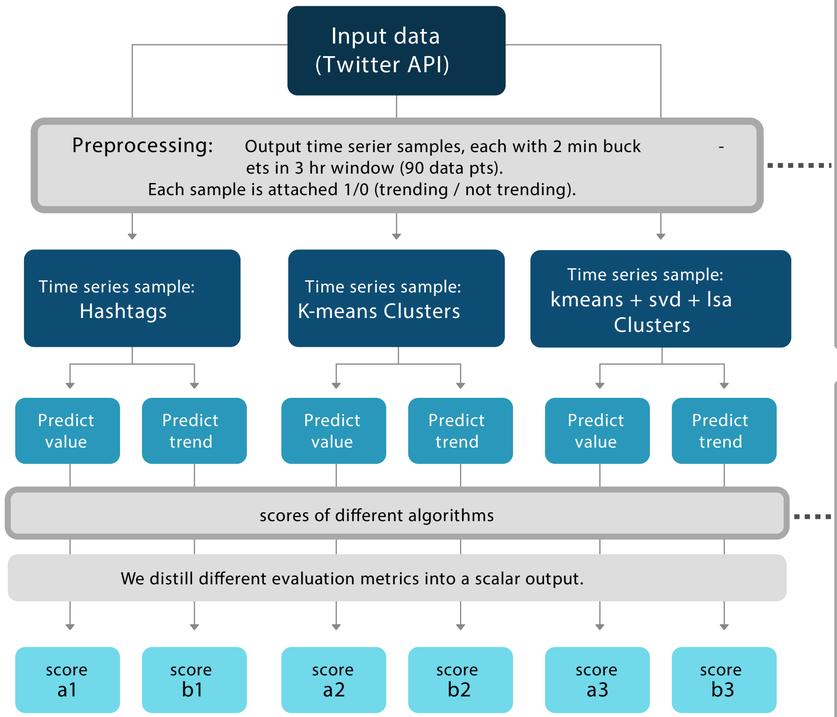
Predicting trends of Interests in Twitter

Luk Wing San (20193803)
Math 4432 Project 3 Self-Proposed Topic

OVERVIEW

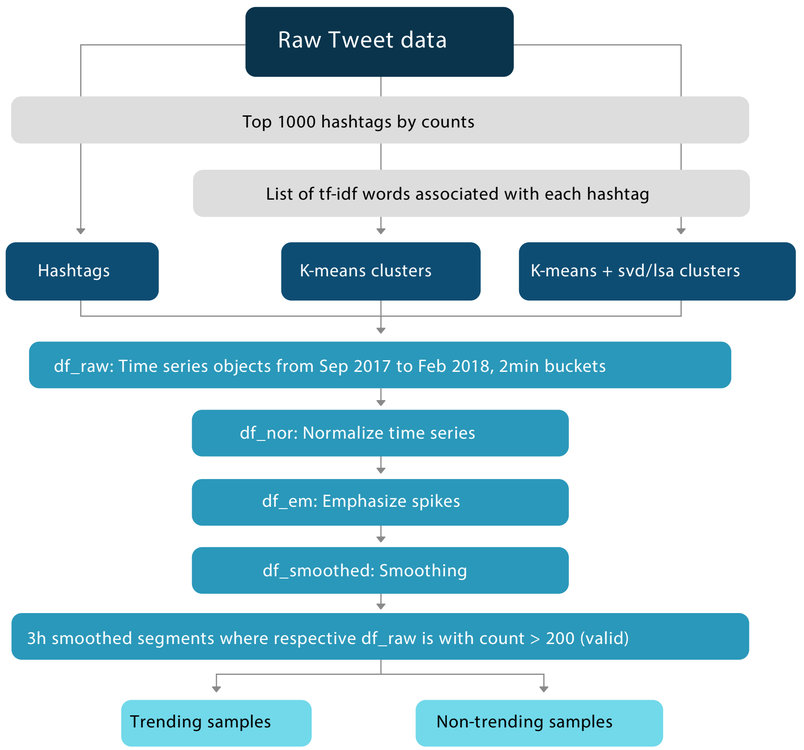
- Literature Review
- Data collection and Feasibility Studies
  Evaluate and improve the quality of data as input to our research
- Develop and complete Research Framework
  Make use of Machine Learning techniques.
- Build software app to present our analytical results
  Make use of Data Visualization techniques.

RESEARCH FRAMEWORK



3 categories of time series data are analysed to find out:
Whether a certain algorithm is better at predicting value/trend.
Whether by clustering tweets can the time series be better predicted.

PREPROCESSING



ALGORITHMS

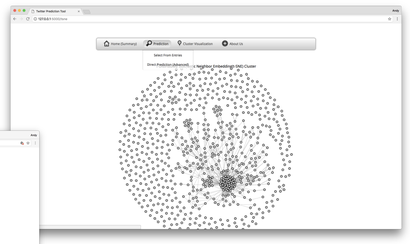
Method \ True Positive rate	hashtag mean	hashtag max	kmeans mean	kmeans max	kmeans_LSA/SVD mean	kmeans_LSA/SVD max
Multi-layer Perceptron Neural Network Model - 11 hidden layers	0.94	1.00	0.56	1.00	0.63	1.00
Gaussian Naive Bayes	0.93	1.00	0.59	0.86	0.51	1.00
Label Propagation classifier	0.30	0.37	0.05	0.17	0.05	0.50

Method \ RMSE	hashtag	kmeans	kmeans_LSA/SVD	mean
Gaussian process regression	344.36	321.28	282.80	316.15
Linear support vector regression	4.46	0.74	0.66	1.95
Kernel ridge regression	4.45	0.76	0.67	1.96

DATA VISUALIZATION AND WEB APPLICATION

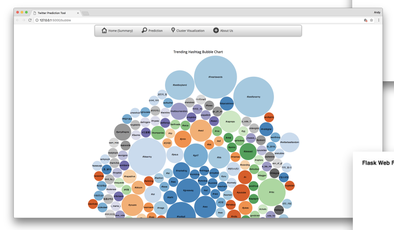
Word Cloud showing the words with highest frequency in different clusters

An illustration of summary page

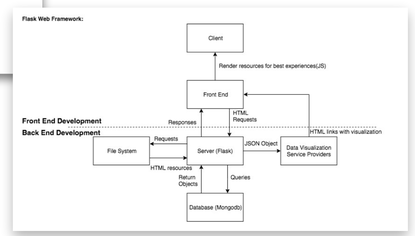


Predictive Summary

Make Prediction for hashtag and cluster



Bubble graphs showing trending hashtags.



Web Application Design

RESULTS

1. Our clustering effectively improve prediction.
2. Multi-layer perceptron with 11 hidden layers on kmeans clustering with SVD and LSA dimensional reduction is best for predicting trend value.