GEOGRAPHICAL UNDERSTANDING OF TWITTER HEALTH RELATED TOPICS BY USING TOPIC MODELLING ALGORITHMS

There is a need for understanding the health conditions worldwide with respect to every country. Most of the diseases and their symptoms are visible only after it has affected a substantially large population of a developed country. People must know about the diseases prevailing in a particular country and the precautions to be taken in case of being affected. An application that gets the information of a particular country health status must be developed. This project does not have any dataset and relies on live data extracted from twitter. Proper text data analysis can be done by using topic modelling (TM). Analysing topics is a critical issue that permits an unreliable number of TM topics that address the dismal outcomes in HRC (Health Related Clustering) in data sources. One of the proposed Distributed topic models Latent Dirichlet Allocation (LDA) is a credible approaches for balancing and clipping to the direction of health topics from various perspective data sources in health statistics clustering. LDA assigns the probabilities to the words basing on their occurence in the given group of documents which here are tweets. A feature extraction technique called tf-idf vectorization must be followed for LDA to produce an efficient output. This service provides a concise description of the nation's hash tag system for good public health and analyzes the development of the most significant tweets about health in order to provide a preliminary public recommendation. This will most effectively improve the health care systems, median age of the world's population.

Group 16

Team Members:

T. Srichakranath [19241A1255]
Shaik Sohel [19241A1251]
P. Venkata Anirudh [19241A1238]
P. Hruthvik Reddy [19241A1240]

Project Guide

Dr. N. Rajasekhar
Professor