International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 3117–3123

Stock Prediction: NLP and Deep Learning Approach

Mr. Yogesh Bodkhe[1], Prof. Rushali A. Deshmukh[2]

Dept. of Computer Engineering JSPM's Rajarshi Shahu College of Engineering Tathawade. Pune – 411033., India
Dept. of Computer Engineering JSPM's Rajarshi Shahu College of Engineering Tathawade. Pune – 411033., India
yogeshbodkhe@gmail.com, radesh19@gmail.com

Abstract

People tend to analyze existing strategies and so planned new strategies for inventory prediction. We have used Sentiment evaluation and Technical evaluation through NLP and Deep mastering approach. To take advantage of sentiment analysis on enterprise associated inventory, we have proposed a machine that will use the sentiment analysis on twits associated with special sectors (e.g. IT sector, Banking sector, Pharmaceutical sector, Automobile sector, Infrastructure sector.) which might be extracted from twits. These twits are extracted from twitter for calculating polarity. The rating of sentiment analysis is calculated here by using algorithm. According to the sector, we have taken 20 groups—the top four performer businesses of every sector. Using the polarity score, we will finalize pinnacle ten groups with great sentiment ratings. We will download the CSV facts of historical share charge of the top ten organizations that we have selected. Then downloaded CSV records are used to build a CNN version to predict in addition stock movement of these pinnacle ten companies.

Keywords: *stock prediction, Natural Language processing, Deep Learning, Price forecasting, sentiment analysis, CNN, SVM, ANN, NSE, BSE, financial market, Data mining*

1. Introduction

Financial analysts are investing in stocks usually, but they are unaware of inventory markets place conduct. People who are an analyst are going through the uncertainties of stocks trading because they cannot understand which stocks they should buy or which stocks they should recommend to other people for maximum earnings on stocks. Nowadays, all required information which is related to inventories and their markets is available easily. Studying all these records, in my opinion, is not simple. So, for that there should have a method that needs to get automated. Here we can say that Data mining techniques come into the picture. Most of the investors are using different learning curves to forecast or predict the inventory market's conduct. So, because of this, we can say that traders or analysts can implement the method which can conduct the inventory studies for people who might be enthusiastic about working accordingly. Here we are giving the historical data input to our gadget. So proper data could be used to find out the stock fee movements. So here we can say that the forecasted values will show us a picture of shares price movement, and according to that, any analyst or trader can book a profit on provided analysis of that stock.

ISSN: 2233-7857 IJFGCN

2. Key take away of Deep Learning

- Deep learning is an AI work that mirrors the activities of the human mind in preparing information for use in basic leadership.
- Deep learning can learn from the data that is from both unstructured and unlabeled source.
- Deep learning is machine learning's subset that can be used to help to detect fraud or money laundering.

3. Literature Survey

Rakhi Batra et al. [1] had used a technique of sentiment analysis for stock tweets, which was related to a different type of Apple product; for this, she had extracted stocks related twits from different social networking sources for eight years of time duration. Apart from shares data, these people had decided to use stocks related data from the Yahoo Finance source for that time duration. They had used the SVM technique to find out the polarity of those tweets. So because of this, they were able to differentiate the tweets as Positive or Negative. After that, they had used polarity results, which was based on sentiment analysis and stock data, to create an SVM model and to forecast a subsequent day's share price movements.

Yaojun Wang et al. [2] used social media sites to gather data for their research. In this research paper, their focus was on the share price movements in the market. For the forecasting of the stocks along with mining techniques, they had used other relevant information. Their result showed that they had calculated the stocks' polarity for better prediction of the stock's price.

Ashish Sharma et al. [3] had gone through the stock market data in regressive manner. So that they got a good amount of stock data for their research from the share market. The motto of their research study was to help the stockbrokers and investors for investing money in the stock market.

Ze Zhang et al. [4] had used one system to find out the opening value of stocks in the financial market. However, their developed system was self-learner so that they were able to predict the opening value of the market. They had given the stocks data to their developed system to find out the forecasted value. Last, they developed another network system and compared both the system with each other to predict the starting day value of the stock.

Dev Shah et al. [5] had studied the news and based on that news; they had done the sentiment analysis. In this paper, they had found out the polarity for the pharma sector. Mainly their focus was on the stock's prices movement, which was based on the polarity dictionary.

Du Peng [6] had mainly studied the market volatility and worked on the people's sentiments to find the relation between share price and the traders' sentiments. For this, he had studied different indexes for the news.

Muhammad Firdaus et al. [7] had used the ANN algorithm for the share market prediction. Based on their studies of ANN, they had claimed that they had achieved a high percentage of accuracy while predicting the values in the stock market. For this, they had studied different methods, and after studying that they had found out the accurate and proper results.

Research work of Nonita Sharma et al. [8] had given the focus on to make predictions of the share prices by studying the historical data. For that, they had taken decade data from the two well know indexes like NSE and BSE. For this, they had developed a model by using the SVM algorithm. For the predicted value, they had considered the closing value of the share. Also, they had forecasted the share values around more than 35 days.

ISSN: 2233-7857 IJFGCN

4. Proposed Methodology

Here we tend to area unit planned system that may work with an improved level of recommendation. The system will be developed with NLP of computer science and the Convolutional Neural Network (CNN) of Deep Learning. Natural Language Processing technology can help the system search out companies with excellent news in terms of live performance in the market. That may facilitate to create a selection of best entertainers in the market. NLP will classify news in positive and negative sets and can provide a performance graph of the selected organization. We will get a sense of the best-performing company. Natural Language Processing provides to system NLP (Natural Language Processing) that will work on our news for detection merchandise and unhealthy of its impact.

4.1. Architecture

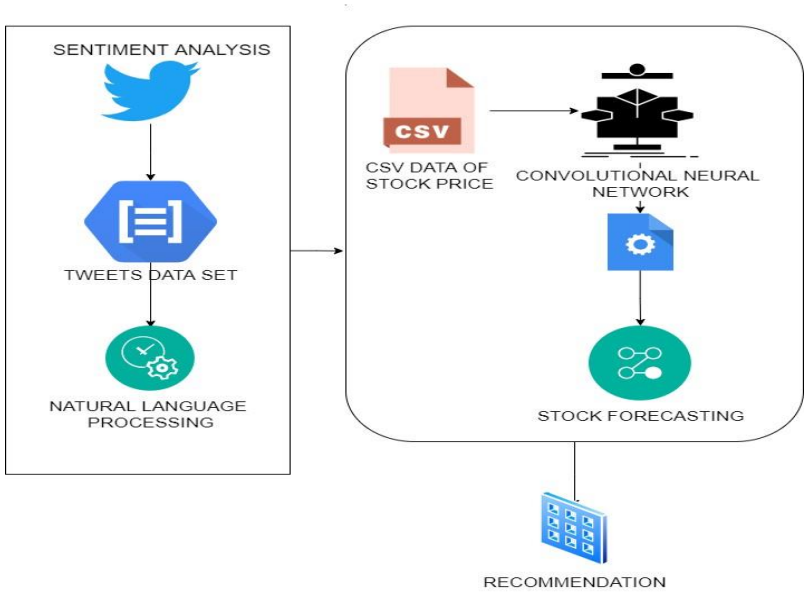

Figure 1. Proposed System architecture

Features of Proposed System:

The proposed system has the advantage of multiple platforms for input files for model development. The system has been developed with over one algorithmic rule; thence Prediction guarantees are magnified. Live updates area unit concerned in prediction thence it is often used for live recommendation.

4.2. Algorithms

First 1D CNN layer:

The primary layer defines a filter (or conjointly known as feature detector) of height ten (also known as kernel size). Solely shaping one filter would enable the neural network to be told one feature within the initial layer. This won't be comfortable so that we will outline N variety of filters. This permits North American country to coach N completely different options on the primary layer of the network. The output of the first neural network layer is in a very somatic cell-matrix type.

ISSN: 2233-7857 IJFGCN

Second 1D CNN layer:

The outcome from the first CNN will be sustained into the second CNN layer. We will again characterize 100 unique channels to be prepared on this level. Following a similar rationale as the primary layer, the yield grid will be of size 62 x 100.

Maxpooling layer:

A pooling layer is frequently utilized after a CNN layer so as to lessen the unpredictability of the yield and forestall overfitting of the information. In our model, we picked a size of three. This implies the size of the yield lattice of this layer is just 33% of the info network. Third and fourth 1D CNN layer: Another arrangement of 1D CNN layers follows to learn more significant level highlights. The yield lattice after those two layers is a 2 x 160 grid.

Average pooling layer:

One all the more pooling layer to additionally abstain from overfitting. This time not the most extreme worth is taken but rather the normal estimation of two loads inside the neural system. The yield lattice has a size of 1 x 160 neurons. Per include identifier, there is just one weight staying in the neural system on this layer.

Dropout layer:

The dropout layer will haphazardly dole out 0 loads to the neurons in the system. Since we picked a pace of 0.5, half of the neurons will get a zero weight. With this activity, the system turns out to be less touchy to respond to littler varieties in the information. Along these lines, it should additionally expand our precision on concealed information. The yield of this layer is as yet a 1 x 160 network of neurons. Fully connected layer with SoftMax activation: The last layer will decrease the vector of stature 160 to a vector of six since we need to anticipate ("Jogging", "Sitting", "Strolling", "Standing", "Upstairs", "Ground floor") because we have six classes. Another lattice augmentation finishes this decrease. Softmax is utilized as the initiation work. It powers every one of the six yields of the neural system to summarize to one. The yield worth will hence speak to the likelihood for every one of the six classes.

Sentiment Intensity analyzer

Business: In advancing field firms use it to build up their strategies, to know clients' sentiments towards product or entire, people answer their battles or item dispatches and why customers don't get some product [1] [5].

VADER Sentiment Analysis:

VADER content opinion investigation utilizes a human-driven methodology, consolidating substance examination, and exact approval by exploiting human raters and hence the information on the gathering.

Five Easy Heuristics

1. Lexical alternatives aren't the sole things inside the sentence that affect the opinion. Their territory unit elective talk segments, similar to accentuation, capitalization, and modifiers that conjointly give feeling. VADER's conclusion examination thinks about these by thinking about five simple heuristics. The aftereffect of those heuristics zone unit, once more, measured exploitation human raters. For Ex.

[1] I Like that. [2] I Like that!!!

2. VADER assumption examination mulls over this by enhancing the sentence's slant score relative to the quantity of shout focuses and question marks finishing the sentence. VADER first figures the estimation score of the sentence. If the score is certain, VADER includes a specific experimentally acquired sum for every accentuation mark (0.292) and accentuation (0.18). On the off chance that the score is negative, VADER subtracts.

ISSN: 2233-7857 IJFGCN

1) The second heuristic is capitalization.
[1] amazing work.
[2] AMAZING work.

Thus, VADER takes this under consideration by incrementing or decrementing the word's sentiment score by zero.733, betting on whether or not the word is positive or negative, severally.

3. The third heuristic is that the use of degree modifiers. View example “effing cute” and “sort of cute”. The modifier's result within the 1st sentence is to extend the intensity of cute, whereas within the second sentence, it's to decrease the intensity. VADER maintains a booster wordbook that contains a collection of boosters and dampeners. The result of the degree modifier conjointly depends on its distance to the word it’s modifying. Farther words have a comparatively smaller exacerbating result on the bottom word. One modifier adjacent to the base word adds or subtracts zero.293 to the slant score of the sentence, wagering whether the base word is sure. A second modifier from the base word includes/subtract ninety-fifth of zero.293, and a third includes/subtracts ninetieth.

4. The fourth heuristic is that the shift in polarity thanks to “but”. In many cases, "but" interfaces 2 provisions with contrastive conclusions. The prevailing opinion, in any case, is that the last one. for example, " I like it, but I don’t wish to use that anymore " the essential provision "I like it" is sure, the other VADER actualizes a "but" checker. Fundamentally, all conclusion bearing words before the "but" have their valence decreased to five hundredth of their qualities, though those when the "but" increment to a hundred and fiftieth of their qualities.

5. The fifth heuristic is looking at the tri-gram before a feeling loaded lexical component to get extremity nullification. Here, a tri-gram alludes to an assortment of 3 lexical choices. VADER keeps up a stock of useless words. Refutation is caught by increasing the opinion score of the assessment loaded lexical component by partner degree experimentally decided cost - 0.74.

A. *1D CNN Algorithm:*
The Algorithm of a 1D-CNN is formed through the following important steps:

Input: Dataframe (train_data , test_data)
Process: Build 1D CNN Model def Model ():

1. define model
2. add filter (kernel) size to each layer model.add(layers) model.add(kernel size)
3. add pooling layer add dropout value
4. model activation layer
5. fit model with training and testing data Model summary

Output: Prediction = model.predict(test data)
Accuracy = (accuracy_score(Y_test,Y_pred)*100)

5. Results and Discussions

Figure 2 displays the base window of the proposed system, where it contains a control panel with buttons that are bind with events to load tweets from twitter according to the selected company. Scraped tweets are displayed in the canvas window at the right side of the base window. Scraped tweets are being displayed in text form in the canvas frame. So here, we can say that users will be able to display all live tweets of a selected company.

ISSN: 2233-7857 IJFGCN

So for our project, we have taken ICICI bank's official twitter handler's tweets as an example.

Figure 3 displays a score of polarity on tweets that were scraped from Twitter. VADER library that supported our proposed system in terms of getting a sentiment of tweets has given the polarity report with Positive, Compound, Neutral, and Negative in the form of a percentage of each sentiment. So, according to the polarity score we will get an idea about the positive and negative tweets.

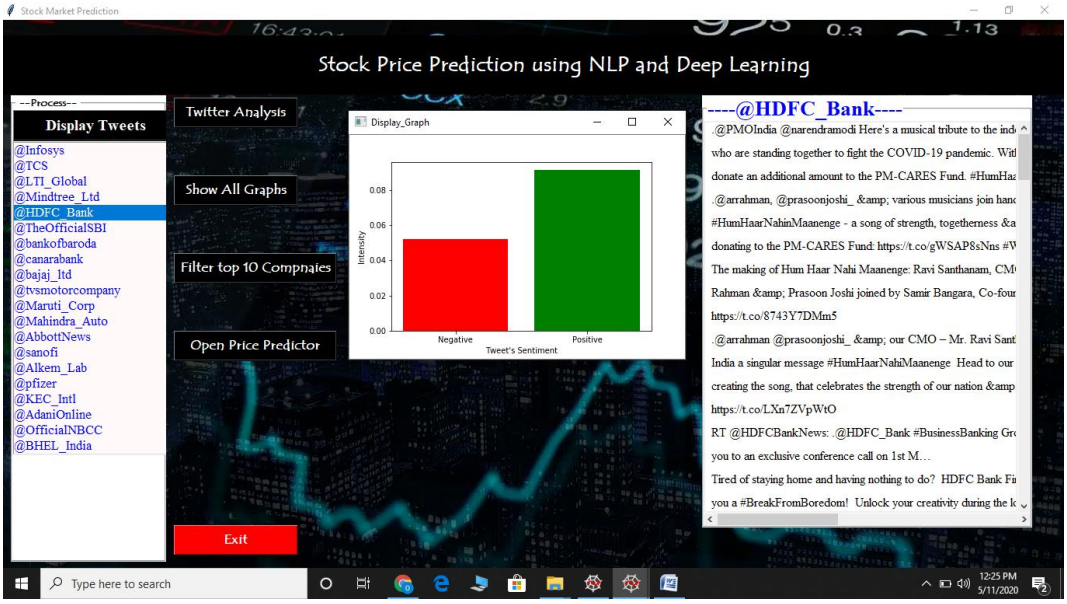

Figure 2. Display Tweet window

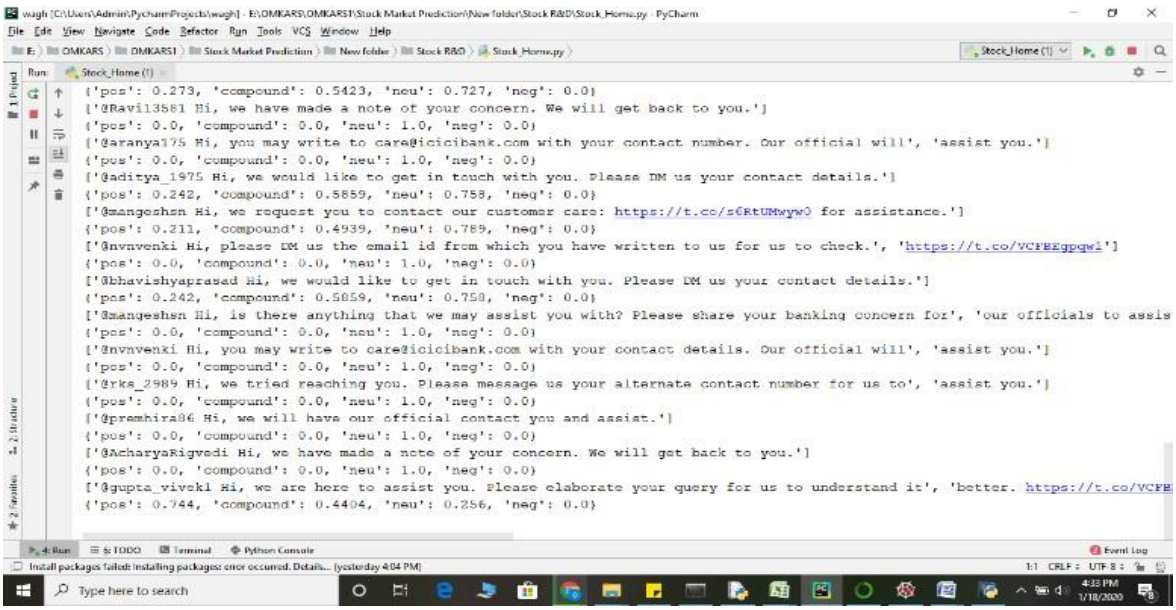

Figure 3. Polarity Result

ISSN: 2233-7857 IJFGCN

6. Conclusion

In previously happened researches on stock prediction were on the trial and error basis as well as it was based on machine learning. This prediction is related to numbers and support vector machine. However, here we have considered the literacy of the people and their study about the market while checking the prices of the stocks and their respective price movement in the financial market. Effectively we have used stock data of companies of different sectors to do the sentiment analysis using NLP technology. So according to that, we have found out the polarity of companies and those companies which are having high polarity are taken for the next level. In that stage, we have developed a CNN based model for the prediction and forecast values. So, at last, we have verified our predicted values of that stocks and their actual values. In the above diagrams of results, we can say that we have shown a polarity, and based on this polarity, we measured the top ten well-performing companies in given sectors. In the future, we will attempt to execute more calculations and all the more new methods planning to give a live proposal to securities exchange financial specialists. Additionally, our emphasis will be on the entire securities exchange for forecasting.

References

[1] Batra, Rakhi, and Sher Muhammad Daudpota. "Integrating StockTwits with sentiment analysis for better prediction of stock price movement." In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), IEEE, **(2018)**, pp. 1-5.

[2] Wang, Yaojun, and Yaoqing Wang. "Using social media mining technology to assist in price prediction of stock market." In 2016 IEEE International Conference on Big Data Analysis (ICBDA), IEEE, **(2016)**, pp. 1-4.

[3] Sharma, Ashish, Dinesh Bhuriya, and Upendra Singh. "Survey of stock market prediction using machine learning approach." In 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), vol. 2, IEEE, **(2017)**, pp. 506-509.

[4] Zhang, Ze, Yongjun Shen, Guidong Zhang, Yongqiang Song, and Yan Zhu. "Short-term prediction for opening price of stock market based on self-adapting variant PSO-Elman neural network." In 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), IEEE, **(2017)**, pp. 225-228.

[5] Shah, Dev, Haruna Isah, and Farhana Zulkernine. "Predicting the Effects of News Sentiments on the Stock Market." In 2018 IEEE International Conference on Big Data (Big Data), IEEE, **(2018)**, pp. 4705-4708.

[6] Peng, Du. "Analysis of Investor Sentiment and Stock Market Volatility Trend Based on Big Data Strategy." In 2019 International Conference on Robots & Intelligent System (ICRIS), IEEE, **(2019)**, pp. 269-272.

[7] Firdaus, Muhammad, Swelandiah Endah Pratiwi, Dionysia Kowanda, and Anacostia Kowanda. "Literature review on Artificial Neural Networks Techniques Application for Stock Market Prediction and as Decision Support Tools." In 2018 Third International Conference on Informatics and Computing (ICIC), IEEE, **(2018)**, pp. 1-4..

[8] Sharma, Nonita, and Akanksha Juneja. "Combining of random forest estimates using LSboost for stock market index prediction." In 2017 2nd International Conference for Convergence in Technology (I2CT), IEEE, **(2017)**, pp. 1199-1202..

ISSN: 2233-7857 IJFGCN