MUCS@LT-EDI2023: Learning Approaches for Hope Speech Detection in Social Media Text

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
[d]hlsrekha@mangaloreuniversity.ac.in

Abstract

Hope exerts a substantial influence on human cognition and behavior, yet hope related content remains under explored in the realm of social media data analysis. Investigating this content unveils valuable insights into users' emotions, aspirations, and anticipations, offering researchers and analysts a richer comprehension of hope's impact on digital-era individuals' well-being, choices, and actions. Further, this area is rarely explored even for high-resource languages. To address the identification of hope text in social media platforms, this paper describes the models submitted by the team MUCS to "Hope Speech Detection for Equality, Diversity, and Inclusion (LT-EDI)" shared task organized at Recent Advances in Natural Language Processing (RANLP) - 2023. This shared task aims to classify a comment/post in English and code-mixed texts in three languages (Bulgarian, Spanish, and Hindi) into one of the two predefined categories: "Hope" or "Non Hope". Two models: i) Hope_BERT - Linear Support Vector Classifier (LinearSVC) model trained by concatenating Bidirectional Encoder Representations from Transformers (BERT) embeddings and Term Frequency-Inverse Document Frequency (TF-IDF) of character n-grams with word boundary (char_wb) for English and ii) Hope_mBERT - LinearSVC model trained by concatenating Multilingual BERT (mBERT) embeddings and TF-IDF of char_wb for Bulgarian, Spanish, and Hindi code-mixed texts, are proposed for the shared task to classify the given text into Hope or Non-Hope categories. The proposed models obtained 1st, 1st, 2nd, and 5th ranks by exhibiting macro F1 scores of 0.61, 0.75, 0.67, and 0.44 for Spanish, Bulgarian, Hindi, and English texts respectively.

1 Introduction

Social media has a profound impact on society, providing a platform for individuals to express their opinions and communicate with others effectively at a much faster rate. It also enables access to diverse opinions, facilitates connection with different individuals, promotes art and culture, and provides a platform for marginalized voices. Social media platforms are also being used effectively to spread awareness and support various causes, for instance: inequality, human rights violations, discrimination, health and wellness, environmental concerns, etc. (Chakravarthi and Muralidaran, 2021; Balouchzahi et al., 2021b). While constructive criticism have fostered healthy discussions, the misuse of freedom of speech on social media has become a prevalent issue (Hegde et al., 2021b). Trolling and online bullying have become unfortunate consequences of this freedom, causing significant harm to individuals' mental well-being. Numerous studies have consistently highlighted the detrimental effects of heavy social media usage, including increased risk of depression, anxiety, loneliness, self-harm, and even suicidal thoughts (Hegde et al., 2022c). Efforts can be made to reduce these negative thoughts by promoting more positive and supportive hope content on social media. Hence, analyzing hope content/speech in social media is considered as an essential determinant for the well-being of users which can also motivate users in a positive way and provide valuable insights into the trajectory of goal-directed behaviors, persistence in the face of misfortunes, and adjusting to positive or negative changes in life (Balouchzahi et al., 2023; Ghanghor et al., 2021).

Hope speech detection refers to the analysis of social media content for the detection of inspirational text/posts with positive vibes. However, hope speech detection has rarely been experimented even for high-resource languages (Chakravarthi et al., 2022). Social media text is often code-mixed and the analysis of code-mixed texts in low-resource languages has been the focus of several studies

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion, pages 279–286, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_043

Language	Sample Text	English Translation	Label
Bulgarian	Искам да ви попитам един важен въпрос който ме мъчи но ако може да е на лично	I want to ask you one important question that bothers me but if it can be done personally	**Not-Hope**
English	And might not be in a position to come out safely	And might not be in a position to come out safely	**Hope**
Hindi	ye koi bimary nahi hai natural state hai apni apni sabko choice honi chahiye apni marzi se jeeene ki	This is not a disease, it is a natural state, everyone should have their own choice to live according to their wish	**Hope**
Spanish	Ser lgtb y zurdo al mismo tiempo es ser un reverendo imbécil, basta con ver como viven estas comunidades en los países zurdos y ver como viven en los países capitalistas para darse cuenta que el capitalismo es el verdadero camino	If you are LGBT and you are an imbecile at the same time, you just have to see how these communities live in poor countries and how they live in capitalist countries to see that capitalism is the true way	**Not-Hope**

Table 1: Sample comments from Hope Speech detection dataset along with the English translation

and workshops (Fake News Detection (Hegde and Shashirekha, 2021), Sentiment Analysis (Hegde et al., 2022a), Word Level Language Identification (Balouchzahi et al., 2022a), Machine Translation (Hegde et al., 2022b; Hegde and Lakshmaiah, 2022), Threatening Language Detection (Hegde and Shashirekha, 2022) and Offensive Language Identification (Hegde et al., 2021a)). Processing code-mixed texts is challenging due to mixing languages within the same utterance or text. These challenges include tokenization, language identification, linguistic variation, and unavailability of pretrained models trained to represent code-mixed text. To address these challenges, "Hope Speech Detection for Equality, Diversity, and Inclusion" shared task[1] at RANLP 2023[2] aims to classify comment/post in English and code-mixed texts in Spanish, Bulgarian, and Hindi, into one of the two predefined categories, namely: "Hope" or "Non hope". The sample comments/posts from the shared task dataset along with their English translations are shown in Table 1.

In this paper, we team MUCS, describe the two binary classification models: Hope_BERT and Hope_mBERT, submitted to "Hope Speech Detection for Equality, Diversity, and Inclusion" shared task (Kumaresan et al., 2023). While Hope_BERT uses a combination of TF-IDF of char_wb and BERT embeddings extracted from $\text{BERT}_{\text{base}}$ models, to train LinearSVC for hope speech detection in English, Hope_mBERT makes use of Multilingual BERT (mBERT) embeddings combined with TF-IDF of char_wb to train LinearSVC for hope speech detection in Bulgarian, Spanish, and Hindi code-mixed texts. The code to reproduce the proposed models is available in github[3].

The rest of paper is organized as follows: while Section 2 describes the recent literature on code-mixed text processing and hope speech detection, Section 3 focuses on the description of the proposed models submitted to the shared task followed by the experiments and results in Section 4. Conclusion and future works are included in Section 5.

2 Related Work

Hope is a positive state of mind that is based on an expectation of confident outcomes with respect to an occurrence of any event in one's life. Researchers have explored many algorithms to detect the hope speech in text and few of the relevant works are described below:

Balouchzahi et al. (2021a) describes the models to detect hope speech in English, Tamil-English and Malayalam-English code-mixed texts. The authors proposed three distinct models: i) CoHope-ML - ensemble of Machine Learning (ML) classifiers (eXtreme Gradient Boosting (XGB), and Logistic Regression (LR), and MultiLayer Perceptron (MLP)) with hard voting, ii) CoHope-NN - based on keras Neural Network (NN) and iii) CoHope-TL - Bidirectional Long Short Term Memory (BiLSTM) with 1 Dimensional Convolutional Neural Network (1DCNN) trained with BERT embeddings. Both, CoHope-ML and CoHope-NN models are

[1]https://codalab.lisn.upsaclay.fr/competitions/11076
[2]http://ranlp.org/ranlp2023/

[3]https://github.com/hegdekasha/Hope_speech

trained with character n-grams in the range (3, 6) and syntactic n-grams in the range (2, 3). CoHope-ML model outperformed the other models and obtained weighted F1 scores of 0.85, 0.92, and 0.59 for Malayalam-English, English and Tamil-English texts respectively. Balouchzahi et al. (2023) created a dataset to identify hope content in code-mixed Spanish-English tweets and implemented several baselines based on ML, Deep Learning (DL), and Transfer Learning (TL) approaches to benchmark their dataset. Their work consists of two subtasks: subtask 1 - a binary classification and subtask 2 - a multiclass classification. TF-IDF of word uni-grams are used to train ML models (Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), LR, XGB, MLP, Categorical Boosting (CB)), Global Vectors for Word Representation (GloVe) and fastText embeddings are used to train the DL models (Long Short Term Memory (LSTM), Bidirectional LSTM, and Convolutional Neural Network (CNN)) and TL based models are trained using BERT, Robustly Optimized BERT Approach (RoBERTA), mBERT, and MLNet features. Among all the learning models, LR and CB classifiers outperformed the other models obtaining macro F1 scores of 0.80 and 0.79 for binary and 0.64 and 0.54 for multiclass classifications respectively. The learning models submitted by Gowda et al. (2022) aims to classify the given comments in English into 'Hope' or 'Not-Hope' using 1DCNN with LSTM model trained with keras embeddings features. Using Synthetic Minority Oversampling Technique (SMOTE) to handle data imbalance in the dataset, they obtained macro F1 score of 0.55 and weighted F1 score of 0.860. Balouchzahi et al. (2022b) presents the ensemble model (two DT classifiers and one Random Forest (RF) classifier) with soft voting to select the best word and character n-grams to train keras NN for hope speech detection. Their models obtained weighted F1 scores of 0.870 and 0.790 for English and Spanish texts respectively. Vijayakumar et al. (2022) presented fine-tuning of A Lite BERT (ALBERT) - a transformer-based model to detect hope speech in code-mixed Dravidian languages (Malayalam and Kannada) and English. During fine-tuning ALBERT model, they used Adam optimizer and obtained weighted average F1 scores of 0.880, 0.740, and 0.750 for English, Malayalam, and Kannada languages respectively. Hande et al. (2021) created the hope speech dataset with 6,176 user-generated comments in code-mixed Kannada language scraped from YouTube and manually annotated them as 'Hope' or 'Not-hope' to detect hope speech. They benchmarked their dataset with ML (LR, k-Nearest Neighbors (k-NN), DT, RF, and Naive Bayes), DL (LSTM, BiLSTM, and CNN), and TL (BERT, mBERT, RoBERTa, RoBERTa-mBERT, Cross Lingual Language Model RoBERTa, and Dual-Channel BERT (DC-BERT4HOPE)) based approaches. Among all the models, RF and DC-BERT4HOPE with RoBERTa-mBERT models outperformed other models with weighted F1 scores of 0.706 and 0.752 respectively.

From the above literature, it is clear that hope speech detection task is not even explored for high-resource languages. It also indicates that there is enough space for developing models and tools for detecting hope speech content in code-mixed low-resource languages.

3 Methodology

The framework of the proposed methodology visualized in Figure 1 includes the following steps:

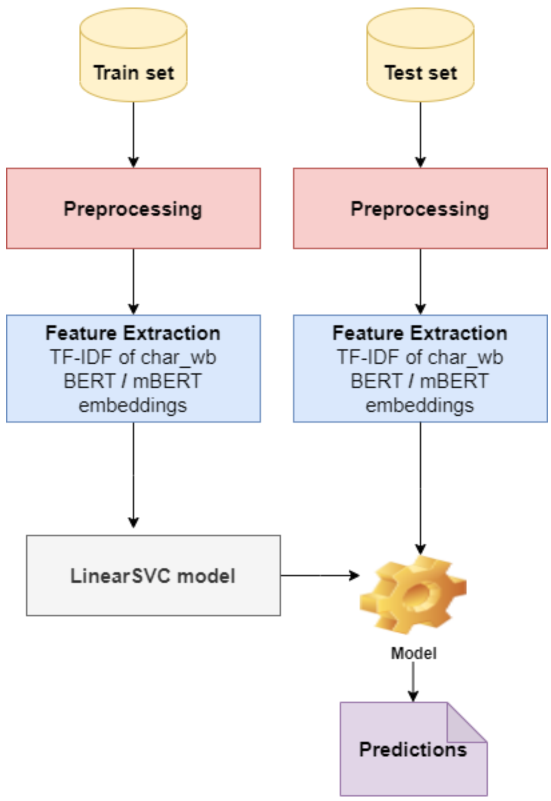

Figure 1: The framework of the proposed methodology

Configuration	Values
WordPiece Vocab size (BERT)	30,522
WordPiece Vocab size (mBERT)	1,19,547
attention heads	12
layers	6
dimension	768
max_length	100

Table 2: Configurations and their values used in BERT and mBERT models

Hyperparameters	Values
penalty	l2
C	1.0
class_weight	balanced
max_iter	max_iter
random_state	100
loss	squared_hinge

Table 3: Hyperparameters and their values used in LinearSVC model

Language	Train set		Development set	
	Hope	Not-Hope	Hope	Not-Hope
Bulgarian	223	4,448	75	514
English	1,562	16,630	400	4,148
Hindi	343	2,219	45	275
Spanish	691	621	100	200

Table 4: Statistics of the Train and Development sets

3.1 Preprocessing

Preprocessing steps involve converting emojis to their corresponding text, eliminating punctuation, digits, and unwanted characters (such as !()-[];:'" ¡¿./?$=%+@* ', etc.) and lowercasing the text. Further, Bulgarian, Hindi, and Spanish stopwords list available at github[4] and English stopwords available at Natural Language Tool Kit[5] are used as references to remove stopwords. These steps help to reduce the irrelevant textual content and improve the performance of the learning models.

3.2 Feature Extraction

Feature extraction is a crucial step which helps to extract features in the given data and the distinguishing features helps to improve the performance of the learning models (Hegde et al., 2022d). A fusion of TF-IDF of char_wb, and BERT/mBERT embeddings (BERT embeddings for English text and mBERT embeddings for Spanish, Bulgarian, and Hindi texts) are used as features to train Hope_BERT/Hope_mBERT models in the proposed approach.

- **TF-IDF of char_wb** - character sequences of length 1 to 3 are extracted using char_wb[6] n-grams and represented as TF-IDF vectors.
- **BERT and mBERT embeddings** - are pretrained models trained on huge unlabeled text data for word representations. BERT is trained on Toronto Book Corpus and Wikipedia and exclusively used for tasks involving English texts, whereas mBERT is trained on wikipedia data and blogs that belong to more than 104 languages and exclusively used for tasks that includes multiple languages. These pretrained models provide tokenizers and for each token/word they provide embeddings which encode the semantic information.

3.3 Model Description

Language Model (LM) analyzes large collections of text data to gain insights into the relationships between words and generate accurate predictions based on the context of the input text. Inspired by the LM, the framework described by Balouchzahi et al. (2021a) is adopted for the proposed models.

3.3.1 Hope_BERT

This model makes use of pretrained BertTokenizer[7] and TFBertModel[8] modules for tokenization and loading the BERT LM respectively for English text. BertTokenizer is a pretrained tokenizer trained on a large amount of English text. It tokenizes the words based on WordPiece[9] tokenizer. Further, 'TFBertModel' is a class in the huggingface transformers library that loads the pretrained BERT LM for English text. This LM can predict the next word in a sentence by considering both the left and right context of the input text. The steps involved in designing Hope_BERT model are described below:

- Tokenization - BertTokenizer is loaded and fine-tuned on the English text provided by the shared task organizers
- Creating features - a pre-trained BERT LM is loaded with a WordPiece vocabulary of

[4]https://github.com/stopwords-iso/

[5]https://pythonspot.com/nltk-stop-words/

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

[7]https://huggingface.co/docs/transformers/main_classes/tokenizer

[8]https://huggingface.co/docs/transformers/model_doc/bert

[9]https://huggingface.co/learn/nlp/course/chapter6/6?fw=pt

Language	Development set		Test set	
	With imbalanced data	With balanced data	With imbalanced data	With balanced data
Hope_mBERT				
Spanish	0.76	0.79	0.6	**0.61**
Bulgarian	0.80	0.81	0.73	**0.75**
Hindi	0.70	0.73	0.65	**0.67**
Hope_BERT				
English	0.65	0.67	0.42	0.44

Table 5: Results of the proposed models

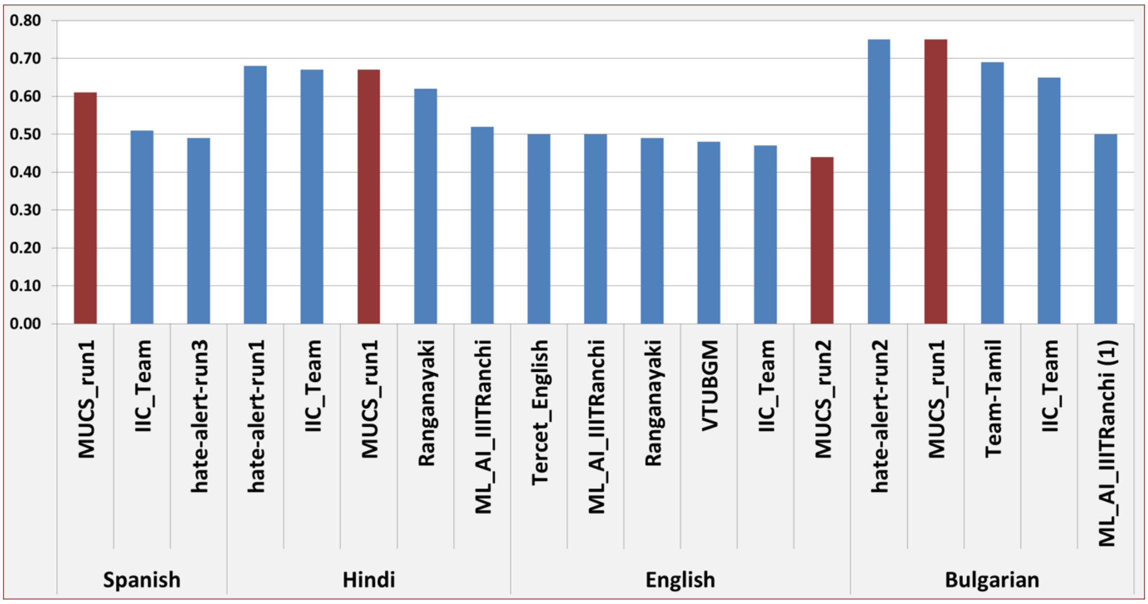

Figure 2: Comparison of macro F1 scores of the participating teams in the shared task

size 30,522 and fine-tuned on the English dataset provided by the organizers, allowing the model to generate feature vectors. These vectors are then used to train the LinearSVC model

The configuration and their values used in constructing Hope_BERT model are shown in Table 2.

3.3.2 Hope_mBERT

Hope_mBERT model utilizes pretrained multilingual BertTokenizer (mBERT tokenizer) and multilingual TFBertModel[10] (mBERT LM) modules to load the pretrained multilingual tokenizer and mBERT LM respectively for the multilingual texts. The steps, configurations and the corresponding values to build Hope_mBERT model are the same as used in constructing the Hope_BERT model. The resulting multilingual word embeddings contains WordPiece vocabulary of size 1,19,547. Hope_mBERT LM is fine-tuned on Spanish/ Bulgarian/ Hindi languages, to build the models for the respective languages. This fine-tuning enables the model to create feature vectors, which are subsequently employed to train the LinearSVC models.

3.3.3 Classifier Construction

LinearSVC is a supervised ML algorithm typically used for classification tasks. It works by mapping data points to a high-dimensional space and then finding the optimal hyperplane that divides the data into consequent classes (Fung and Mangasarian, 2001). This algorithm aims to maximize the margin between the classes, allowing for better generalization of unseen data. Hyperparameters and their values used to train LinearSVC model are shown in Table 3. The hyperparameters which are not mentioned in Table 3 are used with their default values. As the dataset provided by the organizers is imbal-

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

Language	Comments	Actual Label	Predicted Label	Remarks
English	I dont respect LGBT community.. and because of this video I unsubscribe your channel	Not-Hope	Hope	Removing the stopwords (I, don't, and, because, of, this, you) results in the content words (respect, LGBT, community, video, unsubscribe, channel). Among these content words, the word 'respect' speaks about hope and hence the comment is classified as 'Hope'.
	excelent and comprehensive video for understand the LGBTQ topic...	Not-Hope	Hope	The content words 'excellent' and 'comprehensive' speaks about hope. This indicates the incorrect annotation of the comment.
Hindi	@LIMITED GAMER usne glt kya kha. Ye to mujhe bhi odd lga. But doesn't mean ki main culture tabahh kr rhi hu.	Hope	Not-Hope	The words 'glt', 'odd', and 'tabah' used in the comments are associated with 'Not-Hope' class and hence, the comment is classified as 'Not-Hope'.
	गज़ब खबरे okkk but Mene Bola apko Bura to nhi लगा न अगर लगा हो तो sorry pr Boone Ka hak to sbko मिलना चाहिए	Not-Hope	Hope	The content words 'गज़ब', 'खबरे', 'sorry', 'Boone' and 'Bura' are seen in Train set with 'Hope' class and because of this, the comment is classified as 'Hope'.

Table 6: Sample misclassified comments along with the probable reasons

anced, class_weight ='balanced' hyperparameter is used during training the LinearSVC to handle the data imbalance. This hyperparameter value automatically adjusts class weights based on their frequencies, resolving the data imbalance issue to some extent without manual intervention.

4 Experiments and Results

Statistics of the datasets provided by the organizers of the shared task is shown in Table 4 (Chakravarthi, 2020). Hope_BERT and Hope_mBERT models are evaluated using the Test sets and the predictions are submitted to the organizers for evaluation in terms of macro F1 score. The performance of the proposed models are shown in Table 5. Hope_BERT model exhibited a macro F1 score of 0.44 securing 5th rank in the shared task for English text and Hope_mBERT models exhibited macro F1 scores of 0.61, 0.75, and 0.67 securing 1st, 1st, and 2nd ranks for Spanish, Bulgarian, and Hindi code-mixed texts respectively. The low macro F1 score for English may be due to severe class imbalance in the English train set. Few misclassified comments along with the actual and predicted labels (obtained from Hope_BERT and Hope_mBERT models evaluating on the English and Hindi Test sets respectively), and the probable reasons for misclassification are shown in Table 6. From Table 6, it is clear that, removing stopwords and incorrect annotations have shown the impact in deciding the polarity of the comments. Figure 2 gives the comparison of macro F1 scores of all the participating teams for the shared task.

5 Conclusion

In this paper, we team MUCS, presented the description of the proposed models for the "Hope Speech Detection for Equality, Diversity, and Inclusion-LT-EDI" shared task at RANLP-2023. The proposed models: Hope_BERT - trained with a combination of TF-IDF of char_wb and BERT embeddings for English texts exhibited a macro F1 score of 0.44 securing 5th rank and Hope_mBERT trained with a combination of TF-IDF of char_wb and mBERT embeddings for code-mixed Spanish, Bulgarian, and Hindi texts, exhibited macro F1 scores of 0.61, 0.75, and 0.67 securing 1st, 1st, and 2nd ranks for Spanish, Bulgarian, and Hindi respectively. Suitable features that could efficiently capture the contextual information from the code-mixed text will be explored further.

References

Fazlourrahman Balouchzahi, BK Aparna, and HL Shashirekha. 2021a. MUCS@ LT-EDI-EACL2021: CoHope-Hope Speech Detection for Equality, Diversity, and Inclusion in Code-Mixed Texts. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 180–187.

Fazlourrahman Balouchzahi, Sabur Butt, A Hegde, Noman Ashraf, HL Shashirekha, Grigori Sidorov, and Alexander Gelbukh. 2022a. Overview of CoLI-Kanglish: Word Level Language Identification in Code-mixed Kannada-English Texts at ICON 2022. In *Proceedings of the 19th International Conference on Natural Language Processing (ICON): Shared Task on Word Level Language Identification in Code-mixed Kannada-English Texts*, pages 38–45.

Fazlourrahman Balouchzahi, Sabur Butt, Grigori Sidorov, and Alexander Gelbukh. 2022b. CIC@LT-EDI-ACL2022: Are Transformers the only Hope? Hope Speech Detection for Spanish and English Comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 206–211, Dublin, Ireland. Association for Computational Linguistics.

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, Grigori Sidorov, and Alexander Gelbukh. 2023. PolyHope: Two-level Hope Speech Detection from Tweets. In *Expert Systems with Applications*, page 120078. Elsevier.

Bharathi Raja Chakravarthi. 2020. HopeEDI: A Multilingual Hope Speech Detection Dataset for Equality, Diversity, and Inclusion. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 41–53, Barcelona, Spain (Online). Association for Computational Linguistics.

Bharathi Raja Chakravarthi and Vigneshwaran Muralidaran. 2021. 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*, pages 61–72.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, Subalalitha Cn, John McCrae, Miguel Ángel García, Salud María Jiménez-Zafra, Rafael Valencia-García, Prasanna Kumaresan, Rahul Ponnusamy, Daniel García-Baena, and José García-Díaz. 2022. 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*, pages 378–388, Dublin, Ireland. Association for Computational Linguistics.

Glenn Fung and Olvi L Mangasarian. 2001. Proximal Support Vector Machine Classifiers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 77–86.

Nikhil Ghanghor, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IIITK@ LT-EDI-EACL2021: Hope Speech Detection for Equality, Diversity, and Inclusion in Tamil, Malayalam and English. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 197–203.

Anusha Gowda, Fazlourrahman Balouchzahi, Hosahalli Shashirekha, and Grigori Sidorov. 2022. MUCIC@LT-EDI-ACL2022: Hope Speech Detection using Data Re-Sampling and 1D Conv-LSTM. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 161–166, Dublin, Ireland. Association for Computational Linguistics.

Adeep Hande, Ruba Priyadharshini, Anbukkarasi Sampath, Kingston Pal Thamburaj, Prabakaran Chandran, and Bharathi Raja Chakravarthi. 2021. Hope Speech Detection in Under-Resourced Kannada Language. In *arXiv preprint arXiv:2108.04616*.

Asha Hegde, Mudoor Devadas Anusha, Sharal Coelho, and Hosahalli Lakshmaiah Shashirekha. 2021a. MUM at ComMA@ICON: Multilingual Gender Biased and Communal Language Identification Using Supervised Learning Approaches. In *Proceedings of the 18th International Conference on Natural Language Processing: Shared Task on Multilingual Gender Biased and Communal Language Identification*, pages 64–69, NIT Silchar. NLP Association of India (NLPAI).

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. 2021b. 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, Shubhanker Banerjee, Bharathi Raja Chakravarthi, Ruba Priyadharshini, Hosahalli Shashirekha, John Philip McCrae, et al. 2022b. Overview of the Shared Task on Machine Translation in Dravidian Languages. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*, pages 271–278.

Asha Hegde, Sharal Coelho, Ahmad Elyas Dashti, and Hosahalli Shashirekha. 2022c. 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, Sharal Coelho, and Hosahalli Shashirekha. 2022d. MUCS@DravidianLangTech@ACL2022:

Ensemble of Logistic Regression Penalties to Identify Emotions in Tamil text. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*, pages 145–150, Dublin, Ireland. Association for Computational Linguistics.

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. 2021. Urdu Fake News Detection Using Ensemble of Machine Learning Models.

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.

Prasanna Kumar Kumaresan, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, Miguel Ángel García-Cumbreras, Salud María Jiménez-Zafra, José Antonio García-Díaz, Rafael Valencia-García, Momchil Hardalov, Ivan Koychev, Preslav Nakov, Daniel García-Baena, and Kishore Kumar Ponnusamy. 2023. Overview of the Third Shared Task on 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.

Praveenkumar Vijayakumar, Prathyush S, Aravind P, Angel S, Rajalakshmi Sivanaiah, Sakaya Milton Rajendram, and Mirnalinee T T. 2022. SSN_ARMM@ LT-EDI -ACL2022: Hope Speech Detection for Equality, Diversity, and Inclusion Using ALBERT Model. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 172–176, Dublin, Ireland. Association for Computational Linguistics.