PREDICTION OF STOCK MARKET USING KALMAN FILTER

Mumtaz Ahmed[1], Krishan Chopra[2], Mohd Asjad[3]
[1,2,3]Department of Computer Engineering Jamia Millia Islamia,

Abstract

Market forecasting has always been a subject of numerous case studies and researches given its role in the macroeconomics of a nation. The fickleness in the market is well known. Of all the major factors that affect the Stocks, the non mathematical are very crucial such as the economic stance of the ruling government or elections to choose the government. In this paper, we trace the effect of the political shift after the elections through that entire period consisting of pre and post-election results by the tweets posted in regards to the Bombay Stock Exchange (BSE) behaviour. The prediction is put forth by implementing Kalman Filter algorithm that is fed on the sentiment analysis of the tweets we scraped during this period. The available Twitter API for tweet scraping works on a period of current week thus the tweets from the period of 1 April 2014 till 31 May 2014 have been scraped via advanced twitter search and sentiment analysis is done on a daily basis. The Kalman Filter has been fed on these sentiments to make a prediction based on a real time factor rather than considering purely regressive approach of mathematical factors.

Index Terms: Kalman Filter, Sentiment Analysis, Stock Markets, Time Series, Twitter

I. INTRODUCTION

Stock Markets have always been a very crucial factor in the economic system. The uncertain and fickle nature of stock markets has always been hard to write down in form of a definite set of equations. Especially since it is highly prone to changes with non-measurable factors on which our general regression techniques fail. One among such crucial factors includes the political uncertainty in the nation concerned, since the policies opted by the ruling government towards any particular firm may affect the investor's portfolio and his dividends [.4] Such political events usually have abrupt impact on the stock market. Usually the stock market fluctuates because of political announcements such as regulation promulgation, change of economic policies by the ruling government and during the national elections. In this paper, we analyze such factors and take into consideration the changing nature of political ground and its immediate effect on the stock market. For our research, we take into account the Indian general elections of 2014 and track the behaviour of stocks through the period via twitter feeds and make our predictions based on it.

The idea of measuring the social media feeds in capturing the sentiments of the masses (in this case, the investors) through their social media posts is a recent yet a very popular technique to measure the public response to any particular event. We measure the investor sentiments through their twitter feeds posted during the elections and post analysis, predict the inclination of the stock market.

Twitter is one among the social media which is highly accepted in the financial community. [1][4]

This have been employed in capturing the investor sentiment regarding the proceedings of the stock market from the period consisting both pre and post-election result declaration. The Sentiment Polarity of the twitter feeds is used to train Kalman filter in order to make a prediction

based on an exogenous factor that affects the stock market prices i.e. Political turbulence.

II. DATASET

For the purpose of getting the public sentiments on stock market, Twitter happens to be a good source specially due to its real time nature. [2][3] The data required included the tweets from the specified period of time. Although twitter does provide API for the purpose of scraping tweets but it works for the past week from the day of request and our period of analysis is beyond that duration. Thus, tweets had to be scraped using advanced twitter search. Those tweets include data from trusted finance authorities officiated by twitter as well as general sentiment of the masses with regard to the various upheavals of the stock market.

The Financial data used for training is from Yahoo Finance and is available in the form of a CSV (comma separated values) files. Yahoo Finance is dedicated for the purpose of maintaining the historical record of the transaction taking place among the prominent stock markets. The data consists of a record of the Sensex for the time period mentioned above, which is one among the major Indian Stocks.

III. DATA PREPROCESSING AND METHODOLOGY

Data preprocessing primarily involved the raw data from the internet to be bought in the form where the algorithms could be directly applied.

The tweets had to be filtered, so that only relevant tweets from the duration were taken into consideration. Fortunately, this was done directly via the advanced twitter search.

An important step in the preprocessing of data included taking an inner join between the sentiment scores and the closing price of stocks considering dates of trading as a common factor since Stocks do not trade on weekends and on civic holidays but the sentiment polarity was recorded for each day of the specified duration.

Once the data was preprocessed, Kalman Filter was implemented for getting a prediction of the closing prices with sentiment scores being fed as the Gaussian noise for each day.

The following is the flow graph of our study.

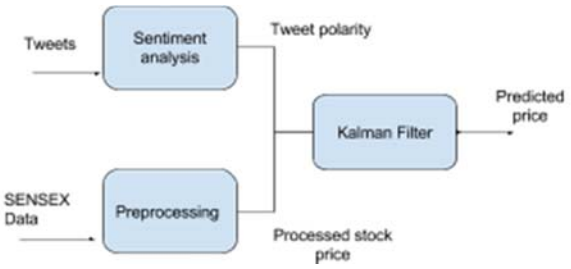

Fig. 1. Flow graph of the study

IV. PROBLEM FORMULATION

The core purpose of taking the problem of stock market prediction is that very few of the previous researches in this field have taken into account the odd behavior of stock market in certain exceptional cases involving the real world problems. In this paper, we put forth the validation of hypothesis that the existing "Public Sentiment" has an effect on the "Market Sentiment". And thus the two have a direct correlation with each other. Sentiment analysis of the available twitter feeds was performed to get the polarity of the tweets and to capture market opinion. Since this gives a clear insight on the extent of fluctuation the Stock is likely to suffer. Sentiment polarity of the scraped tweets from the month including pre and post the result declaration of the General election 2014, were supplied to the Kalman Filter. The following graph depicts the deviation of the closing prices with that of the public sentiment. The correlation factor for the parameters is 0.6572, which clearly indicates a correlation between the two parameters.

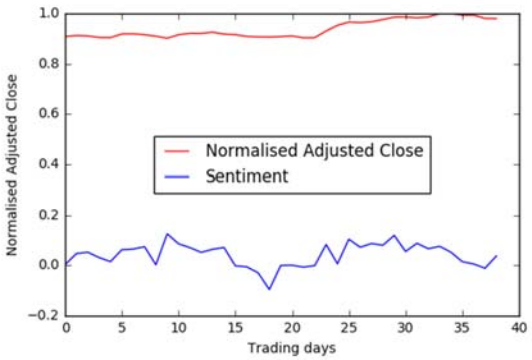

Fig2. Plot of Correlation between Normalized Adjusted Close and Sentiment

V. RELATED WORK

The central idea behind the paper was to include real world scenarios that affect Stock Markets and prices. The idea to work with the data scraped from the social media for Stock Price prediction is influenced by the work of Arpit Goel and Anshul Mittal's research that dealt with the similar problem, during their tenure at Stanford University. They tackled the problem with a different set of algorithms.

Similar to theirs' was the study conducted by the Manipal School of Management that did a predictive analysis of Stocks based on the Social Media Feeds.

Another Study which significantly influenced our research is the "Empirical study on Effect of Lok Sabha elections on Stock Market Performance (BSE SENSEX)" by team of researchers at VIT Business School. We were motivated by their idea of measuring market polarity through sentiment analysis of tweets.

Various other researches target this problem mostly via Machine Learning Algorithms but none proves to be as efficient as Kalman Filter under certain peculiar cases. Of all the known algorithms known to work on Gaussian space and noise, Kalman Filter is not well known but it proved to be very efficient in this use case. Its high efficiency is evident from the fact that it had been used by National Aeronautics and Space Administration (NASA) for its space missions. NASA Technical Memorandum 86847 titled "Discovery of the Kalman Filter as a Practical Tool for Aerospace and Industry" [14] specifies vividly the scope of the problems that can be targeted via this algorithm.

VI. SENTIMENT ANALYSIS

The sentiment polarity for day t S_t cannot be exclusively calculated by considering sentiment polarity of tweets on day t only. This has been deduced from the fact that market sentiment builds over course of time and sentiment of previous days also affect sentiment on any day t[4][7]. The following equation has been used to calculate sentiment polarity on day t [12].

$$S_t = \Phi_1 P_t + \Phi_2 S_{t-1} \quad (1)$$

S_t is the cumulative sentiment polarity of tweets of day t only. P_t is the average sentiment polarity of tweets of day t only. S_{t-1} is the cumulative sentiment polarity of day t-1. Φ_1 and Φ_2 are constants whose values have been calculated. Calculation of sentiment of tweets is a time series process which is recurring in nature.

The calculation of aggregate sentiment polarity for day (t) takes weighted contribution from sentiments of day t, t-1, t-2,.. ,t-k,.. 2,1. The contribution from day t has the highest weight. This is based on fact that the opinion and knowledge of traders are built over a time period and does not depend only on knowledge acquired on a particular day only.

$$S_t = \Phi_1 P_t + \Phi_2(\Phi_1 P_{t-1} + \Phi_2 S_{t-1}) \quad (2)$$

This recurrence relation continues from day t to day 1, So market sentiment of day t also includes sentiment value on day t, day t-1, t-2, till day 1. In this relation, the contribution from day t carries maximum weight and the weight of a particular day decreases as recurrence recede backwards from day t, t-1 till day 1. The opinion of traders is an autoregressive time series and news of events like mergers, acquisitions, elections affect opinion of traders and movement of market providing exogenous shock to the system thus affecting opinion of traders in stock market.

This recurrence can be solved for selecting optimum values of Φ_1 and Φ_2. The value of Φ_1 has been set to 0.5121 and of Φ_2 to 0.4834 in our experiment.

VII. KALMAN FILTER

The Kalman Filter [8][9][10][11] is a linear state space model that acts recursively on noisy input data and produces statistically optimal estimation of the system state. Due to dynamic nature of stock markets which are also affected by noise in the market, application of Kalman filter can help us find a statistically optimal estimate in such system.

The Observation equation[8][10][11][12] we have used to predict the price of stock on next day is

$$P_{t+1} = P_t * \ln(1 + S_t) + P_t \quad (3)$$

The given equation above is a linear regression between the adjusted close of stock on day (t) P_t and sentiment polarity of tweets on day (t) S_t to predict the adjusted close of stock on day (t+1) P_{t+1} .The given equation takes into account P_t and S_t to predict the price P_{t+1}. To predict P_{t+1}, we also calculate a bullishness factor for sentiment of day t. The bullishness factor is calculated on a scale of -1 to +1. The idea of bullishness has been borrowed from a paper by Zhang, Fuehres and Gloor [13]. When the sentiment on day t is less than 0 then the bullishness factor for the day is less than 0. Thus the predicted price P_{t+1} would be lesser than price on day t P_t by a value of $P_t * \ln(1 + S_t)$. If the market sentiment is negative, then the prediction for the next day decreases. But if the market sentiment is positive the Bullishness factor $\ln(1 + S_t)$is positive and the contribution from bullishness factor is positive so the predicted price for day (t + 1) P_{t+1} is more than price on day t. P_t by value of $P_t * \ln(1 + S_t)$. As per this equation as market sentiment becomes positive the predicted price for day (t+1) is more than the price on day t and if the market sentiment is negative then the predicted price for next day is lesser than price on day t P_t. So this equation aptly adapts with changing market sentiment.

Stock Markets are volatile and are affected by factors like traders' opinion [7]. The opinion of traders is affected by various events like elections, mergers and acquisitions. The trader opinion is expressed using various channels of which twitter is a good medium.

The effect of Indian General Elections in 2014 on Bombay Stock Exchange during the election period have been studied in this study [3][5][6]. During this phase a rapid change in market sentiments and rapid developments in political sphere of the country have been noted which may have affected the movement of stocks during this period. Application of Kalman filter in this scenario is appropriate as it can predict a better price in such noisy systems. Kalman filter had been used by NASA for orbit determination of satellites and finds wide use in Apollo, Ranger and Mariner missions.

Kalman Filter was used for trajectory prediction of space probes in case the path gets affected by various exogenous factors which are a source of noise while estimating the path of the space vehicle.

The Kalman filter takes into account both the recorded value and the noise in market sentiment to make a prediction. If the noise is less, the predicted value is closer to the recorded value and if the noise is more, then predicted value differs more from the recorded value in the direction of noise.

VIII. STATE SPACE MODEL OF EXPERIMENT

State Space Models [9][12] are a way to represent state variables. It allows users to model an observed time series to be expressed as vector of state variables which are driven by processes which are stochastic in nature.

For Our Study we have chosen State Space Model with two states. $X_k = [\ S_k\ P_k\ ]$ where S_k is the market sentiment on day k and P_k is the price of stock on day k. The general form of Kalman Filter State Space Model consists of two equations, a transition equation and an observation equation. The observation equation has been discussed earlier. The Transition Equation [8][10][11][12] is

$$X_{k+1} = A_k X_k + W_k \quad (4)$$

This Transition Equation updates the states in the state space model with time. It is an auto regression between the values of states and the noise W_k. In this equation W_k is noise which affects the prediction of state and in our experiment this noise is the variance in stock market sentiment during the time period of study. The volatility with which the market sentiment changes induces uncertainty in market and it has been considered as source of noise in the experiment. The Transition Equation is an autoregressive time series gives a better prediction of the states A_k. A_k is the transition matrix and has been set as a 2 dimensional Identity Matrix in the experiment. X_k is also called hidden state vector which is calculated by the transition equation The model has been tested on data from 1 April 2014, to 31 May 2014. The General Elections were conducted in various phases and results were declared during this time

period. The study aims to explore the effect of change in market sentiment on the movement of stock in the market during this time period.

IX. OBSERVATIONS

The test was run on adjusted close of SENSEX and tweets scrapped during the period 1 April 2014 to 31 May 2014 and the following observations were recorded.

Table I: Correlation Coefficient Table

Coefficient	Value
Pearson's Correlation	0.913

The above table displays the Pearson's Product-Moment Correlation between the actual adjusted close and the predicted adjusted close. The high value of Pearson's coefficient testifies that the learner is able to predict the adjusted close for the next day with good accuracy.

Table II: Mean Values

Parameter	Value
Actual Adjusted Close	23202.947
Market Sentiment	0.042

Table III Error Metrics

Error Metric	Value
Mean Absolute Error	574.130
Mean Absolute Error in %	2.461
Median Absolute Error	647.450

Table IV Variance Table

Parameter	Value
Actual Adjusted Close	858.891
Market Sentiment	0.045

The above table depict high variance in Market Sentiment during the period of study. During the period of study, General Elections were being conducted in India, and during this period high variance in market sentiment has been observed.

The following plot represents the plot of actual adjusted close v/s predicted adjusted close.

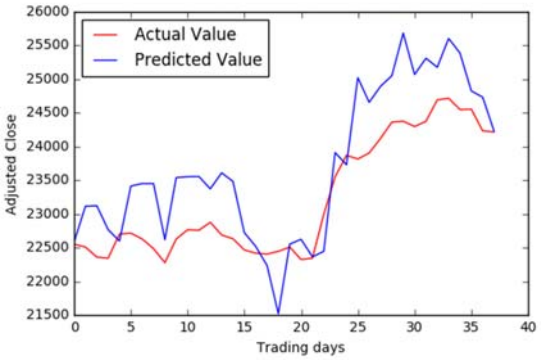

Fig. 3. Graphical representation of Actual Value vs Predicted Value

A similar study was conducted by Kumar Deva, Sophia Sharon, Jucunda Evelyn Maria. To study effects of the Lok Sabha Election on Stock Market Performance (BSE SENSEX)[3]. The study analyzed tweets for 30 days during the election period and declaration of results and employed regression techniques and inferred that the negative emotions in the market affect the market behaviour and the effect is significant. Profits and losses incurred by the company is also affected by the change in the economic decision.They prepared a portfolio optimization of various companies listed on SENSEX to let investors choose companies which are fit for their investment.

X. CONCLUSION AND FUTURE WORK

We have concluded causative relation between market sentiment as measured from the sentiment analysis of tweets scraped from twitter.com and adjusted close of SENSEX. Firstly, we can decipher from our observations that an accurate estimation market sentiment can be captured from the tweets of relevant users by application of Natural Language Processing techniques. We have scraped tweets from official twitter handle of Bombay Stock Exchange and prestigious news agencies. We have also observed that the market sentiment varies in response to exogenous factors like elections, mergers or acquisitions.

Secondly, this should be noted that stock's movement is greatly affected by market sentiment. We observed a significant correlation between market sentiment and stock price and measured value of correlation factor as 0.6572.

Volatility of Stock Price have been induced by rapid change in market sentiment.

Thirdly, sudden changes in market sentiment and adjusted close have been observed. Periods of high volatility of market sentiment have been observed in the study thus proving that market sentiment is prone to sudden change. This change in sentiment also affects the price of Stock.

This have been observed that after the election results were declared there was a rapid change in market sentiment and stock prices. The market showed bullish trend after the declaration of results as evident by rise in adjusted close after declaration of results.

Fourthly application of Kalman filter to predict market movement is beneficial as Kalman filter is designed to predict a more correct value in a noisy system. This can be deciphered from low mean absolute error of 2.461% in our observations. Thus the it can be concluded that prediction of adjusted close of stocks by capturing market sentiment through sentiment analysis of tweets and application of Kalman Filter can be done with high accuracy.

Finally, this paper does not encompass certain avenues which may be studied in future work. Firstly, the market sentiment was measured from the tweets in English only. India is a diverse nation with many languages, and with news agencies increasingly supplying financial news in local languages and increasing number of people subscribing to it, it is an interesting avenue to analyze tweets in local languages.

Secondly exogenous events like influence of international events on Indian stock market must be studied in future work.

REFERENCES

[1] M. Sarah, C. Abdur, H. Gregor, L. Ben, and M. Roger, "Twitter and the Micro-Messaging Revolution," technical report, O'Reilly Radar, 2008.

[2] Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo, Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development. IEEE Transactions on knowledge and data engineering, vol. 25, no. 4, April 2013

[3] Kumar Deva, Sophia Sharon, Jucunda Evelyn Maria. Empirical Study on Effects of the Lok Sabha Election on Stock Market Performance (BSE SENSEX). ISCA Research Journal of Management Sciences Vol. 4(2), 1-9, February (2015)

[4] Bollen J., Pepe A and Mao H., Modeling public mood and emotion: Twitter sentiment and socio-economic socio phenomena (2009).

[5] Ling-Chun Hung, The Presidential Election and the Stock Market in Taiwan, Journal of Business and Policy Research (2011)

[6] Liu, Ling-Fang, "An Empirical Study of the Presidential Elections Effect on Stock Market in Taiwan, South Korea, Singapore, Philippine, and Indonesia," *OpenTheses@Unsyiah_Lib*, accessed June 10, 2017, http://uilis.unsyiah.ac.id/openetheses/items/show/1311.

[7] J Johan Bollen, Huina Mao, "Twitter Mood as a Stock Market Predictor", IEEE Computer Society, vol. 44, no. , pp. 91-94, Oct. 2011, http://doi.ieeecomputersociety.org/10.1109/MC.2011.323

[8] Yan Xu, Guosheng Zhang, Application of Kalman Filter in the Prediction of Stock Price. International Symposium on Knowledge Acquisition and Modeling (KAM 2015)

[9] Wong Jeffrey, Market Predictions using Sentiment Analysis and State Space Models December 2010 http://cs229.stanford.edu/proj2010/WongSentimentAnalysisForMarketMovements.pdf

[10] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Transactions of ASME, Series D, Journal of Basic Engineering, Vol. 82, No. 1, 1960, pp. 35-45.

[11] Welch Greg, Bishop Gary, An introduction to the Kalman Filter. UNC-Chapel Hill, TR 95-041, November 13, 2000

[12] Kunst Robert, State Space Models and the Kalman Filter. 40461 Vektoautoregressive Methoden, January 2007.

[13] Zhang Xue, Fuehres Hauke, Gloor Peter, Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear". COINs 2010

[14] Discovery of the Kalman Filter as a Practical Tool for Aerospace and Industry, NASA Technical Memorandum 86847.