

Sensex Prediction Using Sentiment Analysis

Aarti Deshpande¹, Rucha Paturkar² and Sonali Pimpale³

^{1,2,3}PES Modern College of Engineering, Pune

Abstract: Behavioural economics tells us that emotions can affect individual behaviour and decision making immensely. Predicting market and Sensex indices is of key concern for all and has been a research topic for many financial researchers since more than a decade. Media, with growing technology, has really changed the way of trading and thus influenced the mood of investors by providing platform to share views and thoughts that in turn influence the index values. Social networks such as Twitter and Facebook are considered precious research sources of collecting user mood measurements. Large scale twitter feeds obtained from twitter API are classified by using Naive Bayes classifier. The classified data is fed to tools like Natural Language Tool kit (NLTK) which derive collective mood states that are correlated to value of Bombay Stock Exchange (BSE) index. A feed forward Artificial Neural Network is used to investigate the hypothesis that public mood states are predictive of behaviour of BSE index.

Keywords: Sensex, BSE, Sentiment analysis, Twitter, Naive Bayes, NLTK, ANN

1. INTRODUCTION

Almost everybody in today's world is investing in stock market. Everyday millions of shares are bought and sold throughout the world. The economy of the world affects everyone either directly or indirectly, thus predicting Sensex index has become a growing necessity. For any company to grow the thoughts of investors towards that company is the most important criteria. If investors are happy they will invest more and eventually the shares will grow. Even the opposite of this is true. Financial figures are influenced heavily by the psychology of the people i.e. the mood of the people. Social media and moods of the people are closely related. Nowadays social media sites such as twitter have become a place where everyone expresses their moods and feelings and in turn influence the index values. If these sentiments are correctly classified and analyzed they can be used to predict the future of the market. Sentiment analysis helps in recognizing the attitude of the investors and its correlation with actual index values. Earlier prediction systems were based on only previous index values of Sensex data. Using sentiment analysis for prediction may increase the efficiency of the system. Successful prediction can yield profits for the investors.

2. METHODOLOGY

The system consists of three major steps which lead to the prediction of Sensex value. First, Tweets are downloaded from Twitter by giving the hashtag. Second, these tweets are classified into either positive or negative tweets. The ratio of positive and negative tweets is calculated and finally the previous day's closing value is correlated with this ratio to predict future closing value of Sensex.

A. Extracting tweets

Twitter provides a way through which Tweets can be downloaded and stored into our systems. When a specific hashtag is given to the system, tweets related to this hashtag are downloaded and stored in a log file. While downloading pre processing is performed on these tweets and only the tweeted content is stored in the file.

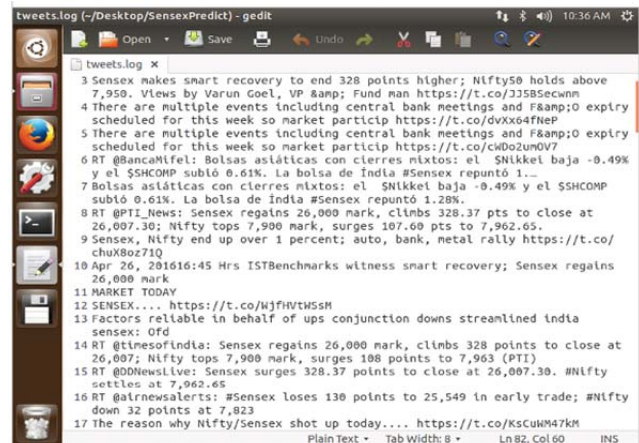


Figure 1: Tweets stored in log file

B. Classification of tweets

The downloaded tweets are then classified in positive or negative tweets. This classification is performed using the Naive Bayes algorithm and NLTK library provided by Python. Each and every tweet downloaded is either classified as positive or negative. A ratio of positive and negative tweets is found.

C. Correlation

The final step is correlation of previous Sensex value with the ratio of tweets. Previous closing values are stored in a CSV file which is downloaded from www.bseindia.com. These Sensex values are correlated with the ratio using a formula to give the final Sensex prediction for the day.

Date	Open	High	Low	Close
1-February-2016	24982.22	25002.32	24788.58	24824.83
2-February-2016	24868.21	24928.75	24460.53	24539
3-February-2016	24393.59	24409.26	24187.54	24223.32
4-February-2016	24386.45	24514.01	24224.74	24358.43
5-February-2016	24360.36	24672.9	24345.79	24616.97
8-February-2016	24637.41	24698.95	24196.84	24287.42
9-February-2016	24076.85	24111.19	23919.47	24020.98
10-February-2016	23938.32	23938.32	23636.72	23758.9
11-February-2016	23758.46	23758.46	22909.12	22951.83
12-February-2016	23060.39	23161.15	22600.39	22986.12
15-February-2016	23223.43	23622.64	23197.67	23554.12
16-February-2016	23688.61	23692.08	23164.54	23191.97
17-February-2016	23237.23	23434.91	23200.94	23381.87
18-February-2016	23536.47	23735.35	23448.21	23649.22
19-February-2016	23640.32	23774.48	23508.36	23709.15
22-February-2016	23783.47	23855.04	23674.86	23788.79
23-February-2016	23850.41	23851.51	23361.94	23410.18
24-February-2016	23332.94	23338.89	23057.45	23088.93
20-February-2016	23105.16	23142.96	22948.1	22976
21-February-2016	23141.08	23227.91	23021.94	23154.3
22-February-2016	23238.5	23343.22	22494.61	23002
23-March-2016	23153.32	2321.49	2313.18	23779.35
24-March-2016	24044.96	24280.42	24043.89	24242.98

Figure 2: Historical Sensex data in CSV file

3. ARCHITECTURAL BLOCK DIAGRAM

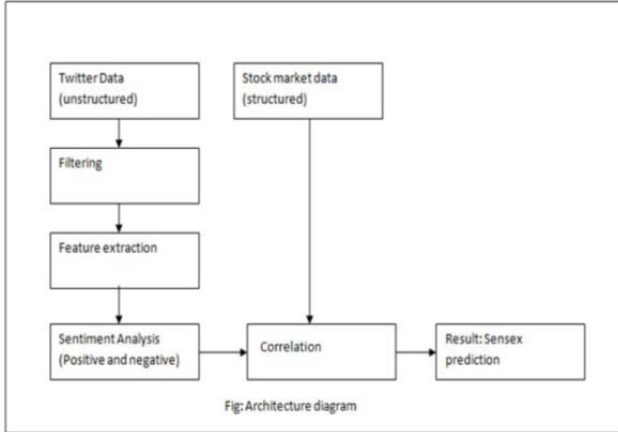


Figure 3: Architectural block diagram

4. ALGORITHM

- 1) Authentication: Access key and Access token.
- 2) Connect to Twitter API.
tweepy.API(auth)
- 3) Search for specified hashtag.
for i in api.search(search_term, count=1000):
- 4) Extract text part from tweet.
feature.extract_feature(str(i.text.encode('ascii', errors='ignore')))
- 5) Append these tweets in array.
tweets.append(i.text)
- 6) Call the classifier and pass the tweets to the classifier.
- 7) Print number of positive and negative tweets.
- 8) Assign weights to positive and negative tweets.
- 9) Fetch live sensex value. Access sensex.csv for calculating movement.
- 10) Calculate movement of Sensex.
- 11) Calculate next value
next_value=abs(np.mean(df[:, 'Close'])) + movement
print(int(upcount*10), int(downcount/2), abs(np.mean(df[:, 'Close'])), int(next_value))
- 12) Print predicted value.

5. RESULTS

The final result is displayed on single window which shows the number of positive and negative tweets, the current Sensex value and the predicted Sensex value. The user must provide hashtag as the input based on which the tweets are downloaded and prediction takes place.



Figure 4: Result window

6. CONCLUSION

This system explicitly aims to optimize the prediction of Sensex behavior. Sentiment analysis on Twitter data generates weights for positive and negative moods classified using Bayes classifier. ANN is trained to correlate the net mood of the day with the behavior of Sensex of the same day. The system focuses on BSE index values but it can also be extended to other indices such as Nifty.

REFERENCES

- [1] Sentiment Analysis to Predict Bombay Stock Exchange Using Artificial Neural Network. Sunil Kumar Khatri, Himanshu Singhal, Prashant Johri Amity Institute of Information Technology, Amity University Uttar Pradesh, Noida, India Galgotias University, Gautam Budh Nagar, Noida, India
- [2] Stock Market Forecast using Sentiment Analysis. Rajat Ahuja UG, BVCOE New Delhi, India Harshil Rastogi UG, BVCOE New Delhi, India Arpita Choudhari UG, BVCOE New Delhi, India Bindu Garg HOD, BVCOE New Delhi, India.
- [3] Will Twitter Make You A Better Investor? A Look At Sentiment, User Reputation And Their Effect On The Stock Market. Eric D. Brown Dakota State University
- [4] Application of Bayesian Network for Nikkei Stock Return Prediction Yi Zou, Masaaki Harada, Takao Mizuno, Eisuke Kita
- [5] Stock Market Trend Prediction Using a Sparse Bayesian Framework Ivana P. Markovi, Milo B. Stojanovi and Milo M. Boi.
- [6] Stock Trend Prediction Relying on Text Mining and Sentiment Analysis with Tweets Phayung Meesad and Jiajia Li Faculty of Information Technology Mongkuts University of Technology North Bangkok Bangkok, Thailand