

Syndromic Surveillance Framework Using Tweets Data Analytics

Authors : David Ming Liu, Benjamin Hirsch, Bashir Aden

Abstract : Syndromic surveillance is to detect or predict disease outbreaks through the analysis of medical sources of data. Using social media data like tweets to do syndromic surveillance becomes more and more popular with the aid of open platform to collect data and the advantage of microblogging text and mobile geographic location features. In this paper, a Syndromic Surveillance Framework is presented with machine learning kernel using tweets data analytics. Influenza and the three cities Abu Dhabi, Al Ain and Dubai of United Arab Emirates are used as the test disease and trial areas. Hospital cases data provided by the Health Authority of Abu Dhabi (HAAD) are used for the correlation purpose. In our model, Latent Dirichlet allocation (LDA) engine is adapted to do supervised learning classification and N-Fold cross validation confusion matrix are given as the simulation results with overall system recall 85.595% performance achieved.

Keywords : Syndromic surveillance, Tweets, Machine Learning, data mining, Latent Dirichlet allocation (LDA), Influenza

Conference Title : ICDMA 2020 : International Conference on Data Mining and Applications

Conference Location : Miami, United States

Conference Dates : March 12-13, 2020