

IIT Dhanbad @LT-EDI-ACL2022- Hope Speech Detection for Equality, Diversity, and Inclusion

Vishesh Gupta

Department of Computer
Science and Engineering ,
Indian Institute of Technology
(Indian School of Mines),
Dhanbad, India

me.guptavishesh@gmail.com

Ritesh Kumar

Department of Computer
Science and Engineering,
National Institute of
Technology Jamshedpur, India

ritesh.cse@nitjsr.ac.in

Rajendra Pamula

Department of Computer
Science and Engineering ,
Indian Institute of Technology
(Indian School of Mines),
Dhanbad, India

rajendrapamula@gmail.com

Abstract

Hope is considered significant for the well-being, recuperation and restoration of human life by health professionals. Hope speech reflects the belief that one can discover pathways to their desired objectives and become roused to utilise those pathways. Hope speech offers support, reassurance, suggestions, inspiration and insight. Hate speech is a prevalent practice that society has to struggle with everyday. The freedom of speech and ease of anonymity granted by social media has also resulted in incitement to hatred. In this paper, we work to identify and promote positive and supportive content on these platforms. We work with several machine learning models to classify social media comments as hope speech or non-hope speech in English. This paper portrays our work for the Shared Task on Hope Speech Detection for Equality, Diversity, and Inclusion at LT-EDI-ACL 2022.

1 Introduction

Nowadays, social media has become a significant part of our lives and just like everything it has its pros and cons. Various benefits of social media come with several challenges including hate speech, offensive and profane content getting published targeting an individual, a group or a society. Social media has become a breeding ground for hate speech and cyberbullying (Chakravarthi, 2020; Chakravarthi and Muralidaran, 2021). Offensive content in online socialization have seriously affected daily life of people (Priyadharshini et al., 2021; Kumaresan et al., 2021; Chakravarthi et al., 2020). Social media companies such as, YouTube, Facebook, and Twitter have their own approaches to eliminate the hate speech content or anything which negatively affects the society. However, detecting such objectionable content at the earliest to curb the menace of spreading such news online is still a major challenge faced by social media companies and researchers (Chakravarthi

et al., 2021). It is very essential to detect such behaviour. The amount of data generated on social media sites can be estimated from the fact that, every second, on average, around 6,000 tweets are generated (Sakuntharaj and Mahesan, 2021, 2017, 2016; Thavareesan and Mahesan, 2019, 2020a,b, 2021). Content moderation of such a huge data is difficult to achieve exclusively through man power. Social networking sites are struggling with content moderation (Sampath et al., 2022; Ravikiran et al., 2022; Chakravarthi et al., 2022; Bharathi et al., 2022; Priyadharshini et al., 2022). Our work aims to change the prevalent way of thinking by moving away from a preoccupation with discrimination, loneliness or the worst things in life to building the confidence, support and good qualities based on comments by individuals.

In this paper, we have explored several machine learning models for classification of social media comments as hope speech or non-hope speech in English.

2 Related Works

Several works have been proposed to detect hope speech across social platforms. (Puranik et al., 2021) proposed a work with several transformer-based models to classify social media comments as hope speech or not hope speech in English, Malayalam and Tamil languages. (Ghanghor et al., 2021) have used the transformer-based pretrained models along with the customized versions of those models for detecting hope and not hope speech for equality, diversity and inclusion in Dravidian languages. (Upadhyay et al., 2021) experimented with two approaches. They used contextual embeddings to train classifiers using logistic regression, random forest, SVM, and LSTM based models. They also used 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.

(Saumya and Mishra, 2021) proposed various machine learning and deep learning-based models (such as support vector machine, logistics regression, convolutional neural network, recurrent neural network) are employed to identify the hope speech in the given YouTube comments. The YouTube comments are available in English, Tamil, and Malayalam languages.

(Vijayaraghavan et al., 2021) proposed a deep neural multi-modal model that can: (a) detect hate speech by effectively capturing the semantics of the text along with socio-cultural context in which a particular hate expression is made, and (b) provide interpretable insights into decisions of their model. (Gomez et al., 2020) target the problem of hate speech detection in multimodal publications formed by a text and an image. They gather and annotate a large scale dataset from Twitter, MMHS150K, and propose different models that jointly analyze textual and visual information for hate speech detection, comparing them with unimodal detection.

(Chang, 1998) shows the influence of high versus low hope on problem-solving ability of college students. It shows that high-hope students were found to have greater problem-solving abilities than low-hope students. (Youssef and Luthans, 2007) shows the impact of hope, optimism, and resilience in the workplace. The outcomes of their work include performance, job satisfaction, work happiness, and organizational commitment. (Snyder and P) shows development and validation of an individual-differences measure of hope.

3 Task and Dataset Description

Here we have described the dataset and task provided by Hope Speech Detection for Equality, Diversity, and Inclusion challenge.

This is a comment / post level classification task. In this, Youtube comments are given and the systems submitted by us should classify it into 'Hope speech' and 'Not hope speech'. (shown in Table 1).

Here training, development and test data is given in English. Distributions of these data is shown in Table 2. The distributions of imbalanced classes in training data is shown in Table 3.

- Hope Speech (HS): Posts that offer support, reassurance, suggestions, inspiration and insight.

- Non Hope Speech (NHS): Posts that explicitly seeks violence and uses gender-based insults.

4 Methodology

4.1 Data Preprocessing

We have performed following steps in data preprocessing :-

- Punctuations, links and numbers removal.
- Lower the letter case.
- Tokenization.
- Turning the texts into sequences.
- Pad the sequences to have the same size.
- Balancing the given imbalanced dataset.

We have used Tokenizer class in TensorFlow for handling above process. The unknown token (UNK) is used when what remains of the token is not in the vocabulary, or if the token is too long. We have used post padding to pad the sequences. We have balanced the imbalanced classes of training data using Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002) which uses KNN for balancing minority classes. Balanced training data is shown in Table 4.

4.2 Models Proposed

We have used various machine learning algorithms, namely- Logistic Regression (Wright, 1995), Multinomial Naive Bayes classifier (Kibriya et al., 2004), Random forest classifier (Liaw et al., 2002) and XGBoost (Ren et al., 2017). We have used the scikit-learn library for logistic regression, MultinomialNB and Random forest classifier. We have used the following values of the parameter :

- In Random Forest, we have used n estimators=1000 and random state=42.
- In XGBoost, we have used learning rate=0.01, max depth=50 and n estimators=300.

All the models have used balanced pre-processed training data for training and we have tested the models on the test data provided in challenge.

Text	Category
@Champions Again He got killed for using false money	Non hope speech
It's not that all lives don't matter	Non hope speech
she is not 60. He is 60	Non hope speech
I'm still hiding my gender to my parents and they don't know I'm dating someone.	Hope speech
Sasha Dumse God accepts everyone.	Hope speech
all lives matter .without that we never have peace so to me forever all lives matter.	Hope speech

Table 1: Examples of hope speech or not hope speech

Type	English
Training	22739
Development	2841
Test	2843
Total	28423

Table 2: Train-Development-Test Data Distribution

Classes	Counts
Non hope Speech	20777
Hope Speech	1962
Total	22739

Table 3: Imbalanced classes distribution in training data

5 Result and Discussions

The results of task are represented in terms of Accuracy, Macro-F1, Micro-F1 and Weighted-F1 (shown in Table 5). The best score as Macro-F1 for the task we get is 0.6130. The XGBoost system have performed better than all other models. There is imbalance between the classes of test data due to which there is more differences between accuracy and Macro-F1 score of each system.

6 Conclusions and Future Work

We have completed the task using various classification algorithms and evaluated the performance of different classification algorithms for Hope Speech Detection for Equality, Diversity, and Inclusion shared task. Our overall best score is 0.6130. We look forward to experimenting with different advance algorithm or neural network models. We are also looking forward to work on random multi model classification algorithm for better accuracy and classification. Also, fine tuning the parameters of the algorithm can help in improvement of the overall performance. We shall be exploring these tasks in the coming days.

Classes	Counts
Non hope Speech	20777
Hope Speech	20777
Total	41554

Table 4: Balanced classes distribution in training data

References

- B Bharathi, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, N Sripriya, Arunaggi 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, Kyiv. Association for Computational Linguistics.
- Bharathi Raja Chakravarthi, Ruba Priyadarshini, Thenmozhi Durairaj, John Phillip McCrae, Paul Buitaleer, Prasanna Kumar Kumaresan, and Rahul Ponnusamy. 2022. 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 Priyadarshini, Vigneshwaran Muralidaran, Shardul Suryawanshi, Navya Jose, Elizabeth Sherly, and John P McCrae. 2020. Overview of the track on sentiment analysis for Dravidian languages in code-mixed text. In *Forum for Information Retrieval Evaluation*, pages 21–24.

Algorithms	Accuracy(in percent)	F1-score-weighted	F1-score-micro	F1-score-macro
XGBoost	84.78	0.8608	0.8478	0.6130
Random Forest	86.15	0.8677	0.8615	0.6110
Multinomial NB	78.63	0.8181	0.7863	0.5503
Logistic Regression	81.06	0.8316	0.8106	0.5504

Table 5: Result

- 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 transphobia in multilingual YouTube comments. *arXiv preprint arXiv:2109.00227*.
- E C Chang. 1998. Hope, problem-solving ability, and coping in a college student population: some implications for theory and practice. *J Clin Psychol*, 54(7):953–962.
- Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.
- Nikhil Ghanghor, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. Iiitk@ It-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.
- Raul Gomez, Jaume Gibert, Lluís Gomez, and Dimosthenis Karatzas. 2020. Exploring hate speech detection in multimodal publications. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*.
- Ashraf M Kibriya, Eibe Frank, Bernhard Pfahringer, and Geoffrey Holmes. 2004. Multinomial naive bayes for text categorization revisited. In *Australasian Joint Conference on Artificial Intelligence*, pages 488–499. 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.
- Andy Liaw, Matthew Wiener, et al. 2002. Classification and regression by randomforest. *R news*, 2(3):18–22.
- 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@ It-edi-eacl2021-hope speech detection: there is always hope in transformers. *arXiv preprint arXiv:2104.09066*.
- 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.
- Xudie Ren, Haonan Guo, Shenghong Li, Shilin Wang, and Jianhua Li. 2017. A novel image classification method with cnn-xgboost model. In *International Workshop on Digital Watermarking*, pages 378–390. Springer.
- Ratnasingam Sakuntharaj and Sinnathamby Mahesan. 2016. A novel hybrid approach to detect and correct spelling in Tamil text. In *2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 1–6.
- Ratnasingam Sakuntharaj and Sinnathamby Mahesan. 2017. Use of a novel hash-table for speeding-up suggestions for misspelt Tamil words. In *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, pages 1–5.
- Ratnasingam Sakuntharaj and Sinnathamby Mahesan. 2021. Missing word detection and correction based on context of Tamil sentences using n-grams. In *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 42–47.

- 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 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, Kyiv. Association for Computational Linguistics.
- Harris C. Anderson J. R. Holleran S. A. Irving L. M. Sigmon S. T. Yoshinobu L. Gibb J. Langelle C. Snyder, C. R. and Harney P. [The will and the ways: Development and validation of an individual-differences measure of hope](#). *journal of personality and social psychology*, 60(4), 570–585.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2019. [Sentiment analysis in Tamil texts: A study on machine learning techniques and feature representation](#). In *2019 14th Conference on Industrial and Information Systems (ICIIS)*, pages 320–325.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2020a. [Sentiment lexicon expansion using Word2vec and fastText for sentiment prediction in Tamil texts](#). In *2020 Moratuwa Engineering Research Conference (MERCOn)*, pages 272–276.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2020b. [Word embedding-based part of speech tagging in Tamil texts](#). In *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, pages 478–482.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2021. [Sentiment analysis in Tamil texts using k-means and k-nearest neighbour](#). In *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 48–53.
- Ishan Sanjeev Upadhyay, Nikhil E, Anshul Wadhawan, and Radhika Mamidi. 2021. [Hopeful_men@lt-edi-eacl2021: Hope speech detection using indic transliteration and transformers](#). *CoRR*, abs/2102.12082.
- Prashanth Vijayaraghavan, Hugo Larochelle, and Deb Roy. 2021. [Interpretable multi-modal hate speech detection](#). *CoRR*, abs/2103.01616.
- Raymond E Wright. 1995. Logistic regression.
- Carolyn M. Youssef and Fred Luthans. 2007. [Positive organizational behavior in the workplace: The impact of hope, optimism, and resilience](#). *Journal of Management*, 33(5):774–800.