

UNDERSTANDING INVESTORS' BEHAVIOR DURING STOCK PRICE MANIPULATION: A CASE OF INDONESIA'S STOCK MARKET

Riznaldi Akbar

Business School, University of Western Australia

Email: riznaldi.akbar@research.uwa.edu.au

Abstract: This study attempts to search textual patterns that are associated with investor behavior and sentiment during stock price manipulation in Indonesia's Stock Market. We propose text mining analysis to demonstrate the process of extracting information from news and media release of Tweeter over the period of 2000 to 2014. We took three sample firms of PT AGIS Tbk (TMPI), PT Garda Tujuh Buana Tbk (GTBO) and PT Bumi Resources Tbk (BUMI) that have been indicated experiencing stock price manipulation. These firms experienced an abnormal stock price appreciation and followed with a sharp decline of about 90 per cent within short period of time. The extracted information from Tweeter and sentiment analysis provide evidence that news and media releases tend to yield common negative key terms such as "plunge", "drop", and "fall"; as well as terms related to income statements such as "loss", "profit", "net", "sales" and "revenue". It shows us that most investors would have negative bias and suffer more when the stock price experienced a significant decline. It also reveals that investors are more short-sighted on the profit and loss statements rather than analyzing the big picture of corporate fundamental valuation as a whole. We thus argue that these knowledge bases and key terms could be used to reflect common characteristics of investor behavior during stock price manipulation in the Indonesia's stock market.

Keywords: Text mining, stock price manipulation, fraud, early warning system

Abstrak: Penelitian ini mencoba untuk mencari pola tekstual yang berkaitan dengan perilaku investor dan sentimen selama manipulasi harga saham di Bursa Efek Indonesia. Kami mengusulkan analisis text mining untuk menunjukkan proses penggalan informasi dari berita dan media rilis Tweeter selama periode 2000 sampai 2014. Kami mengambil tiga sampel perusahaan dari PT AGIS Tbk (TMPI), PT Garda Tujuh Buana Tbk (GTBO) dan PT Bumi Resources Tbk (BUMI) yang telah ditunjukkan mengalami manipulasi harga saham. Perusahaan-perusahaan ini mengalami harga saham apresiasi abnormal dan diikuti dengan penurunan tajam dari sekitar 90 persen dalam waktu singkat. Informasi yang diekstrak dari Tweeter dan analisis sentimen memberikan bukti bahwa berita dan media rilis cenderung menghasilkan umum istilah kunci negatif seperti "terjun", "drop", dan "jatuh"; serta istilah yang terkait dengan laporan laba rugi seperti "kehilangan", "keuntungan", "bersih", "penjualan" dan "pendapatan". Ini menunjukkan bahwa sebagian besar investor akan memiliki bias negatif dan menderita lebih ketika harga saham mengalami signifikan menurun. Hal ini juga mengungkapkan bahwa investor lebih cupet pada laporan laba rugi daripada menganalisis gambaran besar dari penilaian dasar perusahaan secara keseluruhan. Dengan demikian kita berpendapat bahwa basis pengetahuan ini dan istilah kunci dapat digunakan untuk mencerminkan karakteristik

umum dari perilaku investor selama manipulasi harga saham di pasar saham Indonesia.

Kata kunci: Text mining, manipulasi harga saham, penipuan, sistem peringatan dini

INTRODUCTION

There are vast empirical studies investigating stock price manipulation in the stock market. The stock price manipulation is considered as one of type of frauds in the financial market. The growth of fraud in financial market increased significantly in recent years with more sophisticated cases involving many market players. The Securities Exchange Act 1934 classifies stock price manipulation into two main categories. The first can be described as action-based manipulation, that is, manipulation based on action that changes the actual or perceived value of the assets. The second category can be described as information-based manipulation, that is, manipulation based on releasing false information or spreading false rumor. Allen and Gale (1992) included trade-based manipulation in which traders attempt to manipulate stock prices simply by buying and then selling, without making any publicly observable actions to alter the value of the firm or releasing false information to change the price.

Some studies have examined stock market manipulation using various methods (Liu et al., 2013; Imisiker and Tas, 2013; Diaz et al., 2011; Gerace et al., 2014). Liu et al. (2013) studied the manipulation of stock market prices by fund managers in the presence of potential future fund flows. Imisiker and Tas (2013) investigated which firms in the Istanbul Stock Exchange are more susceptible to successful manipulation over the period of 1998 through 2006. The authors used probit regression and results showed that small firms with less free float rate and a higher leverage ratio are more prone to stock price manipulation. Diaz et al. (2011) applied data mining techniques to detect stock price manipulations specifically using decision tree techniques. In recent studies, Gerace et al. (2014) empirically examined 40 cases of stock market manipulation on the HongKong Stock Exchange from 1996 to 2009. The results found that markets appear incapable of efficiently responding to the presence of manipulators and are characterized by information asymmetry.

There are urgent needs to develop an approach that are able to help regulators and relevant authority to address stock price manipulation. The development of financial market monitoring and early warning systems to capture the fraud is essential to reduce stock price manipulation practices in the financial market. To the best knowledge of author, the adoption of text mining analysis in Indonesia's stock exchange has been very limited. The aim of this study is to incorporate text mining approach to identify common keywords in the stock price manipulation. This study will provide empirical evidence of how morphology analysis could help the authority to identify stock price manipulation.

This study uses morphological analysis to extract information from annual report and media releases to identify text patterns and common keywords of stock price manipulation in Indonesia's Stock Exchange. The advantage of this approach is that we would be able to identify common textual characteristics of firm that has stock price manipulation. Using annual reports and media releases; we investigate how morphological analysis could provide insight of common keywords when stock price manipulation is taking place.

The remainder of this paper is structured as follows. Section 2 reviews related

literature. Section 3 describes methods and data sources. Section 4 discusses text mining analysis. Finally, Section 5 draws conclusion.

LITERATURE REVIEW

This section provides a review on existing market monitoring surveillance systems and fraud detection studies using text mining analysis to detect fraudulent behaviors and potential stock price manipulation in the stock market. These systems could support authority to detect unusual transactions where market abuse is suspected.

Stock exchanges have different monitoring system to detect fraudulent behavior in the stock market. NASDAQ has an Advanced Detection System (ADS) as a fraud detection system to monitor trades and detect any suspicious trading behaviors. The Australian Securities and Investment Commission have used a Nasdaq OMX/SMARTS monitoring system to monitor fraudulent and unusual trading behavior in the Australian stock market. UK Financial Conduct Authority has also tools to monitor markets and fraudulent activities in the stock market.

Some studies show text mining analysis has been used to predict stock price movement. (Mittermayer, 2004; Gidófalvi and Elkan, 2001; Fung et al., 2003; Schumaker and Chen, 2009). Mittermayer (2004) used press releases information to predict stock price trends and found that press releases could provide additional information to forecast stock price trends. Gidófalvi and Elkan (2001) adopted a naïve Bayesian text classifier and showed that short-term stock price movements could be predicted using financial news articles. Fung et al. (2003) used mining textual documents and time series to predict the movements of stock prices based on news articles. Schumaker and Chen (2009) examined a predictive machine learning approach for financial news articles analysis using support vector machine to estimate a discrete stock price twenty minutes after a news article was released. They found that the model containing both article terms and stock price at the time of article release had an ability to predict actual future stock price.

The stock fraud detection using text mining techniques has brought a great attention to scholars in recent years. Zaki and Theodoulidis (2014) analyzed stock market fraud cases using a linguistics-based text mining approach and found that text mining could be integrated with the financial fraud ontology to improve the efficiency and effectiveness of extracting financial concept. Zaki et al. (2011) showed an exemplar case study of text mining and data mining to analyze the impact of stock-touting spam emails and misleading press releases on trading data a real case from the over-the-counter market.

Shirata and Sakagami (2009) used text mining with morphological approach to analyze text data in the annual reports of 21 bankrupt Japanese companies and 24 non-bankrupt Japanese companies. The authors extracted keywords to discriminate between the two groups and found that the dividend section of the annual report contained a unique explanation of the company's financial position. The findings showed that terms such as dividends, profit appropriation, and retained earnings are among those with prominent differences in appearance frequencies between the two groups. However, there is a distinct lack of research on securities and stock market frauds using text mining analysis (Ngai et al. 2011).

This research contributes to the development text mining analysis in two folds. First, the contribution of this study is to provide an additional context for text mining analysis in

which how data sources such as news and contents from Tweeter could be analyzed using text mining approach to provide insightful information about stock price manipulation. Second, this study uses news and contents from Tweeter for Indonesian stocks as the data source that has not been addressed in previous studies.

METHODS

Text Mining Design. This section demonstrates the design of text mining, which is constructed to identify stock price manipulation. This study adapted the information news extraction to search common key terms of the stock price manipulation based on the news and contents from Twitter. The flowchart of text mining analysis for our study is shown in the Figure 1.

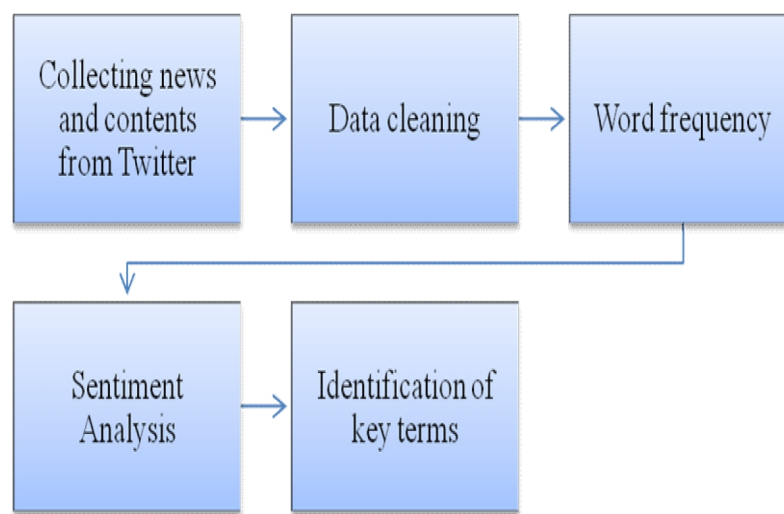


Figure 1. The flowchart of text mining analysis

This study uses text mining analysis to provide underlying framework for the extraction of news and contents from Tweeter to capture common key terms with respects to the stock price manipulation. The text mining has a comprehensive tool that could be used detecting fraud purposes including stock price manipulation in the stock market. The role of the text mining is to identify the knowledge that lies in the news and contents in the Tweeter to answer questions similar to those asked by other users reading the news and contents themselves.

Data Source. In this study, we take several samples of firms in the Indonesia stock exchange that have indicated experienced stock price manipulation. We take the top three firms that experienced a significant price appreciation followed by subsequent huge price drop. For example, PT AGISTbk (TMPI) stock price has increased 7067 per cent and followed by a sharp price decline of about 88.2 per cent. PT Garda TujuhBuana (GTBO) price accelerated to all-time high reached 9815 per cent return and followed by a huge drop of about 96.5 per cent. Similarly, PT Bumi Resources Tbk (BUMI) price has a tremendous increase of about 41000 per cent, but it plunged 98.9 per cent subsequently (Table 1).

The authority has also released multiple UMA (Unusual Market Activity) reports for those firms (Table 2). UMA release is a report produced by the Indonesia's stock exchange to inform market players with regards to unusual trade activities and/or price movements of particular stock within a certain period of time which could potentially disrupt or distort the holding of the stock itself.

This study uses different textual sources of media releases from Tweeter over the period the 2000 through 2014. All tweets data are imported using R library "tweeter".

Table1. Stock price of suspected firms

Firm	Price increase (low to all time high)	Period	Price decrease (all time high to low)	Period
TMPI	7067%	Jan 2003 – May 2007	88.2%	May 2007 – July 2013
GTBO	9815%	Nov 2010 – March 2013	96.5%	March 2013- July 2014
BUMI	40891%	Dec 2002 – June 2008	98.9%	June 2008 – Dec 2014

Table2. Unusual Market Activity (UMA) releases

Firm	Number of UMA releases	Date
TMPI	6 releases	21 May 2008 11 May 2009 12 April 2010 13 April 2010 19 September 2011 6 July 2012
GTBO	6 releases	2 November 2011 18 April 2012 27 September 2012 01 May 2013 6 December 2013 14 August 2014
BUMI	1 release	14 November 2014

Source: Indonesia's Stock Exchange

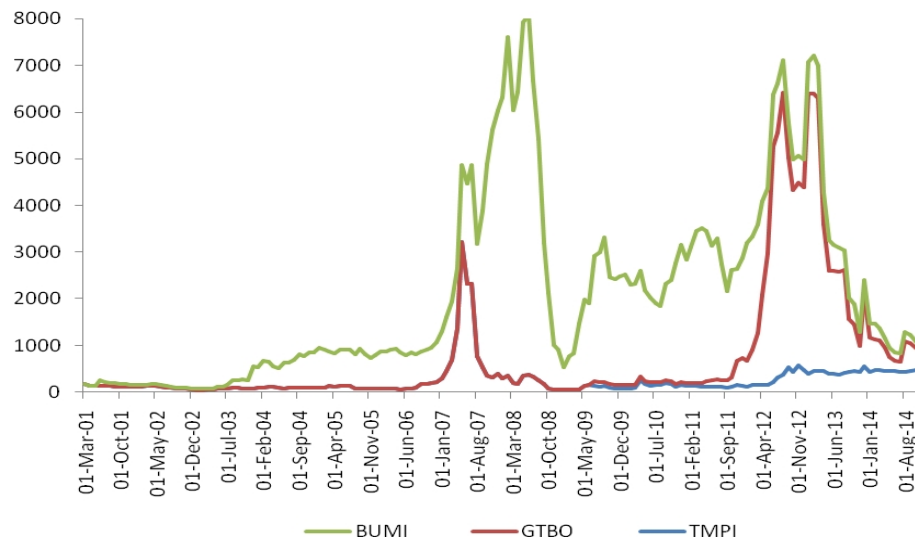


Figure 2. Stock price chart pattern of TMPI, GTBO and BUMI
Source: Indonesia's stock exchange

TEXT MINING ANALYSIS

This study used R programming to develop text mining to search key terms with regards to the stock price manipulation (see Appendix for R code – text mining analysis). The text-mining analysis aims to extract key metadata information from the data source of Tweeter tweets. The text-mining in R used some of the predefined libraries incorporated in the R code, such as R library of “tm” and “tweeter”. The “tm” library provides a framework for text mining applications within R. A “twitterR” library is an R package which provides access to the Twitter API. These libraries are able to capture all related contents and financial information within the document.

Table 3 explains the word frequency and patterns produced by the R results. From the tweets, the “tweeter” library produces several common key terms such as “plunge (rugi in Bahasa)”, “drop (jatuh)”, and “fall (turun)” with high frequency of occurrence. It also shows that terms related to financial statements such as “loss (rugi)”, “profit (laba)”, “net (bersih)”, “sales (penjualan)” and “revenue (pendapatan)” also have a high number of occurrences in our search. It tells us that most investors more concern on the profit and loss statement rather than analyzing the big picture of corporate fundamental valuation as a whole.

In this study, we exclude the variable name itself such as “bumi”, “gtbo”, “tmpi”, “agis”, “resources”, or “gardatujuhbuana” as they are self-explanatory and already have high frequency in our datasets. We also remove “stopwords” such as “on”, “the”, “a”, “between”, “through” as they do not provide meaningful information into our analysis.

Table 3. Frequent words and key terms

TMPI	GTBO	BUMI
Anjlok (plunge)	Anjlok (plunge)	Anjlok (plunge)
Laba	Rugi	Rugi
(profit)	(loss)	(loss)
	Penjualan (sales)	Jatuh
		(fall)
		Penurunan
		(drop)
		Pendapatan (revenue)

Source: R results

Despite the large size of the documents, the R code is sufficiently well enough to demonstrate how text mining analysis could be used to generate common key words of stock price manipulation.

Sentiment Analysis. We also perform sentiment studies based on R code written by Breen (2012). The sentiment analysis measures polarity bias that allows us to classify some texts as positive or negative. For each tweet, we count total number of positive and negative words and calculate sentiment score as the net sum of positive and negative words. In other words, when sentiment score is positive, it indicates an overly optimism on the stock and vice versa. The lists of positive and negative words are collected from opinion lexicon English database (Liu, 2012).

Our results show that investors tend to have negative bias when the stock price decreases significantly as we expect. We found 4783 negative words and 2006 positive words within our observation periods from January 2000 to December 2014. The ratio of negative words to total number of words is about 70.5 per cent indicating the negative bias outweighing positive sentiment.

CONCLUSION

This study attempts to search textual patterns that are associated with investor behavior and sentiment during stock price manipulation in Indonesia's Stock Exchange. We propose text mining analysis to demonstrate the process of extracting information from news and media release of Tweeter over the period of 2000 to 2014. We took three sample firms of PT AGIS Tbk (TMPI), PT. Garda Tujuh Buana Tbk (GTBO) and PT Bumi Resources Tbk (BUMI) that have been indicated experiencing stock price manipulation. These firms experienced an abnormal stock price appreciation and followed with a sharp decline of about 90 per cent within short period of time. The extracted information from Tweeter and sentiment analysis provide evidence that news and media releases tend to yield common negative key terms such as "plunge", "drop", and "fall"; as well as terms related to income statements such as "loss", "profit", "net", "sales" and "revenue". It shows us that most investors would have negative bias and suffer more when the stock price experienced a significant decline. It also reveals that investors are more short-sighted on the profit and loss statements rather than analyzing the big picture of corporate fundamental valuation as a whole. We thus argue that these knowledge bases

and key terms could be used to reflect common characteristics of investor behavior during stock price manipulation in the Indonesia's stock market.

For future works, it is anticipated to expand the cases to evaluate the text mining model by extracting information based on contents from financial information such as annual reports, financial statements and corporate presentations. It is expected that extracted information from financial reports would provide more insightful information for the identification of investor behaviour towards the upside or downside of the stock price in the stock market.

REFERENCES

- Allen, F., and Gale, D. (1992) "Stockprice manipulation", *Review of Financial Studies*, vol. 5, issue 3, pp. 503-529
- Breen, J. O. (2012) Mining Twitter for airline consumer sentiment. *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, pp. 133
- Diaz, D., Theodoulidis, B., and Sampaio, P. (2011) Analysis of stock market manipulations using knowledge discovery techniques applied to intraday trade prices, *Expert Systems with Applications*, vol. 38, issue 10, pp.12757-12771
- Fung, G. P. C., Yu, J. X., and Lam, W. (2003) Stock prediction: Integrating text mining approach using real-time news, In *Computational Intelligence for Financial Engineering 2003 Proceedings*, pp. 395-402
- Gerace, D., Chew, C., Whittaker, C., and Mazzola, P. (2014) "Stock Market Manipulation on the Hong Kong Stock Exchange", *Australasian Accounting, Business and Finance Journal*, vol. 8, issue 4, pp. 105-140
- Gidófalvi, G., and Elkan, C. (2001) Using news articles to predict stock price movements, *Department of Computer Science and Engineering, University of California, San Diego*
- Imisiker, S., and Tas, B. K. O. (2013) "Which firms are more prone to stock market manipulation?" *Emerging Markets Review*, vol. 16, pp. 119-130
- Liu, X., Liu, Z., and Qiu, Z. (2013) "Stock Market Manipulation in the Presence of Fund Flows", *Annals of Economics and Finance*, vol. 14, issue 2, 481-489
- Liu, B. (2012) Sentiment analysis and opinion mining, *Synthesis Lectures on Human Language Technologies*, vol. 5, issue 1, pp. 1-167
- Mittermayer, M. A. (2004) Forecasting intraday stock price trends with text mining techniques", In *System Sciences, 2004, Proceedings of the 37th Annual Hawaii International Conference*, pp. 1-10
- Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., and Sun, X. (2011) The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature, *Decision Support Systems*, vol. 50, issue 3, pp.559-569
- Schumaker, R. P., and Chen, H. (2009) Textual analysis of stock market prediction using breaking financial news: The AZFin text system, *ACM Transactions on Information Systems (TOIS)*, vol. 27, issue 2, pp. 1-12
- Shirata, C. Y., and M. Sakagami. (2009) "An analysis of the 'going-concern assumption': Text mining from Japanese financial reports", *Journal of Emerging Technologies in Accounting*, vol. 6, pp.1-16

- Zaki, M., Theodoulidis, B., and DíazSolís, D. (2011) "Stock-touting through spam e-mails: a data mining case study", *Journal of Manufacturing Technology Management*, vol. 22, issue 6, pp. 770-787
- Zaki, M., and Theodoulidis, B., (2014) Analysing stock market fraud cases Using a Linguistics Based Text Mining Approach, Paper presented at the Finance and Economics on the Semantic Web (FEISW 2014) conference, Crete (Greece), 25-29 May 2014.

Appendix

R-code for Twitter sentiment analysis

Load the required R libraries

```
library(twitterR)
library(ROAuth)
library(RCurl)
library(tm)
library(stringi)
library(wordcloud)
library(RColorBrewer)
library(ggplot2)
library(plyr)
library(stringr)
```

#Setup twitter API

```
setup_twitter_oauth('consumerKey',
'consumerSecret', 'accessToken', 'accessTokenSecret')
```

#Search tweets

```
stock_tweets = searchTwitter("your keywords",
n=1000)
```

Convert tweets into a data frame

```
df <- do.call("rbind", lapply(stock_tweets,
as.data.frame))
dim(df)
```

Build a corpus, which is a collection of text documents

VectorSource specifies that the source is character vectors.

```
myCorpus <- Corpus(VectorSource(df$text))
```

#Data cleaning

```
myCorpus <- tm_map(myCorpus,
content_transformer(tolower), mc.cores=1)
myCorpus <- tm_map(myCorpus,
removePunctuation)
myCorpus <- tm_map(myCorpus, removeNumbers)
myCorpus <- tm_map(myCorpus, stripWhitespace)
myCorpus <- tm_map(myCorpus, removeWords,
stopwords("english"))
myCorpus <- tm_map(myCorpus,
PlainTextDocument)
```

#Build document term matrix(tdm)

```
tdm <- TermDocumentMatrix(myCorpus, control =
list(minWordLength = 1))
```

Define tdm as matrix

```
m = as.matrix(tdm)
```

Get word counts in decreasing order

```
word_freqs = sort(rowSums(m), decreasing=TRUE)
```

```
# Create data frame with words and their frequencies
dm = data.frame(word=names(word_freqs),
freq=word_freqs)
```

Plot wordcloud

```
wordcloud(dm$word, dm$freq,
random.order=FALSE, colors=brewer.pal(8,
"Dark2"))
```

#Sentiment analysis

#Search tweets

```
tweets.text =
lapply(stock_tweets,function(t)t$getText())
```

#Scan positive and negative words

```
pos = scan('c:/positive-words.txt', what='character',
comment.char=';')
```

```
neg = scan('c:/negative-words.txt', what='character',
comment.char=';')
```

```
score.sentiment = function(sentences, pos.words,
neg.words, .progress='none')
```

```
{
  require(plyr)
  require(stringr)
  # W got a vector of sentences. plyr will handle a
  list
  # or a vector as an "l" for us
  # We want a simple array ("a") of scores back, so
  we use
  # "l" + "a" + "ply" = "lapply":
  scores = lapply(sentences, function(sentence,
pos.words, neg.words) {
```

```
  # Clean up sentences with R's regex-driven
  global substitute, gsub
```

```
():
```

```
  sentence = gsub('[[:punct:]]', "", sentence)
```

```
  sentence = gsub('[[:cntrl:]]', "", sentence)
```

```
  sentence = gsub("\\d+", "", sentence)
```

```
  #and convert to lower case:
```

```
  sentence = tolower(sentence)
```

```
  # Split into words. str_split is in the stringr
  package
```

```
  word.list = str_split(sentence, "\\s+")
```

```
  # Sometimes a list() is one level of hierarchy too
  much
```

```
  words = unlist(word.list)
```

```
  #Compare our words to the dictionaries of
  positive & negative terms
```

```
  pos.matches = match(words, pos.words)
```

```
  neg.matches = match(words, neg.words)
```

```
  # Match() returns the position of the matched
  term or NA
```

```
  # We just want a TRUE/FALSE:
```

```
  pos.matches = !is.na(pos.matches)
```

```
  neg.matches = !is.na(neg.matches)
```

```
  # TRUE/FALSE will be treated as 1/0 by
  sum():
```

```
    score = sum(pos.matches) - sum(neg.matches)
    return(score)
}, pos.words, neg.words, .progress=.progress )
scores.df = data.frame(score=scores,
text=sentences)
return(scores.df)
}
```

```
#Result of sentiment analysis
analysis = score.sentiment(tweets.text, pos, neg)
table(analysis$score)
mean(analysis$score)
qplot(analysis$score)
```