

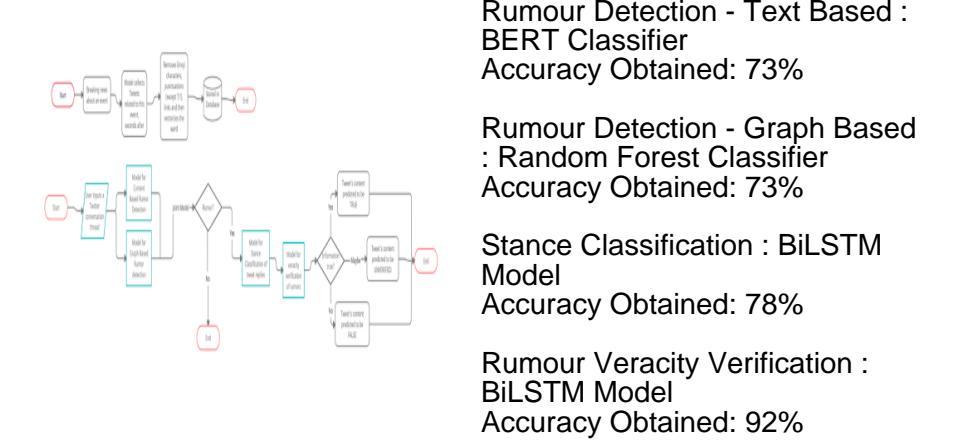
Problem Statement

Rumour detection and veracity verification in Twitter using socio-linguistic data and social graphs

Building a joint model utilizing both features of a rumour graph as well as content features from Tweets

We only deal with rumours, and not fake news

Design Approach / Methods



Background

Existing research identifies rumours based on either content based feature or features of the underlying social graph, not a combination of both

Dataset and Features / Project Requirements / Product Features

SemEval 2019 & Pheme Dataset

Pre-processing:
Count of negative words
Presence of punctuations
Relation to source tweet and previous tweet
Emotions associated with the tweet

Feature Extraction From Social Graph
Feature dictionary consists of 5 features:
User verification
Number of followers
Number of following
Number of retweets
Favourite count

Conversation tree built by creating feature dictionary for each branch of the conversation using structure.json

Results and Discussion

Result obtained after combining both the detection models. They were combined using confidence scores.

	precision	recall	f1-score	support
0	0.96	0.93	0.95	1241
1	0.89	0.93	0.91	687
accuracy			0.93	1928
macro avg	0.92	0.93	0.93	1928
weighted avg	0.93	0.93	0.93	1928

Result obtained for final veracity

	precision	recall	f1-score	support
0	0.91	0.96	0.94	187
1	0.91	0.87	0.89	82
2	1.00	0.89	0.94	64
accuracy			0.92	333
macro avg	0.94	0.91	0.92	333
weighted avg	0.93	0.92	0.92	333

The results obtained were as expected. The combination of two models has given us high accuracies.

Summary of Project Outcome

Newly constructed model can classify a rumour using its two facets: sentiment relayed through text and behaviour identified in social graphs.

Combined the loss functions of the linguistic based model and graph based model to generate a final model.

Once combined, we performed a final analysis on this model with the firm belief that it shall surpass other existent models.

As observable from the decreasing training and validation loss, as well as the improving confusion matrix values through all the epochs, we can confidently claim that this model works efficiently.

Conclusions and Future Work

Extracted features from the dataset that help gauge the presence of an individual in a social network.

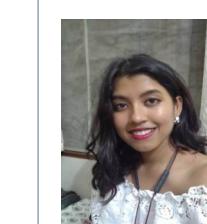
Combined loss functions of the linguistic based model and graph based model to generate a final model.

Once combined, we performed a final analysis on this model with the firm belief that it does surpass other existent models.

Future Work
Extend this application to include other social media websites such as Reddit or Instagram

References

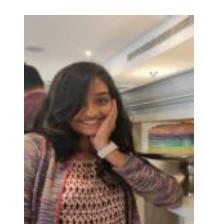
- [1] Zubiaga, Arkaitz & Liakata, Maria & Procter, Rob. (2017). Exploiting Context for Rumour Detection in Social Media. 10.1007/978-3-319-67217-5_8.
- [2] Ma, Jing & Gao, Wei & Mitra, Prasenjit & Kwon, Sejeong & Jansen, Jim & Wong, Kam-Fai & Cha, Meeyoung. (2016). Detecting Rumors from Microblogs with Recurrent Neural Networks.



Sukanya H



Sirisha L



Prajna G



Prof. Bhaskarjyoti D.