBERT (Kakwani et al., 2020)) and BERT available monolingual datasets of Gujarati[3], Kannada[4], and Telugu[5].

Parameter	Value
Batch size	128
epochs	50
learning rate	0.001
train/test/dev	70%/15%/15%
optimizer	Adam

Table 3: Hyper-parameters for training

4.1. Settings for the experiment

The training was carried out in Python and used the packages scikit-learn (Pedregosa et al., 2011) and PyTorch (Paszke et al., 2019), to get final classification reports and training, testing of the models. For training of the Transformer models, we have used the Hugging Face[6] transformers package (Wolf et al., 2020) to access pre-trained models and tokenizers.

[3] https://huggingface.co/l3cube-pune/gujarati-bert

[4] https://huggingface.co/l3cube-pune/kannada-bert

[5] https://huggingface.co/l3cube-pune/telugu-bert

[6] https://huggingface.co/

Homophobia
Gujarati: તમારી સમલૈંગિકતા તમને સામાજિકપણે સ્વીકાર્ય બનાવશે નહીં. **Translation:** Your homosexuality will not make you socially acceptable.
Kannada: ಅವರು ಸಲಿಂಗಕಾಮಿ ಆಗಲು ಹೇಗೆ ಆಯ್ಕೆ ಮಾಡಿದರು? ಅದು ಸಹಜವಾಗಿ ಬರುವುದು ಅಲ್ಲ. **Translation:** How did he choose to be gay? It does not come naturally.
Telugu: స్వలింగ సంపర్కాలు అనేవి కేవలం జరుగుతున్న అనైతిక కార్యకలాపాలు. **Translation:** Homosexual acts are simply immoral activities.

Transphobia
Gujarati: ટ્રાન્સજેન્ડર લોકો સેન્સેક્સુઆલ કામગીરીમાં હાનિકારક છે. **Translation:** Transgender people are disadvantaged in syn-sexual functioning.
Kannada: ನನ್ನ ಮಗಳನ್ನು ಟ್ರಾನ್ಸ್ಜೆಂಡರ್ ಹುಡುಗಿಗೆ ಹೊಂದಿಸುವುದಿಲ್ಲ, ಅದು ಅವಳಿಗೆ ಕೇಡಾಗುವುದು. **Translation:** My daughter will not be matched with a transgender girl, it will harm her.
Telugu: మీరు ట్రాన్స్‌జెండర్గా మారడానికి మీ ముద్దుల మాతా పితలు ఎంత దు:ఖించారు! **Translation:** How sad your lovely parents were when you became a transgender!

Figure 1: Examples for each language for Homophobia and Transphobia labels

After tokenization and obtaining encoded outputs from the pre-trained models, we used a Classification Head, consisting of two Linear Layers and a dropout in between. The output layer has a size of 3, for each of the three classes in the dataset, namely homophobia, none of the categories, and transphobia, which are labeled encoded as 0, 1, and 2 respectively. The model is supervised with CrossEntropyLoss applied on the output, which is a standard loss function used in deep learning for classification tasks. During training the dev set was used to save the best checkpoint of the model as per the loss obtained on the dev set to prevent any over-fitting. On average the models take between 25-30 minutes on a Nvidia 3060 RTX GPU to train on the proposed dataset. The loss shows an overall good downward trend over the epochs as shown by the loss epoch on the Gujarati dataset as an example.

The size of the support set for each of the three classes in the train, test, and dev set of each dataset is provided in Table 2. Hyperparameters for the training are provided in Table 3. The availability of a GPU for training encouraged us to go for 50 epochs to get a better-tuned version of the model than lower epochs. To be able to train for the higher number of epochs and to take advantage of GPU training higher batch size is preferable, however, to also prevent overfitting and memory constraints, we chose a batch size of 128. We have chosen Adam optimizer as it has been shown to perform well on all kinds of models and is easy to implement, with betas taken as (0.9, 0.999) to prevent noise in the gradients computed, especially in a complex model like transformer. The experiments and the dataset for three languages can be accessed through the GitHub repository[7].

5. Results and Discussion

The results of our experiment encompass various performance metrics, including accuracy, precision, recall, and F1-Score (both Weighted and Macro). These metrics were calculated using the Sci-kit-learn package and are presented in Table 4 for the development set and Table 5 for the test set. In the case of the Telugu dataset, all three models exhibited notably high accuracy and F1 scores, with particular prominence in the IndicBERT and BERT models. Notably, the Telugu BERT model, specially trained on Telugu data, achieved a remarkable macro F1 score of 0.96 in the development set and 0.95 in the test set. Similarly, in the Kannada language dataset, the BERT model delivered an outstanding performance, consistently achieving a macro F1 score of 0.92 on both the test and development sets. This robust performance can be attributed to the model's training on Kannada data, aligning with the specific dataset used here—KannadaBERT.

Moving on to the Gujarathi dataset, the IndicBERT model also demonstrated impressive results with a macro F1 score of 0.95. The BERT model, while not far behind, achieved comparable results, in its utilization of the GujarathiBERT variant. As with the previous languages, the success of these models in the Gujarathi dataset can be attributed to their specialized training. The variation, in BERT performance across Telugu, Kannada, and Gujarati languages may be influenced

[7] https://github.com/Prasanna-04/LREC-COLING-2024-Homo-Trans

Development sets								
Languages	**Model**	**Acc**	**mP**	**mR**	**mF1**	**wP**	**wR**	**wF1**
Telugu	xlm-RoBERTa	**0.95**	0.95	0.95	0.95	0.95	0.95	0.95
	IndicBERT	**0.95**	0.95	0.96	**0.96**	0.96	0.96	0.96
	BERT	**0.95**	0.95	0.96	**0.96**	0.96	0.96	0.96
Kannada	xlm-RoBERTa	0.90	0.90	0.91	0.91	0.91	0.90	0.91
	IndicBERT	**0.92**	0.91	0.92	0.91	0.91	0.91	0.91
	BERT	**0.92**	0.92	0.92	**0.92**	0.92	0.92	0.92
Gujarati	xlm-RoBERTa	0.90	0.90	0.90	0.90	0.92	0.91	0.91
	IndicBERT	**0.95**	0.95	0.95	**0.95**	0.96	0.96	0.96
	BERT	**0.95**	0.94	0.95	**0.95**	0.95	0.95	0.95

Table 4: Classification report on development sets for Telugu, Kannada and Gujarati

Test sets								
Languages	**Model**	**Acc**	**mP**	**mR**	**mF1**	**wP**	**wR**	**wF1**
Telugu	xlm-RoBERTa	0.94	0.94	0.95	**0.95**	0.95	0.94	0.94
	IndicBERT	**0.95**	0.95	0.95	**0.95**	0.95	0.95	0.95
	BERT	**0.95**	0.95	0.96	**0.95**	0.95	0.95	0.95
Kannada	xlm-RoBERTa	0.90	0.91	0.91	0.91	0.91	0.91	0.91
	IndicBERT	**0.92**	0.92	0.92	**0.92**	0.92	0.92	0.91
	BERT	0.91	0.92	0.92	**0.92**	0.92	0.92	0.92
Gujarati	xlm-RoBERTa	0.90	0.91	0.91	0.90	0.92	0.91	0.91
	IndicBERT	0.95	0.95	0.95	**0.95**	0.96	0.95	0.95
	BERT	**0.96**	0.94	0.94	0.94	0.95	0.95	0.95

Table 5: Classification report on test sets for Telugu, Kannada, and Gujarati

more by the differences in each language than the quality of data used. Telugu and Kannada seem to align with BERT design making it easier for the model to accurately predict these languages. As a result, BERT can be effective. However, Gujarati poses challenges that are more difficult for BERT to grasp. This could explain the difference in its performance. Gujarati syntax or semantics may be more complex in ways that BERT is not as adept at handling. This highlights the importance of adapting models to suit the characteristics of each language.

In summary, our analysis reveals that transformer models consistently yield excellent results, with a consistent achievement of over 90% in all three datasets. This underscores the significance of training models on data from specific languages, exemplified by the outstanding performance of TeluguBERT, KannadaBERT, and GujarathiBERT on their respective datasets. These results emphasize the effectiveness of utilizing pre-trained models tailored to the linguistic characteristics of the target dataset, contributing to the broader discourse on natural language processing and model specialization.

6. Limitations

Our dataset mainly consists of YouTube comments, potentially not fully representing the broader landscape of online homophobic/transphobic hate speech. The expert-labeled dataset's accuracy depends on annotator interpretations, introducing subjectivity. Online language evolves rapidly, challenging the models' adaptability. Detecting hate speech often involves intricate contextual nuances, a difficulty for machine learning models. Pre-trained models like BERT may inherit biases. Deploying homophobic/Transphobic hate speech detection tools should be approached cautiously to avoid potential unintended consequences, including censorship and privacy issues, and while the models excel at identifying hate speech, capturing user intent may remain a challenge.

7. Conclusion

We present a dataset for three languages with the expert-labeled annotation for the text classification Homophobia, Transphobia, and none of the categories. This is the first dataset for three languages Telugu, Kannada, and Gujarati with a high amount of comments in this particular task. We provided baseline experiments with the pre-trained transformer models. In future work, we would like to expand the dataset into fine-grained levels and use large language models to implement classifiers for homophobia and transphobia detection.

8. Acknowledgements

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Artificial Intelligence under Grant No. 18/CRT/6223, supported in part by a research grant from the Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2(Insight_2) and also supported by the College of Science and Engineering, University of Galway.

9. References

Judith Jeyafreeda Andrew. 2023. JudithJeyafreeda@LT-EDI-2023: Using GPT model for recognition of homophobia/transphobia detection from social media. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 78–82, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Segun Taofeek Aroyehun and Alexander Gelbukh. 2018. Aggression detection in social media: Using deep neural networks, data augmentation, and pseudo labeling. In *Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018)*, pages 90–97, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Bharathi Chakravarthi. 2023. Detection of homophobia and transphobia in YouTube comments. *International Journal of Data Science and Analytics*, pages 1–20.

Bharathi Raja Chakravarthi. 2020. HopeEDI: A multilingual hope speech detection dataset for equality, diversity, and inclusion. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 41–53.

Bharathi Raja Chakravarthi, Adeep Hande, Rahul Ponnusamy, Prasanna Kumar Kumaresan, and Ruba Priyadharshini. 2022a. How can we detect Homophobia and Transphobia? Experiments in a multilingual code-mixed setting for social media governance. *International Journal of Information Management Data Insights*, 2(2):100119.

Bharathi Raja Chakravarthi, Prasanna Kumaresan, Ruba Priyadharshini, Paul Buitelaar, Asha Hegde, Hosahalli Shashirekha, Saranya Rajiakodi, Miguel Ángel García, Salud María Jiménez-Zafra, José García-Díaz, Rafael Valencia-García, Kishore Ponnusamy, Poorvi Shetty, and Daniel García-Baena. 2024. Overview of third shared task on homophobia and transphobia detection in social media comments. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 124–132, St. Julian's, Malta. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Rahul Ponnusamy, Malliga S, Paul Buitelaar, Miguel Ángel García-Cumbreras, Salud María Jimenez-Zafra, Jose Antonio Garcia-Diaz, Rafael Valencia-Garcia, and Nitesh Jindal. 2023. Overview of second shared task on homophobia and transphobia detection in social media comments. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 38–46, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Thenmozhi Durairaj, John Philip McCrae, Paul Buitelaar, Prasanna Kumaresan, and Rahul Ponnusamy. 2022b. Overview of the shared task on homophobia and transphobia detection in social media comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 369–377.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. In *Proceedings of the international AAAI conference on web and social media*, volume 11, pages 512–515.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Sarah Diefendorf and Tristan Bridges. 2020. On the enduring relationship between masculinity and homophobia. *Sexualities*, 23(7):1264–1284.

Kimberly Glasgow and Ronald Schouten. 2014. Assessing violence risk in threatening communications. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 38–45, Baltimore, Maryland, USA. Association for Computational Linguistics.

Raviraj Joshi. 2022. L3cube-mahacorpus and mahabert: Marathi monolingual corpus, marathi bert language models, and resources. *ArXiv*, abs/2202.01159.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, NC Gokul, Avik Bhattacharyya, Mitesh M Khapra, and Pratyush Kumar. 2020. IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4948–4961.

Prasanna Kumar Kumaresan, Rahul Ponnusamy, Elizabeth Sherly, Sangeetha Sivanesan, and Bharathi Raja Chakravarthi. 2022. Transformer Based Hope Speech Comment Classification in Code-Mixed Text. In *International Conference on Speech and Language Technologies for Low-resource Languages*, pages 120–137. Springer.

Kaustubh Lande, Rahul Ponnusamy, Prasanna Kumar Kumaresan, and Bharathi Raja Chakravarthi. 2023. KaustubhSharedTask@LT-EDI 2023: Homophobia-transphobia detection in social media comments with NLPAUG-driven data augmentation. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 71–77, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Nikola Ljubešić, Ilia Markov, Darja Fišer, and Walter Daelemans. 2020. The LiLaH emotion lexicon of Croatian, Dutch and Slovene. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 153–157, Barcelona, Spain (Online). Association for Computational Linguistics.

Thomas Mandl, Sandip Modha, Anand Kumar M, and Bharathi Raja Chakravarthi. 2021. Overview of the HASOC Track at FIRE 2020: Hate Speech and Offensive Language Identification in Tamil, Malayalam, Hindi, English and German. In *Proceedings of the 12th Annual Meeting of the Forum for Information Retrieval Evaluation*, FIRE '20, page 29–32, New York, NY, USA. Association for Computing Machinery.

Ilan H Meyer. 2003. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. *Psychological bulletin*, 129(5):674.

Dean Ninalga. 2023. Cordyceps@LT-EDI: Patching language-specific homophobia/transphobia classifiers with a multilingual understanding. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 185–191, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Kathiravan Pannerselvam, Saranya Rajiakodi, Sajeetha Thavareesan, Sathiyaraj Thangasamy, and Kishore Ponnusamy. 2024. SetFit: A robust approach for offensive content detection in Tamil-English code-mixed conversations using sentence transfer fine-tuning. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 35–42, St. Julian's, Malta. Association for Computational Linguistics.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12:2825–2830.

Rahul Ponnusamy, Prasanna Kumar Kumaresan, Kishore Kumar Ponnusamy, Charmathi Rajkumar, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2023. Team_Tamil at HODI: Few-Shot Learning for Detecting Homotransphobia in Italian Language. In *Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), CEUR. org, Parma, Italy*.

Kogilavani Shanmugavadivel, Sai Haritha Sampath, Pramod Nandhakumar, Prasath Mahalingam, Malliga Subramanian, Prasanna Kumar Kumaresan, and Ruba Priyadharshini. 2022. An analysis of machine learning models for sentiment analysis of Tamil code-mixed data. *Computer Speech & Language*, 76:101407.

Malliga Subramanian, Ramya Chinnasamy, Prasanna Kumar Kumaresan, Vasanth Palanikumar, Madhoora Mohan, and Kogilavani Shanmugavadivel. 2022. Development of Multi-lingual Models for Detecting Hope Speech

Texts from Social Media Comments. In *International Conference on Speech and Language Technologies for Low-resource Languages*, pages 209–219. Springer.

Shwetha Sureshnathan, Samyuktaa Sivakumar, Priyadharshini Thandavamurthi, Thenmozhi D., Bharathi B, and Kiruthika Chandrasekaran. 2023. Tercet@LT-EDI-2023: Homophobia/transphobia detection in social media comment. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 266–271, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Sajeetha Thavareesan and Sinnathamby Mahesan. 2019. Sentiment analysis in tamil texts: A study on machine learning techniques and feature representation. In *2019 14th Conference on Industrial and Information Systems (ICIIS)*, pages 320–325.

Sajeetha Thavareesan and Sinnathamby Mahesan. 2020. Sentiment lexicon expansion using word2vec and fasttext for sentiment prediction in tamil texts. In *2020 Moratuwa Engineering Research Conference (MERCon)*, pages 272–276.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Zeerak Waseem and Dirk Hovy. 2016. Hateful symbols or hateful people? predictive features for hate speech detection on Twitter. In *Proceedings of the NAACL Student Research Workshop*, pages 88–93, San Diego, California. Association for Computational Linguistics.

Jamaal Hay Wenpeng Yin and Dan Roth. 2019. Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach. In *EMNLP*.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Sidney Wong, Matthew Durward, Benjamin Adams, and Jonathan Dunn. 2023. cantnlp@LT-EDI-2023: Homophobia/transphobia detection in social media comments using spatio-temporally retrained language models. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 103–108, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Hsiao-Han Wu and Shu-Kai Hsieh. 2017. Exploring Lavender Tongue from Social Media Texts [In Chinese]. In *Proceedings of the 29th Conference on Computational Linguistics and Speech Processing (ROCLING 2017)*, pages 68–80.