Sentiment Analysis of Tweets
Team num : 15

Progress so far

- **Data collection**
  - Downloaded data with 1.6M tweets(50% +ve and 50% -ve) from http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip and analyzed data.

- **Data processing**
  - Processed data with beautifulsoup library for decoding HTML and regex for :
    - Remove username
    - Convert string to lowercase
    - Removing hyperlinks
    - Removing hash of hashtags
    - Remove words with length <= 2
    - Removing punctuations
    - Removing whitespaces, newline characters
    - Removing stopwords(sklearn)
    - Stemming [*used porter stemmer algo to implement stemmer*]

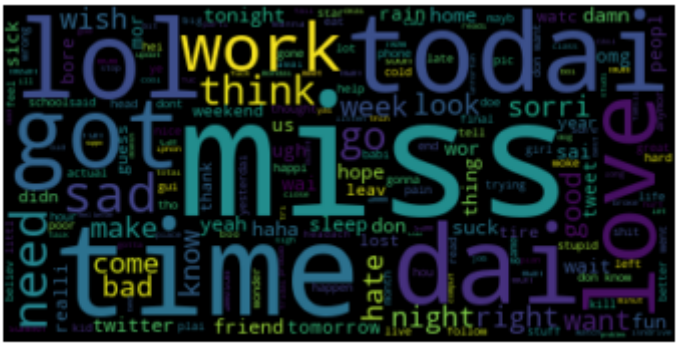
- **Data caching**
  - Splitted data in 60 : 20 : 20 for trani : validation : test
  - Stored processed data to avoid repeated processing.

- **Visualization**
  - Created wordles for whole dataset, only +ve and only -ve.

Whole dataset

Sentiment Analysis of Tweets
Team num : 15

Positive tweets



Negative tweets

- **Currently working on**
  - SVM implementation using bag of words of unigram
  - Logistic Regression implementation using TF-IDF of unigram
  - Naive bayes implementation using bag of words of unigram

What else is left

- SVM - [CountVec TFIDF] - bigram
- Logistic Regression - [CountVec, TFIDF] - bigram
- Naive Bayes - [CountVec, TFIDF] - bigram