Overview of the shared task on Hope Speech Detection for Equality, Diversity, and Inclusion

Prasanna Kumar Kumaresan[1], **Bharathi Raja Chakravarthi**[2],
Subalalitha Chinnaudayar Navaneethakrishnan [3], **Miguel Ángel García-Cumbreras**[4],
Salud María Jiménez-Zafra[4], **José Antonio García-Díaz**[5], **Rafael Valencia-García**[5],
Momchil Hardalov[6], **Ivan Koychev**[7], **Preslav Nakov**[8], **Daniel García-Baena**[4],
Kishore Kumar Ponnusamy[9], **Blake Preston**[10],
[1] Insight SFI Research Centre for Data Analytics, University of Galway, Ireland,
[2] Insight SFI Research Centre for Data Analytics, School of Computer Science,
University of Galway, Ireland, [3] SRM Institute Of Science And Technology, Tamil Nadu
[4] Universidad de Jaén, Spain, [5] Universidad de Murcia, Spain, [6] Amazon, Barcelona, Spain,
[7] Sofia University St. Kliment Ohridski, Bulgaria,
[8] Mohamed bin Zayed University of Artificial Intelligence Masdar City, UAE,
[9] Guru Nanak College, Tamil Nadu, India,
[10] School of Computer Science, University of Galway, Ireland.

Abstract

Hope serves as a potent driving force that motivates individuals to persist in the face of life's unpredictable nature. The Hope Speech poses a substantial challenge to online information credibility, mainly due to rapid content dissemination on social media. This article offers a concise overview of the "Hope Speech Detection for Equality, Diversity, and Inclusion- LT-EDI-RANLP 2023" shared task[1]. The task's objective is to classify social media posts as hopeful or not, with a specific focus on four languages: English, Hindi, Bulgarian, and Spanish. Numerous teams participated in the shared task, presenting a range of methodologies including machine learning techniques, transformer-based models, and resampling methods.

1 Introduction

Hope serves as a powerful driving force that encourages individuals to persevere in the face of the unpredictable nature of human existence. It instills motivation within us to remain steadfast in our pursuit of important goals, regardless of the uncertainties that lie ahead (Chakravarthi, 2020). In today's digital age, platforms such as Facebook, Twitter, Instagram, and YouTube have emerged as prominent social media outlets where people freely express their views and opinions. These platforms have also become crucial for marginalized individuals seeking online assistance and support (García-Baena et al., 2023; García-Díaz et al., 2020; Jiménez-Zafra et al., 2023a). The outbreak of the pandemic has exacerbated people's fears around the world, as they grapple with the possibility of losing loved ones and the lack of access to essential services such as schools, hospitals, and mental health facilities. As a result, people have turned to online forums as a means to fulfill their informational, emotional, and social needs. Through social networking sites, individuals can connect with others, experience a sense of social inclusion, and cultivate a feeling of belonging by actively participating in online communities (Kumaresan et al., 2022). The presence of these factors has a profound impact on both physical and psychological well-being, as well as mental health (Jiménez-Zafra et al., 2023b).

By leveraging the power of hope and utilizing online platforms, individuals are able to find solace, support, and resources during challenging times (Hande et al., 2021). These digital spaces offer an avenue for people to seek guidance, share their experiences, and foster connections with others who may be going through similar struggles. Through virtual networks, individuals can combat feelings of isolation, gain access to valuable information, and receive emotional support, all of which contribute to their overall well-being and mental resilience (Ghanghor et al., 2021). It is important to acknowledge the pivotal role that hopes and online platforms play in providing a lifeline to individuals in need. As we navigate through unpredictable circumstances, these digital resources serve as beacons of light, reminding us that we are not alone

[1] https://codalab.lisn.upsaclay.fr/competitions/11076

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion, pages 47–53, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_007

in our struggles and that there is a collective effort to support one another (Puranik et al., 2021). Through the convergence of hope, technology, and human connection, we can weather the storm and find strength in our shared experiences, ultimately emerging stronger and more resilient as a global community.

2 Related work

Despite the optimistic nature of Hope's speech, it has garnered relatively little attention within the research community. This lack of research interest could be attributed to the absence of available labeled datasets. In recent years, there has been a notable increase in attention towards this issue, thanks in part to the efforts of the organizers of the LT-EDI-EACL2021 workshop who shared a labeled dataset. Several frameworks were submitted during this workshop to address the challenge of detecting Hope Speech (Roy et al., 2022). While numerous research endeavors have focused on filtering out hateful and offensive comments from social media posts, the identification of hopeful comments has received comparatively less consideration (Chakravarthi, 2020).

Counter-narratives, denoting informed textual responses, have surfaced as a notable strategy, drawing recent attention from experimenters (Chung et al., 2019). This approach to counter-narratives aims to balance the preservation of freedom of speech while preventing excessive content blocking. (Mathew et al., 2019) took the initiative to construct and release a counterspeech dataset using YouTube comments. However, the central concept of directly intervening with textual responses can inadvertently escalate hostility. Although it is beneficial for content creators to comprehend why their comments or posts were removed or blocked, and subsequently adjust their discourse and attitudes favorably, this intervention can sometimes exacerbate tensions. This has directed our research focus toward the exploration of positive content, such as messages of hope, and promoting such constructive activities.

3 Task description

The hope speech in our context refers to comments or posts on YouTube that provide support, comfort, recommendations, motivation, and understanding. While a comment or post in the dataset could consist of multiple sentences, the average sentence length across the corpus is one. Annotations in the dataset are done at the level of individual comments or posts. Participants were provided with datasets via the CodaLab website[2] for the specific languages Bulgarian, English, Hindi, and Spanish for development, training, and testing purposes.

4 Dataset

The shared task's dataset comprises comments in four distinct languages: English, Spanish, Bulgarian, and Hindi, totaling 27,545, 5,859, 3,203, and 2,159 comments, respectively. These comments are sourced from social media platforms like YouTube and Twitter. For the English language subset, we utilized the HopeEDI dataset introduced (Chakravarthi, 2020). This dataset focuses on socially significant subjects, including Equality, Diversity, and Inclusion, addressing topics like LGBTIQ issues, COVID-19, women in STEM, Dravidian languages, Black Lives Matter, and more. The inter-annotator agreement was assessed using Krippendorf's alpha. These dataset details are shown in Table 1. This method is the same for all the rest of the languages and the statistics were shown int he Table 2 for Bulgarian, and Table 3 for Hindi.

Table 1: Statistics for the English Dataset (HS stands for Hope Speech, and NHS for Non-Hope Speech

Set	HS	NHS	Total
Train	1,562	16,630	18,192
Development	400	4,148	4,548
Test	21	4,784	4,805
Total	1,983	25,562	27,545

Table 2: Statistics for the Bulgarian Dataset (HS stands for Hope Speech, and NHS for Non-Hope Speech

Set	HS	NHS	Total
Train	223	4,448	4,671
Development	75	514	589
Test	150	449	599
Total	448	5,411	5,859

The Spanish dataset is an improved and extended version of the SpanishHopeEDI dataset (García-Baena et al., 2023). The SpanishHopeEDI dataset was improved by manual revision of the annotations, as some annotation errors were found in the error analysis of the baseline experiments conducted with the dataset (García-Baena et al., 2023).

[2]https://codalab.lisn.upsaclay.fr/competitions/11076

Table 3: Statistics for the Hindi Dataset (HS stands for Hope Speech, and NHS for Non-Hope Speech

Set	HS	NHS	Total
Train	343	2,219	2,562
Development	45	275	320
Test	53	268	321
Total	441	2,762	3,203

It consists of LGTB-related tweets that were collected with the Twitter API (June 27, 2021, to July 26, 2021) and using a lexicon of LGBT-related terms, such as #OrgulloLGTBI or #LGTB, as seed for the search. This dataset was extended with a set of tweets collected using the UMUCorpus-Classifier tool (García-Díaz et al., 2020), which allows defining different search criteria such as keywords, accounts, and geolocation. The key-

Table 4: Statistics for the Spanish Dataset (HS stands for Hope Speech, and NHS for Non-Hope Speech)

Set	HS	NHS	Total
Train	691	621	1,312
Development	100	200	300
Test	300	247	547
Total	1,091	1,068	2,159

words used to collect the tweets were related to transphobia and homophobia, such as #transfobia (*#transphobia*), transexual (*transsexual*), identidad de género (*gender identity*), #homofobia (*#homophobia*), homosexual (*homosexual*), #AlertaHomofobia (*#HomophobiaAlert*), or #StopLGTBIfobia *#StopLGTBIphobia*. It should be mentioned that all the tweets of this dataset were manually labeled by the organizers of the shared task marking a tweet as HS (hope speech) if the text: i) explicitly supports the social integration of minorities; ii) is a positive inspiration for the LGTB community; iii) explicitly encourages LGTB people who might find themselves in a situation; or iv) unconditionally promotes tolerance. On the contrary, a tweet was marked as NHS (non-hope speech) if the text: i) expresses negative sentiment towards the LGTB community; ii) explicitly seeks violence; or iii) uses gender-based insults. Table 4 shows the distribution of the dataset considering the number of samples for each label and set. It should be noted that this dataset was also used, but with a different test set, in the *HOPE: Multilingual Hope Speech Detection* shared task (Jiménez-Zafra et al., 2023a) at IberLEF 2023 workshop (Jiménez-Zafra et al., 2023b).

5 Methodology

In this shared task, there are eight teams actively participated and implemented their models. They evaluated their model's performance on our Hope Speech Detection shared task:

hate-alert: (Das et al., 2023) The participants employed two types of transformer-based models, namely mBERT and XLMR-base, in their study. In the first run, they conducted fine-tuning on the mBERT model. For the second and third runs, they took the CLS embeddings from both the mBERT and XLMR models, subjected them to two Dense layers, and ultimately utilized a classification head. Notably, the utilization of cutting-edge transformer-based models for the classification task is a noteworthy aspect of their approach. The results have demonstrated the superiority of these transformer-based models over earlier deep-learning models like LSRM and CNN-GRU.

MUCS: (Hegde et al., 2023) The MUCS team's journey began with enhancing models for a pivotal project. Armed with determination, they fused BERT embeddings with syllable TF-IDF, promising innovative insights. Venturing further, they intertwined BERT embeddings with TF-IDF and resampling. Across three runs, they refined their approach, each iteration marking a step towards precision. This dedication culminated in a triumphant unveiling, a testament to ingenuity, poised to reshape their field.

Team-Tamil: (Ponnusamy et al., 2023) In our pursuit of enhancing model performance, we embarked on a journey fueled by innovation. With the creative fusion of MPNet Embeddings, we engineered a powerful and distinct form of representation. As the threads of our endeavor wove together, an H2O model emerged, infused with the essence of our collective efforts. This union of cutting-edge techniques resulted in a transformative leap forward, poised to unravel new insights and pave the way for future advancements.

IIC_Team: (Vajrobol et al., 2023) The method employed in this task involved utilizing Bidirectional LSTM (Long Short-Term Memory) and BiLSTM with embeddings. Additionally, XLM-ROberta models were employed for each language. Since the datasets used in this task exhibited significant class imbalances, various techniques were applied to address this issue. To balance the datasets, a combination of undersampling and other augmentation methods was implemented.

ML_AI_IIITRanchi: (Kumari et al., 2023) This team implemented text classification, the bag-of-words (BoW) model is used, and there are multiple steps in the procedure. The labeled dataset was first prepared by being divided into a training set and a testing set. The text data should be next preprocessed by reducing noise, standardizing the text format, and deleting stopwords. Then, they used a vectorization method, such as CountVectorizer, to construct the BoW representation. This method turns each document into a numerical vector based on word frequencies. To train the classifier, divide the BoW representation into features and labels. Then Select a classification algorithm—Random Forest and AdaBoost, for example—and train the model using the training data. Lastly, make predictions on the testing set to assess the model.

Tercet: (Sivakumar et al., 2023) The method that they have employed for this task is Support Vector machines (SVM). Given a higher precision, F1 score, and weighted averages as compared to models such as random forest, logistic regression, and naïve Bayes models, SVM was a good fit for classifying the given test datasets. The process starts with preprocessing of the text data such as removing punctuation, emoticons, and stop words. To convert the text data into a form that is usable by the model, they used a tf-idf vectorizer algorithm and utilized the data in the SVM model.

Ranganayaki: (EM et al., 2023) The data set is preprocessed to translate emojis, convert text to lowercase, username removal, and extra space removal. The vocabulary of English and Hindi data is formed to correct spelling mistakes in training and testing data using Levenshtein distance. Preprocessed data is converted into fastText embedding of dimension 100x100. The dataset is oversampled using ADASYN oversampling to handle class imbalance. The data is then fed into a capsule network, to form the model, which is used to make predictions.

VTU_BGM: (Sanjana M. Kavatagi and Biradar, 2023) Employing layer differential tuning, the team harnessed the ULMFiT model for advanced feature generation. ULMFiT was ingeniously designed to overcome limited labeled data challenges for specific tasks, unfolding through pretraining and fine-tuning stages. The initial pretraining phase involved training a language model on an extensive, unlabeled text corpus like Wikipedia, grasping universal language patterns and semantics. The proposed method embraced layered fine-tuning of ULMFiT, capitalizing on its core principle: first, pre-trained on general language data, then fine-tuned on task-specific datasets. This approach facilitated adaptation to task intricacies, leading to enhanced performance across a spectrum of NLP endeavors.

6 Result

The submissions received for the classification of English, Bulgarian, Hindi, and Spanish datasets were 6, 5, 5, and 3, respectively. Among these, the team "hate-alert" secured the top rank for both Bulgarian and Hindi languages, achieving a macro average F1 score of 0.75 and 0.68, respectively. They employed transformer-based models, specifically mBERT and XLMR-base, demonstrating the effectiveness of these models. Their approach involved fine-tuning mBERT, utilizing CLS embeddings from both mBERT and XLMR and employing Dense layers along with a classification head. This innovative use of cutting-edge transformer-based models showcased their superiority over earlier deep-learning models like LSRM and CNN-GRU, as depicted in Table 5 and Table 7.

The team "MUCS" also secured the first rank and in the Spanish languages with the macro F1 score of 0.61, which is shown in Table 8. Their journey began with enhancing models for a pivotal project, where they combined BERT embeddings with syllable TF-IDF, followed by integration with TF-IDF and resampling techniques. Through iterative refinement across three runs, their approach demonstrated a continuous progression towards precision, resulting in a notable advancement in the field.

For the English language classification, two teams, namely "Tercet-English" and "ML_AI_IIITRanchi," both achieved the top rank with a macro F1 score of 0.50. "Tercet-English" adopted Support Vector Machines (SVM) as their method of choice, benefitting from higher precision, F1 scores, and weighted averages compared to other models such as random forest, logistic regression, and naive Bayes. Their preprocessing involved text data cleaning and conversion using a TF-IDF vectorizer algorithm, subsequently utilized in the SVM model. On

Table 5: Bulgarian Rank List

Team name	MF1	Rank
hate-alert-run2 (Das et al., 2023)	0.75	1
MUCS_run1 (Hegde et al., 2023)	0.75	1
Team-Tamil (Ponnusamy et al., 2023)	0.69	2
IIC_Team (Vajrobol et al., 2023)	0.65	3
ML_AI_IIITRanchi(1) (Kumari et al., 2023)	0.50	4

Table 6: Engilsh Rank List

Team name	MF1	Rank
Tercet_English (Sivakumar et al., 2023)	0.50	1
ML_AI_IIITRanchi (Kumari et al., 2023)	0.50	1
Ranganayaki (EM et al., 2023)	0.49	2
VTUBGM (Sanjana M. Kavatagi and Biradar, 2023)	0.48	3
IIC_Team (Vajrobol et al., 2023)	0.47	4
MUCS_run2 (Hegde et al., 2023)	0.44	5

the other hand, "ML_AI_IIITRanchi" employed text classification using the bag-of-words (BoW) model. Their process included dividing the labeled dataset into training and testing sets, preprocessing text data, constructing BoW representations using vectorization methods like CountVectorizer, and training the classifier using classification algorithms such as Random Forest and AdaBoost.

These results collectively highlight the effectiveness of various techniques employed by different teams across languages, showcasing advancements in hope speech classification methodologies.

7 Conclusion

This paper provides an overview of the Hope Speech Detection shared task conducted during LT-EDI-RANLP 2023, with a specific focus on four languages: Bulgarian, English, Hindi, and Spanish. The task attracted participation from eight teams, each submitting predictions for evaluation. The resulting rank list, featuring macro F1 scores as outlined in the result section, is presented. In essence, this paper succinctly captures the essence of the LT-EDI-RANLP 2023 Hope Speech Detection shared task. It emphasizes the diverse range of strategies adopted by participating teams and underscores the prominence of machine learning and transformer-based methods, which have contributed to notable enhancements in performance.

References

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.

Yi-Ling Chung, Elizaveta Kuzmenko, Serra Sinem Tekiroglu, and Marco Guerini. 2019. Conan–counter narratives through nichesourcing: a multilingual dataset of responses to fight online hate speech. *arXiv preprint arXiv:1910.03270*.

Mithun Das, Shubhankar Barman, and Subhadeep Chatterjee. 2023. Hope speech detection using transformer-based models. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Ranganayaki EM, Abirami Murugappan, Lysa Packiam R S, and Deivamani M. 2023. Hope speech detection using capsule networks. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Daniel García-Baena, Miguel García-Cumbreras, Salud María Zafra, José García-Díaz, and Rafael Valencia-García. 2023. Hope speech detection in spanish. *Language Resources and Evaluation*, pages 1–28.

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

Table 7: Hindi Rank List

Team name	MF1	Rank
hate-alert-run1 (Das et al., 2023)	0.68	1
IIC_Team (Vajrobol et al., 2023)	0.67	2
MUCS_run1 (Hegde et al., 2023)	0.67	2
Ranganayaki (EM et al., 2023)	0.62	3
ML_AI_IIITRanchi (Kumari et al., 2023)	0.52	4

Table 8: Spanish Rank List

Team Name	Macro F1	Rank
MUCS (Hegde et al., 2023)	0.61	1
IIC_Team (Vajrobol et al., 2023)	0.51	2
hate-alert (Das et al., 2023)	0.49	3

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.

Adeep Hande, Ruba Priyadharshini, Anbukkarasi Sampath, Kingston Pal Thamburaj, Prabakaran Chandran, and Bharathi Raja Chakravarthi. 2021. Hope speech detection in under-resourced kannada language.

Asha Hegde, Kavya G, Sharal Coelho, and Hosahalli Lakshmaiah Shashirekha. 2023. Learning approaches for hope speech detection in social media text. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Salud María Jiménez-Zafra, Miguel Ángel García-Cumbreras, Daniel García-Baena, José Antonio García-Díaz, Bharathi Raja Chakravarthi, Rafael Valencia-García, and L. Alfonso Ureña-López. 2023a. Overview of HOPE at IberLEF 2023: Multilingual Hope Speech Detection. *Procesamiento del Lenguaje Natural*, 71.

Salud María Jiménez-Zafra, Francisco Rangel, and Manuel Montes-y Gómez. 2023b. Overview of IberLEF 2023: Natural Language Processing Challenges for Spanish and other Iberian Languages. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org*.

Prasanna Kumar Kumaresan, Rahul Ponnusamy, Elizabeth Sherly, Sangeetha Sivanesan, and Bharathi Raja Chakravarthi. 2022. Transformer based hope speech comment classification in code-mixed text. In *International Conference on Speech and Language Technologies for Low-resource Languages*, pages 120–137. Springer.

Kirti Kumari, Shirish Shekhar Jha, Zarikunte Kunal Dayanand, and Praneesh Sharma. 2023. Identification of hope speech of youtube comments in mixed languages. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Binny Mathew, Punyajoy Saha, Hardik Tharad, Subham Rajgaria, Prajwal Singhania, Suman Kalyan Maity, Pawan Goyal, and Animesh Mukherjee. 2019. Thou shalt not hate: Countering online hate speech. In *Proceedings of the international AAAI conference on web and social media*, volume 13, pages 369–380.

Rahul Ponnusamy, Malliga Subramaniam, Sajeetha Thavareesan, and Ruba Priyadharshini. 2023. Teamtamil@lt-edi: Automatic detection of hope speech in bulgarian language using embedding techniques. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

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. *arXiv preprint arXiv:2104.09066*.

Pradeep Kumar Roy, Snehaan Bhawal, and Chinnaudayar Navaneethakrishnan Subalalitha. 2022. Hate speech and offensive language detection in dravidian languages using deep ensemble framework. *Computer Speech & Language*, 75:101386.

Rashmi R. Rachh Sanjana M. Kavatagi and Shankar S. Biradar. 2023. Hope speech identification using layered differential training of ulmfit. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

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

Vajratiya Vajrobol, Nitisha Aggarwal, and Karanpreet Singh. 2023. Leveraging pre-trained transformers for fine-grained depression level detection in social media. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.