International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 10 | Oct -2020 ISSN: 2582-3930

Stock Price Prediction Using NLP and Deep Learning

Sneha M. Phutane, Dept. of Computer Engineering, MES College of Engineering, Pune, India.

Komal K. Gawali, Dept. of Computer Engineering, MES College of Engineering, Pune, India.

Sanjana R. Pawar, Dept. of Computer Engineering, MES College of Engineering, Pune, India.

Sonali S. Gadegaonkar, Dept. of Computer Engineering, MES College of Engineering, Pune, India.

Revati M. Wahul, Dept. of Computer Engineering, MES College of Engineering, Pune, India.

Abstract

Earlier, the prediction of stock values was just as good as a random guess. Even experienced people were not sure enough about their judgement and investing in stock market was a gamble. But this has changed due to the advancement in technology as loads of recorded data can be analysed and used for making decisions that are well directed based on experience, trends, and past knowledge. Today, there is a lot of information that can be shared and accessed instantly via the Internet, therefore huge amount of stock market data is gathered from different companies and locations daily. Analysing all this information individually or manually is tremendously difficult. This is where machine learning techniques help. Understanding that analysis of numerical time series gives close results, intelligent investors can use machine learning techniques in predicting the stock market behaviour. This will allow financial analysts to foresee the behaviour of the stock that they are interested in and thus act accordingly. The analysis of existing methods and then proposal of new methods for stock market prediction along with constant improvement in the efficiency of machine learning model is very important. We are proposing that our AI model will perform better than the existing systems and will also provide easy interactive user interface to the clients. Innovation of new methods and technologies opens doors to better implementations and approaches. Three different approaches that can be enhanced are fundamental analysis, technical analysis, and better machine learning techniques.

Keywords: Share Market; Data Analysis; Stock Prediction; Machine Learning Technology; Neural networks; CNN algorithm; Pandas library; PyTorch, matplotlib

1. Introduction

Investment in stock is one of the most rated businesses for making money for investors and traders. Company's share price is most

important point for investors which always fluctuates between high and low on a daily basis. Eyes always need to be on the online price of share market and instant decision making is necessary to prevent loss of money and eventually to gain money. For this an individual has to study different company's financial history and future agenda. But there are limitations to study as one cannot be sure whether the studies and analyses are correct. Company's market history, tendency of maintaining business in any period or slack, policies and announcements are the key points of stock rate. It is a difficult field of work and it needs a lot of experience to be a successful investor. In this paper, we will present a structure in which we coordinate client expectations into the present AI calculation, utilizing open verifiable information to improve our outcomes. The purpose of system is to build an AI model so that it will do analysis by itself and express predictions. Stock price prediction is a classic and important problem. With a fruitful model for stock forecast, we can pick up knowledge about market conduct, identifying patterns that may be hidden or gone unnoticed. With the inexorably computational intensity of today's computers, AI will be an efficient technique to take care of this issue. Nonetheless, people in general stock dataset are sometimes restricted for AI calculations to display results quickly and request for more computationally intensive specific features that may increase the budget of work on a day to day basis. The spurred thought is that, in the event that we know all data about today's stock exchanges (of all specific merchants), the value is predicted close enough. With the development of the Internet, informal organizations, and online social associations, getting day by day client forecasts is practical and verifiable so that the information gained is correct and establishes confidence in the system. In this way, our intension is to structure an open help consolidating critical information and clients forecasts to make a more grounded model that will benefit everybody. Presenting analysed data to the client is as much important as analysing itself. The client should feel comfortable in understanding the information present on the screen. Comparison of stocks which are represented using time series forecasting graphs should be easily distinguishable for different companies.

2. Literature Survey

The main studies are the specific mechanism of investor sentiment affecting stock market volatility. With the help of Pollet and Wilson's theory of volatility decomposition, it performs a comparative analysis based on big data sources. It collects the data of web news emotion index, web search volume, social network emotion index, social network heat index, and establishes corresponding analysis index. After correlation analysis and Granger Causality tests, it extracts the indicators which have significant correlation with the financial market and brings them into forecasting analysis. The model constructs market volatility index and analyses the correlation between investor sentiment and stock price changes.

The effects of news sentiments on the stock market is also retrieved, extracted, and analysed. Stock price is one of the intricate non-linear dynamic system. News which is relevant to any share market listed company is effective on further movement of stock. Recent studies have shown that the vast

amount of online information in the public domain such as Wikipedia usage pattern, news stories from the mainstream media, and social media discussions can have an observable effect on investors' opinions towards financial markets.

Prediction plays a very important role in stock market business which is very complicated and a challenging process. regression analysis beats traditional methods, like fundamental and technical analysis, which may not be reliable. Efficient regression approach predicts the stock market price from stock market data base. In future, the results of multiple regression approach could be improved using a greater number of variables and the interlinkages and dependencies of these variables which affect the weightage of parameters considered while forming prediction results. Paying close attention on how certain variable dependencies have affected the results in the past is crucial for analysing the future.

Verification of models by some stock prices and their comparison with Back Propagation (BP) neural network and Elman neural network is carried out so as to draw the result that shows the difference. Typically, Elman neural network is a local recurrent neural network, having one context layer that memorizes the past states, which is quite fit for resolving time series issues. Given this, it takes Elman network to predict the opening price of stock market. Considering that Elman network is limited, adoption of self-adapting variant Particle Swarm Optimization (PSO) algorithm to optimize the weights and thresholds of Convolutional Neural Networks (CNN) or Recurrent Neural Network (RNN) plays an important role in better outputs and hence better prediction results.

3. Methodology

Stock market normally opens at 9:15 am and closes at 3:30 pm. Starting price, peak price, and ending price of all company's stock is taken as input and stored in CSV format. Yahoo Finance is the site that is being used to fetch these values and these fetched values are made to go through algorithms to get the desired results.

- **Convolutional Neural Network (CNN):**

Convolutional neural network (CNN) is part of the family of neural networks (NN) which is a variation of a multilayer perceptron (MLP). CNN consists of an input layer, several hidden layers and an output layer like any other NNs. Input layer is a representation of identity function, $f(x) = x$. Output layer which makes decisions, passes previously calculated weights through a linear function. Hidden layers are either convolutional, pooling, dropout or fully connected. In addition, all layers have activation functions at the end which gives additional functionality e.g. normalization. sigmoid, tanh and RELU are examples of these activation functions.

Weights of convolution layers can be seen as 2D-filters and they apply convolution operation with these filters. Convolution operation is a process which sums the point-wise multiplications of given two functions while sliding the operation window. Pooling layer generalizes the elements in window frame while sliding this window.

Dropout selects several neurons, that feed the input of next layers and reduces overfitting., In order to optimize weights of CNN, an adaptive learning rate method (ADADELTA) optimizer is used. Apart from the fact that CNNs give noteworthy performance, they require much more data compared to other types of models. Merging makes it possible to not stick to one set of stocks. The model trained with this dataset is be universal for all kinds of stocks rather than a particular stock.

- **Feedforward Neural Network:**

A feedforward neural network is inspired by classification algorithm. In this network, neuron like processing units are organized in layers. Every unit in a layer is connected with all the units in the previous layer. These units are called nodes. Each connection might have the same amount of strength and weight or different amounts. Encoding the network knowledge is performed by these weights.

Data enters at the inputs, passes through the network layer by layer, and eventually arrives at the output. There is no feedback between layers when data acts like a classifier. Hence the name - Feedforward Neural Networks. The network possesses no cycles or loops.

- **Twitter Analysis:**

Social media is a significant platform for marketing and advertising. A good amount of interest is developed and customer base is created and attracted. Comments are gathered to understand the popularity of the new product/scheme.

Twitter is the social platform that companies and people use to promote their products and schemes. Launching of new products is promoted on twitter with all the users commenting and giving their opinions. Twitter page is helpful in predicting the future rise and fall of shares.

Sentiment analysis is done on these comments on the companies. Sentiment analysis can help us find the polarity of comments, helping us classify whether they are positive or negative.

The remarks and comments on Twitter page are analyzed in the model to predict the possibility of the future endeavors of the company becoming a success or a failure. The investors can use this analysis to invest in the most trending and positive prospect shares for better returns in the future. The sentiment analysis of negative comments gives an idea of the business falling short of expectations and hence prevents the loss of investors.

The combination of both previous deep learning analysis and future prospects helps to make decisions with confidence and better success rate.

4. Proposed System Design

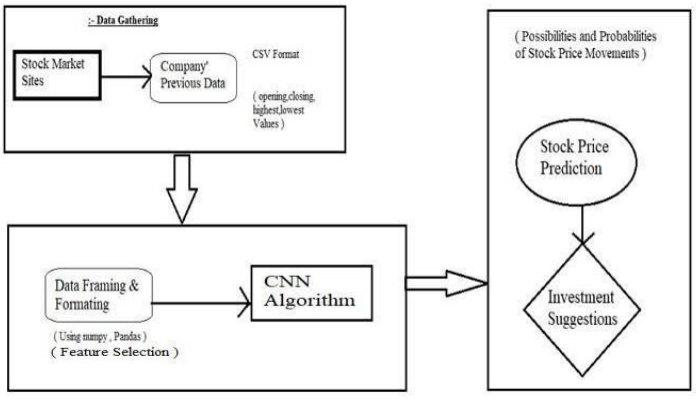

Fig 1: System Architecture

The model will be trained on huge data of different companies and time stamps in Comma Separated Values (CSV) format. CNN will be used for time series forecasting on the data analyzed from the CSV files and model output using deep learning methods. Time Series Forecasting is the base of the system as the model uses data of stock values on number of days with date format. The model is well trained on previous stock market data and can make predictions and handle queries of the client.

The surveyed results according to the present numbers are displayed to the client. The user will be provided with predicted value of stock of the company they choose and on the basis of that, recommendations are given.

Here, we propose that our system will work with improved level of recommendation. The system is being developed in Python and will be designed in modules where errors can be detected and fixed easily. The administrator will have access to the database of each individual user and will be responsible to maintain the security and integrity of the data. All details of the user are confidential and the user needs to go through authentication process to prove his validity. Data will be modified by the authorized person in authorized manner only.

5. Expected Result

The system should make considerably better predictions mainly based on the previous day values and be able to capture ongoing trends in the market through analysis of previous datasets given as input to the model and prospects of companies. The model aims to give a result with more precision, recommend trending companies to the users, and provide necessary and appropriate graphs where necessary to give a precise idea of rise and fall of the companies. The system displays top 2 companies from each department, and top 10 companies overall. This will help the user track the consistency of a company based on the rankings it has achieved.

A colourful and an interactive display, real time updates of stock prices, and predicted and actual values of various companies along with comparisons between them should be displayed with at most accuracy. The system should strive to make the interface and understanding of stock market easier for the users with these features.

6. Implementation

The main screen of the system provides with a simple interface from where users can select any one of the options – Twitter Analysis, Show All Graphs, Filter Top 10 Companies, Open Price Predictor. Tweets of different companies can also be accessed, with the company names on the left side and the corresponding tweets based on the company selected on the right side.

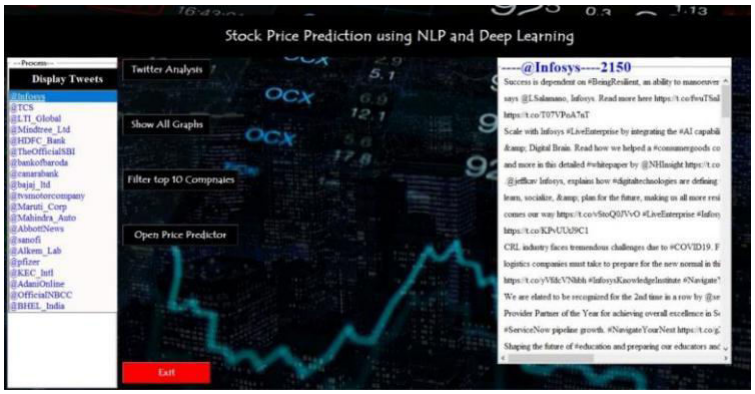

Fig. Main Screen

- **Twitter Analysis**

Sentiment Analysis is used to know whether the comments fall under positive or negative with regards to stock market. The green bar shows the amount of positive comments found regarding the stock value of the company whereas the red bar shows the amount of negative comments found. The graph is plotted with "Intensity" on Y-axis and "Twitter Sentiment" on X-axis. "Twitter Sentiment" gives an idea of negative and positive comments, whereas "Intensity" gives an idea of the extent of positivity and negativity in the overall comments.

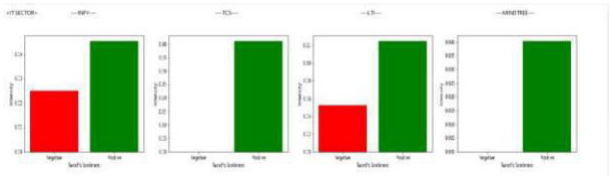

Fig. Twitter Analysis

- **Displaying Company Graphs**

A graph showing "Price" on Y-axis and "Date" on X-axis will be displayed for each company. Each company's rise and fall will be shown on this graph. This will give the users a better idea of how well the company has been doing, and give some information to predict by acknowledging the consistency of that company.

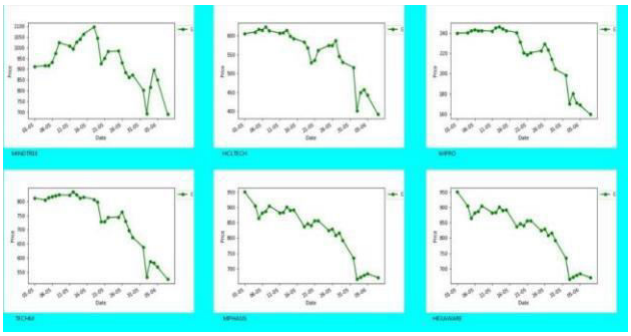

Fig. Graphs of Companies

- **Displaying Top 2 Companies from Each Sector**

Top 2 companies from each sector will be displayed. This will give the users a better idea on which companies did the best in stock market in the last 24 hours based on their sector.

Fig. Top 2 Companies from Each Sector

- **Actual and Predicted Values**

CNN algorithm will provide with the probability for companies to gain a particular amount of stock and based on this probability, predictions will be carried out. These predications will give the users a better idea on which stock to invest in. Predictions are rarely 100% true, but the best has been tried to achieve.

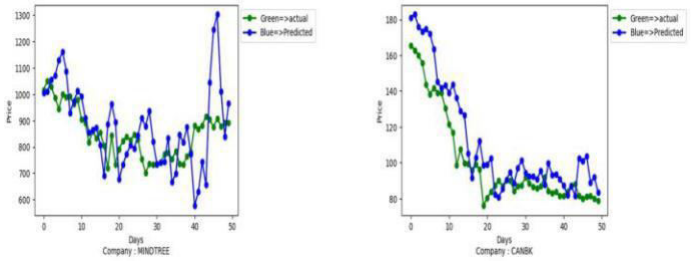

Fig. Actual and Predicted Values Graph

7. Conclusion

Despite its prevalence, stock market prediction remains a secretive and empirical art. A chief goal of this work is to add to the academic understanding of stock market prediction. The hope is that with a greater understanding of how the market moves, investors will be better equipped to prevent another financial crisis. Practically, there is a

need to analyse data on a vast level with multiple companies. Hence, we can take help of machine learning algorithms. Use of CNN algorithm and regression analysis can have much better prediction and accuracy. We are here covering multiple aspects in terms of getting better and updated results from our research-based system. Multiple country data can be tracked for international investments. Even National Level Effective Decisions can be tracked for early prediction.

8. References

[1] Peng, Du, 2019. Analysis of Investor Sentiment and Stock Market Volatility Trend Based on Big Data Strategy. IEEE International Conference on Robots Intelligent System ICRIS.

[2] Dev Shah, Haruna Isah and Farhana Zulkernine, 2018. Predicting the Effects of News Sentiments on the Stock Market. IEEE International Conference on Big Data ICBDA.

[3] Ashish Sharma, Dinesh Bhuriya and Upendra Singh, 2017. Survey of Stock Market Prediction using Machine Learning Approach. IEEE 7th Joint International Information Technology and Artificial Intelligence Conference ICECA.

[4] Ze Zhang, Yongjun Shen, Guidong Zhang, Yongqiang Song and Yan Zhu, 2017. Short-term Prediction for Opening Price of Stock Market based on Self-Adapting Variant PSO Elman Neural Network. IEEE 8th International Conference on Software Engineering and Service Science ICSESS.

[5] Wang and Yaoqing, 2016. Using Social Media Mining Technology to Assist in Price Prediction of Stock Market. IEEE International Conference on Big Data Analysis ICBDA.

[6] Nonita Sharma, Akanksha Juneja, 2017. Combining of Random Forest Estimates using LSboost for Stock Market Index Prediction. IEEE 2nd International Conference for Convergence in Technology I2CT.

[7] Muhammad Firdaus, Swekandiah Endah Pratiwi, Dionysia Kowanda and Anacostia Kowanda, 2018. Literature Review on Artificial Neural Networks Techniques Application for Stock Market Prediction and as Decision Support Tools. Third International Conference on Informatics and Computing ICIC.

[8] Muhammad Wagar, Hassan Dawood, Ping Guo, Muhammad Bilal Shahnawaz and Mustansar Ali Ghazanfar. Prediction of Stock Market by Principal Component Analysis. 13th International Conference on Computational Intelligence and Security CIS, 2017.

[9] Lei Zhao, Lin Wang, 2015. Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm. IEEE 5th International Conference on Big Data and Cloud Computing.

[10] Shashank Tiwari, Akshay Bharadwaj and Sudha Gupta, 2017. Stock Price Prediction using Data Analytics. International Conference on Advances in Computing, Communication and Control ICAC3.