SSN_ARMM@ LT-EDI -ACL2022: Hope Speech Detection for Equality, Diversity, and Inclusion Using ALBERT model

PraveenKumar V , Prathyush S , Aravind P, Angel Deborah S, Rajalakshmi S, Milton R S, Mirnalinee T T

Department of Computer Science and Engineering,
Sri Sivasubramaniya Nadar College of Engineering,
Chennai, India
vpraveenkumar0211@gmail.com, aravind14110@gmail.com, prathyushsunil2510@gmail.com
angeldeborahs@ssn.edu.in, rajalakshmis@ssn.edu.in, miltonrs@ssn.edu.in, mirnalineett@ssn.edu.in

Abstract

In recent years social media has become one of the major forums for expressing human views and emotions. With the help of smartphones and high-speed internet, anyone can express their views on Social media. However, this can also lead to the spread of hatred and violence in society. Therefore it is necessary to build a method to find and support helpful social media content. In this paper, we studied Natural Language Processing approach for detecting Hope speech in a given sentence. The task was to classify the sentences into 'Hope speech' and 'Non-hope speech'. The dataset was provided by LT-EDI organizers with text from Youtube comments. Based on the task description, we developed a system using the pre-trained language model BERT to complete this task. Our model achieved 1st rank in the Kannada language with a weighted average F1 score of 0.750, 2nd rank in the Malayalam language with a weighted average F1 score of 0.740, 3rd rank in the Tamil language with a weighted average F1 score of 0.390 and 6th rank in the English language with a weighted average F1 score of 0.880.

1 Introduction

Social media has become an essential part of our lives. People tend to reflect on their inner selves through their online conversations. There is a huge increase in the number of individuals looking for support through the internet. In recent times, there has been a surge in these online support sources. Gowen et al. (2012) Online support groups help people going through similar disabilities, health problems, etc and overcome their difficulties together. Recently researchers Ganda and Madison (2014) have found out that social media network and online support groups have a great impact on people's self-understanding. YouTube is a Social media platform which connects billions of users across the internet. It has gained outstanding popularity across the globe (Sakuntharaj and Mahesan, 2021, 2017, 2016; Thavareesan and Mahesan, 2019, 2020a,b, 2021). With commenting options available, one can easily manipulate different people through this. As social media is used predominantly in day to day life, it is crucial, not only to protect users from harmful content but also to spread and encourage hope and optimism in this society (Sampath et al., 2022; Ravikiran et al., 2022; Chakravarthi et al., 2022b; Bharathi et al., 2022; Priyadharshini et al., 2022). In recent days, NLP has gained many architectural advancements and gained better results than state of art methods Wang et al. (2019).

The task focuses on the classification of Hope speech in multiple languages with each language having different class imbalances. Hope speech detection can uplift the amount of positive content on social media and helps to build a peaceful world. In our task, we used Hope Speech dataset for Equality, Diversity, and Inclusion in English, Tamil, Malayalam, and Kannada (Chakravarthi, 2020; Chakravarthi and Muralidaran, 2021; Chakravarthi et al., 2021, 2022a). Dravidian languages are a group of 250 million people who speak mostly in southern India, north-east Sri Lanka, and south-west Pakistan. The Dravidian languages are classified as South, South-Central, Central, and North, with each category subdivided into 24 subgroups. The Indian constitution recognizes four main literary languages: Telugu, Tamil, Malayalam, and Kannada. Tamil is one of the world's longest-surviving classical languages (Subalalitha, 2019; Srinivasan and Subalalitha, 2019; Narasimhan et al., 2018). Tamil is a member of the southern branch of the Dravidian languages, a group of about 26 languages indigenous to the Indian subcontinent. It is also classed as a member of the Tamil language family, which contains the languages of around 35 ethno-linguistic groups, including the Irula and Yerukula languages (Anita and Subalalitha, 2019b,a; Subalalitha and Poovammal, 2018).

Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, pages 172 - 176

The remainder of this paper is organized as follows: the next section includes related work followed by Dataset Description in section 3. Section 4 contains the Methodology. Results are presented in Section 5 and the Conclusion is presented in Section 6.

2 Related Works

Hope speech detection has been one of the important areas of research in recent years. Researchers have developed a wide range of tools, datasets, and models for Text Classification problems. In recent years many researchers have developed automatic methods for hope speech detection in social media. These methods rely on popular technologies like Machine Learning and Natural Language Processing. (Zhang et al., 2018) Did hate speech analysis for short text such as tweets. The proposed DNN method helps in identifying features useful for classification. They evaluated their model with the Twitter dataset and obtained good results. (Ribeiro et al., 2018) characterized hate speech in Online Social Networks with the help of n DeGroot's learning model. They found how hateful users are different from normal users using centrality measures and user activity patterns. (Ghanghor et al., 2021) Carried out hope speech detection task with various models and found out that mBERTcased model gave the best results. They employed zero short cross-lingual model transfer which is used to fine-tune the model evaluation. They found out that degradation of the model performance was due to freezing of base layers of transformer model . Muralidhar et al. (2018) focused on YouTube sentiment analysis. The researchers analyzed these data to find their trends and it was found that real-life events are influenced by user sentiments. Hope speech can also be termed as the opposite of hate speech. Hate speech includes offensive and bad comments on a particular work or a particular person. (Chakravarthi et al., 2020) These offensive comments create a bad impact on this society. Work done by (Puranik et al., 2021) includes analyzing the corpus of data collected from Youtube comments.

3 Dataset Description

In this work, we made use of the datasets provided by the Association for Computational Linguistics for Hope Speech Detection for Equality, Diversity, and Inclusion competition. These are multi-lingual datasets constructed by Chakravarthi (2020). It consists of comments made by users from the social media platform YouTube with 28,424, 17715, 9918, and 6176 comments in English, Tamil, Malayalam, and Kannada respectively, manually labeled. In these datasets, the comments are classified into two different categories as Hope-speech and Non-hope-speech. The distribution of each language dataset is shown in the table 1.

Language	Train	Test	Dev
Tamil	14199	1761	1755
English	22740	2841	2841
Malayalam	7873	1071	974
Kannada	4940	618	618

Table 1: Summary of Dataset

4 Methodology

We have applied a transformer-based approach to detect hope speech for multi-lingual Dravidian language comments. In our implementation of the code, we have used the Simple Transformers library which is built upon the transformers library by huggingface. ALBERT model has similar architecture as the BERT model, but the ALBERT model takes 18x fewer parameters compared to the BERT model. The Transformer-based neural network gives us another advantage through a technique called parameter-sharing where they use the same parameters for different independent layers. The architecture diagram of ALBERT model is given in figure 1. The transformers are non-sequential and always are processed in batches or as a whole sentence. We have also used a bi-directional approach so not only the previous words but the words from the right can also help in tokenizing.

We are using the IndicBERT model for tokenizing the Input Samples. We are tokenizing each input to convert the input sequence to tokens. IndicBERT is an ALBERT model that is pre-trained on 12 major Indian languages with a huge corpus of roughly 9 billion tokens. It's trained by choosing a single model for all languages to learn the relationship between languages. We have employed ai4bharat/Indic-bert model for both tokenizing and classifying model. We are tokenizing each sentence with a maximum length of 400 and truncation is enabled. The special tokens are included in this model to capture multi-lingual tokens and padding is not done for each sentence but is done

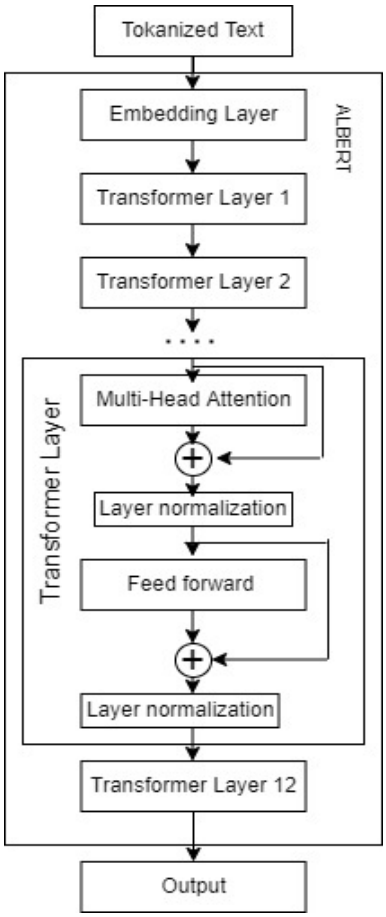

Figure 1: Architecture of ALBERT model

in batches later. Now the tokens have different sequence lengths and the inputs are now sorted based on the sequence length. This helps us in creating smart batches. Then the smart batches with a batch size of 16 are created and then the padding is done in those batches and attention masks are added. The resulting token count from padding in smart batches is 93.9% lesser than Fixed padding while not discarding any important token.

Now we are using transfer learning to import the ALBERT model for sequence classification and its configurations. The AdamW optimizer is used. Adam optimizer updates the weights based on stochastic gradient descent with an adaptive estimation based on first and second-order moments. AdamW optimizer does the same but it is additionally decoupled with the weight decay of the variable. We have also used a dynamic learning rate that is updated in each step by the linear scheduler function. The training is done in batches so the loss function is the average loss over the batch taken. Our hyperparameters are present in table 2.

Hyperparameters	Value
epoch	4
batch size	16
learning rate	$2e^{-5}$
max_length	400
activation	tanh
optimizer	AdamW
Adam_epsilon	$1e^{-8}$
Truncate	Enabled
Padding	Smart Batches
Learning rate	Dynamic

Table 2: Hyper-parameters of the model

5 Experimental Result

This section presents our experimental results and their analysis. In the results announced by the organizers. Multilingual comments are subjected to vary because people tend to write it in code-mixed data or in their native language which can be misinterpreted. A variation of such comments between train, test, and validation can impact the result of the model. Our model got 1st rank in Kannada, 2nd rank in Malayalam, 3rd rank in Tamil, and 6th rank in English. Our evaluation panel used the F1 score weighted average as a result indicator. The performance of our model is given in table 3. This result is due to the adoption of a pre-trained language model for this shared task. The ALBERT model is based on the Transformer model that has great potential for capturing global information Vaswani et al. (2017).

Language	Precision	Recall	F1 score
English	0.880	0.890	0.880
Tamil	0.370	0.420	0.390
Malayalam	0.700	0.780	0.740
Kannada	0.740	0.760	0.750

Table 3: The results of our model

6 Conclusion

Due to the pandemic there has been a sudden increase in active social media users which has led to abundant online content. There is a need to promote and motivate positive content to spread peace and knowledge in this society. In this paper, we proposed a transformer-based approach for Hope speech detection in 4 different languages (English, Tamil, Malayalam, Kannada). We used the AL-

BERT model with AdamW optimizer for classification. Our model got an F1 score of 0.880, 0.390, 0.740, and 0.750 in English, Tamil, Malayalam, and Kannada. In future work, techniques like Data augmentation can be used to fine-tune the model on more data.

References

R Anita and CN Subalalitha. 2019a. An approach to cluster Tamil literatures using discourse connectives. In *2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP)*, pages 1–4. IEEE.

R Anita and CN Subalalitha. 2019b. Building discourse parser for Thirukkural. In *Proceedings of the 16th International Conference on Natural Language Processing*, pages 18–25.

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, Kyiv. Association for Computational Linguistics.

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, Vigneshwaran Muralidaran, Ruba Priyadharshini, and John Philip McCrae. 2020. Corpus creation for sentiment analysis in code-mixed Tamil-English text. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 202–210, Marseille, France. European Language Resources association.

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*.

Ganda and Madison. 2014. Social media and self: Influences on the formation of identity and understanding of self through social networking sites. page 55. University Honors Theses.

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, Kyiv. Association for Computational Linguistics.

Kris Gowen, Matthew Deschaine, Darcy Gruttadara, and Dana Markey. 2012. Young adults with mental health conditions and social networking websites: Seeking tools to build community. *Psychiatric rehabilitation journal*, 35:245–50.

Skanda Muralidhar, Laurent Nguyen, and Daniel Gatica-Perez. 2018. Words worth: Verbal content and hirability impressions in YouTube video resumes. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 322–327, Brussels, Belgium. Association for Computational Linguistics.

Anitha Narasimhan, Aarthy Anandan, Madhan Karky, and CN Subalalitha. 2018. Porul: Option generation and selection and scoring algorithms for a tamil flash card game. *International Journal of Cognitive and Language Sciences*, 12(2):225–228.

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.

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, 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.

Manoel Horta Ribeiro, Pedro H. Calais, Yuri A. Santos, Virgílio A. F. Almeida, and Wagner Meira Jr. 2018. "like sheep among wolves": Characterizing hateful users on twitter. *CoRR*, abs/1801.00317.

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.

R Srinivasan and CN Subalalitha. 2019. Automated named entity recognition from tamil documents. In *2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP)*, pages 1–5. IEEE.

C. N. Subalalitha. 2019. Information extraction framework for Kurunthogai. *Sādhanā*, 44(7):156.

CN Subalalitha and E Poovammal. 2018. Automatic bilingual dictionary construction for Tirukural. *Applied Artificial Intelligence*, 32(6):558–567.

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.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *CoRR*, abs/1706.03762.

Chenguang Wang, Mu Li, and Alexander J. Smola. 2019. Language models with transformers. *CoRR*, abs/1904.09408.

Ziqi Zhang, David Robinson, and Jonathan Tepper. 2018. Detecting hate speech on twitter using a convolution-gru based deep neural network. In *The Semantic Web*, pages 745–760, Cham. Springer International Publishing.