MUCS@LT-EDI2023: Homophobic/Transphobic Content Detection in Social Media Text using mBERT

Asha Hegde[a], **Kavya G**[b], **Sharal Coelho**[c],
Hosahalli Lakshmaiah Shashirekha[d]
Department of Computer Science, Mangalore University, Mangalore, India
{[a]hegdekasha,[b]kavyamujk,[c]sharalmucs}@gmail.com
[e]hlsrekha@mangaloreuniversity.ac.in

Abstract

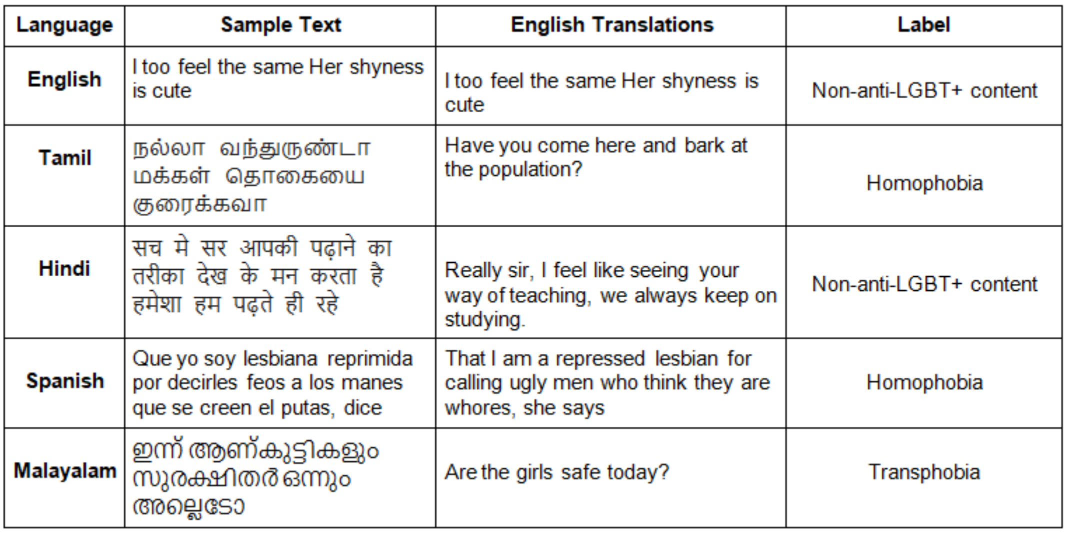
Homophobic/Transphobic (H/T) content includes hate speech, discrimination text, and abusive comments against Lesbian, Gay, Bisexual, Transgender, and Queer (LGBTQ) individuals. With the increase in user generated text in social media, there has been an increase in H/T content also. Further, most of the text data on social media is code-mixed and this poses challenges for efficient analysis and detection of H/T content on social media. The complex nature of code-mixed text necessitates the development of advanced tools and techniques to effectively tackle this issue on social media platforms. Hence, in this paper, we - team MUCS, describe the transformer based models submitted to ”Homophobia/Transphobia Detection in social media comments” shared task in Language Technology for Equality, Diversity and Inclusion (LT-EDI) at Recent Advances in Natural Language Processing (RANLP)-2023. The proposed methodology makes use of oversampling technique to handle data imbalance in the given Train set and this oversampled data is used to fine-tune the Transfer Learning (TL) based Multilingual Bidirectional Encoder Representations from Transformers (mBERT) models. These models obtained weighted F1 scores of 0.91, 0.91, 0.95, 0.94, and 0.81 securing 11th, 5th, 3rd, 3rd, and 7th ranks for English, Tamil, Malayalam, Spanish, and Hindi languages respectively in Task A and weighted F1 scores of 0.14, 0.82, and 0.85 securing 8th, 2nd, and 2nd ranks for English, Tamil, and Malayalam languages respectively in Task B.

1 Introduction

Social media platforms provide a means for users to express their views, ideas, reviews, comments, opinions, and emotions, freely and instantaneously without any barriers of the language and content. This has given raise to the creation and sharing of useful content as well as unhealthy posts, such as offensive, abusive, and hatred content, targeting a person, a group, or a community (Hegde et al., 2021; Balouchzahi et al., 2021b). H/T content is one such content that expresses the hatredness towards LGBTQ community on their sexual orientation or gender identity (Chakravarthi et al., 2021). LGBTQ individuals face various forms of textual violence and discrimination such as, hate speech, cyberbullying, exculsion and isolation, online shaming, and misgendering in online environment or on social media platforms. They also become targets of threats and abuse, leading to significant mental health issues (Hegde et al., 2022b). Hence, identifying and removing H/T content on social media platforms is a crucial aspect in order to promote equality, diversity, and inclusion in the society. By implementing these measures, it is possible to create a safer online environment for the LGBTQ community and support their well-being and mental health (Mandl et al., 2020).

Identifying H/T content in social media text poses challenges due to the complex nature of code-mixed text prevalent on these platforms (Chakravarthi, 2023; Hegde and Shashirekha, 2022). Social media text often includes the mixing of local or regional languages such as Hindi, Malayalam, Tamil, etc., with English, at sub-word, word and sentence level leading to code-mixed content. (Jose et al., 2020; Hegde et al., 2022a; Balouchzahi et al., 2022). This code-mixing makes the identification and analysis of H/T content more difficult, as traditional language processing models may fail to accurately interpret and classify such mixed-language texts. To effectively address this challenge, Natural Language Processing (NLP) techniques need to be explored for code-mixed text, taking into consideration the linguistic nuances and variations in different languages. By developing robust algorithms and models to handle code-mixed H/T text, it is possible to identify and combat H/T

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion, pages 287–294, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_044



Language	Sample Text	English Translations	Label
English	I too feel the same Her shyness is cute	I too feel the same Her shyness is cute	Non-anti-LGBT+ content
Tamil	நல்லா வந்துருண்டா மக்கள் தொகையை குரைக்கவா	Have you come here and bark at the population?	Homophobia
Hindi	सच मे सर आपकी पढ़ाने का तरीका देख के मन करता है हमेशा हम पढ़ते ही रहे	Really sir, I feel like seeing your way of teaching, we always keep on studying.	Non-anti-LGBT+ content
Spanish	Que yo soy lesbiana reprimida por decirles feos a los manes que se creen el putas, dice	That I am a repressed lesbian for calling ugly men who think they are whores, she says	Homophobia
Malayalam	ഇന്ന് ആണ്കുട്ടികളും സുരക്ഷിതർ ഒന്നും അല്ലെടോ	Are the girls safe today?	Transphobia

Table 1: Sample text and their English Translations for Task A

content in a better way ensuring a safer and more inclusive online environment for all users.

To address the challenges of H/T content identification in social media text, in this paper, we - team MUCS, describe the models submitted to ”Homophobia/Transphobia Detection in Social media Comments” shared task[1] at RANLP-2023[2]. The shared task consists of two subtasks: i) Task A - a comment-level polarity classification task to identify H/T content in English, Tamil, Hindi, Malayalam, and Spanish languages, with 3 labels (Non-anti-LGBT+, Homophobia, and Transphobia) and ii) Task B - to identify H/T content in English, Tamil, and Malayalam texts, with 7 labels (None-of-the-above, Hope-speech, Counter-speech, Homophobic-derogation, Homophobic-Threatening, Transphobic-derogation, and Transphobic-derogation) (Chakravarthi et al., 2023). Sample comments from the datasets provided by the organizers of the shared task for Task A and Task B are shown in Table 1 and Table 2 respectively. As there are more than two classes in the dataset, the shared task is modeled as a multi-class text classification problem. The proposed methodology includes oversampling the Training set as the given data is imbalanced and fine-tuning the BERT models for both Task A and Task B.

The rest of the paper is structured as follows: Section 2 contains related works and Section 3 explains the methodology. Section 4 describes the experiments and results and the paper concludes in Section 5 with future work.

2 Related work

Several researchers have explored H/T content detection, offensive language identification and hate speech and offensive content detection in various languages and few of the relevant ones are described below:

Hegde and Shashirekha (2022) describe the learning models to perform Sentiment Analysis (SA) and H/T content detection in code-mixed Dravidian languages as Task A (Malayalam and Kannada) and Task B (Tamil and romanized Tamil (Tamil-English)) respectively. Using conventional preprocessing of converting emojis to text and removal of digits and stopwords, these models make use of Dynamic Meta Embedding (DME) to train Long Short Term Memory (LSTM) model for SA and H/T content identification in code-mixed Dravidian languages. These models obtained macro F1 scores of 0.61 and 0.44 for Malayalam and Kannada languages respectively in Task A and 0.58 and 0.74 for Tamil and Tamil-English texts respectively in Task B. Singh and Motlicek (2022) presented TL approach for fine-tuning Cross Lingual Language Models Robustly Optimized BERT (XLM ROBERTA) model in Zero-Shot learning framework for the detection of H/T contents in English and Tamil-English and obtained macro F1 scores of 0.89 and 0.85 for English and Tamil-English texts respectively.

Ashraf et al. (2022) proposed the Machine Learning (ML) models (Support Vector Machines (SVM), Random Forest (RF), Passive Aggressive

[1] https://codalab.lisn.upsaclay.fr/competitions/11077/
[2] http://ranlp.org/ranlp2023/

Language	Sample Text	English Translation	Label
English	You are so pretty, you really look like a girl	You are so pretty, you really look like a girl	None
	Pro please Boxing prank part2	Pro please Boxing prank part2	Hope-Speech
	Best movie and people not understand relationship feeling I miss my life	Best movie and people not understand relationship feeling I miss my life	Hope-Speech
Tamil	இந்திய ஒன்றியம் நூறு வருஷத்துக்கு பின் தங்கியுள்ளோம் இது இயற்கை அனைத்தது உயிரினங்களிடம் உள்ளது	Union of India is behind a hundred years and it is all about nature	Counter-speech
	இன விருத்தி எப்படி செய்வது என்பதையும் அந்த நீதிபதி சொல்லியருக்க வேண்டும்	That judge should also have told how to do ethnic development	Homophobic-derogation
Malayalam	കുണ്ടൻ polayadi മക്കൾ അണ്ടി ചെത്തി കളയണം	Kundan polayadi children should be undressed	Homophobic-Threatening
	ഒന്ന് പോടാ ശിഖണ്ഡി ഒമ്പതുകളെ ത്ഫൂ	Shikhandi tfoo nines if not one	Transphobic-Threatening
	പിടിച്ച് കല്ലെറിഞ്ഞു കൊല്ലലാണ് ഇസ്ലാം മതം ഇവർക്ക് വിധിച്ചിട്ടുള്ളത്	Islam has condemned them to be caught and stoned to death	Transphobic-Threatening

Table 2: Sample texts and their English Translations for Task B

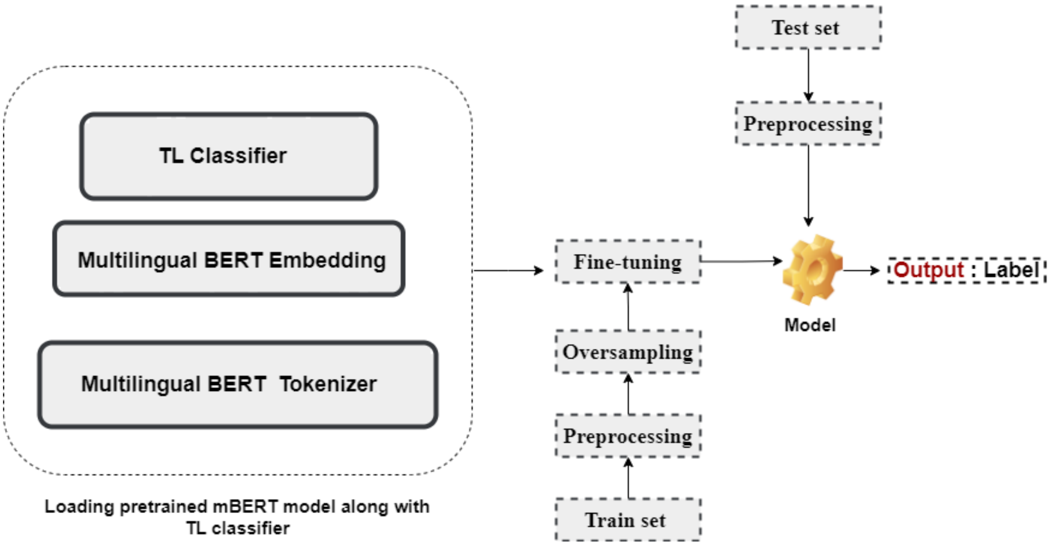

Test set
Preprocessing
TL Classifier
Multilingual BERT Embedding
Multilingual BERT Tokenizer
Loading pretrained mBERT model along with TL classifier
Fine-tuning
Model
Output : Label
Oversampling
Preprocessing
Train set

Figure 1: The framework of the proposed model

Classifier (PA), Gaussian Naive Bayes (GNB), Multi-Layer Perceptron (MLP)) to detect H/T content in three languages (English, Tamil and Tamil-English) trained with Term Frequency-Inverse Document Frequency (TF-IDF) of word bigrams. Among these models, SVM outperformed all other classifiers with weighted F1 scores of 0.91, 0.92, 0.88 for English, Tamil and Tamil-English languages respectively.

Two distinct models: COOLI-Ensemble - a Voting Classifier with three estimators (Multi Layer Perceptron (MLP), eXtreme Gradient Boosting (XGB) and Logistic Regression (LR)) and COOLI-Keras - a Keras dense neural network architecture model, described by Balouchzahi et al. (2021a) aims to classify code-mixed texts in Kannada-English, Malayalam-English, and Tamil-English language pairs into six predefined categories and Malayalam-English language pair into five categories for identifying offensive content. Character and word sequences extracted are vectorized using CountVectorizer[3] and the relevant fea-

[3] https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html

	Train set			Development set		
Language	**Non-anti-LGBT+content**	**Homophobia**	**Transphobia**	**Non-anti-LGBT+content**	**Homophobia**	**Transphobia**
English	2,978	179	7	748	42	2
Tamil	2,064	453	145	507	118	41
Hindi	2,423	45	92	305	2	13
Malayalam	2,468	476	170	937	197	79
Spanish	450	200	200	150	43	43

Table 3: : Classwise distribution of labels in the dataset for Task A

tures are selected using feature selection algorithms (Chi-Square test, Mutual Information (MI), and F test). COOLI-Ensemble model performed better and obtained weighted F1 scores of 0.97, 0.75, and 0.69 for Malayalam-English, Tamil-English and Kannada-English language pairs respectively. Balouchzahi and Shashirekha (2020) proposed three distinct models: ensemble of ML classifiers (Random Forest Classifier, LR, and Support Vector Classifier (SVC)) with hard voting, TL classifier using Universal Language Model Fine-tuning (ULMFiT) model, and ML-TL - an ensemble of ML and TL models with hard voting, to detect hate speech and offensive content in English, German and Hindi languages. Among all the models, ensemble of ML models exhibited better macro F1 score of 0.5044 for German language and ML-TL model obtained better macro F1 score of 0.5182 for Hindi language.

From the literature review, it is clear that identification of H/T content in low-resource languages like, Tamil, Malayalam, and Hindi are rarely explored. Hence, there is lot of scope in this direction for further research.

3 Methodology

The proposed methodology includes preprocessing, resampling, and classifiers construction using mBERT models to address the challenges of Task A and Task B of the shared task. The framework of the proposed methodology is visualized in Figure 1 and the steps involved in the methodology are given below:

3.1 Preprocessing

Preprocessing plays a crucial role in preparing text for further processing. Using a preprocessing pipeline, URLs, punctuation, digits, unrelated characters, and stopwords (English, Tamil, Hindi and Spanish languages) are removed as these elements do not contribute to the classification task. Additionally, as emojis - a visual representation of emotions, objects, and symbols carry valuable information, they are converted into corresponding English text allowing their content to be utilized along with the textual data.

3.2 Resampling

Data imbalance refers to the situation where the number of instances belonging to different classes vary significantly (Srinivasan and Subalalitha, 2021). Because of this, learning models become biased towards majority class exhibiting poor performance for minority class. This biased training could be resolved to some extent using resampling techniques. Resampling is a technique commonly used to address data imbalance in classification tasks. There are two types of resampling: i) Oversampling - duplicates samples in the minority class and adds them to the Train set until it get balanced and ii) Undersampling - deletes the samples in the majority class.

The proposed work utilizes oversampling the Train set for both Task A and Task B and the description of the parameters used in oversampling is given below:

- replace = True - indicates whether sampling should be done with replacement. When set to True, it allows the same sample to be selected more than once
- n_samples = n_samples - specifies the number of samples in the majority class for the resampling process
- random_state = None - determines the random seed used for sampling. If set to None, the random seed is not fixed and will vary for each resampling

This technique creates a balanced dataset which is further used for training purposes, ensuring that the model receives an equal representation of both

Label	Train set			Development set		
	English	Tamil	Malayalam	English	Tamil	Malayalam
None-of-the -above	2,240	1,634	2,247	553	395	848
Hope-Speech	436	218	69	111	52	29
Counter-Speech	302	212	152	84	60	60
Homophobic-derogation	162	416	419	41	107	181
Homophobic-Threatening	12	37	57	1	11	16
Transphobic-derogation	6	111	163	2	31	75
Transphobic-Threatening	1	34	7	-	10	4

Table 4: : Classwise distribution of labels in the dataset for Task B

Hyperparameters	**Values**
Layers	6
Dimension	768
Attention heads	12
Learning Rate	2e-5
Batch Size	32
Maximum Sequence Length	128
Dropout	0.3

Table 5: Hyperparameters and their values used in mDistilBERT model

the classes potentially addressing issues related to class imbalance.

3.3 Model construction

TL involves training a model on one task and utilizing the learned knowledge to improve the performance on a similar task. Instead of creating a model from the scratch, the pre-trained knowledge obtained from the source task is transferred to accelerate learning and enhance the performance of the target task. This approach leverages the generalizable features and representations learned from a large dataset in the source task, allowing for efficient adaptation to the target task even for potentially less amount of labeled data (Fazlourrahman et al., 2022; Hegde and Lakshmaiah, 2022). mBERT is a variant of the BERT model that has been trained on multilingual data using the pretraining strategy similar to that used for pretraining BERT, viz. Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) (Pires et al., 2019). This model leverages the power of transformer architecture to learn contextualized representation of words in multiple languages.

mBERT model works in two stages: pretraining and fine-tuning. During pretraining, the model is trained on a large corpus of text from different languages. It learns to predict the next word in the sentences using the MLM objective. Further, it learns to predict if two sentences are consecutive in a document using the Next Sentence Prediction (NSP) objective. This process enables the model to capture both word-level and sentence-level contextual information (Yasaswini et al., 2021). After pretraining, the model is fine-tuned for specific downstream tasks, such as SA, hate speech detection, offensive language detection, and opinion mining. This involves training the model on task-specific labeled data, for tasks such as sentiment analysis, hate speech detection, or named entity recognition. During both pretraining and fine-tuning, the model utilizes attention mechanisms to process the input text. It considers the context of each word by attending to its surrounding words, capturing long-range dependencies effectively. Thus, mBERT model is designed to provide a powerful and flexible framework for multilingual NLP tasks such as SA, text categorization, named entity recognition, and language identification, leveraging its pretrained knowledge and ability to handle code-mixed text effectively with the multilingual support (Chen and Kong, 2021).

4 Experiments and Results

Statistics of the dataset provided by the organizers of the shared task for the identification of H/T content in social media text for Task A and Task B are shown in Tables 3 and 4 respectively (Chakravarthi et al., 2022). From the tables, it is clear that the datasets provided by the organizers are highly imbalanced. To overcome this, oversampling is carried out for both the tasks using oversampling methods provided by the sklearn library[4].

bert-base-multilingual-cased[5] - a mBERT model from the huggingface repository is used to extract

[4] https://scikit-learn.org/stable/modules/generated/sklearn.utils.resample.html

[5] https://huggingface.co/bert-base-multilingual-cased

Language	Development set		Test set	
	Before Oversampling	After Oversampling	Before Oversampling	After Oversampling
Task A				
English	0.82	0.93	0.81	**0.91**
Tamil	0.69	0.85	0.69	**0.91**
Malayalam	0.79	0.95	0.76	**0.95**
Hindi	0.81	0.93	0.81	**0.94**
Spanish	0.83	0.84	0.80	**0.81**
Task B				
English	0.15	0.15	0.13	0.14
Tamil	0.68	0.83	0.67	**0.82**
Malayalam	0.66	0.85	0.60	**0.85**

Table 6: Performance of the proposed models before and after oversampling for both Task A and Task B on Development and Test set

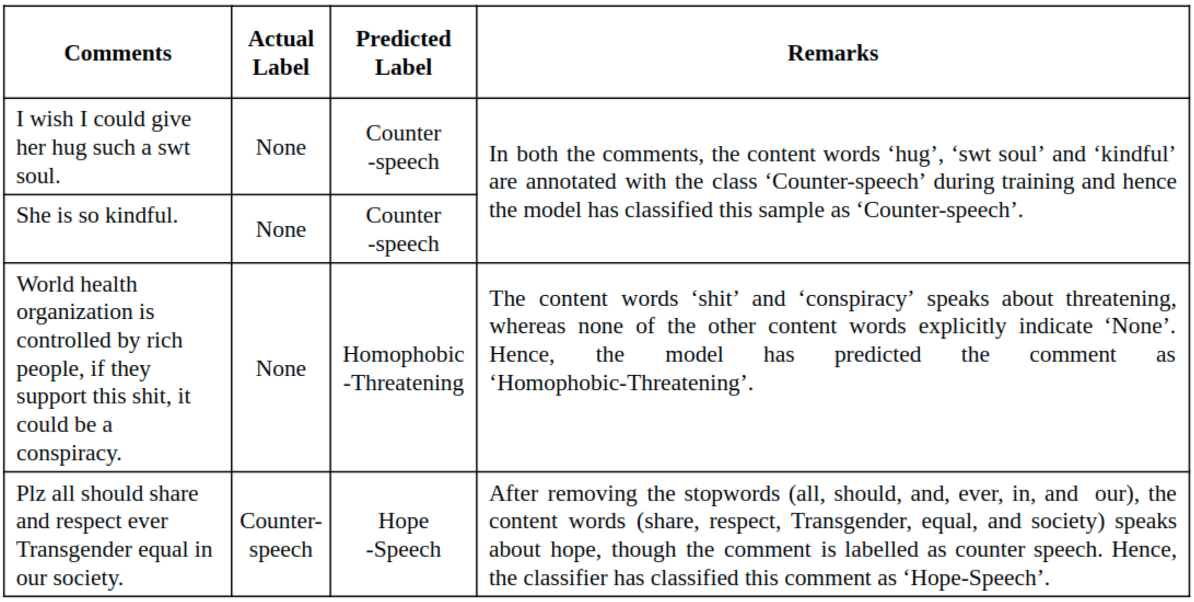
Language	Comments	Actual Label	Predicted Label	Remarks
English	Stefanie Suhashini This is much worse than male commenters who talk cheap and call them prostitutes and insult them. Women are not helping them at all. This needs to change.	None	Homophobia	After removing stopwords (This, is, much, than, etc.) the content words, ‘worse’, ‘cheap’, ‘prostitutes’ ‘insult’ are associated with homophobia class and hence, the model has classified this comment as ‘Homophobia’.
	I accept that you are a lesbian. but the body language and the way you given answers to the question is like a rowdi. Change your attitude it might help you in future.	None	Transphobia	The content words ‘lesbian’ and ‘body language’ are associated with Transphobia class and hence the model has labeled this comment as ‘Transphobia’ as it fails to capture the rest of the information that indicates the class ‘None’.
Hindi	Kya gay.	None	Homophobia	The word ‘gay’ speaks about homophobia and hence this comment is classified as ‘Homophobia’.
	मोन्त्री रोय कि तो बात अलग ह		None	Transphobia	The word ‘मोन्त्री’ is present in the comments belonging to ‘Transphobia‘ class during training and hence the model has predicted the label of this comment as ‘Transphobia‘.

Table 7: Samples of misclassification in Task A for English and Hindi language datasets

the feature vectors. After loading the pretrained mBERT model with its default parameter values, the model is frozen to prevent further updates to its weights. ClassificationModel[6] - a transformer-based classifier is employed to make predictions and the hyperparameters and their values used in the model are shown in Table 5. The hyperparameters which are not mentioned in Table 5 are used with their default values.

The models are evaluated based on weighted F1 scores by incorporating class weights. Performance of the proposed models before and after oversampling for both Task A and Task B on Development and Test sets are reported in Table 6. The proposed models obtained weighted F1 scores of 0.91, 0.91, 0.95, 0.94, and 0.81 securing 11th, 5th, 3rd, 3rd, and 7th ranks for English, Tamil, Malayalam, Hindi, and Spanish languages respectively in Task A and weighted F1 scores of 0.14, 0.82, and 0.85 securing 8th, 2nd, and 2nd ranks for English, Tamil, and Malayalam languages respectively in Task B. From the table, it is clear that the mBERT models with oversampling has exhibited comparatively better weighted F1 scores over the mBERT models without oversampling. Though oversampling technique is used to resolve the data imbalance issues, the proposed methodology has still exhibited low weighted F1 score for English text. As the oversampling technique increases the number of instances in the minority classes by duplicating the samples

[6] 10https://simpletransformers.ai/docs/classificationmodels/



Comments	Actual Label	Predicted Label	Remarks
I wish I could give her hug such a swt soul.	None	Counter -speech	In both the comments, the content words 'hug', 'swt soul' and 'kindful' are annotated with the class 'Counter-speech' during training and hence the model has classified this sample as 'Counter-speech'.
She is so kindful.	None	Counter -speech	
World health organization is controlled by rich people, if they support this shit, it could be a conspiracy.	None	Homophobic -Threatening	The content words 'shit' and 'conspiracy' speaks about threatening, whereas none of the other content words explicitly indicate 'None'. Hence, the model has predicted the comment as 'Homophobic-Threatening'.
Plz all should share and respect ever Transgender equal in our society.	Counter-speech	Hope -Speech	After removing the stopwords (all, should, and, ever, in, and our), the content words (share, respect, Transgender, equal, and society) speaks about hope, though the comment is labelled as counter speech. Hence, the classifier has classified this comment as 'Hope-Speech'.

Table 8: Samples of misclassification in Task B for English language dataset

in the minority classes, the model becomes too specialized on the minority class and fails to generalize well to unseen data resulting in over-fitting.

Table 7 shows the sample text from English and Hindi labeled Test sets, the actual and predicted labels (obtained for the Test sets after evaluating mBERT models fine-tuned with oversampled Train sets) along with the remarks for Task A and Table 8 shows the sample text from English labeled Test set, the actual and predicted labels along with the remarks for Task B. It can be observed that most of the wrong classifications are due to lack of context. The data presented in Tables 7 and 8 highlights a noticeable inconsistency in the usage of content words within the classes of the Train set. This inconsistency could potentially be responsible for erroneous classifications, as the lack of uniformity in the content word usage might be leading the misclassification model.

5 Conclusion and Future work

This paper describes the models submitted by our team - MUCS, to the shared task ”Homophobia/-Transphobia Detection in social media comments” at RANLP 2023 for the identification of H/T content in social media text. TL model with mBERT are proposed for both Tasks A and B along with oversampling. These models secured 11th, 5th, 3rd, 3rd, and 7th ranks for English, Tamil, Malayalam, Spanish, and Hindi respectively in Task A and 8th, 2nd, and 2nd ranks for English, Tamil, and Malayalam respectively in Task B. Data augmentation techniques for handling imbalanced classes with effective feature extraction techniques will be explored in future.

References

Nsrin Ashraf, Mohamed Taha, Ahmed Abd Elfattah, and Hamada Nayel. 2022. Nayel@ lt-edi-acl2022: Homophobia/Transphobia Detection for Equality, Diversity, and Inclusion using SVM. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 287–290.

Fazlourrahman Balouchzahi, Aparna B K, and H L Shashirekha. 2021a. MUCS@DravidianLangTech-EACL2021:COOLI-Code-Mixing Offensive Language Identification. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 323–329, Kyiv. Association for Computational Linguistics.

Fazlourrahman Balouchzahi and H. Shashirekha. 2020. LAs for HASOC-Learning Approaches for Hate Speech and Offensive Content Identification. pages 145–151.

Fazlourrahman Balouchzahi, Hosahalli Lakshmaiah Shashirekha, and Grigori Sidorov. 2021b. HSSD: Hate Speech Spreader Detection using N-grams and Voting Classifier. In *CLEF (Working Notes)*, pages 1829–1836.

Fazlourrahman Balouchzahi, Hosahalli Lakshmaiah Shashirekha, Grigori Sidorov, and Alexander Gelbukh. 2022. A Comparative Study of Syllables and Character Level N-grams for Dravidian Multi-Script and Code-Mixed Offensive Language Identification.

In *Journal of Intelligent & Fuzzy Systems*, pages 1–11. IOS Press.

Bharathi Raja Chakravarthi. 2023. Detection of Homophobia and Transphobia in YouTube Comments. *International Journal of Data Science and Analytics*, pages 1–20.

Bharathi Raja Chakravarthi, Adeep Hande, Rahul Ponnusamy, Prasanna Kumar Kumaresan, and Ruba Priyadharshini. 2022. 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*, pages 100–119.

Bharathi Raja Chakravarthi, Rahul Ponnusamy, Malliga Subramanian, Paul Buitelaar, Miguel Ángel García-Cumbreras, Salud María Jiménez-Zafra, José Antonio García-Díaz, Rafael Valencia-García, and Nitesh Jindal. 2023. Overview of Second Shared Task on Homophobia and Transphobia Detection in English, Spanish, Hindi, Tamil, and Malayalam. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Kayalvizhi Sampath, Durairaj Thenmozhi, Sathiyaraj Thangasamy, Rajendran Nallathambi, and John Phillip McCrae. 2021. Dataset for Identification of Homophobia and Transophobia in Multilingual YouTube Comments. In *arXiv preprint arXiv:2109.00227*.

Shi Chen and Bing Kong. 2021. cs@ DravidianLangTech-EACL2021: Offensive Language Identification based on Multilingual BERT Model. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 230–235.

B Fazlourrahman, BK Aparna, and HL Shashirekha. 2022. CoFFiTT-COVID-19 Fake News Detection Using Fine-Tuned Transfer Learning Approaches. In *Congress on Intelligent Systems: Proceedings of CIS 2021, Volume 2*, pages 879–890. Springer.

Asha Hegde, Mudoor Devadas Anusha, Sharal Coelho, Hosahalli Lakshmaiah Shashirekha, and Bharathi Raja Chakravarthi. 2022a. Corpus Creation for Sentiment Analysis in Code-Mixed Tulu Text. In *Proceedings of the 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages*, pages 33–40.

Asha Hegde, Mudoor Devadas Anusha, and Hosahalli Lakshmaiah Shashirekha. 2021. Ensemble Based Machine Learning Models for Hate Speech and Offensive Content Identification. In *Forum for Information Retrieval Evaluation (Working Notes)(FIRE), CEUR-WS. org*.

Asha Hegde, Sharal Coelho, Ahmad Elyas Dashti, and Hosahalli Shashirekha. 2022b. MUCS@ Text-LT-EDI@ ACL 2022: Detecting Sign of Depression from Social Media Text using Supervised Learning Approach. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 312–316.

Asha Hegde and Shashirekha Lakshmaiah. 2022. Mucs@ mixmt: Indictrans-based Machine Translation for Hinglish Text. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 1131–1135.

Asha Hegde and Hosahalli Lakshmaiah Shashirekha. 2022. Leveraging Dynamic Meta Embedding for Sentiment Analysis and Detection of Homophobic/-Transphobic Content in Code-mixed Dravidian Languages.

Navya Jose, Bharathi Raja Chakravarthi, Shardul Suryawanshi, Elizabeth Sherly, and John P McCrae. 2020. A survey of Current Datasets for Code-Switching Research. In *2020 6th international conference on advanced computing and communication systems (ICACCS)*, pages 136–141. IEEE.

Thomas Mandl, Sandip Modha, Anand Kumar M, and Bharathi Raja Chakravarthi. 2020. Overview of the hasoc Track at Fire 2020: Hate Speech and Offensive Language Identification in Tamil, Malayalam, Hindi, English and German. In *Forum for information retrieval evaluation*, pages 29–32.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How Multilingual is Multilingual BERT? In *arXiv preprint arXiv:1906.01502*.

Muskaan Singh and Petr Motlicek. 2022. IDIAP Submission@ LT-EDI-ACL2022: Homophobia/Transphobia Detection in Social Media Comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 356–361.

R Srinivasan and CN Subalalitha. 2021. Sentimental Analysis from Imbalanced Code-mixed Data using Machine Learning Approaches. In *arXiv preprint arXiv:1906.01502*.

Konthala Yasaswini, Karthik Puranik, Adeep Hande, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IIITT@ DravidianLangTech-EACL2021: Transfer Learning for Offensive Language Detection in Dravidian Languages. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 187–194.