

Overview of the Shared Task on Sentiment Analysis and Homophobia Detection of YouTube Comments in Code-Mixed Dravidian Languages^{*}

Kogilavani Shanmugavadivel¹, Malliga Subramanian¹, Prasanna Kumar Kumaresan², Bharathi Raja Chakravarthi³, B Bharathi⁴, Subalalitha Chinnaudayar Navaneethakrishnan⁵, Lavanya Sambath Kumar⁵, Thomas Mandl⁶, Rahul Ponnusamy⁷, Vasanth Palanikumar⁸ and Manoj J Balaji⁹

¹Kongu Engineering College, Tamil Nadu, India

²Indian Institute of Information Technology and Management-Kerala, India

³Insight Centre for Data Analytics, National University of Ireland, Galway

⁴SSN College of Engineering, Tamil Nadu, India

⁵SRM Institute of Science and Technology, Chennai, Tamil Nadu, India

⁶University of Hildesheim, Germany

⁷Techvantage Analytics, Kerala, India

⁸Chennai Institute of Technology, Tamil Nadu, India

⁹WorldQuant University, New Orleans, Louisiana

Abstract

Sentiment analysis is the task of identifying subjective opinions or emotional responses about a given topic. Sentiment analysis of social media posts, which are primarily code-mixed for Dravidian languages, is becoming more and more popular. Homophobia detection is the task of identifying homophobia, transphobia, and non-anti-LGBT+ content from social media YouTube comments. In this paper, we report an overview of the findings and results from the shared task on sentiment analysis and homophobia detection in Code-Mixed Dravidian languages organized as a part of FIRE 2022. For shared task-A, the participants were provided with development, training, and test dataset code-mixed text in Dravidian languages (Tamil-English, Malayalam-English, and Kannada-English). The goal of the shared task-A is to classify code-mixed YouTube comments into positive, negative, neutral, or mixed emotions. For shared task B, the participants were provided with development, training, and test dataset in English, Malayalam, and Tamil languages. The goal of the shared task B is to classify the text as homophobic, transphobic, or not. A total of 95 participants registered for the shared task, 13 teams finally submitted their results for task-A, and 10 teams submitted their results for task B. The performance of the systems submitted was evaluated in terms of macro-F1 score. The datasets for this challenge are openly available on the competition website¹.

Keywords

Sentiment analysis, Homophobia detection, Code-Mixed Dravidian Language, Machine Learning, Deep Learning,

¹<https://codalab.lisn.upsaclay.fr/competitions/5310>

1. Introduction

It is now possible for a greater number of people than ever before to exercise their right to freedom of expression thanks to the proliferation of social media platforms like Twitter and YouTube, as well as the anonymity afforded to users of these platforms. This leads to an increase in user-generated content, which can include opinions, sentiments, reviews about products and movies, likes and dislikes regarding an event or news, and much more. Due to objectionable content, such as threats and remarks directed at individuals, groups, or organizations, this, on the other hand, leads to the exploitation of these platforms in order to spread violence [1, 2, 3, 4]. Naturally, comments, posts, and articles have a tendency to imply a variety of things to a wide variety of people all over the world.

People frequently take advantage of this freedom to make comments that promote hatred and toxic behavior. Due to the ease with which users can share content (videos, posts, and shots), as well as like, share, and comment on said content, YouTube has become an extremely popular platform. The negative impact of this is that it makes more room for overt forms of cyberbullying and online harassment to occur [5]. This frequently has a significant influence on the lives of the individuals and communities that are impacted [6].

The field of Natural Language Processing (NLP) has seen an increase in the use of shared tasks [7, 8] in an effort to identify such exploits. Researchers and academicians have become interested in developing models for these shared tasks. This article summarizes the research works submitted for the shared task on Sentiment Analysis and Homophobia detection of YouTube comments in Code-Mixed Dravidian Languages [9]. Sentiment analysis and identifying homophobic and transphobic comments make up the two subtasks included in this shared task. Below we provide a brief explanation of these subtasks.

Sentiment analysis is a subtask of NLP that uses computational methods to analyze, process, and better understand a user's emotions behind a text or interaction [10]. It sorts the opinions of its users into different groups. It lets organizations learn from a large amount of unstructured data and change their strategies to suit their target market better. Sentiment analysis has a subtask to find subjective opinions or emotional reactions to a given topic [11]. In the last two decades 20 years, both academia and industry have been doing research in this area.

There is an increasing demand for sentiment detection on social media texts which are largely code-mixed for Dravidian languages. Tamil, Malayalam, and Kannada belongs to Dravidian

Shared task on Sentiment Analysis and Homophobia detection of YouTube comments in Code-Mixed Dravidian Languages, FIRE 2022

*You can use this document as the template for preparing your publication. We recommend using the latest version of the ceurart style.

*Corresponding author.

†These authors contributed equally.

✉ kogilavani@kongu.ac.in (K. Shanmugavadivel); mallinishanth72@gmail.com (M. Subramanian); prasanna.mi20@iiitm.ac.in (P. K. Kumaresan); bharathi.raja@insight-centre.org (B. R. Chakravarthi); bharathib@ssn.edu.in (B. Bharathi); subalalitha@gmail.com (S. C. Navaneethakrishnan); sklavyasambath@gmail.com (L. S. Kumar); mandl@uni-hildesheim.de (T. Mandl); rahulponnusamy160032@gmail.com (R. Ponnusamy); vasantpcse2019@citchennai.net (V. Palanikumar); manojbalaji1@gmail.com (M. J. Balaji)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

languages [12, 13]. In the shared task, sentiment analysis aims to figure out how the text makes people feel and put them into predefined groups. This shared subtask includes YouTube comments written in Tamil, Malayalam, and Kannada and labeled as "positive," "negative," "mixed," or "unknown." The Tamil dataset has an extra class called "Non-Tamil" for comments that are not written in Tamil [14]. The researchers are also given test datasets to evaluate the proposed classification models.

The term "LGBTQ+ community" [15] designates a group or community of individuals who identify as lesbian, gay, bisexual, transgender, or queer, including all gender identities and sexual orientations not expressly covered by LGBTQ. The term "homophobia" describes the hostility toward those who identify as homosexual, transgender, or queer. LGBTQ individuals may experience significant psychological stress due to homophobia and transphobia, which will prevent them from participating in typical social activities and could result in severe mental illness. To clear cyberspace, build a friendly and healthy online community, and increase awareness of the unfair treatment of LGBTQ groups, it is crucial to identify and remove homophobia and transphobia as soon as they appear [16]. An effort to spread positivity about the LGBTQIA community by building a Tamil dataset about the community and identifying the offensive and Non-Offensive terminology in the dataset [17][18].

The second subtask of identifying such unpleasant comments has been made to aid in this issue. The datasets for detecting Homophobic and transphobic comments in three languages, Tamil, English, and Malayalam, are presented as a part of this shared sub-task. In addition, a code-mixed dataset with Tamil and English has also been proposed. The datasets have the class labels such as Homophobic, Transphobic, and Non-anti-LGBT+ content.

Participants in the shared task were given access to the training and validation data, complete with labels, and the test data, which did not contain any labels. These participants in the shared task built machine learning and deep learning models for the two subtasks and then submitted their predictions for the labels that should be applied to the test data.

This article presents an overview of the shared task on Sentiment Analysis and Homophobia detection of YouTube comments in Code-Mixed Dravidian Languages. This work discusses the various models submitted to the shared task and the results of the participating teams. The rest of the article is orchestrated as follows: Section 2 describes the shared task. Section 3 discusses about the dataset. Section 4 provides information about task setting. Section 5 summarizes the systems and the methodologies used in each participating team for both the shared sub-tasks and highlights the features of each model. The analysis of the results and findings of the methodologies submitted by the participants are presented in Section 6. Concluding remarks are presented in Section 7.

2. Task Description

The goal of the proposed shared task is to perform Sentiment Analysis and Homophobia detection of social media comments in Code-Mixed Dravidian Languages. Sentiment analysis is the task of identifying subjective opinions or emotional responses about a given topic. Homophobia and Transphobia detection identifies homophobia, transphobia, and non-anti-LGBT+ content from the given corpus.

2.1. Shared Task-A

Shared task-A aims to identify sentiment polarity such as positive, negative, neutral, or mixed emotions of the code-mixed dataset of comments/posts in Tamil-English, Malayalam-English, and Kannada-English collected from social media. The participants were provided training, development, and test datasets in three code-mixed languages (Tamil-English, Malayalam-English, and Kannada-English). The annotations of the datasets were made at the comment/post level. A comment/post may contain more than one sentence, but the average sentence length of the corpus is one. The participants could choose to take part in classifying one or more code-mixed languages. Leader-board results were published for each code-mixed language. Some sample sentences from the datasets and their annotations are provided below.

2.2. Shared Task-B

Shared task B aims to identify homophobia, transphobia, or not from the given corpus. Homophobia and transphobia are toxic languages directed at LGBTQ+ individuals described as hate speech. In this shared task, participants were provided with comments extracted from social media platforms, and they had to predict whether it was homophobic/transphobic in nature. The seed data for this task is the Homophobia/Transphobia Detection dataset [9], a collection of comments from YouTube. This dataset consists of manually annotated comments indicating whether the text is homophobic/transphobic or not. The participants were provided with development, training, and test datasets in English, Malayalam, and Tamil. Some sample sentences from the datasets and their annotations are provided below.

3. Datasets

The corpus in this shared task-A consists of 67,554 social media comments in three different code-mixed languages. There are 40,267 comments in Tamil-English, 19,616 in Malayalam-English, and 7,671 in Kannada-English. The corpus provided in this shared task B consists of 20,150 social media comments in Tamil, English, Malayalam, and English+Tamil. There are 3,977 comments in Tamil, 4,946 in English, 5,193 in Malayalam, and 6,034 in English+Tamil.

Table 1 shows the corpus statistics for task A, and Table 2 represents the corpus details of task B in terms of language. The annotated datasets were divided into training, development, and test sets to contain approximately 80%, 10% and 10% of the total number of comments. The corpus statistics were calculated using the NLTK tool [19]. There are more non-hope speech comments than hope speech [20, 21, 22]. This makes the datasets imbalanced and skewed more towards one class than the other, which the participants had to consider when developing their classification systems.

Table 1Number of comments in each dataset used for **Task A**

	Task - A			
	Train	Dev	Test	Total
Tamil	35656	3962	650	40268
Malayalam	15888	1766	1962	19616
Kannada	6212	691	768	7671

Table 2Number of comments in each dataset used for **Task B**

	Task - B			
	Train	Dev	Test	Total
Tamil	2662	666	649	3977
Malayalam	3114	866	1213	5193
English	3164	792	990	4946
Tamil-English	3861	966	1207	6034

4. Task Setting

4.1. Training Phase

During the training phase, we provided participants with labeled training and development data that they could use to train and validate their models. We released the data for all the languages, and the participants had to decide whether they could participate in developing models for more than one language. The goal of this phase was to provide the participants with sufficient data that they could use to perform cross-validation for their preliminary evaluations and hyper-parameter setting. This ensured that participants were ready for the assessment before releasing the unlabeled test data. 95 participants were registered for the shared task and downloaded the datasets in this phase.

4.2. Testing Phase

During the testing phase, the participants were given test data without labels. Each participating team was allowed as many submissions as possible, from which the best result was considered for preparing the leaderboard ranking. The submission outputs were compared with the gold standard labels. The ranking list was based on the best performance measured on the macro F1-scores. For the shared task-A phase, 8,11, and 13 participants submitted their results for Tamil-English, Malayalam-English, and Kannada-English, respectively. For the shared task-B stage, there were 8,8,9, and 8 participants submitted their results for Tamil, English, Malayalam, and English+Tamil, respectively.

5. Systems

The systems used by the participants include a broad spectrum of machine learning algorithms, deep learning algorithms, and transformer-based models. Machine learning algorithms have Lo-

gistic Regression, Passive Aggressive, Support Vector Machine, Naive Bayes, Gradient Boosting, Stacking Ensemble, Gradient Boosting Classifier, Random forest, and Voting Ensemble models [23][24]. Count vectorizer and TF-IDF have been used for feature representation. Bidirectional LSTM, Multi-Layered Perceptron, and fastText+LightGBM are the deep learning algorithms opted for by the participants. Transformer-based models like MPNet, SBERT, XLM RoBERTa, Indic BERT, and LaBSE model have been experimented with for the given task [25][26][27][28].

While analyzing the performance of the top three systems of Task A for the Tamil-English dataset, Bidirectional LSTM has performed well when compared to that of the machine learning, and transformer paradigms [29]. For the Malayalam- English and Kannada-English datasets, the transformer models, mBERT and XLMRoBERTa, have better performance compared to the other models [30].

On the other hand, while analyzing the top three systems of Task B, for Tamil-English, LSTM has performed well, followed by XLMRoBERTa [31][32][33]. For Malayalam-English, again deep learning model has performed better when compared to that of the other models followed by machine learning models and FastText [34]. For English-only texts, fastText, XLMRoBERTa, and Indic BERT have performed well.

6. Results and Discussion

The rank list obtained by the participant’s language wise for Task A is represented in Tables 3, 4, and 5. Table 3 shows the rank list of task A in Tamil language, Table 4 presents the rank list of task A in Malayalam language, and Table 5 depicts the rank list for task A in English track. The rank list obtained by the participant’s language wise for task B is presented in Tables 6, 7, 8, and 9. Table 6 shows the rank list of task B in Tamil language, Table 7 presents the rank list of task B in English language, Table 8 shows the rank list of task B in Malayalam language and Table 9 depicts the rank list for task B in Tamil-English track.

Table 3
Rank list for **Task A: Tamil** track

Team Name	Precision	Recall	F1-score	Rank
SRMNLP	0.340	0.330	0.270	1
BharathNLP	0.190	0.220	0.190	2
bilstm	0.220	0.190	0.190	3
SSN-CSE	0.220	0.260	0.170	4
Sentiment	0.240	0.220	0.170	5
MUCS	0.240	0.190	0.160	6
Fnet	0.150	0.130	0.130	7
JPMCAI	0.020	0.160	0.020	8

It can be observed that the performance of the top-ranked systems, the precision, and recall values have not seen high scores. This indicates the need for much more robust pre-processing methods and models to interpret and classify the code-mixed data in Dravidian languages. While looking at models used for the Tamil language for Task A and task B, the transformer models have yielded low precision scores compared to deep learning models like LSTM. The

Table 4Rank list for **Task A: Malayalam** track

Team Name	Precision	Recall	F1-score	Rank
IRLAB	0.670	0.670	0.660	1
Fnet	0.660	0.620	0.640	2
Sentiment	0.620	0.630	0.630	3
MUCS	0.610	0.610	0.610	4
NITK	0.600	0.600	0.600	5
SRMNLP	0.610	0.550	0.570	6
lone_warrior	0.520	0.590	0.520	7
bilstm	0.490	0.580	0.500	8
BharathNLP	0.160	0.270	0.200	9
JPMCAI	0.340	0.200	0.140	10
SSN-CSE	0.090	0.140	0.110	11

Table 5Rank list for **Task A: Kannada** track

Team Name	Precision	Recall	F1-score	Rank
IRLAB	0.560	0.560	0.550	1
Sentiment	0.520	0.500	0.510	2
lone_warrior	0.470	0.510	0.480	3
NITK	0.480	0.500	0.480	4
Fnet	0.500	0.490	0.480	5
AI Defenders	0.490	0.480	0.480	6
SRMNLP	0.540	0.440	0.460	7
JPMCAI	0.550	0.430	0.450	8
MUCS	0.470	0.460	0.440	9
bilstm	0.480	0.500	0.430	10
QWERTY	0.460	0.350	0.350	11
BharataNLP	0.290	0.330	0.300	12
SSN-CSE	0.120	0.170	0.110	13

Table 6Rank list for **Task B: Tamil** track

Team Name	F1-score	Rank
mucs [33]	0.366	1
fnet [25]	0.327	2
CITK	0.290	3
IRLab@IITBHU [30]	0.289	4
qwerty [23]	0.234	5
SSN-CSE-2022 [24]	0.234	5
BharataNLP [28]	0.234	5
nlpzip [26]	0.228	6

transformer models generally outperform deep learning algorithms like LSTM due to its multi-headed self-attention mechanism through which it tries to understand the code mixed data

Table 7Rank list for **Task B: English** track

Team Name	F1-score	Rank
BharataNLP [28]	0.493	1
fnet [25]	0.486	2
nlpzip [26]	0.462	3
mucs [33]	0.374	4
IRLab@IITBHU [30]	0.337	5
qwerty [23]	0.332	6
SSN-CSE-2022 [24]	0.322	7
kongu.eng-21MSR002 [27]	0.319	8

Table 8Rank list for **Task B: Malayalam** track

Team Name	F1-score	Rank
Nitk [34]	0.974	1
qwerty [23]	0.943	2
BharataNLP [28]	0.942	3
CITK	0.860	4
mucs [33]	0.750	5
fnet [25]	0.696	6
nlpzip [26]	0.542	7
IRLab@IITBHU [30]	0.427	8
SSN-CSE-2022 [24]	0.296	9

Table 9Rank list for **Task B: Tamil-English** track

Team Name	F1-score	Rank
mucs [33]	0.580	1
fnet [25]	0.555	2
CITK	0.477	3
nlpzip [26]	0.393	4
qwerty [23]	0.344	5
IRLab@IITBHU [30]	0.333	6
SSN-CSE-2022 [24]	0.316	7
BharataNLP [28]	0.316	8

better than the LSTM-based models [17], but surprisingly the transformer-based models have not given good results.

This may be alleviated if the model is fine-tuned with more number of code-mixed data and complimented with robust pre-processing strategies like transliteration, translation, spell checking, etc., to handle the code-mixed data. A detailed error analysis of the test set will give a more precise idea of the type of scenarios and the pr-processing strategy to be chosen. Furthermore, it can also be observed that the values spitted by the top models are also not that high as the top-ranked team has got only a precision score of 0.34. Since the top-ranked

models are LSTM-based models, the values can be pushed up by choosing a suitable embedding mechanism instead of relying on TF-IDF and count vectorizer for feature representations.

While looking at Malayalam and Kannada, the transformer models have outperformed the deep learning for Task A, which reflects the current state of the art, and again in Task B, for Malayalam, the transformer-based models have not outperformed the deep learning models. This can likewise be dealt with as mentioned earlier for Task A.

7. Conclusion

To summarize, this shared task has two subtasks, sentiment analysis and homophobia/transphobia detection. The shared task aims to promote the research work on Dravidian Languages. Sentiment analysis aims at classifying the text that makes people feel and puts them into predefined groups. The second sub-task focuses on detecting hateful comments against the LGBTQ community. There were 13 submissions out of which 10 submissions focused on detecting comments against LGBTQ. The participants have developed various models based on machine learning and deep learning. The submissions were ranked based on the performance of the models. When compared to the performance of the other systems, it was found that the transformer models exhibited significantly higher levels of performance.

References

- [1] A. A. Siegel, online hate speech v2, 2019. URL: https://alexandra-siegel.com/wp-content/uploads/2019/08/Siegel_Online_Hate_Speech_v2.pdf.
- [2] S. Thavareesan, S. Mahesan, Sentiment lexicon expansion using word2vec and fasttext for sentiment prediction in tamil texts, in: 2020 Moratuwa Engineering Research Conference (MERCon), IEEE, 2020, pp. 272–276.
- [3] S. Thavareesan, S. Mahesan, Sentiment analysis in tamil texts: A study on machine learning techniques and feature representation, in: 2019 14th Conference on Industrial and Information Systems (ICIIS), IEEE, 2019, pp. 320–325.
- [4] S. Thavareesan, S. Mahesan, Word embedding-based part of speech tagging in tamil texts, in: 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), IEEE, 2020, pp. 478–482.
- [5] R. Priyadharshini, B. R. Chakravarthi, C. N. Subalalitha, T. Durairaj, M. Subramanian, K. Shanmugavadivel, S. U. Hegde, P. K. Kumaresan, Findings of the shared task on Abusive Comment Detection in Tamil, in: Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages. Association for Computational Linguistics, 2022.
- [6] B. R. Chakravarthi, R. Priyadharshini, R. Ponnusamy, P. K. Kumaresan, K. Sampath, D. Thenmozhi, S. Thangasamy, R. Nallathambi, J. P. McCrae, Dataset for identification of homophobia and transphobia in multilingual youtube comments, arXiv preprint arXiv:2109.00227 (2021).
- [7] B. R. Chakravarthi, V. Muralidaran, R. Priyadharshini, C. N. Subalalitha, J. P. McCrae, M. Á. García, S. M. Jiménez-Zafra, R. Valencia-García, P. Kumaresan, R. Ponnusamy, et al.,

Overview of the shared task on hope speech detection for equality, diversity, and inclusion, in: Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, 2022, pp. 378–388.

- [8] B. R. Chakravarthi, V. Muralidaran, Findings of the shared task on hope speech detection for equality, diversity, and inclusion, in: Proceedings of the first workshop on language technology for equality, diversity and inclusion, 2021, pp. 61–72.
- [9] K. Shanmugavadivel, M. Subramanian, P. K. Kumaresan, B. R. Chakravarthi, B. Bharathi, C. N. Subalalitha, S. K. Lavanya, T. Mandl, R. Ponnusamy, V. Palanikumar, B. Manoj J, Overview of the Shared Task on Sentiment Analysis and Homophobia Detection of YouTube Comments in Code-Mixed Dravidian Languages, in: Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation, CEUR, 2022.
- [10] A. Sampath, T. Durairaj, B. R. Chakravarthi, R. Priyadharshini, C. N. Subalalitha, K. Shanmugavadivel, S. Thavareesan, S. Thangasamy, P. Krishnamurthy, A. Hande, et al., Findings of the shared task on Emotion Analysis in Tamil, in: Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages, 2022, pp. 279–285.
- [11] K. Shanmugavadivel, S. H. Sampath, P. Nandhakumar, P. Mahalingam, M. Subramanian, P. K. Kumaresan, R. Priyadharshini, An analysis of machine learning models for sentiment analysis of Tamil code-mixed data, *Computer Speech & Language* (2022) 101407.
- [12] R. Anita, C. Subalalitha, An approach to cluster tamil literatures using discourse connectives, in: 2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP), IEEE, 2019, pp. 1–4.
- [13] C. Subalalitha, E. Poovammal, Automatic bilingual dictionary construction for tirukural, *Applied Artificial Intelligence* 32 (2018) 558–567.
- [14] B. R. Chakravarthi, V. Muralidaran, R. Priyadharshini, J. P. McCrae, Corpus creation for sentiment analysis in code-mixed Tamil-English text, in: Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL), European Language Resources association, Marseille, France, 2020, pp. 202–210. URL: <https://aclanthology.org/2020.sltu-1.28>.
- [15] B. R. Chakravarthi, R. Priyadharshini, T. Durairaj, J. P. McCrae, P. Buitelaar, P. Kumaresan, R. Ponnusamy, 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, 2022, pp. 369–377.
- [16] N. Moyano, M. del Mar Sanchez-Fuentes, Homophobic bullying at schools: A systematic review of research, prevalence, school-related predictors and consequences, *Aggression and violent behavior* 53 (2020) 101441.
- [17] S. K. Lavanya, C. N. Subalalitha, Building Tamil Text Dataset on LGBTQIA and Offensive Language Detection using Multilingual BERT, in: 2022 International Conference on Inventive Computation Technologies (ICICT), IEEE, 2022, pp. 489–496.
- [18] M. Subramanian, R. Ponnusamy, S. Benhur, K. Shanmugavadivel, A. Ganesan, D. Ravi, G. K. Shanmugasundaram, R. Priyadharshini, B. R. Chakravarthi, Offensive language detection in Tamil YouTube comments by adapters and cross-domain knowledge transfer, *Computer Speech & Language* 76 (2022) 101404.
- [19] S. Bird, Nltk: the natural language toolkit, in: Proceedings of the COLING/ACL 2006

Interactive Presentation Sessions, 2006, pp. 69–72.

- [20] B. R. Chakravarthi, Hope speech detection in youtube comments, *Social Network Analysis and Mining* 12 (2022) 1–19.
- [21] B. R. Chakravarthi, Multilingual hope speech detection in english and dravidian languages, *International Journal of Data Science and Analytics* 14 (2022) 389–406.
- [22] B. R. Chakravarthi, A. Hande, R. Ponnusamy, P. K. Kumaresan, R. Priyadharshini, 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 (2022) 100119.
- [23] S. Saumya, V. Jha, S. Biradar, Sentiment and Homophobia Detection on YouTube using Ensemble Machine Learning Techniques, in: *Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation*, CEUR, 2022.
- [24] J. Varsha, B. Bharathi, A. Meenakshi, Sentiment Analysis and Homophobia detection of YouTube comments in Code-Mixed Dravidian Languages using machine learning and transformer models, in: *Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation*, CEUR, 2022.
- [25] F. Nilsson, S. S. Al-Azzawi, G. Kovács, Leveraging Sentiment Data for the Detection of Homophobic/Transphobic Content in a Multi-Task, Multi-Lingual Setting Using Transformers, in: *Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation*, CEUR, 2022.
- [26] S. Venkatesan, S. Donepudi, P. P, T. Durairaj, Homophobia and Transphobia Detection of Youtube Comments in Code-Mixed Dravidian Languages using Deep learning, in: *Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation*, CEUR, 2022.
- [27] D. Manikandan, M. Subramanian, K. Shanmugavadeivel, A System For Detecting Abusive Contents Against LGBT Community Using Deep Learning Based Transformer Models, in: *Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation*, CEUR, 2022.
- [28] M. B. J, C. Hs, A Study on Sentimental Analysis, Homophobia-Transphobia Detection for Dravidian Languages, in: *Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation*, CEUR, 2022.
- [29] S. K. Lavanya, F. N. Muhammad, Sentiment Analysis of YouTube comments in Dravidian Code-Mixed Language using Deep Neural Network, in: *Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation*, CEUR, 2022.
- [30] S. Chanda, A. Mishra, S. Pal, Sentiment Analysis and Homophobia detection of Code-Mixed Dravidian Languages leveraging pre-trained model and word-level language tag, in: *Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation*, CEUR, 2022.
- [31] S. K. Lavanya, S. Sivaprasath, Homophobia, Transphobia Detection in Tamil, Malayalam, English Languages using Logistic Regression and Code-Mixed Data using AWD_LSTM, in: *Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation*, CEUR, 2022.
- [32] S. K. Lavanya, A. A. Samuel, A Sequential DNN for Sentiment Analysis of Dravidian Code-Mixed Language Comments on YouTube, in: *Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation*, CEUR, 2022.
- [33] A. Hegde, H. Shashirekha, Leveraging Dynamic Meta Embedding for Sentiment Analysis and Detection of Homophobic/Transphobic Content in Code-mixed Dravidian Languages, in: *Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation*, CEUR, 2022.

- [34] S. Ugursandi, A. Kumar M, Sentiment Analysis and Homophobia detection of YouTube comments, in: Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation, CEUR, 2022.
- [35] R. Sakuntharaj, S. Mahesan, A novel hybrid approach to detect and correct spelling in tamil text, in: 2016 IEEE international conference on information and automation for sustainability (ICIAFS), IEEE, 2016, pp. 1–6.
- [36] R. Sakuntharaj, S. Mahesan, Use of a novel hash-table for speeding-up suggestions for misspelt tamil words, in: 2017 IEEE international conference on industrial and information systems (ICIIS), IEEE, 2017, pp. 1–5.
- [37] H. Visuwalingam, R. Sakuntharaj, R. G. Ragel, Part of speech tagging for tamil language using deep learning, in: 2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS), IEEE, 2021, pp. 157–161.