International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 08, 2020
ISSN: 1475-7192

Stock Prediction Using Twitter Sentiment Analysis

[1]Sai Prasad Sashank Urlam, [2]Srijeeta Mandal, [3]S. Poornima

Abstract--*Prediction and analysis of stock market data are very important in today's day and age. Since the economic interactions are too complex for shallow neural networks this paper implements Long Short Term Memory (LSTM) neural networks. LSTM is chosen as it helps to vectorize the data and thus give better predictions. This paper agrees that longer horizon predictions e.g. a month are more useful than shorter horizon e.g. a day. A very important factor is the mood of the people. A person's emotions have the power to influence the stock market. Sentiment analysis on twitter is used to find a correlation between the future of the stock and the general public's mood. Our paper works on comparing the sentiment analysis and the predicted stock value and showing that the two are rather similar and that people's emotions affects the future of the stock prices and to do a comparison between prediction with and without using the results of the sentiment analysis to further prove the motion.*

Key words--*Long ShortTerm Memory, Recurrent Neural Network, Stock markets, Sentiment analysis, Twitter, LSTM*

I. INTRODUCTION

There have been many attempts in the past where multiple people have tried to predict stock trends as accurately as possible. These attempts have been made my companies, researchers, hobbyists etc. It is imperative to be able to have an insight into how the stock market would behave. There's always money involved in stock market. Regardless of how small or large the sum of money is, investors must have a good enough idea as to when to in- vest their money and when to sell. Our paper looks into the public mood and opinion and tries to show the correlation between the general mood of the public and the future movement of stock market. We use the following parameters of the stock market to make our predictions: open, close, high, low.

A. Stock Market

The opening price is the price at which a security first trades upon the opening of an ex- change on a trading day; for example, the New York Stock Exchange (NYSE) opens at precisely 9:30 a.m. Eastern time. The price of the first trade for any listed stock is its daily opening price.

A stock's closing price is the standard benchmark used by investors to track its performance over time. The closing price is the last price at which the stock traded during the regular trading day.

[1]*ComputerScience & Engg., SRMInstitute of Science and Technology Chennai, India, uspsashank_rajasekhar@srmuniv.edu.in*
[2]*ComputerScience & Engg., SRMInstitute of Science and TechnologyChennai, India, srijeetamandal_biswajyoti@srmuniv.edu.in*
[3]*ComputerScience & Engg., SRMInstitute of Science and TechnologyChennai, India, poornims@srmist.edu.in*

DOI: 10.37200/IJPR/V24I8/PR280113
Received: 21 Jan 2020 | Revised: 08 Feb 2020 | Accepted: 14 Mar 2020

The high value of a stock market basically means the highest value that a stock has attained for that particular trading day. The low value of a stock market basically means the lowest value that a stock has attained for that particular trading day.

B. Twitter Sentiment Analysis

Sentiment analysis is a method to understand public attitude toward a topic or product. With machine learning we can learn to classify thousands of posts without having to read them manually.

Sentiment analysis is an important part of our paper's solution as the output of this module is used for our learning model. While there has been a lot of research going on in classifying a piece of text as either positive or negative. In this project, we use three mood classes, namely, positive, neutral and negative. The algorithm decides if a particular tweet is bullish or bearish. Bullish means a rise in the market and bearish means a drop in the market.

The tweets used are gathered for the same day and 1 hour before and after the stock prices of that company are closed on that day. This leads us to truly quantify and analyse the effect the tweets have on the trends of the stock prices.

II. STATE OF THE ART

Our work is based on Bollen et al's strategy which received widespread media coverage recently. They also at- tempted to predict the behavior of the stock market by measuring the mood of people on Twitter. The authors considered the tweet data of all twitter users in 2008 and used the Opinion- Finder and Google Profile of Mood States (GPOMS) algorithm to classify public sentiment into 6 categories, namely, Calm, Alert, Sure, Vital, Kind and Happy. They cross validated the resulting mood and time series by comparing its ability to detect the public's response to presidential elections and Thanksgiving day in 2008. They also used causality analysis to investigate the hypothesis that public mood states, as measured by the Opinion- Finder and GPOMS mood time series, are predictive of changes in DJIA close values. The authors used Self Organizing Fuzzy Neural Networks to predict DJIA values using previous values. Their results show a remarkable accuracy of nearly 87% in predicting the up and down changes in the closing values of Dow Jones Industrial Index (DJIA). Our paper however uses the Indian index i.e. the BSE to find the company's stock price trends as the dataset and aims to prove the correlation between the stock market and the public mood based off of their tweets.

III. PROPOSED WORK

Our paper uses the Indian index i.e. the SENSEX as the dataset and aims to prove the correlation between the stock market and the public mood based off of their tweets. The first half is sentiment analysis and the next half is the prediction of the stock values. Sentiment analysis is done using live tweets received using the Twitter Search API and removing retweets, stop words, punctuations, emoticons and other unwanted parts. After sentiment analysis, granger causality is used to prove the stark correlation between the sentiment and the actual stock prices by plotting a graph. Hence, we prove that public mood affects the stock prices. The next half is stock prediction done using Long

Short Term Memory (LSTM) networks. The prediction is done on a static dataset and the LSTM network contains 4 layers.

IV. IMPLEMENTATION

A. Stock Prediction

a) **LSTM:** Long short term memory is a type of recurrent neural network that can learn that can learn order dependence in sequence prediction problems. This is a behavior required in complex problem domains like machine translation, speech recognition, and more. Our paper utilizes Long Short Term Memory(LSTM) network to take in data stamps. The data is scraped from SENSEX.

b) Import data set to neural network: The data set which is acquired from a link is imported to the neural network. It then undergoes feature scaling followed by re- shaping.

c) Building RNN: Building RNN involves creating 4 layers namely, initializing the RNN, adding 3 drop out regularization layers and finally a 5th output layer.

d) Making prediction and visualizing result: The final module produces the output of the predicted stock. These predictions are made by taking in 60 time stamps. The final graph comparing real stock price and final stock price is produced.

B. Sentiment Analysis

a) Data Collection: Tweets on Amazon are extracted from the Twitter Search API. The tweets will have been collected from public users ’timelines that contain the word “amazon” either as text or a hash tag using Twitter API. Not only the opinion of public about the company’s stock but also the opinions about products and services offered by the company. The reason amazon is used as a keyword is to make sure that the tweets are extracted in such a way that they represent the exact emotions of public about Amazon over a period of time. Stock opening and closing prices of Amazon are obtained from Yahoo! Finance.

b) Data pre-processing: Tweets consists of many acronyms, emoti- cons and unnecessary data like pictures and URL’s. So, tweets are pre-processed to rep- resent correct emotions of public. For pre- processing of tweets, we employed three stages of filtering: Tokenization, stop words removal and regex matching for removing special characters.

1) Tokenization: Tweets are split into individual words based on the space and irrelevant symbols like emoticons are removed. We form a list of individual words removed. Form a list of individual words for each tweet. Multiple words like “heyyyy” is converted to “hey”

2) Stop word Removal: Words that do not express any emotion are called Stop words. After splitting a tweet, words like a, is, the, with etc. are removed from the list of words.

3) Regex Matching for special character Removal: Regex matching in Python is performed to match URLs and are replaced by the term URL.

DOI: 10.37200/IJPR/V24I8/PR280113
Received: 21 Jan 2020 | Revised: 08 Feb 2020 | Accepted: 14 Mar 2020

4) Removal of hashtags and digits: The hashtag symbol and numeric digits are removed as they are unnecessary character and hinder the sentiment analysis.

***c*) Sentiment Analysis**

Sentiment analysis task is very much field specific. Tweets are classified as positive, negative and neutral based on the sentiment present. Out of the total tweets are examined by humans and annotated as 1 for Positive, 0 for Neutral and 2 for Negative emotions. For classification of nonhuman annotated tweets, a machine learning model is trained whose features are extracted from the human annotated tweets. Before sentiment analysis is done, we check the correlation between the two datasets; the stock prices and sentiment polarities to show that there is a stark correlation between the two. The correlation test is a statistical method to find correlations. Then, each tweet is assigned a polarity of positivity/negativity ranging from -1 (very negative) to 1 (very positive) as a float value using Support Vector Machines. For each day, the polarity of the tweets for the given day are averaged and used as a feature in the LSTM stock prediction model. The other features include, the same averaged value of the polarity of the tweets one hour before and after the stock market closes every day. On prediction with the new features, the result is compared with that of prediction with only the stock prices as features and it is found as to which method is better thus aiming to prove if tweets about a company do affect its stock trends and if yes, by how much.

V. RESULTS DISCUSSION

Stock market trends are highly unpredictable and affected by numerous factors. With the rapid onset of social media, namely Twitter, this paper finds the relation between Tweets i.e. mood of the people with the stock market values. A correlation test is used to determine whether one time series is useful in forecasting another. Using the results from the sentiment analysis as a feature in the LSTM, we have found a stark correlation between the moods discovered using sentiment analysis and the stock values as compared to using LSTM for stock prediction without the additional features. This analysis is graphed for future reference.

Then we use LSTM to predict the stock market close value. The values are jotted down on a graph. Next step is repeating the LSTM prediction after including the sentiment analysis polarity values as a feature and plotting it and we can see or deduce from the two graphs after a comparative analysis as to which is better for prediction and if tweets about a company do affect its stock trends and if the knowledge about the same can be of help to the company for making smart business decisions, finding market demographics, finding customer greivances and potential areas of improvement and more.

VI. CONCLUSION

We have investigated the causative relation between public mood as measured from a large scale collection of tweets from twitter.com and the Dataset taken from BSE. Our results show that public mood can indeed be captured from the large-scale Twitter feeds by means of simple natural language processing techniques, as indicated by the responses towards a variety events that occur.

DOI: 10.37200/IJPR/V24I8/PR280113
Received: 21 Jan 2020 | Revised: 08 Feb 2020 | Accepted: 14 Mar 2020

Secondly, among the observed dimensions of moods, only positive moods are Granger causative of the SENSEX by 3-4 days. We've witnessed how positive and negative moods can affect the stock market as this is pitted against the prediction made by an LSTM algorithm. It comes to show that the public opinion and public mood can influence the stock market to large degrees.

REFERENCES

1. Amin Hedayati Moghaddama, Moein Hedayati Moghaddamb, Morteza Esfandyari, 2018. 'Stock market index prediction using artificial neural network. 'Journal of Economics, Finance and Administrative Science, 21 (2016) ,89–93.
2. Hiransha Ma , Gopalakrishnan E.A , Vijay Krishna Menon, Soman K.P. 'NSE Stock Market Prediction Using Deep-Learning Models. 'Procedia Computer Science 132 (2018) 1351–1362, 2018.
3. Xiongwen Pang · Yanqiang Zhou · Pan Wang · Weiwei Lin · Victor Chang. 'An innovative neural network approach for stock market prediction. 'Springer Nature, 2018.
4. Johan Bollen, Huina Mao, ,Xiao-Jun Zeng. Twitter mood predicts the stock market
5. J. Bollen and H. Mao. Twitter mood as a stock market predictor. IEEE Computer, 44(10):91–94.
6. C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011.
7. G. P. Gang Leng and T. M. Mc Ginnity. An on-line algorithm for creating self-organizing fuzzy neural networks. Neural Networks, 17(10):1477–1493.
8. A. Lapedes and R. Farber. Nonlinear signal processing using neural network: Prediction and system modeling. In Los Alamos National Lab Technical Report.
9. Hasan, Md. Nazmul , N. M. Mahmudul Alam Bhuiya, Mohammed Kamrul Hossain, and . "In silico molecular docking, PASS prediction and ADME/T analysis for finding novel COX-2 inhibitor from Heliotropium indicum." Journal of Complementary Medicine Research 10 (2019), 142-154. doi:10.5455/jcmr.20190525051057
10. Nag, T., Ghosh, A. Cardiovascular disease risk factors in Asian Indian population: A systematic review(2013) Journal of Cardiovascular Disease Research, 4 (4), pp. 222-228. DOI: 10.1016/j.jcdr.2014.01.004

DOI: 10.37200/IJPR/V24I8/PR280113
Received: 21 Jan 2020 | Revised: 08 Feb 2020 | Accepted: 14 Mar 2020