

Using Twitter to predict Stock Market Returns (August 2015)

Dinesh Sonachalam, 2nd Year, Velammal Institute of Technology, Chennai

Abstract—Twitter is an online social networking service that enables users to send and read short 140-character messages called "tweets". Twitter generates 8TB data per day! This huge dataset can be analyzed using data mining techniques to find interesting trends, combat crime, etc. It also proves to be an interesting source for predicting stock market trends using the emotions expressed by twitter users. Sentiment Analysis is a data mining technique which aims to determine the attitude of a speaker or a writer with respect to some topic. We know that people are easily attracted towards emotional contents, gossips, trends, fashion, etc. When it comes to social networking sites, people express their perspective and their opinions about the things they experience. Normally, if we are planning to go for a movie we see the reviews of that movie to check it will be worth to spend time watching the movie. Using the same principle, we can predict the future of an industry to predict gain or loss in stock market..

Index Terms—Twitter, Sentiment Analysis, Stock Market, Forecasting, Time Series Analysis.

I. INTRODUCTION

Twitter is an online social networking service that enables users to send and read short 140-character messages called "tweets". Twitter generates 8TB data per day! This huge dataset can be analyzed using data mining techniques to find interesting trends, combat crime, etc. It also proves to be an interesting source for predicting stock market trends using the emotions expressed by twitter users.

II. BIG DATA

Big data is a term given to large datasets where traditional data processing approaches don't work. Systems like Hadoop are built to handle huge volumes of data. Loading data into databases and in-memory applications like excel, notepad are not sufficient when handling huge volumes of data. These huge datasets can be analyzed using data mining techniques to find interesting trends, spot diseases, find cures, combat crime, etc.

Companies like Amazon use Big Data and Data mining to suggest products to consumers. Facebook and LinkedIn use Big Data to suggest friends.

Big data is a \$125 billion dollar industry. Many companies like Oracle, IBM have developed own Big Data tools to analyze and predict data.

III. TWITTER

Twitter is an online social networking service that enables users to send and read short 140-character messages called "tweets". The service rapidly gained worldwide popularity, with more than 100 million users who in 2012 posted 340 million tweets per day. In 2013 Twitter was one of the ten most-visited websites, and has been described as "the SMS of the Internet". As of May 2015, Twitter has more than 500 million users. Twitter generates 8TB data per day!

IV. SENTIMENT ANALYSIS

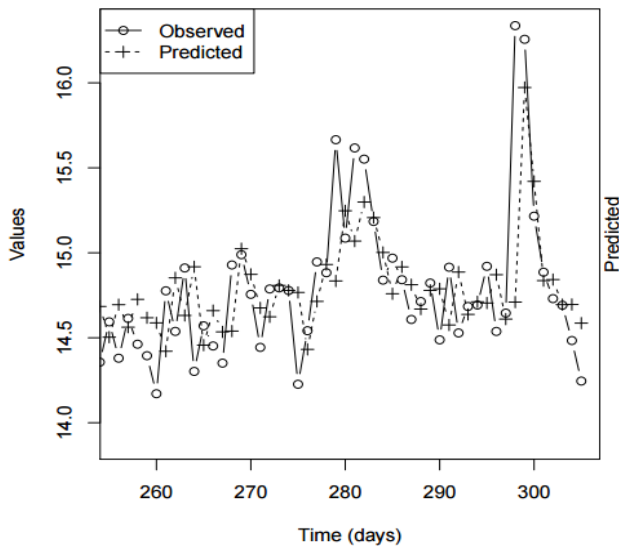
Sentiment Analysis is a data mining technique which aims to determine the attitude of a speaker or a writer with respect to some topic. In other words, it determines whether a piece of writing is positive, negative or neutral. Sentiments express deep feeling and mental attitude of a person.

Many websites now provide free sentiment analysis based on input text.

V. OUR APPROACH

People share both their good and bad emotions about a company in Twitter. This is considered the base for trading. The trading and investing community on Twitter also share a lot of ideas and trades. Using automated methods, these Twitter messages are gathered, stored and analyzed for sentiment. The sentiments found in these messages are then combined to create various signals. The signals are considered as signs to trade. For example if there are large number of positive tweets on Infosys, a positive signal is generated indicating that the stock is a good investment option.

VI. RESULTS



Using the approach mentioned a correlation between Twitter sentiment data and stock market movement can be seen. This analysis has been proven to be effective in predicting market data up to 3 days. This lays the foundation for further research in algorithm trading methods that incorporate sentiment data to increase accuracy and, ultimately, profits.

VII. CONCLUSION

This proves that social media and social networking sites reflect the current emotions of users and can be used as a potential indicator to predict ever changing stock prices.

REFERENCES

- [1] Sentiwordnet, an enhanced lexical resource for sentiment analysis and opinion mining.
- [2] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. 2010.
- [3] Albert Bifet and Eibe Frank. Sentiment knowledge discovery in twitter streaming data. 2011.
- [4] J. Bollen and H. Mao. Twitter mood as a stock market predictor. *IEEE Computer*, 44(10):91–94, 2010.
- [5] A. Bromberg. Sentiment analysis in python. 2013.
- [6] Eric D. Brown. Will twitter make you a better investor? a look at sentiment, user reputation and their effect on the stock market. *SAIS Proceedings*, 2012.
- [7] Ray Chen and Marius Lazer. Sentiment analysis of twitter feeds for the prediction of stock market movement. 2012.
- [8] Hal Daum. Notes on cg and lm-bfgs optimization of logistic regression. 2004.

[9] Shangkun Deng, Takashi Mitsubuchi, Kei Shioda, Tatsuro Shimada, and Akito Sakurai. Combining technical analysis with sentiment analysis for stock price prediction. In *Proceedings of the 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, pages 800–807, IEEE Computer Society, Washington, DC, USA., 2011. DASC '11 IEEE Computer Society.

[10] S. Elson, D. Yeung, R. Parisa, S. . R. Bohandy, and A. Nader. Using social media to gauge iranian public opinion and mood after the 2009 election. 2012.

[11] Ronen Feldman. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89, 2013.

[12] E. Hurwitz and T. Marwala. Common mistakes when applying computational intelligence and machine learning to stock market modelling. 2012.

[13] Anshul Mittal and Arpit Goel. Stock prediction using twitter sentiment analysis. 2012.

[14] Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. *Proceedings of LREC*, 2010.

[15] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. pages 79–86, 2002.

[16] DongSong Zhang and Lina Zhou. Discovering golden nuggets: Data mining in financial application. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 34(4):513–22, 2004.