

## **Opinion mining on various aspects of health through social media analytics using collective sentiment feature analysis and deep neural networks**

**T.M.M.B.B. Abeysinghe, B.J. Gunasekara, and R.D. Nawarathna**\*

*Department of Statistics and Computer Science, University of Peradeniya, Sri Lanka*  
*\*ruwand@pdn.ac.lk*

Social media such as Twitter and Facebook provide a perfect platform to create awareness in the areas such as health, current affairs, etc., because of their dynamic behavior. In particular, they encourage users to post ideas, views and random details of their everyday life. Most of those messages (i.e., wall posts and tweets) contain a less informational value, but a collection of messages can generate important knowledge which provides valuable insight on various aspects of life. According to recent studies, even in developed countries, the ‘health literacy’ is found to be considerably low. Typically, people have their own views and experiences regarding certain diseases and treatments. For instance, in the social media, there can be tweets and wall posts on numerous real-life experiences about treatments for some diseases such as common flu. In fact, those views show the real picture about the effectiveness of those treatments. Hence, comprehensive analysis of large volume of such social media content may lead to interesting conclusions.

One of the key steps in making health awareness using social media analytics is the sentiment analysis of health related messages. Sentiment is the attitude, opinion or feeling toward something which indicates the contextual polarity. This study presented a general framework for opinion mining on various aspects of health, through the analysis of typical public reactions towards health and well-being in Twitter media. The proposed framework was developed based on “collective sentiment feature analysis” and deep neural networks. This novel collective sentiment feature analysis method aims to map a given tweet to a feature space that captures the sentiment of the tweet as a whole taking the relationship of the keywords in the sentence into consideration. First, the tweets mentioning about selected health issues were collected, preprocessed and labeled. Each sentence was represented by a “collective sentiment feature vector” that is learned using a classification algorithm. In order to make a better generalization in the polarity classification of tweets, a state-of-the-art deep neural network is trained. The main advantage of the proposed method over the existing methods is its ability to generalize even large data sets as a result of the trained deep neural network. The performance of the proposed framework was evaluated by conducting an experimental study on twitter posts extracted using “Tweepy” a python wrapper for Twitter REST API. Three data sets containing tweets about dengue disease, H1N1 influenza and other general health topics were used for the evaluation. A promising average F1-score value of 0.73 was obtained during testing.