

# Commenting on twitter rumour detection in the health domain

[Business](#), [Company](#)



## **The general recommendation**

We will follow a continuous approach to review the selected paper, starting with general recommendation and followed by specific comments.

### **Topic: Twitter Rumour Detection in the Health Domain**

The paper talks about the role of social media specially the micro blogs like twitter for spread of news content. With increasing cases of mob lynching and other such incidents there is a need to determine which posts are rumors. So the topic in itself is something that requires immediate look upon and hence the paper was a selected topic for review.

In the section of Background and Motivations the authors have summarized all the works done in the field of twitter rumour detection. This literature review tell us about all the researches done starting from the first paper which was published on twitter rumour detection to the most recent paper. The author mentioned the papers, the authors of those papers and then clearly explains the underlying architecture used in that paper. Author also talks about how it has been improved compared to the previous paper published in this topic.

Dataset: The total sample size is around 1500 tweets. This consists of two sub samples, taken one month apart from each other containing #zika and zika microcephaly. Only Tweets that generated replies and retweets are considered, since standalone tweets do not cause spreading of a rumour. Each tweet was classified into 3 groups: rumour, non-rumour and unknown.

The rumour detection system is consisting of two phases, the training and the testing phase. The data acquired from twitter is identified in two levels, the user level and the network level and then the features are extracted. From the user level we identify the characteristics of the user and his/her statuses and the network level identifies the interaction between users in the network and the related properties given by retweets and replies. The set of features has been divided into three different groups: referred to as influence potential, personal interest and network characteristics. The authors have selected most representative features.

### **The Specific Comments**

1. The in-depth explanation about the underlying architectures of the previous papers tells that the authors have done extensive research before writing this paper.
2. The comparisons of the change in the architectures used in one paper to the ones used in the other papers was done very well.
3. There is a section about background and motivations, but nowhere in this section it is clearly mentioned as to why the authors were motivated to do twitter rumour detection especially in health domain and any previous work done on rumour detection in health domain is also not mentioned.
4. One main concern is the size of the sample. It may be smaller than ideal.
5. The hashtags used are promising. The choice of disease (zika) is also a good one since it was world news and it happened relatively recently.

6. The sample size may not reflect the trends of the population as a whole.
7. For the evaluation of the dataset the number and type of features chosen were sufficient and appropriate.
8. For feature selection they could have implemented a neural network model as so many type of neural networks like RNN and its variants are available.