Overview of Second Shared Task on Homophobia and Transphobia Detection in Social Media Comments

Bharathi Raja Chakravarthi[1], **Rahul Ponnusamy**[2], **Malliga Subramanian**[3]
Paul Buitelaar[2], **Miguel Ángel García-Cumbreras**[4], **Salud María Jiménez-Zafra**[4],
José Antonio García-Díaz[5], **Rafael Valencia-García**[5] **Nitesh Jindal**[6]
[1] School of Computer Science, University of Galway, Ireland
[2] Insight SFI Research Centre for Data Analytics, University of Galway, Ireland
[3] Kongu Engineering College, Tamil Nadu, India [4] Universidad de Jaén, Spain
[5] Universidad de Murcia, Spain [6] University of Galway, Ireland
bharathi.raja@universityofgalway.ie

Abstract

We present an overview of the second shared task on homophobia/transphobia Detection in social media comments. Given a comment, a system must predict whether or not it contains any form of homophobia/transphobia. The shared task included five languages: English, Spanish, Tamil, Hindi, and Malayalam. The data was given for two tasks. Task A was given three labels, and Task B fine-grained seven labels. In total, 75 teams enrolled for the shared task in Codalab. For Task A, 10 teams submitted systems for English, 7 for Tamil, 4 for Spanish, 7 for Malayalam, and seven teams for Hindi. For Task B, 8 teams were submitted for English, 7 for Tamil, and 6 for Malayalam. We present and analyze all submissions in this paper.

1 Introduction

A victim, an aggressor, and bully-victims are all necessary components of the aggressive behavior known as bullying (Colvin et al., 1998). Students, parents, teachers, and administrators all share a considerable concern with the phenomenon of bullying that is motivated by homophobia (Horn et al., 2009; Wright et al., 1999; Basile et al., 2009). The unfavorable views, attitudes, prejudices, and behaviors that are held towards sexual minorities are what are referred to as homophobia (Hong and Garbarino, 2012). Homophobia is the underlying mentality that plays a contributing role in the practice of discrimination against LGBTQ+ vulnerable individuals (Alichie, 2022).

The fact that the most prevalent definition of homophobia is "an attitude of hostility toward male or female LGBTQI+ vulnerable individuals" implies that this idea is rather narrow and has a tendency to individualize the process of discrimination and rejection(Herek, 1988). The Internet is a tool that LGBTQI+ vulnerable young individuals in regional, remote, and rural areas use to overcome isolation and construct relationships that extend beyond the physical limitations of a geographic area by commenting on YouTube videos or reaching out via Twitter or other means in social media (Venzo and Hess, 2013; Soriano, 2014; Han et al., 2019a; Chakravarthi, 2023). Since social media is used by vulnerable individuals, it should be without homophobic/transphobic bullying or other hate speech against LGBTQ+ vulnerable individuals (Han et al., 2019b; Rokhmansyah et al., 2021; Ștefăniță and Buf, 2021). To tackle homophobia and transphobia in social media, Chakravarthi et al. (2022a) introduced a new dataset in English, Tamil, and Tamil-English. Chakravarthi et al. (2022b) conducted new shared tasks in Tamil, English, and Tamil-English (code-mixed) languages. It received 10 Tamil systems, 13 English systems, and 11 Tamil-English systems. The average macro F1-score for the top systems for Tamil, English, and Tamil-English was 0.570, 0.877, and 0.610, respectively. Chinnaudayar Navaneethakrishnan et al. (2023) conducted a shared task on Sentiment Analysis and Homophobia Detection; in that task, new language data, Malayalam, was added. In our task[1], We conducted two sub-tasks: Task A and Task B. We included English, Tamil, Spanish, Malayalam, and Hindi for Task A and English, Tamil, and Malayalam for Task B. Overall submission of eleven teams that participated in Task A and eight teams that participated in Task B. The Weighted F1 scores of top-performing models for these languages are 0.888, 0.997, 0.979, 0.969, and 0.949. For Task B, we included English, Tamil, and Malayalam; the top-performing model scored with weighted F1 scores of 0.822, 0.884, and 0.865.

[1] https://codalab.lisn.upsaclay.fr/competitions/11077

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion, pages 38–46, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_006

set	H	T	N	Total
Training	179	7	2,978	3,164
Dev	42	748	2	792
Test	55	4	931	990
Total	276	759	3,911	4,946

Table 1: Statistics for the Task A English Dataset (H stands for Homophobia, T for Transphobia, and N for Non-anti-LGBT+ content)

set	H	T	N	Total
Training	453	145	2,064	2,662
Dev	118	41	507	666
Test	152	47	634	833
Total	723	233	3,205	4,161

Table 2: Statistics for the Task A Tamil Dataset (H stands for Homophobia, T for Transphobia, and N for Non-anti-LGBT+ content)

set	H	T	N	Total
Training	476	170	2,468	3,114
Dev	197	79	937	1,213
Test	140	52	674	866
Total	813	301	4,079	5,193

Table 3: Statistics for the Task A Malayalam Dataset (H stands for Homophobia, T for Transphobia, and N for Non-anti-LGBT+ content)

set	H	T	N	Total
Training	92	45	2,423	2,560
Dev	13	2	305	320
Test	10	3	308	321
Total	115	50	3,036	3,201

Table 4: Statistics for the Task A Hindi Dataset (H stands for Homophobia, T for Transphobia, and N for Non-anti-LGBT+ content)

set	H	T	N	Total
Training	200	200	450	850
Dev	43	43	150	236
Test	100	100	300	500
Total	343	343	900	1,586

Table 5: Statistics for the Spanish Dataset (H stands for Homophobic, T for Transphobic, and N for None)

2 Task description

This is a classification task at the level of comments and posts. When presented with YouTube comments, the algorithms that the participants have developed should categorize it. Participants were given sentences in the comment section that were taken from social media. It is the responsibility of the participant's system to determine, given a comment, whether or not it contains any type of homophobia or transphobia. In order to determine whether the text contains homophobia or transphobia, the comments have been manually annotated. We divided the task into two subtasks: A and B.

2.1 Task A

In this task, participants were given a dataset with three labels. As a result, the participants' system must categorize the contents as homophobia, transphobia, or non-anti-LGBT+ content. The training, development, and test datasets for the following languages were given to the participants: English, Spanish, Hindi, Tamil, or Malayalam.

2.2 Task B

In this task, the participants were provided with the dataset with 7 labels. The participants' system needs to categorize the text into Homophobic-derogation, Homophobic-Threatening, Transphobic-derogation, Transphobic-Threatening, Hope-Speech, Counter-speech, and None-of-the-above. The participants were provided with the training, development, and test datasets for the following languages: English, Tamil, and Malayalam.

3 Dataset

3.1 Tamil, Malayalam, Hindi, and English Dataset

The comments were gathered using a tool known as the YouTube Comment Scraper[2]. These comments

[2] https://pypi.org/project/youtube-comment-scraper-python/

set	HD	HT	TD	TT	CS	HS	N	total
Training	167	12	6	1	302	436	2,240	3,164
Dev	41	1	2	0	84	111	553	792
Test	54	1	3	1	100	140	691	990
Total	262	14	11	2	486	687	3,484	4,946

Table 6: Statistics for the Task B English Dataset (HD stands for Homophobic-derogation, HT for Homophobic-Threatening, TD for Transphobic-derogation, TT for Transphob)

set	HD	HT	TD	TT	CS	HS	N	total
Training	416	37	111	34	212	218	1,634	2,662
Dev	107	11	31	10	60	52	395	666
Test	138	14	28	19	64	65	505	833
Total	661	62	170	63	336	335	2534	4,161

Table 7: Statistics for the Task B Tamil Dataset (HD stands for Homophobic-derogation, HT for Homophobic-Threatening, TD for Transphobic-derogation, TT for Transphob)

were utilized by us in the process of manually annotating our datasets. We collected Tamil, Malayalam, Hindi, and English from YouTube videos selected by us. However, we discovered that the text contained a substantial quantity of English in addition to a variety of other languages. The presence of responses written in languages other than the target language made the already challenging task of extracting pertinent text from the comment section more difficult. As part of the preparatory operations for data cleansing, we used langdetect library[3] to distinguish between distinct languages and separate them into their own categories. We separated the data into three distinct sections, including English and Tamil. The remaining code-mixed Tamil and English were maintained. For Hindi and Malayalam, we discarded all other comments, including comments in English; we only took Hindi and Malayalam comments from those videos. To comply with the regulations governing the preservation of user data, we removed all user-related information from the corpus. In order to better prepare for the exam, we eliminated any unnecessary information, including URLs. We manually annotated them into three labels and seven labels according to our guidelines with the help of trained annotators. The data statistics for Task A and B of all languages are shown in Tables 1, 2, 3, 4, 6, 7, and 8.

3.2 Spanish Dataset

The Spanish dataset is composed of a set of tweets collected using the UMUCorpusClassifier tool (García-Díaz et al., 2020), which allows for defining different search criteria such as keywords, accounts, and geolocation. The keywords used to collect tweets related to transphobia were: #transfobia (*#transphobia*), trans (*trans*), transexual (*transsexual*), transgénero (*transgender*), identidad de género (*gender identity*) and androginia (*androgyny*). Regarding homophobia, the words selected were: #homofobia (*#homophobia*), homosexual (*homosexual*), #AlertaHomofobia (*#HomophobiaAlert*), marica (*queer*), lesbiana (*lesbian*), maricones (*fags*), maricona (*fag*), bolleras (*dykes*), gay (*gay*), afeminado (*effeminate*), petar AND (culo OR ojete) (*butt-fucking*) and #StopLGTBIfobia *#StopLGTBIphobia*. In addition, for the latter, words related to the murder of the Samuel Luiz[4] were added: samuel luiz (*samuel luiz*), asesinato de samuel (*samuel murder*), asesinos de samuel (*samuel killers*), muerte de samuel (*samuel death*), #samuel (*#samuel*), el chico de galicia (*the boy from galicia*), #Justiciaparasamuel (*#justiceforsamuel*). In total, it was retrieved 473,191 tweets for homophobia and 451,565 for transphobia. From this collection of tweets, we discarded those tweets with short length and retweets. A subset of the collected tweets was manually labeled by organizers of the shared task to determine which were really related to homophobia, which to transphobia, and which to neither, as it is not possible to rely on keywords in the texts for the annotation. Finally, for the shared task, it was selected a total of 1,586 tweets that were distributed in development, train-

[3] https://pypi.org/project/langdetect/

[4] https://es.wikipedia.org/wiki/Asesinato_de_Samuel_Luiz

set	HD	HT	TD	TT	CS	HS	N	total
Training	419	57	163	7	152	69	2,247	3,114
Dev	181	16	75	4	60	29	848	1,213
Test	129	11	48	4	46	22	606	866
Total	729	84	286	15	258	120	3,701	5,193

Table 8: Statistics for the Task B Malayalam Dataset (HD stands for Homophobic-derogation, HT for Homophobic-Threatening, TD for Transphobic-derogation, TT for Transphob)

Team Name	Run Name	weighted F1	Rank
teamplusone	1	0.9692868	1
SuperNova (Reddy et al., 2023)	1	0.9658864	2
SsnTech2_Run1 (Sivanaiah et al., 2023)	1	0.9582267	3
Tercet_English (Sivakumar et al., 2023)	1	0.9534853	4
Cordyceps (Ninalga, 2023)	2	0.9512845	5
cantnlp (Wong et al., 2023)	1	0.9425137	6
DeepBlueAI	1	0.9416178	7
adsa_nlp_sys2	1	0.9363040	8
MUCS_Run3 (Hegde et al., 2023)	3	0.9198598	9
JudithJeyafreeda (Andrew, 2023)	1	0.8986411	10

Table 9: Task A – English

Team Name	Run Name	weighted F1	Rank
teamplusone	1	0.9793323	1
SuperNova (Reddy et al., 2023)	1	0.9793323	2
Cordyceps (Ninalga, 2023)	2	0.9695374	3
cantnlp (Wong et al., 2023)	1	0.9653340	4
DeepBlueAI	1	0.9591820	5
MUCS_Run3 (Hegde et al., 2023)	3	0.9418278	6
JudithJeyafreeda (Andrew, 2023)	1	0.0185185	7

Table 10: Task A – Hindi

Team Name	Run Name	weighted F1	Rank
Cordyceps (Ninalga, 2023)	2	0.9976971	1
MUCS_Run2 (Hegde et al., 2023)	2	0.9563322	2
DeepBlueAI	1	0.9493561	3
cantnlp (Wong et al., 2023)	1	0.9382083	4
SuperNova (Reddy et al., 2023)	1	0.9318975	5
teamplusone	1	0.8753247	6
JudithJeyafreeda (Andrew, 2023)	1	0.2520196	7

Table 11: Task A – Malayalam

Team Name	Run Name	weighted F1	Rank
Cordyceps (Ninalga, 2023)	2	0.8883174	1
MUCS_Run2 (Hegde et al., 2023)	2	0.8138490	2
SuperNova (Reddy et al., 2023)	1	0.7957093	3
VEL (Kumaresan et al., 2023)	1	0.3000000	4

Table 12: Task A – Spanish

Team Name	Run Name	weighted F1	Rank
Cordyceps (Ninalga, 2023)	1	0.9496857	1
DeepBlueAI	1	0.9424593	2
cantnlp (Wong et al., 2023)	1	0.9264145	3
MUCS_Run2 (Hegde et al., 2023)	2	0.9132474	4
SuperNova (Reddy et al., 2023)	1	0.8942428	5
teamplusone	1	0.8643490	6
JudithJeyafreeda(Andrew, 2023)	1	0.2702824	7

Table 13: Task A – Tamil

ing, and test sets, as can be seen in Table 5.

4 Methods of Participants

The team "teamplusone" submitted a system for Task A and B with an English dataset. They used a pre-trained model called BERT(Bidirectional Encoder Representations from Transformers). They used the default parameter setting for training. With this, they were able to achieve the weighted F1 score of 0.9692868 in Task A and 0.8221297 in Task B.

The "SuperNova" (Reddy et al., 2023) team used Term Frequency-Inverse Document Frequency (TF-IDF) for classifying both tasks. TF measures the frequency of a term within a document. Since Support Vector Machines(SVMs) are known for their ability to handle overfitting, this team used SVM to classify both tasks. This team also claimed that SVMs can perform well even with relatively small training datasets and generalize effectively from limited examples, making them suitable for sentiment analysis applications in various domains.

A team "SSNTech2" (Sivanaiah et al., 2023) classified Task A. For this task, the team first preprocessed and cleaned the dataset and assigned token values to each category. They used the nltk module for preprocessing, such as stop word removal, lemmatizing and normalizing, and removing stop words. For the first test run, the team used the SGD classifier. It produced an accuracy of 0.93, an F1 average score of 0.38, and a weighted score of 0.92. For the second test run, the team used the SVM classifier. It produced better results than the SGD classifier, with an accuracy of 0.94, an F1 average score of 0.42, and a weighted score of 0.94.

SVM is used for classifying the dataset in English under Task A by the team named "Tercet" (Sivakumar et al., 2023). Given a higher precision, F1 score, and weighted averages compared to the random forest, logistic regression, and Naive Bayes models, SVM was a good fit for classifying the test datasets. The team did preprocess the text data, such as removing punctuation, emoticons, and stop words. To convert the text data into a form that is usable by the model, the team also used the TF-IDF vectorizer algorithm. It utilized the extracted features in the SVM classifier. They used the TF-IDF vectorizer algorithm to convert the text data to the model understandable form. Then SVM classifier used the vectorized features for the classification.

The "Cordyceps" (Ninalga, 2023) team classified both tasks using a weight-space ensembling technique. First, they trained a multilingual model on a dataset that included all the languages and then created finetuned models for each language. Ultimately, for each language, they performed linear interpolation between the finetuned and multilingual models' weights. The resulting interpolated model is then used for inference. The selection of the linear interpolation parameter is based on a held-out validation set consisting of samples in the language of the finetuned model that were not encountered during training. The team also observed that weight-space ensembling enhances performance, particularly for low-resource languages. The most interesting aspect of this work is the novel application of weight-space ensembling on code-mixed data, aiming to leverage the strengths of both multilingual and finetuned models for improved performance in analyzing mixed-language text.

A custom pre-trained XLM-RoBERTa transformer-based multilingual model has been developed by the team "CantNLP" (Wong et al., 2023). This team has pre-trained the language model with a random sample of 50,000 tweets (over 50 characters) for each language condition. For the language conditions with Brahmic scripts (Hindi, Malayalam, and Tamil), the team romanized a quarter of the text samples to simulate

Team Name	Run Name	weighted F1	Rank
teamplusone	1	0.8221297	1
SuperNova (Reddy et al., 2023)	1	0.8014732	2
DeepBlueAI	1	0.7219212	3
KaustubhSharedTask (Lande et al., 2023)	1	0.6991867	4
cantnlp (Wong et al., 2023)	1	0.5397906	5
JudithJeyafreeda (Andrew, 2023)	1	0.2255661	6
MUCS_Run2 (Hegde et al., 2023)	2	0.1462137	7
Cordyceps (Ninalga, 2023)	2	0.1113251	8

Table 14: Task B – English

Team Name	Run Name	weighted F1	Rank
cantnlp (Wong et al., 2023)	1	0.8842916	1
MUCS_Run2 (Hegde et al., 2023)	2	0.8595397	2
DeepBlueAI	1	0.8533519	3
teamplusone	1	0.8233696	4
JudithJeyafreeda (Andrew, 2023)	1	0.0639703	5
Cordyceps (Ninalga, 2023)	1	0.0108560	6

Table 15: Task B – Malayalam

Team Name	Run Name	weighted F1	Rank
DeepBlueAI	1	0.8651552	1
MUCS_Run2 (Hegde et al., 2023)	2	0.8219683	2
SuperNova (Reddy et al., 2023)	1	0.8162569	3
cantnlp (Wong et al., 2023)	1	0.8041158	4
teamplusone	1	0.7548580	5
JudithJeyafreeda (Andrew, 2023)	1	0.6547745	6
Cordyceps (Ninalga, 2023)	1	0.0122772	7

Table 16: Task B – Tamil

script-mixing as observed in the comments and finetuned the language model with the training data. The team also over-sampled the training data to reduce class imbalance. Each model was trained with eight epochs, with Adam as the optimizer.

The team "DeepBlueAI" finetuned XLM-RoBERTa as the base model for classifying both tasks. This team has attempted mixing multiple language datasets at different proportions and performed cross-validation.

The team "Adsa_nlp_sys" used SVM in conjunction with TF-IDF Vectorization for classifying the English comments under Task B and used the ADASYN sampling technique. In order to hypertune the model, the team has used TF-IDF Grid. The team claimed that ADASYN, with TF-IDF and Grid search, can find the best model and parameters.

The team "KaustubhSharedTask" (Lande et al., 2023) participated in Task B in the English dataset. Due to class imbalance in task B's training dataset in the English language, they used NLPAUG - a tool to augment the text data and reduce the degree of imbalance. In augmentation of the text data, they tried several parameters like synonym replacement, word insertion, and word substitution to get augmented sentences with the same meaning as the original sentence. They did text preprocessing and applied various transformers models with fine tuning and got the best results on the bilstm model trained on the word embeddings generated from word2vec.

The "MUCS" (Hegde et al., 2023) team has tried using mBERT(Multilingual BERT) and resampling with BERT to classify both tasks. For feature extraction, they used TF-IDF.

A GPT2 model has been used by the team "JudithJeyafreeda" (Andrew, 2023) to finetune the training set for classifying both tasks. For using this model, the team substituted the comments in other languages with English letters.

The team “VEL” (Kumaresan et al., 2023) utilized the ”muril-large-cased” model, which is a variant of the GPT-3.5 architecture developed by OpenAI to classify the Spanish comments under Task A. In addition to using the ”muril-large-cased” model, the team also employed machine learning techniques such as Naive Bayes (NB), Support Vector Machines (SVM), Logistic Regression (LR), Decision Trees (DT), and Random Forests (RF) with count vectorizers.

5 Results and Discussion

Overall, we received a total of 10,7,4,7, and 7 submissions for English, Tamil, Spanish, Malayalam, and Hindi in Task A. For Task B, we received 8,7, and 6 submissions for English, Tamil, and Malayalam in Task B. The Tables 9,13,12,10 and 11 shows the rank list of all languages of Task A, and the tables 14, 16, and 15 shows the rank list of all languages of Task B.

In Task A, the model of the team “teamplusone” achieved the top-performing model in English and Hindi language. They used BERT pretrained model with the default setting for the training and achieved the weighted F1 score of 0.96928 and 0.97933, and the “Cordyceps” model is the top-performing model in Tamil, Malayalam, and Spanish languages. They used the weight space ensembling technique, improving the performance of analyzing the mixed language text. They achieved 0.94968, 0.99769, and 0.88831.

The top-performing models in Task B are the model developed by the team “teamplusone,” ranked 1st in the English language. They used BERT model for training with a default parameter setting. They achieved a weighted F1 of 0.82212. In Tamil language, the “DeepblueAI” team’s model got the 1st rank. They used the XLM-RoBERTa base model and performed cross-validation by combining multiple language datasets in varied amounts, which gained the weighted F1 score of 0.88429. For Malayalam, the “cantnlp” team developed the custom pre-trained XLM-RoBERTa model with 50000 random tweets, and they also oversampled the training to tackle the class imbalance problem. This model achieved 0.86515.

6 Conclusion

We presented the second shared task findings on homophobia/transphobia detection in social media comments in this publication. We got an extensive variety of entries that fulfilled the aims of the shared task. We expect that the shared task on homophobia/transphobia detection will have a long-term impact on the NLP discipline.

Acknowledgments

This work has been partially supported by Project CONSENSO (PID2021-122263OB-C21), Project MODERATES (TED2021-130145B-I00) and Project SocialTox (PDC2022-133146-C21) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR, Project PRECOM (SUBV-00016) funded by the Ministry of Consumer Affairs of the Spanish Government, Project FedDAP (PID2020-116118GA-I00) supported by MICINN/AEI/10.13039/501100011033 and WeLee project (1380939, FEDER Andalucía 2014-2020) funded by the Andalusian Regional Government. It is also part of the research project LT-SWM (TED2021-131167B-I00) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. This work is also part of the research projects LaTe4PoliticES (PID2022-138099OB-I00) funded by MCIN/AEI/10.13039/501100011033 and the European Fund for Regional Development (FEDER)-a way to make Europe and LaTe4PSP (PID2019-107652RB-I00/AEI/ 10.13039/501100011033) funded by MCIN/AEI/10.13039/501100011033. Salud María Jiménez-Zafra has been partially supported by a grant from Fondo Social Europeo and the Administration of the Junta de Andalucía (DOC_01073). The author, Bharathi Raja Chakravarthi, was supported in part by a research grant from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2 (Insight_2).

References

Bridget O. Alichie. 2022. Communication at the margins: Online homophobia from the perspectives of lgbtq+social media users. *Journal of Human Rights*, 0(0):1–15.

Judith Jeyafreeda Andrew. 2023. Using gpt model for recognition of homophobia/transphobia detection from social media. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Kathleen C Basile, Dorothy L Espelage, Ian Rivers, Pamela M McMahon, and Thomas R Simon. 2009. The theoretical and empirical links between bullying behavior and male sexual violence perpetration. *Aggression and Violent Behavior*, 14(5):336–347.

Bharathi Raja Chakravarthi. 2023. Detection of homophobia and transphobia in youtube comments. *International Journal of Data Science and Analytics*.

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, Ruba Priyadharshini, Thenmozhi Durairaj, John 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, Dublin, Ireland. Association for Computational Linguistics.

Subalalitha Chinnaudayar Navaneethakrishnan, Bharathi Raja Chakravarthi, Kogilavani Shanmugavadivel, Malliga Subramanian, Prasanna Kumar Kumaresan, Bharathi, Lavanya Sambath Kumar, and Rahul Ponnusamy. 2023. Findings of shared task on sentiment analysis and homophobia detection of YouTube comments in code-mixed Dravidian languages. In *Proceedings of the 14th Annual Meeting of the Forum for Information Retrieval Evaluation*, FIRE '22, page 18–21, New York, NY, USA. Association for Computing Machinery.

Geoff Colvin, Tary Tobin, Kelli Beard, Shanna Hagan, and Jeffrey Sprague. 1998. The school bully: Assessing the problem, developing interventions, and future research directions. *Journal of behavioral education*, 8:293–319.

José Antonio García-Díaz, Ángela Almela, Gema Alcaraz-Mármol, and Rafael Valencia-García. 2020. Umucorpusclassifier: Compilation and evaluation of linguistic corpus for natural language processing tasks. *Procesamiento del Lenguaje Natural*, 65:139–142.

Xi Han, Wenting Han, Jiabin Qu, Bei Li, and Qinghua Zhu. 2019a. What happens online stays online? —— social media dependency, online support behavior and offline effects for lgbt. *Computers in Human Behavior*, 93:91–98.

Xi Han, Wenting Han, Jiabin Qu, Bei Li, and Qinghua Zhu. 2019b. What happens online stays online?——social media dependency, online support behavior and offline effects for lgbt. *Computers in Human Behavior*, 93:91–98.

Asha Hegde, Kavya G, Sharal Coelho, and Hosahalli Lakshmaiah Shashirekha. 2023. Homophobic/transphobic content detection in social media text using mbert. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Gregory M. Herek. 1988. Heterosexuals' attitudes toward lesbians and gay men: Correlates and gender differences. *The Journal of Sex Research*, 25(4):451–477.

Jun Sung Hong and James Garbarino. 2012. Risk and protective factors for homophobic bullying in schools: An application of the social–ecological framework. *Educational Psychology Review*, 24:271–285.

Stacey S Horn, Joseph G Kosciw, and Stephen T Russell. 2009. Special issue introduction: New research on lesbian, gay, bisexual, and transgender youth: Studying lives in context.

Prasanna Kumar Kumaresan, Kishore Kumar Ponnusamy, Kogilavani S V, SUBALALITHA CN, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2023. Detecting homophobia and transphobia in code-mixed spanish social media comments. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Kaustubh Lande, Rahul Ponnusamy, Prasanna Kumar Kumaresan, and Bharathi Raja Chakravarthi. 2023. Homophobia-transphobia detection in social media comments with nlpaug-driven data augmentation. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Dean Ninalga. 2023. Patching language-specific homophobia/transphobia classifiers with a multilingual understanding. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Ankitha Reddy, Pranav Moorthi, and Ann Maria Thomas. 2023. Homophobia/transphobia detection in social media comments:-(english, tamil, hindi, spanish, malayalam). In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Alfian Rokhmansyah, Widyatmike Gede Mulawarman, and Yusak Hudiyono. 2021. LGBT news on tirto. id online media: Fairclough's critical discourse analysis. In *6th International Conference on Science, Education and Technology (ISET 2020)*, pages 191–197. Atlantis Press.

Samyuktaa Sivakumar, Priyadharshini Thandavamurthi, Shwetha Sureshnathan, Thenmozhi Durairaj, Bharathi B, and G L Gayathri. 2023. Hope speech detection for equality, diversity, and inclusion. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Rajalakshmi Sivanaiah, Vaidhegi D, Priya M, Angel Deborah S, and Mirnalinee ThankaNadar. 2023. Homophobia/transphobia detection in social media comments using linear classification techniques. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Cheryll Ruth Reyes Soriano. 2014. Constructing collectivity in diversity: online political mobilization of a national lgbt political party. *Media, culture & society*, 36(1):20–36.

Paul Venzo and Kristy Hess. 2013. "honk against homophobia": Rethinking relations between media and sexual minorities. *Journal of Homosexuality*, 60(11):1539–1556. PMID: 24147586.

Sidney Wong, Matthew Durward, Benjamin Adams, and Jonathan Dunn. 2023. Homophobia/transphobia detection in social media comments using spatiotemporally retrained language models. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Lester W Wright, Henry E Adams, and Jeffery Bernat. 1999. Development and validation of the homophobia scale. *Journal of psychopathology and behavioral assessment*, 21:337–347.

Oana Ștefăniță and Diana-Maria Buf. 2021. Hate speech in social media and its effects on the lgbt community: A review of the current research. *Romanian Journal of Communication and Public Relations*, 23(1):47–55.