

Analyzing the Stock Market behavior Using Event Study and Sentiment Analysis on Twitter Posts

¹K. Tejashwini, ²B. Saleena, ³B. Prakash and ⁴Sharon Shopia

¹Fikka Technologies, Bangalore.

²School of Computing Science and Engineering,
VIT, Chennai Campus, Vandalur, Chennai.

³School of Computing Science and Engineering,
VIT, Chennai Campus, Vandalur, Chennai.

⁴VIT Business School
VIT, Chennai Campus, Vandalur, Chennai.

Abstract

Users broadcast brief text information within their known public or selected group of people, these updates are known as tweets, and this form of online communication is known as microblogging. Opinions and feelings of people are aggregated by social networking services at a very low cost. The proposed system aims to analyse the stock market variation with respect to twitter posts, event study and financial ratios, using neural network. Investigations have been conducted to study the correlation between the measurements of collective mood states derived from twitter tweets and stock market value of companies. Event study in financial management is considered as measuring the effects of valuation of a corporate event such as earning announcements or merger, by determining the response of the stock market value around the announcement or occurrences of the event. For this study, top 10 Indian IT company tweeter feeds are taken for sentiment analysis and Cumulative abnormal return (CAR) for these 10 companies are computed for this event. The final model also consists of considering the importance of independent variable of financial ratios such as Profit after tax(PAT) net of P&E as % of total income net of P&E, PAT net of P&E as % of net worth, Current ratio, Debt to equity ratio, Interest cover, Quick ratio, Average cost of Funds, Total income/total assets, Total income/compensation to employees. According to researchers, neural network is one of the best methods in predicting stock market value. In Neural network, multilayer perceptron is considered in this paper. Neural network model is computed using three main features: mood states, CAR and financial ratios.

Index Terms: Stock market, Tweets, Sentiment analysis, event study, cumulative abnormal ratio (CAR), multilayer perceptron privacy.

1. Introduction

Stock market prediction is about finding out the particular company's future stock value traded on a financial exchange [12]. Significant profit is obtained by a successful and efficient prediction of stock's future price. Investors are better guided when the direction of the market is successfully predicted. Over the years, many methodologies and tools have been developed for predicting the market value. Before it would take long time to spread any information regarding a company, any good news or rumors can be easily spread all over the world within few minutes through microblogging. The short term performance of financial market such as stocks or micro-economic level is highly influenced on short term sentiments obtained from social media buzz. Earlier, according to Efficient Market Hypothesis (EMH) it was told that stock value movement follows random walk hypothesis and it is mostly unpredictable. Random walk hypothesis states that the price of a particular financial instrument follows random walk.

In recent days, Twitter becomes one of the most accepted social media among financial community. Short messages from twitter called as tweets can be conveniently accessed through application programming interface (API). Many sub forums like Stocktwits and TweetTrader are as part of twitter, acts as a discussion platform among financial advisors and investors. Here, the company tweets have been taken from www.topsy.com, which is a certified twitter partner, social search and analytics company. Sentiment analysis is done using R-studio and emotions of each tweet and polarity is computed. Bullishness, agreement and percentage of each emotion for each company are computed using these emotions and polarity.

Event study is a statistical method to study impact of an event on the value of a financial firm. Cumulative abnormal return (CAR) is computed based on the event occurred during period of time. Sum of all abnormal returns is known as cumulative abnormal returns. Abnormal changes in the stock price are determined for a particular event, here Satya Nadella joining as CEO of Microsoft during 2014 is considered as an event. Neural networks are getting widely used in various domains like medicine, finance, geology, physics and engineering. Here multilayer perceptron is used using SPSS tool. Emotions computed from sentiment analysis and financial ratios obtained from Prowess software for each company acts as independent variables to the neural network model.

Fig: 1.1 depicts the overall idea of the proposed system. The importance of variables from sentiment analysis and financial ratios on stock price for the particular event is obtained. The figure shows three main parts needed for computing the neural network model. Based on the information about importance of each variable, the future stock prices estimation can be made efficiently.

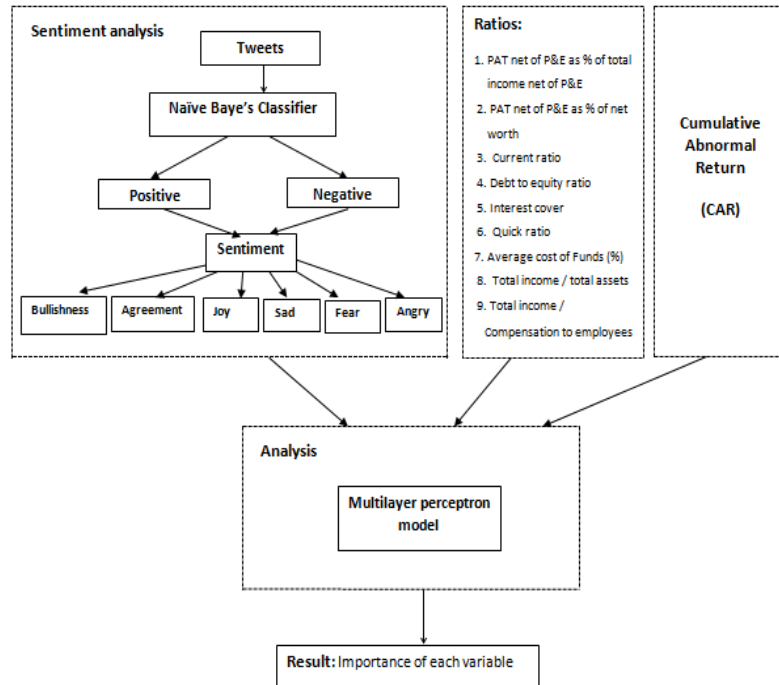


Fig 1.1: Block Diagram of the Proposed System

The investor gets a brief idea about stock price by considering the variables with high importance. The rest of the paper is organized as follows: Section II discusses about the related work, Section III is about the implementation and in section IV the results are evaluated and in section V conclusions are discussed by outlining the scope for future work.

2. Related Works

This section discusses about various research work that has been done for predicting stock market based on twitter posts by different researchers. Sheng Yu et al [1] discussed about mining the social media attributes and contents, that gives an opportunity to explore more on social structure characteristics, analyse qualitatively and quantitatively action patterns. It gives various backgrounds where social media can be predicted. According to Xue Zhang et al [2], hope and fear of the investors are measured on daily basis and the correlation between these emotional factors and stock market indicators are analysed. According to Chris Loughlin et al, [3] they considered tweets from a period of three months and predicted the stock market of four companies by using a linear model to predict daily stock returns. They have considered three estimators, the Bear index pattern, the Bull index pattern, and the Google Trend index pattern. After computing the significance of these estimators cross correlation is created to determine the best estimator in determining stock value. According to

Panagiotis Papaioannoet al [4], forecasting of foreign exchange (intra-day) is computed using twitter posts. They have used time series and trading simulation analysis. It provides valuable evidence that in certain cases the information provided in social platforms such as Twitter can enhance the forecasting efficiency regarding the very short (intra-day) foreign exchange. AbhishekKar, [5] has predicted stock price using artificial neural network. Back propagation algorithm is used in neural network model. Based on the past data of high, low, open, close stock value as inputs, the future close value is predicted. Danfeng Yan et al [10] brought out the relationship between stock market prices and mood of public from blogging.

3. Proposed System Implementation

Sentiment Analysis

Sentiment analysis in general, is a task of identifying whether the opinion expressed in the text is positive or negative about a particular topic or context. Twitter posts for companies are taken from top websites. Companies which are considered and their stock symbol are shown in fig 3.1.1. Twitter posts obtained is given to R-studio. Firstly, preprocessing of the tweets are done, like removing symbols, spaces, URLs. The total number of posts obtained for each company from 20-01-2014 to 20-02-2014 is shown in fig 3.1.2. Naive Baye’s classifier is used to classify the tweets into positive and negative polarity. Tweets can be positive or negative based on the event. Naive Baye’s is a simple probabilistic classifier based on Baye’s theorem. According to Baye’s rule, any event’s outcome (H) can be estimated based on the evidence (E).

Company name	Stock Symbol	Company name	Number of Tweets
Accel Frontline Ltd.	\$AFL	Accel Frontline Ltd.	35
H C L Technologies Ltd.	\$HCLTECH	H C L Technologies Ltd.	54
Infosys Ltd.	\$INFY	Infosys Ltd.	89
I T C Ltd.	\$ITC	I T C Ltd.	94
Mphasis Ltd.	\$MPHASIS	Mphasis Ltd.	79
Oracle Financial Services Software Ltd.	\$ORACLEFIN	Oracle Financial Services Software Ltd.	15
Ramco Systems Ltd.	\$RAMCOSYS	Ramco Systems Ltd.	185
Tata Consultancy Services Ltd.	\$TCS	Tata Consultancy Services Ltd.	54
Tech Mahindra Ltd.	\$TECHM	Tech Mahindra Ltd.	92
Wipro Ltd.	\$WIPRO	Wipro Ltd.	87

Fig 3.1.1: Company Name with its Stock Symbol Fig 3.1.2: Total Number of Tweets

Typically, better classification accuracy is obtained when there are more evidences. Baye's rule for multiple evidences:

$$P(H|(E_1, E_2, \dots, E_n)) = \frac{P(E_1, E_2, \dots, E_n|H) * P(H)}{P(E_1, E_2, \dots, E_n)}$$

The tweets are classified to one of the five emotions, joy, sad, fear, anger and surprise. Fig 3.1.3, shows the graphical representation of emotions in tweets obtained about \$infy. Fig 3.1.4, shows the graphical representation of polarity of tweets obtained for \$infy. Sentiments obtained are converted into numerical form using the below formulae,

(1) Bullishness

$$B_t = \ln \left(\frac{1 + M_t^{Positive}}{1 + M_t^{Negative}} \right)$$

(2) Agreement

$$B_t = 1 - \sqrt{1 - \frac{M_t^{Positive} - M_t^{Negative}}{M_t^{Positive} + M_t^{Negative}}}$$

From agreement, we can determine whether positive or negative has higher influence on the stock value. Further, different human feelings of each tweet are represented in emotional terms such as joy, sad, fear, anger and surprise.

(3) Positive % = (total +ve tweets) / (total no. of tweets) (4) Negative % = (total -ve tweets) / (total no. tweets)

(5) Joy % = (total joy tweets) / (total number of tweets) (6) Sad % = (total sad tweets) / (total number of tweets)

(7) Fear % = (total fear tweets) / (total number of tweets) (8) Anger % = (total anger tweets) / (total number of tweets)

(9) Surprise % = (total surprise tweets) / (total number of tweets)

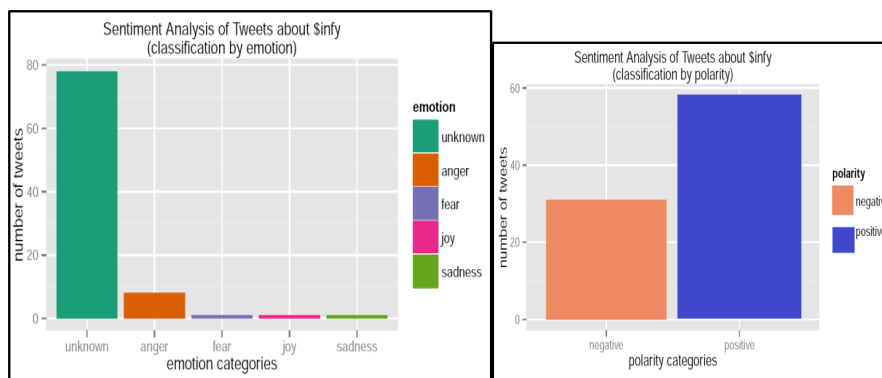


Fig 3.1.3: Classification by Emotion (\$infy) Fig 3.1.4: Classification by Polarity (\$infy)

Event Study

Event considered here is, SatyaNadella was appointed as CEO of Microsoft on 4th February 2014, succeeding Steve Ballmer. Cumulative abnormal return is calculated for +10 days and -10 days for each company.

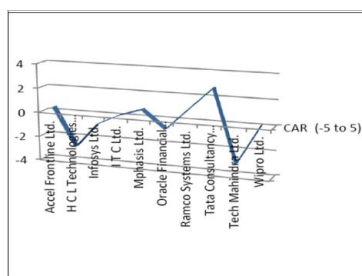


Fig 3.2.1: Trend Analysis

J	A	B	C	D	E	F	G	H	I	J	K
1	Date	Close Price day	Index	daily return	market return	alpha	beta	theoretical return	abnormal return	CAR	
2	20-Jan-14	3749.3	9661.46								
3	21-Jan-14	3753.03	-10	935.29	0.11313742	-0.68499909	0.043885	1.017873	-0.83220234	0.80647957	
4	22-Jan-14	3767.2	9	963.6	0.29497795	0.19512056	0.043885	1.017873	0.23820694	0.0680725	0.572125
5	23-Jan-14	3762.8	-8	959.79	0.4822224	-0.216588195	0.043885	1.017873	-0.176310297	0.838713897	0.93462
6	24-Jan-14	3759.05	-7	9529.51	-0.8898459	-0.65338374	0.043885	1.017873	-0.82366034	-0.266757721	0.591979
7	27-Jan-14	3712.8	-6	9469.09	-0.781632	-0.6383012	0.043885	1.017873	-0.89293495	-0.99646312	-0.3648
8	28-Jan-14	3677	-5	9388.86	-1.4897982	-1.50494962	0.043885	1.017873	-1.03318294	-0.45409016	-0.5348
9	29-Jan-14	3717.65	-4	9421.69	1.10552081	0.56888911	0.043885	1.017873	0.617740367	0.67776638	0.01532
10	30-Jan-14	3704.75	-3	9408.56	-0.3469594	-0.13959237	0.043885	1.017873	-0.397936883	-0.24876513	0.218724
11	31-Jan-14	3695.45	-2	9476.42	-0.1480596	0.73388812	0.043885	1.017873	0.76665944	-0.02112129	-1.17217
12	1-Feb-14	3628.7	-1	9321.09	-1.5124464	-1.64139697	0.043885	1.017873	-1.430316779	-0.28612666	-1.29524
13	4-Feb-14	3566.45	0	9154.99	-1.7997981	-1.78190434	0.043885	1.017873	-1.769888336	0.05609556	-2.2852
14	5-Feb-14	3581.25	1	9242.29	0.4177396	0.95378122	0.07954	0.773398	0.81722308	-0.39946677	-0.33986
15	6-Feb-14	3583.45	2	9234.47	-0.501678	-0.69090733	0.07954	0.773398	-0.02120405	-0.45443216	-0.6388
16	7-Feb-14	3566.7	3	9185.05	0.09120375	-0.624493994	0.07954	0.773398	-0.403247037	0.48470787	0.040605
17	10-Feb-14	3568.5	4	9120.38	0.05946602	-0.421744892	0.07954	0.773398	-0.246728829	0.297187667	0.791954
18	11-Feb-14	3555.8	5	9220.57	0.7602732	0.96626876	0.07954	0.773398	0.342420989	-0.07676868	0.220299
19	12-Feb-14	3599.8	6	9242.79	0.15567885	0.34127327	0.07954	0.773398	0.267640504	-0.15967066	-0.33885
20	13-Feb-14	3586.45	7	9234.57	-0.3031884	-0.688954185	0.07954	0.773398	0.03740713	-0.37620987	-0.53883
21	14-Feb-14	3645.8	8	9363.17	1.4492052	1.97029206	0.07954	0.773398	1.50646721	0.45240476	0.13057
22	17-Feb-14	3688	9	9379.12	0.9446026	1.1704829	0.07954	0.773398	0.21213095	0.76232424	1.204899
23	18-Feb-14	3662.25	10	9377.44	-0.482337	-0.617912128	0.07954	0.773398	0.066883207	-0.54800584	0.184262

Fig 3.2.2: CAR Computation for \$infy

The trend analysis for the companies based on CAR value. The CAR computation for \$infy is shown in the above Fig 3.2.2. From fig 3.2.1, we can infer that companies HCL, Tech Mahindra and Wipro are highly sensitive to the event. Whereas Infosys, Oracle, Ramco are slightly sensitive and other companies are not very sensitive to the event. By this graph, the sensitivity of the companies to the event can be easily analysed. Data needed for calculating CAR are market returns and closing price of the stock. Daily returns or actual returns and theoretical returns should be calculated for each day using the below formulae,

$$\text{Daily returns (actual return)} = (\text{present day closing price} - \text{previous day price}) / (\text{previous day closing price}).$$

$$\text{Theoretical return} = \text{alpha} + \text{beta} * (\text{market return for the day})$$

where, beta is slope and alpha is intercept, which should be calculated for both before event and after event. Abnormal return is calculated using below formula and CAR is the sum of abnormal returns.

$$\text{Abnormal return} = \text{actual return} - \text{theoretical return}.$$

Neural Network and Analysis Using SPSS

Perceptron is a basic linear classifier. By implementing the multilayer perceptron concept, Input x is fed to the input layer, the activation propagates in forward direction. Each hidden unit is a perceptron. The values of hidden units z_h is calculated by applying activation function to its weights. In the below formula activation function considered is sigmoid.

$$z_h = \text{sigmoid}(w_h^T x) = \frac{1}{1 + \exp[-(\sum_{j=1}^d w_{hj} x_j + w_{h0})]}, \quad h = 1, \dots, H$$

The output y, is a perceptron from output layer which takes hidden units (zh) as their inputs.

$$y_i = v_i^T z = \sum_{h=1}^H v_{ih} z_h + v_{i0}$$

SPSS (Statistical package for social sciences) is an IBM product used for statistical analysis. For predictor variable, we include sentiment variables which

are converted to ordinal form. Financial ratios are used as covariates. Activation function used in hidden layer is hyperbolic tangent,

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1} = \frac{1 - e^{-2x}}{1 + e^{-2x}}$$

Error estimation used is cross-entropy. It is more useful when the target is 0 and 1. In output layer, activation function used is Softmax.it is used when all the dependant variables are categorical.

4. Results

Result shows the network information, which gives the details about error used, number of hidden layers, activation function and variables used. It gives the neural network model and model summary which provides the error value. It also gives the independent variable importance and the importance graph.The Network summary obtained through Neural networks describes about the variable used, dependant variable used is CAR(-5 to +5).Activation functions used in hidden layer is Hyperbolic tangent and in output layer is softmax.The neural network is modelled with input layer, hidden layer and output layer. Output layer has two states 0 and 1. Fig 4.1 describes the model summary. Training time is 00:00:00:001. There is 0% of incorrect predictions. Training is stopped when there is no decrease in the error. Cross entropy error for training is 0.006.

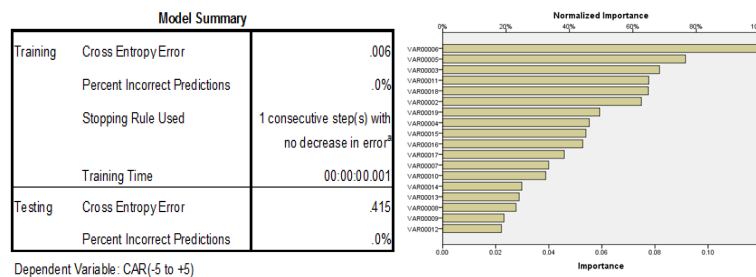


Fig 4.1: Model Summary Fig 4.2: Independent Variable Importance Graph

A sensitivity analysis is conducted where each predictor importance of independent variables are calculated in determining the neural network. The analysis is based on the consideration of both training and testing samples or based only on the training sample if testing sample is unavailable. The independent variable importance is a measure of how much the network’s model-predicted value changes for different values of the independent variable.Fig 4.2, shows the chart of variable importance. Normalized importance is simply the importance values divided by the largest importance values and the results are expressed in percentages.It is observed that among the financial ratios, interest cover plays the major role in determining the stock value. From sentiment analysis, bullishness, joy, sad, fear, surprises is more important and adds a great weightage in determining the stock value.

5. Conclusion and Future Work

Predicting stock market is no more a random walk hypothesis, there are commercial and economical indicators which can be used in analysing the stock value. Analysing stock value using social media provides knowledge about the market behaviour to the investor. And it's very useful to the company in determining changes which can occur due to any event. By studying the company details and knowing about the serious changes made in the company, and adding more events, additional variables to the neural network model can be added. By taking data from more than one social media and doing sentiment analysis, will give an efficient model for analysing stock value.

References

- [1] Sheng Yu, SubhashLal, A Survey of Prediction Using Social Media (2012).
- [2] Zhang X., Fuehres H., Gloor P.A., Predicting stock market indicators through twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences* 26 (2011), 55-62.
- [3] Loughlin C., Harnisch E., *The Viability of StockTwits and Google Trends to Predict the Stock Market*, Springer (2013).
- [4] Papaioannou P., Russo L., Papaioannou G., Siettos C.I., Can social microblogging be used to forecast intraday exchange rates?, *Netnomics: Economic Research and Electronic Networking* 14(1-2) (2013), 47-68.
- [5] Kar A., *Stock prediction using artificial neural networks*, Dept. of Computer Science and Engineering, IIT Kanpur (1990).
- [6] Bollen J., Mao H., Zeng X., Twitter mood predicts the stock market, *Journal of computational science* 2(1) (2011), 1-8.
- [7] Gayo-Avello D., Metaxas P., Mustafaraj, E., Limits of electoral predictions using social media data, *Proceedings of the International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain (2011).
- [8] Ritterman J., Osborne M., Klein E., Using prediction markets and Twitter to predict a swine flu pandemic, *1st international workshop on mining social media* 9 (2009), 9-17.
- [9] Weerkamp W., De Rijke M., Activity prediction: A twitter-based exploration, *SIGIR Workshop on Time-aware Information Access* (2012).
- [10] Yan D., Zhou G., Zhao X., Tian Y., Yang F., Predicting stock using microblog moods, *China Communications* 13(8) (2016), 244-257.

- [11] RAJESH, M. "A SYSTEMATIC REVIEW OF CLOUD SECURITY CHALLENGES IN HIGHER EDUCATION." *The Online Journal of Distance Education and e-Learning* 5.4 (2017): 1.
- [12] Rajesh, M., and J. M. Gnanasekar. "Protected Routing in Wireless Sensor Networks: A study on Aimed at Circulation." *Computer Engineering and Intelligent Systems* 6.8: 24-26.
- [13] Rajesh, M., and J. M. Gnanasekar. "Congestion control in heterogeneous WANET using FRCC." *Journal of Chemical and Pharmaceutical Sciences* ISSN 974 (2015): 2115.
- [14] Rajesh, M., and J. M. Gnanasekar. "Hop-by-hop Channel-Alert Routing to Congestion Control in Wireless Sensor Networks." *Control Theory and Informatics* 5.4 (2015): 1-11.
- [15] Rajesh, M., and J. M. Gnanasekar. "Multiple-Client Information Administration via Forceful Database Prototype Design (FDPD)." *IJRESTS* 1.1 (2015): 1-6.
- [16] Rajesh, M. "Control Plan transmit to Congestion Control for AdHoc Networks." *Universal Journal of Management & Information Technology (UJMIT)* 1 (2016): 8-11.
- [17] Rajesh, M., and J. M. Gnanasekar. "Consistently neighbor detection for MANET." *Communication and Electronics Systems (ICCES), International Conference on. IEEE, 2016.*

