International Journal of Innovative Research in Engineering
Special Issue, Volume 4, Issue 3(May-June 2023), PP: 102-103.
Innovations in Computing, Materials and Communication Technologies (ICICMCT'23) ***ISSN No: 2582-8746***

Dravidian Languages Hope Speech Detection

Hanna Zahrin[1], Dini Davis[2]
[1]P.G. Student, Department of Computer Science & Engineering, IES Engineering College, Thrissur, Kerala, India
[2]Assistant Professor, Department of Computer Science &Engineering, IES Engineering College, Thrissur, Kerala, India.

How to cite this paper:
Hanna Zahrin[1], Dini Davis[2]," Dravidian Languages Hope Speech Detection", IJIRE-V4I03-102-103.

Abstract: *The task of hope speech detection has gained traction in the natural language processing feldowing to the need for an increase in positive reinforcement on line during the COVID-19 pandemic. Hope speech detection focuses on identifying texts among social media comments that could invoke positive emotions in people. Students and working adults alike posit that they experience a lot of work-induced stress further proving that there exists a need for external inspiration which in this current scenario, is mostly found online. In this paper, we propose a multilingual model, with main emphasis on Dravidian languages, to automatically detect hope speech. We have employed as tacked encoder architecture which makes use of language agnostic cross- lingual word embeddings as the dataset consists of code-mixed YouTube comments. Additionally, we have carried out an empirical analysis and tested our architecture against various traditional, transformer, and transfer learning methods.*

Key Word: *Dravidian Language, Language Detection, Hope Speech Detection.*

I.INTRODUCTION

According to studies, the average time a person spends on social media is approxi- mately 2 h and 25 min on a daily basis. It is also a known fact that the majority of teenagers and young adults. Apart from the positive consequences of the advent of social media, there have been several negative impacts too.

Removing punctuations, extra spaces, removing special characters, and lowercasing, stop word removal, converting emoticons to text using the demoji library in the English language, and normalizing contractions are the basic steps performed in preprocessing steps. Contractions are shortened forms of words joined by an apostrophe. For example, couldnt, wont are contractions also.

Moreover, one of the most common methods by which cyberbullie sharass their victims is by commenting on their posts on social media. Considering that young adults are readily exposed to such vitriolic content on social media, it tends to create a snowball efect. Abu- sive comments tend to produce more abusive comments and ultimately result in an avalanche of online aggression. This also works the other way around, with positive comments inspiring more people to leave more positive comments.

II.METHODOLOGY

The proposed system uses a transformer-based approach for identifying hope speech in You Tube comments. We have adopted as tacked transformer based encoder model long within corporating cross-lingual word embeddings. The methodology con- sists of two variations. In the first variant, the original data sets are used to train the model and the flow of our approach can be seen in figure 1. In the second variant ,the split datasets of the two Dravidian languages are used and this approach is depicted in figure1.

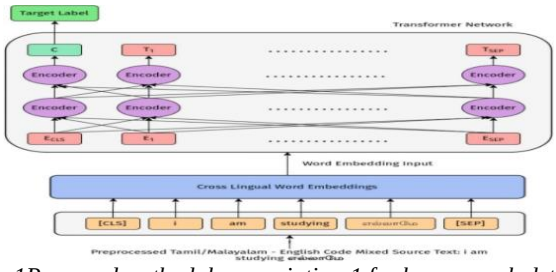

Figure 1Proposed methodology variation-1 for hope speech detection

This model was trained over three epochs with a batch size of 32 using the Dravidian language datasets. The discussion of the validation out comes appears in table below. These findings apply to our transformer model, which was developed using initial datasets.

III.EXPIREMENTAL RESULT

To further ground our approach amongst other pre-existing models for text clas- sifcation, evaluated the learning ability of our models using k-fold (fve-fold) cross validation.

Amongst all the models that implemented, choose the best performing models in each approach as elaborated in Empirical Analysis. Figures 2 and 3 compare the performance of the suggested model to SVMmodel that performs the best when implemented traditionally, and indic-BERTmodel that performs the best when im- plemented using a transformer. Figure 4 contains the performance metrics of the various model for fvefolds.

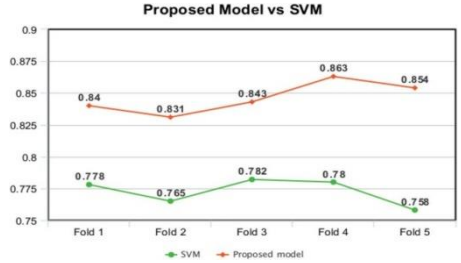

Figure 2 : Comparison of Malayalam language approach against SVM

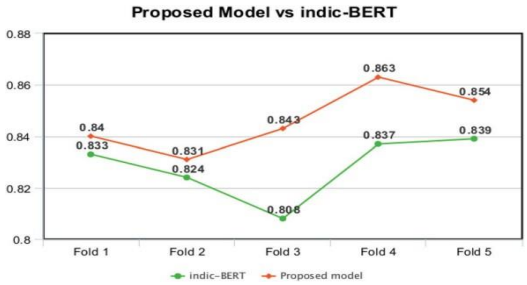

Figure 3 : Comparison of Malayalam language approach against indic-BERT

Model	Fold	Precision	Recall	F1-score
SVM	1	0.797	0.805	0.778
	2	0.791	0.795	0.765
	3	0.805	0.811	0.782
	4	0.809	0.810	0.780
	5	0.790	0.791	0.758
indic-BERT	1	0.831	0.837	0.833
	2	0.822	0.829	0.824
	3	0.807	0.811	0.808
	4	0.836	0.842	0.837
	5	0.838	0.844	0.839
Our model	1	0.840	0.842	0.840
	2	0.831	0.832	0.831
	3	0.843	0.844	0.843
	4	0.862	0.866	0.863
	5	0.853	0.856	0.854

Figure 4 : Cross-validation results for Malayalam language

IV. CONCLUSION

Garbage This project presented a language-agnostic transformer model to detect hope speech in the Dravidian languages datasets. The proposed model makes use of a bidirec- tional dual-encoder to produce the embeddings for texts in our dataset. By utilizing attention mechanisms, the network better understands contextual relationships be- tween the words. This is evidenced by the comparison of our proposed approach with other diferent methodologies as seen in the k-fold paired t test scores.Moreover,this approach requires less data augmentation and reduces the overhead when compared with other traditional and transformer learning methods. This work contributes to the reduction of abrasive content on social media by fltering out negative comments so that only positive and hopeful comments are retained.

References

[1] Chakravarthi BR, "HopeEDI a multilingual hope speech detection dataset for equality, diversity, and inclusion",2020.
[2] Dave B, Bhat S, Majumder P, "Hopespeech detection in code mixed text using TF-IDF char n-grams and MuRIL",2021.
[3] Mujadia V, Mishra P, Sharma DM, "HateSpeech Detection", 2019.
[4] Dowlagar S, Mamidi R, "Pretrainedtransformers with convolutional neural net- works for hope speech detection",2021.
[5] Saumya S, Mishra AK, "Hopespeech detection in YouTube multilingual com- ments",2021.
[6] Huang B, Bai Y, "Hopespeech detection based on pre-trained language model", 2021.
[7] Hossain E, Sharif O, Hoque MM, "Multilingualcode-mixed hope speech detection using cross-lingual representation learner",2021.
[8] Gundapu S, Mamidi R, "Hopespeech detection with BERT transformer model", 2021.