TERCET@LT-EDI-RANLP2023: Hope Speech Detection for Equality, Diversity, and Inclusion

Priyadharshini T
SSN College of Engineering
priyadharshini2210228@ssn.edu.in

Samyuktaa Sivakumar
SSN College of Engineering
samyuktaa2210189@ssn.edu.in

S Shwetha
SSN College of Engineering
shwetha2210210@ssn.edu.in

Durairaj Thenmozhi
SSN College of Engineering
thenid@ssn.edu.in

B. Bharathi
SSN College of Engineering
bharathib@ssn.edu.in

Gayathri G L
SSN College of Engineering
gayathri2010090@ssn.edu.in

Abstract

Hope is a cheerful and optimistic state of mind which has its basis in the expectation of positive outcomes. Hope speech reflects the same as they are positive words that can motivate and encourage a person to do better. Non-hope speech reflects the exact opposite. They are meant to ridicule or put down someone and affect the person negatively. The shared Task on Hope Speech Detection for Equality, Diversity, and Inclusion at LT-EDI - RANLP 2023 was created with data sets in English, Spanish, Bulgarian and Hindi. The purpose of this task is to classify human-generated comments on the platform, YouTube, as Hope speech or non-Hope speech. We employed multiple traditional models such as SVM (support vector machine), Random Forest classifier, Naïve Bayes and Logistic Regression. Support Vector Machine gave the highest macro average F1 score of 0.49 for the training data set and a macro average F1 score of 0.50 for the test data set. We ranked 1st for this task in the English language.

1 Introduction

With the world around us developing at an incomprehensible speed and the advent of the 'Digital Age,' the way we communicate has fundamentally evolved. Technology has become a pivotal role in ensuring connectivity among humans in the form of calls, messages and most importantly, social media. Social media is a platform for people to create, share and collaborate their ideas, information and opinions with each other virtually. The number of users only keeps increasing and these platforms show no signs of slowing down. (Drus and Khalid, 2019) Social media is a platform of the masses, for the masses.

While social media is a great tool to enable, entertain and equip people, it has its own drawbacks. Not every user has positive intentions while using popular platforms like Facebook and Twitter. This may lead to presence of unwanted content and comments that can range from discouragement to targeted hate speech. Detrimental speech on such platforms have an extended psychological impact on the victim of such comments (Gongane et al., 2022). To ensure the cleanliness and positive environment of such platforms, it is quintessential to identify such texts and moderate them.

Applying sentiment analysis using data from social media platforms is a great tool to analyse and understand the feelings of the general population (Chauhan et al., 2021). Sentiment analysis is a process where multiple people's opinions, views, feelings and emotions are analysed based on different projects, topics and general discourse. It is also known as opinion mining (Wankhade et al., 2022). Sentiment analysis is an incredibly useful tool to automate the detection of negative comments on social media by analysing the sentiments from the user-generated text. Both supervised and unsupervised learning can be applied to perform sentiment analysis. The only drawback of performing supervised learning is that high-quality training data with proper labels are required for the model to predict accurately. Due to the non-requirement of a labelled training data, unsupervised learning is more robust and can be used more widely to perform sentiment analysis (Neri et al., 2012).

The given task is related to detecting Hope speech and non-Hope speech from user-generated text on the popular video-sharing platform, YouTube. Hope being a positive sentiment, caters to increasing the general positivity of the social me-

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion, pages 257–261, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_039

dia platform and thus improving the mood of users alike.(Chakravarthi, 2020) Hope speech therefore contributes the same and helps motivate, encourage and inspire individuals on the platform. Non-hope speech being a negative sentiment has an adverse impact on the social media environment by propagating negative feelings through the users. Detecting and moderating these comments is imperative to create an inclusive and safe space for millions of users to share a part of themselves.(Chakravarthi et al., 2022)

The text in the given text is in the form of code-mixed data. Usually, on social media, most data is not grammatical in nature and has some words and phrases from the native language but in non-native scripts. (Chakravarthi et al., 2020) In our data set, the text is code-mixed in English-Tamil. This can be attributed to the ease of typing in the Roman script on social media while conveying the same intended sentiment. (Patra et al., 2018)Being able to write code-mixed text on social media provides users with a wider choice to express themselves freely and more accurately.

The paper is organized as follows: section 2 pertains to related works as per the literature survey; section 3 is related to the task and data description; section 4 pertains to the methodology used to perform this task; section 5 shows the results and analysis of the results and section 6 entails the conclusion.

2 Related Works

Opinion mining or sentiment analysis is a growing field with increasing applications on social media and e-commerce platforms. To cater to the same, extensive research is going on in this field to build the most efficient and robust models which range from Multi-Layer Models (MLMs) to Natural Language Processing(NLP) models. Research conducted by (Vijayakumar et al., 2022) used the transformer model, ALBERT for doing hope speech detection in the English, Tamil, Malayalam and Kannada language.

A new Convolutional Neural Network(CNN) based model was proposed by (Chakravarthi, 2022) that outperformed other traditional models for detecting hope-speech. Both binary hope speech classification as well as multi-class hope speech classification was performed by (Balouchzahi et al., 2022). The binary task involved only two labels whereas the multi-class task involved three labels. Multiple traditional, transformer and deep learning models were applied on the dataset.

A logistic regression classifier was applied by (Palakodety et al., 2019) with a L2 regularization whose results indicated that a hope-speech classifier with good precision and recall can be constructed.

Models based on Long Short Term Memory (LSTM) network, deep learning and hybrid learning on the Tamil and Malayalam language dataset were used by (Saumya and Mishra, 2021). The best performing model on the English dataset was the 2-parallel CNN-LSTM that used GloVe and Word2Vec embeddings. The best performing model on the Malayalam dataset was the 3-parallel Bi-LSTM.

Multiple transformer models for hope speech detection in the English, Tamil and Malayalam languages were applied by (Ghanghor et al., 2021) . The models applied were the multilingual m-BERT-cased, XLM-Roberta(XLMR), and IndicBERT. Among these models, the m-BERT-cased model gave the best F1-score.

Hope detection was done by (Chinnappa, 2021) for three language data sets in Tamil, English and Malayalam. They applied language models that include Compact Language Detector 2, Compact Language Detector 3, langid, textblob language detector, and langdetect. The experimental results from the same showed that detecting hope detection from text is a difficult task especially with code-mixed data.

These papers exhibit the versatility of the models that can be applied to perform sentiment analysis on comments from social media. Based on literature survey, we decided to apply traditional models with the application of a simple transformer model for a more accurate classification of text.

3 Task and Data Description

The shared Task on Hope Speech Detection for Equality, Diversity, and Inclusion at LT-EDI - RANLP 2023(Kumaresan et al., 2023) is intended to determine whether the text was Hope speech or non-Hope speech. The data set was available for four languages, English, Hindi, Bulgarian and Spanish, but we submitted the English data set only. The data set consisted of two fields: Text and Labels which were gathered from YouTube comments. The training data set consisted of around 18191 texts out of which 16630 were labelled as non-Hope speech and 1561 were labelled as Hope speech. The development data set for English consisted of 4547 comments out of which 4148 were labelled as non-Hope speech and 399 were labelled as Hope speech. The test data had 4805 texts out of which 4783 were classified as non-Hope speech and 22 were classified as Hope speech.

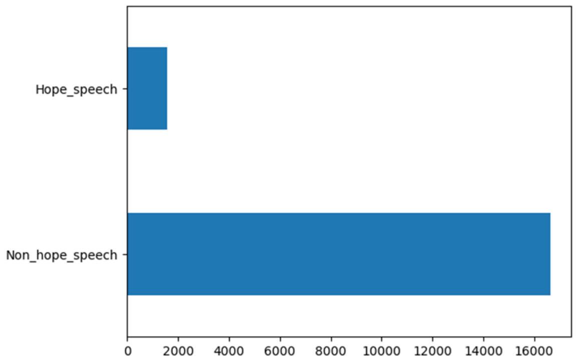

Figure 1: Distribution of data

Label	Example	Instances in Train	Instances in Dev
Hope-Speech	Totally agree! All Lives matter!	1561	399
non-Hope Speech	Sadly, slaves were hunted by rival tribes in Africa	16630	4148

Data Description

Figure 2: Description of Data

4 Methodology

Multiple traditional models were employed to identify which texts were hope speech from the YouTube comments given in the data set.

4.1 Data Preprocessing and Cleaning

To convert the raw data into readable data for the model, we performed various data cleaning processes.

1)Firstly, all punctuation and emoticons were removed from the text. All the text data was then converted to lower case to create uniformity in the data set. It is important to remove signs and emoticons as the model solely focuses on the words themselves and has no requirement for punctuation and emoticons.

2)It is important to remove stop words as they are redundant words that do not have any significant contributions to the sentiments of the analysed texts. Using the nltk library, we removed the English stop words as well as extended the list to include stop words in Tamil. This was done as the text consisted of words in Tamil(code-mixed) in addition to English.

3)Machine learning models generally take mathematical inputs in the form of numbers or 2D-arrays. Considering that the existing data is in the form of raw text, it is necessary to transform them into a vector. The vectorizer we employed is the Term frequency-inverse document frequency (TF-IDF) vectorizer to transform the given texts into its vector form. Here, term frequency refers to how important a specific term is in a document whereas inverse document frequency refers to the weight of the term. The weight is reduced if the term is scattered all over the document.

4.2 Feature Extraction

The Language Agnostic BERT Sentence Embedding model(Feng et al., 2022) is a model released by Google which is based on the BERT model. It focuses on Bi-text mining and sentence embedding tasks. The tokenization algorithm that is used by LaBSE is WordPiece which was developed by Google to pretrain the BERT model. LaBSE's architecture is a dual-encoder model with two encoders that encode source and target sentences independently. The encodings are then passed to a scoring function where it is ranked based on similarity. (Miłkowski et al., 2022) used LaBSE to aid in classification of sentiment polarization. It was proved that language-agnostic representations are quite efficient. The best results were obtained for the LaBSE embeddings and the same was implemented in their online service. We applied the LaBSE model for embedding the preprocessed text. These embeddings were fed to the classifier model for classifying the text based on the labels of emotions associated with them.

4.3 Classification model

To classify the text data, we experimented with multiple traditional models that include SVM, Random Forest, Naïve Bayes and Logistic regression as well as the simple transformer model, LaBSE. After evaluating the metrics of multiple models, we focused on combining the LaBSE feature extraction model along with the SVM classifier. Support Vector Machine is a popular supervised learning algorithm used mainly in classification problems. It operates by creating a decision boundary that separates n-dimensional spaces into classes so that a new data point can be assigned to its relevant category.

5 Results and Analysis

5.1 Performance Metrics:

The sklearn metrics library provides the classification report for evaluation of the performance of the model. It consists of the following metrics:

1) Precision: Precision is defined as the ratio of true positives to sum of true and false positives.

$$Precision = \frac{TP}{TP+FP}$$

2) Recall: Recall is defined as the ratio of true positives to sum of true positives and false negatives.

$$Recall = \frac{TP}{TP+FN}$$

3) F1-Score: The F1 is the weighted harmonic mean of precision and recall. The closer the value of F1 is to 1, better is the performance of the model.

$$F1\text{-}score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

4) Support: Support is the number of actual occurrences of the class in the data set.

5.2 Results

The classification report was applied on the development data after training the model using the training data.

It is observed that the overall accuracy on the development data is 0.90 which is higher compared to the 0.68 accuracy of the Naïve Bayes model and 0.68 accuracy of the Logistic Regression model.

Upon testing the accuracy of the data on the unlabelled test data against the labelled test data (released after evaluation), the accuracy score was found to be 0.99105. The rank list released by the organizers of the task ranked our model at 1st position with a macro F1 score of 0.50.

	precision	recall	f1-score	support
0	0.91	0.98	0.95	4152
1	0.08	0.02	0.03	396
accuracy			0.90	4548
macro avg	0.50	0.50	0.49	4548
weighted avg	0.84	0.90	0.87	4548

Figure 3: Classification Report for SVM on Dev Data

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4784
1	0.00	0.00	0.00	21
accuracy			0.99	4805
macro avg	0.50	0.50	0.50	4805
weighted avg	0.99	0.99	0.99	4805

Figure 4: Classification Report for SVM on Test Data

6 Conclusion

Through the scope of this paper, we have explored and presented a traditional model coupled with a simple sentence transformer model (LaBSE) to perform classification of Hope-speech and non-Hope speech on the given data by DravidianLangTech in the English language. It was noted that SVM gave the best performance metrics against other traditional models with a macro F1 score of 0.50. It is our belief that the classification can be improved by applying deep learning models on the given data set to obtain a higher accuracy.

References

Fazlourrahman Balouchzahi, Grigori Sidorov, and Alexander Gelbukh. 2022. Polyhope: Dataset creation for a two-level hope speech detection task from tweets. *arXiv preprint arXiv:2210.14136*.

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. 2022. Hope speech detection in youtube comments. *Social Network Analysis and Mining*, 12(1):75.

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.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, and John P McCrae. 2020. Corpus creation for sentiment analysis in code-mixed tamil-english text. *arXiv preprint arXiv:2006.00206*.

Priyavrat Chauhan, Nonita Sharma, and Geeta Sikka. 2021. The emergence of social media data and sentiment analysis in election prediction. *Journal of Ambient Intelligence and Humanized Computing*, 12:2601–2627.

Dhivya Chinnappa. 2021. dhivya-hope-detection@LT-EDI-EACL2021: Multilingual hope speech detection for code-mixed and transliterated texts. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 73–78, Kyiv. Association for Computational Linguistics.

Zulfadzli Drus and Haliyana Khalid. 2019. Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, 161:707–714.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic bert sentence embedding.

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.

Vaishali U Gongane, Mousami V Munot, and Alwin D Anuse. 2022. Detection and moderation of detrimental content on social media platforms: current status and future directions. *Social Network Analysis and Mining*, 12(1):129.

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.

Piotr Miłkowski, Marcin Gruza, Przemysław Kazienko, Joanna Szołomicka, Stanisław Woźniak, and Jan Kocoń. 2022. Multi-model analysis of language-agnostic sentiment classification on multiemo data. In *International Conference on Computational Collective Intelligence*, pages 163–175. Springer.

Federico Neri, Carlo Aliprandi, Federico Capeci, and Montserrat Cuadros. 2012. Sentiment analysis on social media. In *2012 IEEE/ACM international conference on advances in social networks analysis and mining*, pages 919–926. IEEE.

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*.

Braja Gopal Patra, Dipankar Das, and Amitava Das. 2018. Sentiment analysis of code-mixed indian languages: An overview of sail_code-mixed shared task@ icon-2017. *arXiv preprint arXiv:1803.06745*.

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.

Praveenkumar Vijayakumar, S Prathyush, P Aravind, S Angel, Rajalakshmi Sivanaiah, Sakaya Milton Rajendram, and TT Mirnalinee. 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.

Mayur Wankhade, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. 2022. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7):5731–5780.