SOA_NLP@LT-EDI-ACL2022: An Ensemble Model for Hope Speech Detection from YouTube Comments

Abhinav Kumar[1], **Sunil Saumya**[2], and **Pradeep Kumar Roy**[3]
[1]Siksha 'O' Anusandhan, Deemed to be University, Bhubanewar, Odisha, India
[2]Indian Institute of Information Technology Dharwad, Karnataka, India
[3]Indian Institute of Information Technology Surat, Gujarat, India
(abhinavanand05,sunil.saumya007,pkroynitp)@gmail.com

Abstract

Language should be accommodating of equality and diversity as a fundamental aspect of communication. The language of internet users has a big impact on peer users all over the world. On virtual platforms such as Facebook, Twitter, and YouTube, people express their opinions in different languages. People respect others' accomplishments, pray for their well-being, and cheer them on when they fail. Such motivational remarks are hope speech remarks. Simultaneously, a group of users encourages discrimination against women, people of color, people with disabilities, and other minorities based on gender, race, sexual orientation, and other factors. To recognize hope speech from YouTube comments, the current study offers an ensemble approach that combines a support vector machine, logistic regression, and random forest classifiers. Extensive testing was carried out to discover the best features for the aforementioned classifiers. In the support vector machine and logistic regression classifiers, char-level TF-IDF features were used, whereas in the random forest classifier, word-level features were used. The proposed ensemble model performed significantly well among English, Spanish, Tamil, Malayalam, and Kannada YouTube comments.

1 Introduction

People have started to spend more time on social media platforms in recent years. As a result, many informed decisions are taken based on the sentiment of the social media community (Sampath et al., 2022; Ravikiran et al., 2022; Chakravarthi et al., 2022b; Bharathi et al., 2022; Priyadharshini et al., 2022). Social media gives us the opportunity to track the activity of our friends and family, just like we do in real life (Priyadharshini et al., 2021; Kumaresan et al., 2021). Additionally, it also allows us to communicate with people we have never met in person across the globe. On these websites, there are mostly two types of vibes: Hope and Hate. Hope is a positive state of mind defined by the expectation of favourable outcomes in one's life events and circumstances. People are motivated to act when they are filled with hope. Hope can be useful for anyone who wishes to maintain a consistent and positive outlook on life (Dowlagar and Mamidi, 2021). We often use hope terms, such as "Well Done!", "Congratulations", "Can be done in better way", "Keep up the good work" and so on to encourage on one's work. Hope contents frequently assist us in a variety of critical situations, including emergency management, photo sharing, video streaming, trip planning, and citizen engagement (Kumar et al., 2020b,c). Hate, on the other hand, is a negative vibe present on an online platform with the intention of harassing an individuals based on their race, religion, ethnic origin, sexual orientation, disability, or gender (Roy et al., 2022; Kumar et al., 2020b, 2021). The ultimate purpose of every social media platform is to reduce hate content while simultaneously promoting hope content.

Although there is much work being done to eradicate negativity from the social media (Priyadharshini et al., 2022; Chakravarthi et al., 2021; Saumya et al., 2021), Hope speech detection focuses on spreading optimism by detecting content that is encouraging, positive, and supporting. There hasn't been much work done in the domain of hope speech detection, although the NLP community has recently shown interest in it (Singh et al., 2021). To reduce hostility, (Chakravarthi, 2020) developed a hope detection methodology for the YouTube platform in 2019. In the year 2020, (Chakravarthi and Muralidaran, 2021) presented the *LT-EDI-EACL2021* shared task[1], which attempted to discover hope speeches in a corpus of English, Tamil, and Malayalam. To identify hope content in YouTube comments, (Thara et al., 2021) developed a bidirectional long short-term memory (BiLSTM) using attention-based technique. For

[1]https://sites.google.com/view/lt-edi-2021/home

Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, pages 223 - 228
May 27, 2022

the same goal, (Gundapu and Mamidi, 2021) presented a transformer-based BERT model. (Sharma and Arora, 2021) employed synthetically generated code-mixed data to train a transformer-based model RoBERTa, which they used with their pre-trained ULMFiT in an ensemble for hope speech categorization.

The second workshop on language technology for Equality, Diversity, and Inclusion (LT-EDI-2022) is proposed in ACL 2022 (Chakravarthi and Muralidaran, 2021; Chakravarthi et al., 2022a), with the shared assignment available in English, Tamil, Malayalam, Kannada, and Spanish. We participated in the LT-EDI-2022 competition and submitted an ensemble model by utilizing char-level features with support vector machine and logistic regression classifiers and word-level features with random forest classifier. The proposed ensemble model placed 8th, 4th, 3rd, 2nd, and 3rd for English, Tamil, Malayalam, Kannada, and Spanish dataset, respectively, among all other submitted models in the competition.

The remaining parts of the paper are organized as follows: Section 2 analyses similar research for hope speech detection, Section 3 examines the datasets and technique used in the study, Section 4 discusses results of the proposed model and Section 5 concludes the study with future directions.

2 Related work

Several studies have been reported by researchers (Kumar et al., 2020b; Saumya et al., 2021; Kumar et al., 2020a) to identify hate and offensive material from social media, but relatively few efforts have been done to identify hope speech from social media (Chakravarthi, 2020; Thara et al., 2021; Gundapu and Mamidi, 2021; Sharma and Arora, 2021).

(Puranik et al., 2021) evaluated different transfer learning-based models for hope speech identification from English, Tamil, and Malayalam social media postings, including BERT, ALBERT, DistilBERT, RoBERTa, character-BERT, mBERT, and ULMFiT. ULMFiT achieved an F_1-score of 0.9356 on English data due to its improved fine-tuning process. On the Malayalam test set, mBERT achieved a weighted F_1-score of 0.8545, whereas distilmBERT achieved a weighted F_1-score of 0.5926 on the Tamil test set. (Balouchzahi et al., 2021) offered three models based on Ensemble of classifiers, Neural Network (NN), and BiLSTM with one dimensional convolution model. The first two models were trained using character and word gram features, while the third model was created using BiLSTM and one dimensional convolution. Finally, classification was carried out in each case. Ensemble of classifiers outperformed the other two models, with F1-scores of 0.85, 0.92, and 0.59 for Malayalam, English, and Tamil datasets, respectively.

The hope speech was identified using a fine-tuned XLM-Roberta model by (Que, 2021; Hossain et al., 2021). (Awatramani, 2021) used a Pre-trained transformers with paraphrasing generation for Data Augmentation for hope content identification. (Thara et al., 2021) used an attention-based strategy to create a bidirectional long short-term memory (BiLSTM), (Gundapu and Mamidi, 2021) offered a transformer-based BERT model, and (Sharma and Arora, 2021) built a transformer-based model RoBERTa with synthetically produced code-mixed data, which they used with their pre-trained ULMFiT in an ensemble for hope speech classification. In accordance with the existing literature, this paper proposes an ensemble model for the hope speech detection from English, Spanish, Tamil, Malayalam, and Kannada YouTube comments.

3 Methodology

Figure 1 depicts the overall flow diagram of the proposed ensemble model. The proposed ensemble model combines three machine learning algorithms: (i) Support Vector Machine (SVM), (ii) Logistic Regression (LR), and (iii) Random Forest (RF). The suggested approach is tested using YouTube comments in five distinct languages: English, Spanish, Tamil, Malayalam, and Kannada. Table 1 shows the total data statistic used to validate the proposed system.

To find the best-suited features and classifiers, we experimented with seven machine learning classifiers such as (i) Support Vector Machine, (ii) Random Forest, (iii) Logistic Regression, (iv) Naive Bayes, (v) K-Nearest Neighbor, (vi) Decision Tree, and (vii) AdaBoost with different combinations of n-gram char-level and word-level Term-Frequency-Inverse-Document-Frequency (TF-IDF). We varied the n-gram range from 1 to 6 for both char-level and word-level features. After performing extensive experiments, we found that 1 to 6-gram char-level

Table 1: Data statistics for English, Spanish, Tamil, Malayalam, and Kannada language comments

Dataset	Label	Non-hope Speech	Hope Speech
English	Train	20778	1962
	Dev	2569	272
Spanish	Train	499	491
	Dev	169	161
Tamil	Train	7872	6327
	Dev	998	757
Malayalam	Train	6205	1668
	Dev	784	190
Kannada	Train	3241	1699
	Dev	408	210

TF-IDF feature with Logistic Regression and Support Vector Machine performed best among all the mentioned classifiers, whereas 1 to 3-gram word-level features performed best for Random Forest classifier. The performance of best-suited classifiers with best-suited features are tabulated in Table 2.

The prediction of all three best-performed machine learning classifiers Support Vector Machine, Logistic Regression, and Random Forest are taken into account and performed a majority voting (see Figure 1) to get the final class value for the data sample.

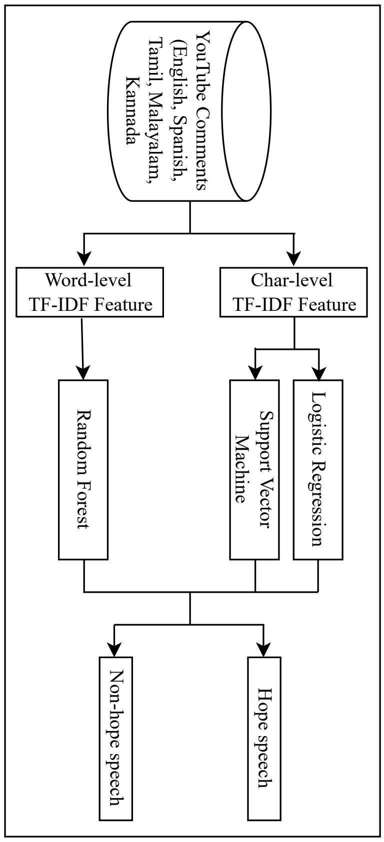

Figure 1: Proposed ensemble model for the hope speech identification

4 Results

All experiments were run on the Google Colab platform[2] with the Sklearn Python library[3] and the default classifier hyper-parameters. The performance of the proposed ensemble model is measured using macro precision, macro recall, macro F_1-score, weighted precision, weighted recall, and weighted F_1-score.

The results of the English, Spanish, Tamil, Malayalam and Kannada language YouTube dataset are listed in Table 3. For the English dataset, the proposed model achieved a macro precision, recall, and F_1-score of 0.460, 0.370, and 0.380, respectively. Similarly, it achieved a weighted precision, recall, and F_1-score of 0.880, 0910, and 0.880, respectively. The suggested model achieved 0.790 macro precision, recall, $F1$-score, weighted precision, recall, and $F1$-score on the Spanish dataset (see Table 3. The suggested model obtained a macro precision of 0.280, a macro recall of 0.320, a macro $F1$-score of 0.290, a weighted precision of 0.360, a weighted recall of 0.430, and a weighted $F1$-score of 0.380 on the Tamil dataset. The suggested model obtained a macro precision of 0.520, a macro recall of 0.480, a macro $F1$-score of 0.480, a weighted precision of 0.720, a weighted recall of 0.790, and a weighted $F1$-score of 0.740 on the Malayalam dataset. The suggested model achieves a macro precision of 0.490, a macro recall of 0.470, a macro $F1$-score of 0.470, a weighted precision of 0.740, a weighted recall of 0.760, and a weighted $F1$-score of 0.750 for the Kannada language.

5 Conclusion

The current work utilized an ensemble strategy that includes a support vector machine, logistic regression, and random forest classifiers to identify hope speech from YouTube comments. The efficiency of different combinations of n-gram char-level and word-level TF-IDF features were also explored in the identification of hope speech from YouTube comments. The use of 1 to 6-gram char-level TF-IDF features with support vector machine and logistic regression performed best, whereas 1 to 3-gram word-level features with random forest classifier

[2]https://colab.research.google.com/
[3]https://scikit-learn.org/

Table 2: Results of the best-suited features ((1-6)-gram TF-IDF char-level feature (Support Vector Machine and Logistic Regression) and (1-3)-gram TF-IDF word-level feature (Random Forest)) with best performed classifiers on development dataset.

Dataset	Class	SVM			Logistic Regression			Random Forest		
		Precision	Recall	F_1-score	Precision	Recall	F_1-score	Precision	Recall	F_1-score
English	Hope speech	0.75	0.25	0.38	0.68	0.23	0.34	0.79	0.19	0.31
	Non-hope speech	0.93	0.99	0.96	0.92	0.99	0.96	0.92	0.99	0.96
	Macro Avg.	0.84	0.62	0.67	0.80	0.61	0.65	0.85	0.59	0.63
	Weighted Avg.	0.91	0.92	0.90	0.90	0.92	0.90	0.91	0.92	0.89
Spanish	Hope speech	0.81	0.71	0.76	0.80	0.66	0.72	0.75	0.72	0.73
	Non-hope speech	0.73	0.83	0.78	0.70	0.83	0.76	0.72	0.75	0.73
	Macro Avg.	0.77	0.77	0.77	0.75	0.74	0.74	0.73	0.73	0.73
	Weighted Avg.	0.77	0.77	0.77	0.75	0.74	0.74	0.73	0.73	0.73
Tamil	Hope speech	0.70	0.47	0.56	0.67	0.51	0.58	0.59	0.52	0.55
	Non-hope speech	0.68	0.84	0.75	0.69	0.81	0.74	0.67	0.73	0.70
	Macro Avg.	0.69	0.66	0.66	0.68	0.66	0.66	0.63	0.62	0.62
	Weighted Avg.	0.69	0.68	0.67	0.68	0.68	0.67	0.63	0.64	0.63
Malayalam	Hope speech	0.84	0.41	0.55	0.85	0.38	0.52	0.72	0.29	0.41
	Non-hope speech	0.87	0.98	0.92	0.87	0.98	0.92	0.85	0.97	0.91
	Macro Avg.	0.86	0.70	0.74	0.86	0.68	0.72	0.79	0.63	0.66
	Weighted Avg.	0.87	0.87	0.85	0.86	0.87	0.84	0.83	0.84	0.81
Kannada	Hope speech	0.73	0.45	0.56	0.74	0.42	0.54	0.68	0.46	0.55
	Non-hope speech	0.76	0.91	0.83	0.76	0.92	0.83	0.76	0.89	0.82
	Macro Avg.	0.74	0.68	0.69	0.75	0.67	0.69	0.72	0.68	0.69
	Weighted Avg.	0.75	0.76	0.74	0.75	0.75	0.73	0.74	0.74	0.73

Table 3: Result of the proposed model for different language datasets

Dataset	English	Spanish	Tamil	Malayalam	Kannada
Macro Precision	0.460	0.790	0.280	0.520	0.490
Macro Recall	0.370	0.790	0.320	0.480	0.470
Macro F_1-score	0.380	0.790	0.290	0.480	0.470
Weighted Precision	0.880	0.790	0.360	0.720	0.740
Weighted Recall	0.910	0.790	0.430	0.790	0.760
Weighted F_1-score	0.880	0.790	0.380	0.740	0.750

performed best among all the three mentioned classifiers. The proposed ensemble model achieved a macro F_1-scores of 0.380, 0.790, 0.290, 0.480, and 0.470 for English, Spanish, Tamil, Malayalam, and Kannada language YouTube comments, respectively. As the use of char-level features performs significantly well, therefore the char-level features can be explored. Deep learning-based models such as BERT, CNN, and auto-encoders can also be explored with proper pre-processing of the texts to achieve better performance.

References

Vasudev Awatramani. 2021. Hopeful NLP@ LT-EDI-EACL2021: Finding hope in YouTube comment section. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 164–167.

Fazlourrahman Balouchzahi, BK Aparna, and HL Shashirekha. 2021. 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.

B Bharathi, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, N Sripriya, Arunaggiri Pandian, and Swetha Valli. 2022. Findings of the shared task on Speech Recognition for Vulnerable Individuals in Tamil. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.

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 Chinnaudayar Navaneethakrishnan, John Phillip McCrae, Miguel Ángel García-Cumbreras, Salud María Jiménez-Zafra, Rafael Valencia-García, Prasanna Kumar Kumaresan, Rahul Ponnusamy, Daniel García-Baena, and José Antonio García-Díaz. 2022a. Findings 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*. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Thenmozhi Durairaj, John Phillip McCrae, Paul Buitaleer, Prasanna Kumar Kumaresan, and Rahul Ponnusamy. 2022b. Findings of the shared task on Homophobia Transphobia Detection in Social Media Comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.

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. *arXiv preprint arXiv:2109.00227*.

Suman Dowlagar and Radhika Mamidi. 2021. EDIOne@ LT-EDI-EACL2021: Pre-trained transformers with convolutional neural networks for hope speech detection. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 86–91.

Sunil Gundapu and Radhika Mamidi. 2021. Autobots@ LT-EDI-EACL2021: One world, one family: Hope speech detection with BERT transformer model. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 143–148.

Eftekhar Hossain, Omar Sharif, and Mohammed Moshiul Hoque. 2021. NLP-CUET@ LT-EDI-EACL2021: multilingual code-mixed hope speech detection using cross-lingual representation learner. *arXiv preprint arXiv:2103.00464*.

Abhinav Kumar, Sunil Saumya, and Jyoti Prakash Singh. 2020a. NITP-AI-NLP@ HASOC-Dravidian-CodeMix-FIRE2020: A machine learning approach to identify offensive languages from Dravidian code-mixed text. In *FIRE (Working Notes)*, pages 384–390.

Abhinav Kumar, Sunil Saumya, and Jyoti Prakash Singh. 2020b. NITP-AI-NLP@ HASOC-FIRE2020: Fine tuned BERT for the hate speech and offensive content identification from social media. In *FIRE (Working Notes)*, pages 266–273.

Abhinav Kumar, Jyoti Prakash Singh, Yogesh K Dwivedi, and Nripendra P Rana. 2020c. A deep multi-modal neural network for informative Twitter content classification during emergencies. *Annals of Operations Research*, pages 1–32.

Gunjan Kumar, Jyoti Prakash Singh, and Abhinav Kumar. 2021. A deep multi-modal neural network for the identification of hate speech from social media. In *Conference on e-Business, e-Services and e-Society*, pages 670–680. Springer.

Prasanna Kumar Kumaresan, Ratnasingam Sakuntharaj, Sajeetha Thavareesan, Subalalitha Navaneethakrishnan, Anand Kumar Madasamy, Bharathi Raja Chakravarthi, and John P McCrae. 2021. Findings of shared task on offensive language identification in Tamil and Malayalam. In *Forum for Information Retrieval Evaluation*, pages 16–18.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, Thenmozhi Durairaj, Malliga Subramanian, Kogilavani Shanmugavadivel, Siddhanth U Hegde, and Prasanna Kumar Kumaresan. 2022. 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.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Sajeetha Thavareesan, Dhivya Chinnappa, Durairaj Thenmozhi, and Rahul Ponnusamy. 2021. Overview of the DravidianCodeMix 2021 shared task on sentiment detection in Tamil, Malayalam, and Kannada. In *Forum for Information Retrieval Evaluation*, pages 4–6.

Karthik Puranik, Adeep Hande, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IIITT@ LT-EDI-EACL2021-Hope Speech Detection: There is always hope in transformers. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 98–106.

Qinyu Que. 2021. Simon @ LT-EDI-EACL2021: Detecting hope speech with BERT. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 175–179, Kyiv. Association for Computational Linguistics.

Manikandan Ravikiran, Bharathi Raja Chakravarthi, Anand Kumar Madasamy, Sangeetha Sivanesan, Ratnavel Rajalakshmi, Sajeetha Thavareesan, Rahul Ponnusamy, and Shankar Mahadevan. 2022. Findings of the shared task on Offensive Span Identification in code-mixed Tamil-English comments. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Pradeep Kumar Roy, Snehaan Bhawal, and C.N. Subalalitha. 2022. Hate speech and offensive language detection in Dravidian languages using deep ensemble framework. *Computer Speech & Language*, page 101386.

Anbukkarasi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, Ruba Priyadharshini, Subalalitha Chinnaudayar Navaneethakrishnan, Kogilavani Shanmugavadivel, Sajeetha Thavareesan, Sathiyaraj Thangasamy, Parameswari Krishnamurthy, Adeep Hande, Sean Benhur, Kishor Kumar Ponnusamy, and Santhiya Pandiyan. 2022. Findings of the shared task on Emotion Analysis in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Sunil Saumya, Abhinav Kumar, and Jyoti Prakash Singh. 2021. Offensive language identification in Dravidian code mixed social media text. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 36–45.

Megha Sharma and Gaurav Arora. 2021. Spartans@ LT-EDI-EACL2021: Inclusive speech detection using pretrained language models. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 188–192.

Pankaj Singh, Prince Kumar, and Pushpak Bhattacharyya. 2021. CFILT IIT Bombay@ LT-EDI-EACL2021: Hope speech detection for equality, diversity, and inclusion using multilingual representation fromtransformers. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 193–196.

S Thara, Ravi teja Tasubilli, et al. 2021. Amrita@ LT-EDI-EACL2021: Hope speech detection on multilingual text. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 149–156.