

Engineering BigData project Document

Team members:

Abhijit Shinde
Reema Dmello
Tushar Goel

Topic/title/idea:

News articles sentimental analysis

Dataset:

Scrapping data from multiple sources like twitter and facebook, will be using R and twitter and facebook APIs to scrape data.

Plan:

Basic idea is to scrape the comments/tweets of the users related to news articles and perform sentimental analysis.
Initially we will be using scrapping data once using R, would convert that data from doc2vec in batches and load complete data in HDFS.
Relational data will be loaded in to Hive tables for processing.
Using Mahout would perform classification and clustering as data after doc2vec would be label data.
Final plan is to get final sentiment and display using reporting tools like tableau or website.

Github link:

<https://github.com/tushargl016/BigDataTeam>