

dhivya-hope-detection@LT-EDI-EACL2021: Multilingual Hope Speech Detection for Code-mixed and Transliterated Texts

Dhivya Chinnappa, Thomson Reuters

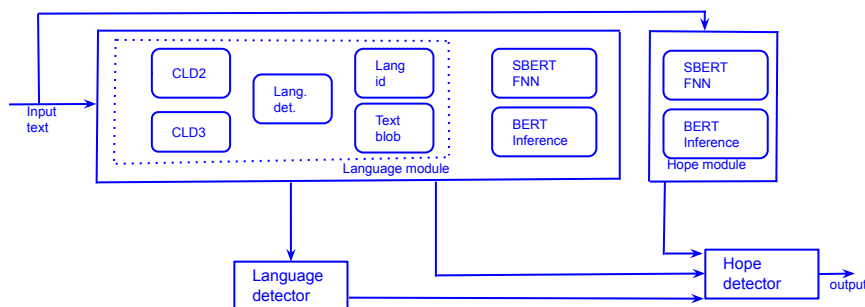
Objective: Build classification models to identify hope speech in youtube comments

Languages: English, Tamil, Malayalam

Labels: Hope, Non hope, Not-*lang*

English	
God accepts everyone.	Hope
the tech industry is tough for everyone. Most people are pushed hard and treated like garbage	Non hope
CASA. LA. FEMME nWesT. ViLLAGE n2008 nNyc	Not English

Architecture:



- Detect language to filter not-*lang* labels
- Binary classification to predict hope or non-hope

Results:

	English			Tamil			Mal		
	P	R	F	P	R	F	P	R	F
Hope	0.56	0.56	0.56	0.56	0.52	0.54	0.63	0.46	0.54
Non hope	0.96	0.96	0.96	0.61	0.66	0.63	0.85	0.92	0.88
Not lang.	0	0	0	0.61	0.54	0.58	0.83	0.66	0.74
W. avg.	0.92	0.92	0.92	0.59	0.59	0.59	0.81	0.82	0.81

Conclusion:

- Tamil and Malayalam datasets include code-mixed data.
- Two phase mechanism to identify hope
 - Identify language
 - Classify hope vs. non hope
- Results show hope detection is challenging