# SSNCSE_NLP@LT-EDI-ACL2022:Hope Speech Detection for Equality, Diversity and Inclusion using sentence transformers

**Dhanya Srinivasan** **Josephine Varsha** **B. Bharathi** **D. Thenmozhi** **B. Senthil Kumar**
SSN College of Engineering
dhanya2010903@ssn.edu.in
josephine2010350@ssn.edu.in
bharathib@ssn.edu.in
theni_d@ssn.edu.in
senthil@ssn.edu.in

## Abstract

In recent times, applications have been developed to regulate and control the spread of negativity and toxicity on online platforms. The world is filled with serious problems like political & religious conflicts, wars, pandemics, and offensive hate speech is the last thing we desire. Our task was to classify a text into 'Hope Speech' and 'Non-Hope Speech'. We searched for datasets acquired from YouTube comments that offer support, reassurance, inspiration, and insight, and the ones that don't. The datasets were provided to us by the LTEDI organizers in English, Tamil, Spanish, Kannada, and Malayalam. To successfully identify and classify them, we employed several machine learning transformer models such as m-BERT, MLNet, BERT, XLMRoberta, and XLM_MLM. The observed results indicate that the BERT and m-BERT have obtained the best results among all the other techniques, gaining a weighted F1-score of 0.92, 0.71, 0.76, 0.87, and 0.83 for English, Tamil, Spanish, Kannada, and Malayalam respectively. This paper depicts our work for the Shared Task on Hope Speech Detection for Equality, Diversity, and Inclusion at LTEDI 2021.

## 1 Introduction

With an evolving and diversifying world filled with worries and uncertainty, people turn to religion and faith to give them hope. With the existence of marginalized communities, such as the LGBTQIA community, and racial and ethnic minorities which focus on having faith and are hopeful for their complete acceptance in society(Puranik et al., 2021), there is an increasing need for positive reinforcement in society. With the wide usage of the internet, now intensified by the ongoing pandemic, more people are seeking the same kind of reinforcement through online forums(Arunima et al., 2021).

Youtube is an online platform connecting billions of users across the globe. It is an application owned by Google, which allows people worldwide to showcase their talents and express opinions, and connect in the comments section. With around 30000 hours of content uploaded every hour(Arunima et al., 2021), each comment under it has people expressing not only their trivial thoughts but also their controversial opinions. This includes but is not restricted to saying hurtful things and shaming communities and groups they harbor ill will against, given the flexibility of speech in most nations in the world. This can be touted as a bane as well leading to rigorous research in Offensive Speech Detection and Hate Speech Detection(Sai and Sharma, 2020)(Wani et al., 2019)(Alsafari et al., 2020).

There has been inadequate research done concentrating on Hope Speech Detection. With this being an era of mental health awareness, it is crucial to develop a solution that recommends uplifting and positive tweets and posts to people, while sidelining the negative, discouraging and disheartening ones.

In our paper, we have approached Hope Speech Detection using pre-existing transformer models trained from the dataset provided by the LT-EDI organizers with the data obtained from tweets in English, Kannada, Malayalam, Spanish, and Tamil. We have used multilingual transformers such as XLM-MLM, BERT, and XLM-ROBERTA, achieving promising results using the BERT multilingual transformer model. Hence the task was implemented using the same.

The rest of this paper is organized as follows. Section 2 discusses the related work on Hope Speech Detection tasks. The dataset for the shared task is discussed in Section 3. Section 4 outlines the features and the methods used for this task. Results are presented in Section 5. Section 6 concludes the paper.



*Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 218 - 222
May 27, 2022 ©2022 Association for Computational Linguistics

## 2 Related Work

Researchers have experimented with a few approaches to deal with Hope Speech Detection in many languages recently, but more so in English. The authors of Hope Speech Detection: A Computational Analysis of the Voice of Peace(Palakodety et al., 2019) take the help of polyglot word-embeddings to discover language clusters and subsequently construct a language identification technique that requires minimal supervision and performs well on short social media texts generated in a linguistically diverse region of the world. In Hope Speech Detection Using Indic Transliteration and Transformers(Upadhyay et al., 2021), the authors used 2 approaches. They used contextual embeddings to train classifiers using logistic regression, random forest, SVM, and LSTM based models. A similar approach is used in Hope Speech Detection in YouTube multilingual comments(Saumya and Mishra, 2021). The second approach involved using a majority voting ensemble of 11 models which were obtained by fine-tuning pre-trained transformer models (BERT, ALBERT, RoBERTa, IndicBERT) after adding an output layer. They found that the second approach yielded better results for English, Tamil, and Malayalam. In A Multilingual Hope Speech Detection for Equality, Diversity, and Inclusion using Context-Aware Embedding(Junaida and Ajees, 2021), the authors present deep learning techniques using context-aware string embeddings for word representations and Recurrent Neural Network (RNN) and pooled document embeddings for text representation. In this paper, however, we use pre-trained multilingual transformer models to determine whether a comment is Hope Speech or not.

## 3 Dataset Analysis and Preprocessing

The dataset provided by the LT-EDI 2021 Chakravarthi et al. (2022) for the five languages, English, Tamil, Spanish, Kannada, and Malayalam consisted of 28424, 17716, 1653, 6176, and 9918 Youtube comments respectively. Refer to the table below 1

Subjects like Hope's speech might raise confusion and disagreement among literates belonging to different groups. Puranik et al. (2021). Chakravarthi et al. (2022)

A Youtube comment may contain acronyms, small words, and emojis. It is required to process the data before training it. Data pre-processing is critical for the success of any machine learning solution. There may be signs of irregularities in the continuity of texts and misspelled words in many YouTube comments. For the cleaning up of the dataset and to normalize these irregularities we go through pre-processing, where all the HTML tags, hashtags, social media mentions, and URLs are removed. It is also required to annotate emojis and emoticons as they play an important role in defining the speech. These are replaced with the text they represent and substituted back into the comment. The text data may contain short words, and these are replaced with their original full word. We resort to a look-up table that replaces the short word with their expanded form, such as: 'what's' with 'what is', 'u' with 'you'. The sequence of texts is then converted to lowercase and the extra unwanted white spaces are removed.Suseelan et al. (2019)Thenmozhi et al. (2019) (Bharathi et al., 2021).

| Language | Training | Development | Test |
|---|---|---|---|
| English | 22740 | 2841 | 2843 |
| Tamil | 14200 | 1755 | 1761 |
| Spanish | 991 | 331 | 331 |
| Kannada | 4940 | 618 | 618 |
| Malayalam | 7873 | 974 | 1071 |

Table 1: Dataset description

## 4 Methodology

For this approach, we used a few models, namely, the XLM-MLM models, BERT models, and XLM-ROBERTA models for Spanish and Malayalam and the BERT models and XLNET models for English, Kannada and Tamil. Out of all these models, the BERT models produced the best results out of all the models, in all the languages.

For English, bert-base-multilingual-uncased is shown to yield the best accuracy of 93%. For Kannada, bert-base-uncased yielded the best accuracy of 87%. For Malayalam, bert-base-multilingual-uncased yielded a result of 84%. For Spanish too, the bert-base-multilingual-uncased model outperformed the others by giving an accuracy of 76%. For Tamil, bert-base-uncased yielded an accuracy of 72%. An analysis of all the results is elaborated in the next section.

### 4.1 bert-base-multilingual-uncased and bert-base-uncased

Introduced in the paper BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding(Devlin et al., 2018), it is a pre-trained model trained on the top 102 languages with the largest Wikipedia using a masked language modeling (MLM) objective. It is a transformer model pre-trained on a large corpus of multilingual data in a self-supervised fashion. BERT uses bi-directional learning to gain context of words from left to right context simultaneously. This bi-directional approach is optimized for Masked Language Modeling(MLM), which includes randomly masking 15% of the words in the input and then running it through the model to predict the masked words. It also helps to optimize Next Sentence Prediction(NSP) which predicts the relationship between two sentences(whether they follow each other or not).

### 4.2 xlnet-base-cased

XLNet is a model pre-trained in the English language. Introduced in the paper XLNet: Generalized Autoregressive Pretraining for Language Understanding(Yang et al., 2019), it is a new unsupervised language representation learning method based on a novel generalized permutation language modeling objective. It employs Transformer-XL as the backbone model, exhibiting exemplary performance for language tasks involving long context. It achieves great performance on downstream tasks such as document ranking, question answering, sentiment analysis, and natural language interference. It is primarily aimed at being fine-tuned on tasks that use the whole sentence (potentially masked) to make decisions, such as sequence classification, token classification, or question answering.

### 4.3 xlm-mlm-tlm-xnli15-1024 and xlm-mlm-100-1280

XLM is a model presented by Facebook AI in the paper Cross-lingual Language Model Pretraining(Lample and Conneau, 2019). It is an improved version of BERT, achieving excellent results in classification and translation tasks in Natural Language Processing. XLM uses a known preprocessing technique (BPE) and a dual-language training mechanism with BERT to learn relations between words in different languages. The model outperforms other models in a cross-lingual classification task and significantly improves machine translation when a pre-trained model is used for the initialization of the translation model. Here, the initial embeddings of the tokens are taken from a pretrained MLM and fed into the translation model.

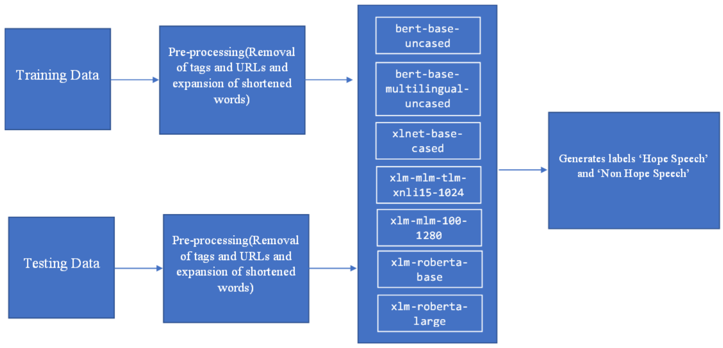


Figure 1: Proposed methodology

These embeddings are used to initialize the tokens of both the encoder and the decoder of the translation model (which uses a transformer)(Lample et al., 2018).

### 4.4 xlm-roberta-base and xlm-roberta-large

First introduced in the paper Unsupervised Cross-lingual Representation Learning at Scale(Conneau et al., 2019), it is a multilingual version of RoBERTa pre-trained on 2.5TB of filtered CommonCrawl data containing 100 languages. It is a transformers model pre-trained on a large corpus in a self-supervised fashion. Similar to BERT, it is pre-trained with the Masked Language Modeling(MLM) objective.

## 5 Observation

In this section, we will be looking into the performance of various machine learning transformer models for the 5 languages (English, Tamil, Spanish, Kannada, Malayalam). The weighted F1 score determines the excellence of the models. The tables below present the evaluation results of all the models on the test dataset. The English dataset has the highest F1 score of 0.92 using the m-BERT model, winning over a slight margin against the BERT model. The BERT model had outperformed by its accuracy 2. In the case of the Tamil language, the BERT model had produced the highest F1-Score of 0.71 with better accuracy than the other models 3. In the case of Spanish, both the m-BERT and the XLM ROBERTA had performed equally well with an F1 score of 0.76 4. BERT model had performed well with Kannada datasets too with an F1 score of 0.87 and an equally good accuracy of 0.87 5. In the case of Malayalam, m-BERT outperformed the other models, and its

| Pre-trained model | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| bert-base-uncased | 0.91 | 0.92 | 0.91 | 0.92 |
| xlnet-base-cased | 0.83 | 0.91 | 0.87 | 0.91 |
| bert-base-mulitingual-uncased | 0.91 | 0.93 | 0.92 | 0.93 |

Table 2: Performance analysis of the proposed system using development data for English

| Pre-trained model | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| bert-base-uncased | 0.72 | 0.72 | 0.71 | 0.72 |
| xlnet-base-cased | 0.31 | 0.55 | 0.40 | 0.55 |
| bert-base-mulitingual-uncased | 0.70 | 0.70 | 0.69 | 0.70 |

Table 3: Performance analysis of the proposed system using development data for Tamil

F1 score was 0.83 6. For all the languages, the BERT model had showcased the best results and outperformed all the other models.

## Conclusion

The need for Hope Speech Detection in social media content is growing to be more and more important every day. With more and more people's lives being ridden by social media content every day, it becomes critical to have a filter to differentiate between the positive and negative content to promote optimism and a can-do attitude instead of a pessimistic and dispirited outlook. Hope Speech Detection models, though proven to be essential, have had insufficient work done on it. In this paper, we use pre-trained multilingual transformer models to detect Hope Speech in 5 languages, namely English, Kannada, Malayalam, Spanish and Tamil. The model submitted for the task is BERT which proved to be the best out of all the transformer models used. It yielded an accuracy of 93%, 87%, 84%, 76% and 72% for English, Kannada, Malayalam, Spanish and Tamil respectively. BERT's capabilities extend to making more accurate predictions when dealing with newer documents even when the type of document differs significantly in key properties such as length and vocabulary. This attribute of BERT makes it the perfect choice when dealing with multiple languages and code-switching within text.(Arunima et al., 2021) This model can be further improved to deal with data with multiple languages in the future.

| Pre-trained model | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| xlm-mlm-tlm-xnli15-1024 | 0.71 | 0.71 | 0.71 | 0.71 |
| xlm-mlm-100-1280 | 0.24 | 0.49 | 0.32 | 0.49 |
| bert-base-mulitingual-uncased | 0.76 | 0.76 | 0.76 | 0.76 |
| xlm-roberta-base | 0.76 | 0.76 | 0.76 | 0.76 |
| xlm-roberta-large | 0.26 | 0.51 | 0.35 | 0.51 |

Table 4: Performance analysis of the proposed system using development data for Spanish

| Pre-trained model | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| bert-base-uncased | 0.87 | 0.87 | 0.87 | 0.87 |
| xlnet-base-cased | 0.73 | 0.74 | 0.73 | 0.74 |
| bert-base-mulitingual-uncased | 0.86 | 0.83 | 0.83 | 0.83 |

Table 5: Performance analysis of the proposed system using development data for Kannada

| Pre-trained model | Precision | Recall | F1-score | Accuracy |
|---|---|---|---|---|
| xlm-mlm-tlm-xnli15-1024 | 0.65 | 0.80 | 0.72 | 0.80 |
| bert-base-multilingual-uncased | 0.83 | 0.84 | 0.83 | 0.84 |
| xlm-roberta-base | 0.80 | 0.83 | 0.81 | 0.83 |

Table 6: Performance analysis of the proposed system using development data for Malayalam

## References

Safa Alsafari, Samira Sadaoui, and Malek Mouhoub. 2020. Hate and offensive speech detection on arabic social media. *Online Social Networks and Media*, 19:100096.

S Arunima, Akshay Ramakrishnan, Avantika Balaji, D Thenmozhi, et al. 2021. ssn_dibertsity@ lt-edi-eacl2021: hope speech detection on multilingual youtube comments via transformer based approach. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 92–97.

B Bharathi et al. 2021. Ssncse_nlp@ dravidianlangtech-eacl2021: Offensive language identification on multilingual code mixing text. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 313–318.

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. 2022. 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.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

MK Junaida and AP Ajees. 2021. Ku_nlp@ lt-edi-eacl2021: a multilingual hope speech detection for equality, diversity, and inclusion using context aware embeddings. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 79–85.

Guillaume Lample and Alexis Conneau. 2019. Cross-lingual language model pretraining. *arXiv preprint arXiv:1901.07291*.

Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018. Phrase-based & neural unsupervised machine translation. *arXiv preprint arXiv:1804.07755*.

Shriphani Palakodety, Ashiqur R KhudaBukhsh, and Jaime G Carbonell. 2019. Hope speech detection: A computational analysis of the voice of peace. *arXiv preprint arXiv:1909.12940*.

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*.

Siva Sai and Yashvardhan Sharma. 2020. Siva@ hasoc-dravidian-codemix-fire-2020: Multilingual offensive speech detection in code-mixed and romanized text. In *FIRE (Working Notes)*, pages 336–343.

Sunil Saumya and Ankit Kumar Mishra. 2021. Iiit_dwd@ lt-edi-eacl2021: hope speech detection in youtube multilingual comments. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 107–113.

Angel Suseelan, S Rajalakshmi, B Logesh, S Harshini, B Geetika, S Dyaneswaran, S Milton Rajendram, and TT Mirnalinee. 2019. Techssn at semeval-2019 task 6: Identifying and categorizing offensive language in tweets using deep neural networks. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 753–758.

D Thenmozhi, Aravindan Chandrabose, Srinethe Sharavanan, et al. 2019. Ssn_nlp at semeval-2019 task 3: Contextual emotion identification from textual conversation using seq2seq deep neural network. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 318–323.

Ishan Sanjeev Upadhyay, Anshul Wadhawan, Radhika Mamidi, et al. 2021. Hopeful_men@ lt-edi-eacl2021: Hope speech detection using indic transliteration and transformers. *arXiv preprint arXiv:2102.12082*.

Abid Hussain Wani, Nahida Shafi Molvi, and Sheikh Ishrah Ashraf. 2019. Detection of hate and offensive speech in text. In *International Conference on Intelligent Human Computer Interaction*, pages 87–93. Springer.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32.