1.125 Arch & Engineering Software Systems
Term Project Assignment

Deliverable #1
Due Date:
Thursday, Nov 10th, 2016
Team:
Aswini Narayana Prasad and Shweta Jindal

A. Project proposal

The goal of this project is to perform sentiment analysis based on the product reviews from the end customers/critics/experts on digital platforms like twitter and provide a summary of the overall impression about the firm and its offerings in the market. The main audience would be the senior leadership of the firm who are interested in getting the first hand information on the general customer impression.

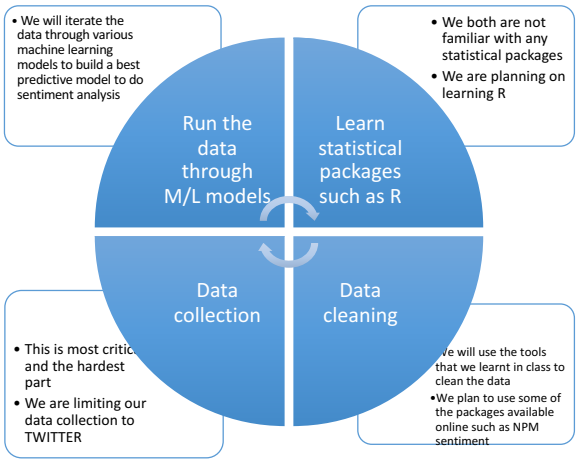
In general these reviews are available in different platform but these data are large in volume and often contradictory and subjective too. Hence the objective is to use machine learning to make meaningful conclusion from these reviews. Accordingly, based on the data collected from the social media, sentiment analysis is done on these reviews based on the semantic structures of these sentences and then tagged as positive/negative/neutral using classification methods followed by linear regression or similar techniques and data visualization to present the overall trend. Current plan is to use the statistical packages in R and Javascript.

B. Competitive Analysis

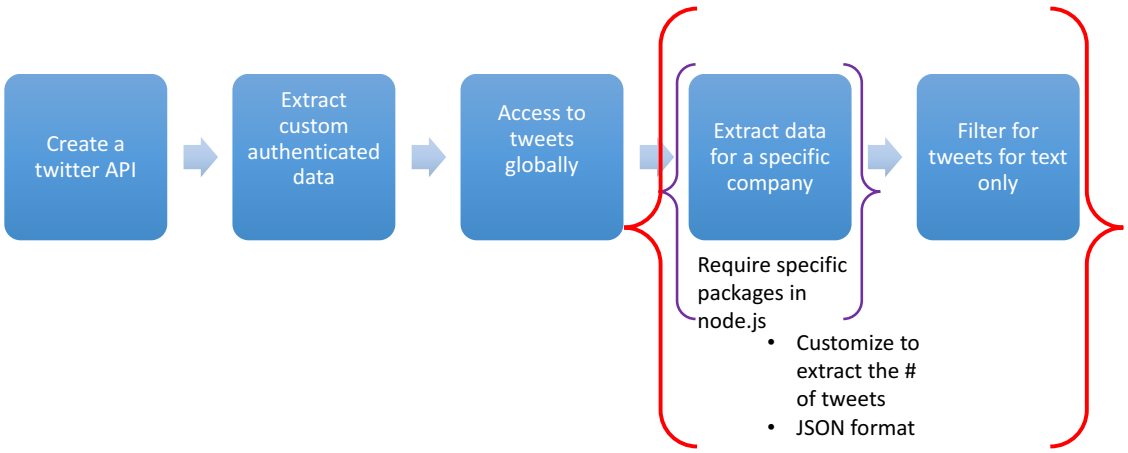
Existing Tools/Codes	Pros	Cons
1. Predicting Stock market prices based on Tweets using psychology tool 'Profile of Mood States'	Accuracy as good as >85% in the tested cases	calmness index is not considered in predicting future prices Tweets are not filtered based on location and hence may not reflect rightly

2.http://twittersentiment.appspot.com - for twitter sentiment analysis based on machine learning	Nimble	
3.http://brands.peoplebrowsr.com - Provides brand analyzer as part of a larger suite of tools.	Includes twitter as well as other social platforms	Paid version
4. Hootsuite – Enables marketing campaigns, schedule posts in advance, identify and grow audiences based on tracking of hashtags, mentions, Twitter lists	More focussed on marketing for a brand among focussed audience	Not strong in sentiment analysis, more for collaborative work
5. http://www.tweetfeel.com	Good with data visualisation	Paid version Based on keyword and not sentence.

C. Storyboard:



Data cleaning flow:



D. Code Artifacts:

Code for Twitter data Extraction:

Below code is written in node.js. A npm package 'Twit' is used for extracting the twitters from a specific keyword.

Below security key related authentication are obtained by creating a dedicated twitter app and hidden here.

Node.js

```
var T = new Twit({

  consumer_key: 'xxx',
  consumer_secret: 'xxx',
  access_token: 'xxx',
  access_token_secret: 'xxx',
  timeout_ms:   60*1000,  // optional HTTP request timeout to apply to all requests.
```

```
})

var fs = require('fs');
var util=require('util');
var logFile=fs.createWriteStream('log.json',{flags:'a'});
var logStdout=process.stdout;
var cleanfile=fs.createWriteStream('log.json');

//  search twitter for all tweets containing the keyword since a given data

T.get('search/tweets', { q: '#Samsung since:2014-01-10',count:100}, function(err, data,
response) {
console.log(data);
});

//  writting the output as log.json

 console.log=function(){

  logFile.write(util.format.apply(null, arguments)+'/n');
  logStdout.write(util.format.apply(null, arguments)+'/n');
 }
console.error=console.log;
```

Code for Filtering the tweets from JSON: (Javascript)

```
<html>
<script src="log.json">
</script>
<script>

var myArray=[];
var i=0;
window.onload=function() {
        document.getElementById("demo").innerHTML = "The required data
is"+myArray;
};
```

```
for (var i=0;i<data.statuses.length;i++)
{
myArray[i]=data.statuses[i].text;
window.onload=function() {
        document.getElementById("demo").innerHTML = "<p>"+myArray;
};
}


</script>
</head>
<body>
<h2 id='demo'>hello</h2>
<div id="page-wrapper">

</div>
</body>
</html>
```

Code for Reading data in R

```
library(RTextTools) //load library packages
library(e1071)

pos_tweets =  rbind(  //defning the positive tweets
  c('I love this car', 'positive'),
  c('This view is amazing', 'positive'),
  c('I feel great this morning', 'positive'),
  c('I am so excited about the concert', 'positive'),
  c('He is my best friend', 'positive')
)
neg_tweets = rbind( //defining the negative tweets
  c('I do not like this car', 'negative'),
```

```
  c('This view is horrible', 'negative'),
  c('I feel tired this morning', 'negative'),
  c('I am not looking forward to the concert', 'negative'),
  c('He is my enemy', 'negative')
)
test_tweets = rbind(  //testing the tweets
  c('feel happy this morning', 'positive'),
  c('larry friend', 'positive'),
  c('not like that man', 'negative'),
  )
tweets = rbind(pos_tweets, neg_tweets, test_tweets) // passing different types of tweets in tweets variable.
```

Next, we will work on R code to create basic ML models to predict the sentiment given a tweet. This involves selecting the training data properly to help train the model well.

E. A timesheet

1.125 Final term project time sheet

1.125 Final term project time sheet			
TimeSheet	Shweta	Aswini	Tasks
3-Nov	4 hr	4 hr	Project proposal work
10-Nov	6 hr	10 hr	Preliminary Study, installation of relevant packages, prelim code
17-Nov			
24-Nov			
1-Dec			