PREDICTING STOCK MARKET TRENDS THROUGH SOCIAL MEDIA

ADITYA SHIRODKAR
A20332644

PROBLEM

1. Can we apply Sentiment Analysis and Machine Learning techniques to find the correlation between "Public Sentiments" and "Market Sentiments"?

2. If yes then, Can we predict the stock market trends using these two sentiments?

APPROACH

- Collected Twitter and Stock Market data.

- Preprocessed the Stock Market data. (filtered tweets that were irrelevant for the experiment)

- The tweets were then fed to the sentiment analysis algorithm to determine whether each tweet is "Bullish" or "Bearish".

- The classified tweets and the stock data are then fed to our model (Logistic Regression) to find correlation between tweets and stock data.

- The model then tries to predict the stock market trend. (Bullish or Bearish)

DATA

1. Twitter
   - Collected tweets over the timespan of Nov 02 to Nov 27 (8:00 EST to 16:00 EST) (only of Trading-Days).
   - The total collected data was 52 GB i.e. about 15 Million Tweets.
   - Only the text and timestamp fields of the collected tweets were used.
   - Twitter Streaming API was used to collect the tweets.

2. YAHOO! FINANCE
   - Standard Stock Data for every Trading-Day from Nov 02 to Nov 27 (4 weeks).

NOTE: Twitter data is available at http://cloud.aditya11.com

RESULTS

- Number of bullish as well as bearish tweets in a day are almost the same.

- Accuracy comes to around 50%

CONCLUSIONS

- Cannot predict stock market from social media using models used in this experiment.

- Use percent change in stock market instead of UP/DOWN.

- Use pair of words as features.

- Instead of bullish or bearish tweets, we can add more dimensions.

- Use a different classification model.