R & D Project Report-1

Academic Year- 2020-21

On

<Stock Market Predictions - First Report >

Submitted by

1. Somanath Vamshi BT18GCS076 CSE
2. Mohammed Faisal BT18GCS067 CSE
3. Arjun Bakshi BT18GEC134 CSE

Supervisor KARTIKAY GUPTA

* Research And Development
** NIIT University

I. INTRODUCTION

There has been extensive research in the field of prediction of stock prices in the past through the use of carefully structured and appropriately tweaked models. If it is correctly modeled, stock prices can be predicted and can give better results compared to human predictions. An approximate prediction of financial data using one or a series of models can help understand the movement of stock market and changes in the financial market at the macroscopic level and provide a good return to the investors, thereby allowing them to maximize profits. Since financial data have complex and fuzzy information it is extremely difficult to predict the movement. Stock prices can be predicted with two methods which are indirectly linked, one is Technical analysis where we use the historical data and some variables and attributes to predict the closing price of the stock using various Machine Learning and Deep Learning techniques. The other is Sentiment analysis which deals with the backend of the stock movement. Basically the news which affects the particular stock or the whole bunch of stocks such as Nifty50, Bank Nifty, Sensex etc. Sentiment analysis is very important since the stock price of a company depends on intrinsic as well as extrinsic attributes. Some bad news can trigger the stock, it may affect a company value by decreasing its stock value. Some of the intrinsic factors could be company's net profit, liabilities, demand stability, competition in market, stakes in raw material supplier and finished product distributors etc. and some of the extrinsic attributes are crude oil price, dollar exchange rate, political stability, government policy decision etc. Note that extrinsic attributes are not in control of the company. Many researchers have tried technical analysis and neglected sentiment analysis which we think is the very essential thing to consider in the prediction. Many different statistical models were applied such as moving average(MA), Weighted moving average, ARIMA, SARIMA etc. Some used deep neural networks like CNN, RNN etc. In this paper we have proposed a hybrid model using Machine Learning, Deep Learning, and Natural Language Processing which will be predicting the closing price of the stock based on the past stock prices and the stock news affecting that company.Sentiment scores for each day will be added as a separate parameter to the dataset.Some researchers included market sentiments but fail to experiment with the lookback periods(Window Size). For example using 2 previous day's sentiment instead of only one day.
Also, a combination of stock news and multivariate time series is taken into consideration in our paper. For example high, low, open ,close prices can all be used instead of only one in combination with sentiment analysis. Finally using a validation set during training and testing data.

II. PROBLEM STATEMENT AND OBJECTIVE

The goal of our work is to collect stock prices of companies listed in NSE of India and scrap the news data related to that stock and predict the closing price of the stock. Prediction accuracy will be expected to be better than the previous research works in the field of stock market prediction by using LSTM and GRU as a robust Deep learning architecture in technical analysis .This will include web scraping and evaluating the keywords to understand the sentiments of the market. More financial data like open value, close value, high, low, volume traded, turnover is collected so that the accuracy will be increased. News will be extracted from google news or money control or screener and then scraped and downloaded. It will be done through python libraries(Beautiful Soup, Selenium and Scrapy), then we can get the sentiment scores by comparing the news with lexicons and then repeat the process for every keyword. These scores will then be added into our dataset.

III. LITERATURE REVIEW

Arjun and Shakya in [1] propose a unique machine learning model combining particle swarm optimization and Least-Square SVMs. Displays results on 13 company datasets. It shows the advantages of incorporating optimization techniques in addition to models, here free parameter combination selections.
The study showed that this combination outperformed even ANNs which are prone to overfitting and local minima problems; these problems are effectively removed by this model. Sentiment analysis was not included in this paper.

The SVM model is used by V Kranthi Sai Reddy[5], to predict the closing date of the NSE. By taking the sharp closing data as the training data to predict the future closing price.The author mainly concentrates on machine learning models for prediction, They try to find the efficiency of SVM in predicting closing price of the stock . By taking the data from different global financial markets. This study indicates that SVM high efficiency of SVM for predicting stock prices.Financial news is considered in this study.

This paper mainly concentrates on comparing machine learning models and deep learning for predicting prices . Hiransha Ma , Gopalakrishnan E.Ab , Vijay Krishna Menonab, Soman K.P They have considered[6] two types of datasets one form NSE and NYSE , In NSE they have taken one company three sectors Automobile,Banking and IT sectors. In NYSE they have taken top two active stocks Bank of America (BAC) and Chesapeake Energy (CHK). For training they used Tata Motors Data and predicted closing prices for each company using ARIMA,MLP,RNN,LSTM,CNN . This study concludes that Deep learning algorithms were more dynamic for time series data and Machine learning algorithms were not dynamic and DL algorithms were outperformed. .Financial news is considered in this study and hybrid networks are not explored.

Arjun Singh and Subarna Shakya [2] performed analysis of the parameter of lookback with variants of RNN for the stock price prediction performance of the two commercial banks listed on Nepal Stock Exchange (NEPSE). It concludes with GRUs getting highest performance(takes into account 7 look back variants). The paper does not provide a variety in comparison of pre-processing, error analysis. Sentiment influence was also not considered.

Sidra Mehtab and Jaydip Sen [7] took the Nifty 50 dataset . The paper mainly concentrates on comparing different machine learning models and LSTM models . Four different LSTM models were used , Univariant with one, two weeks previous opening values as input variables and Multivariate with one,two weeks previous open,close,high,low and volume as input variables. They have many optimization algorithms and error analysis . This study concludes that Deep learning algorithms are more accurate with time series(stock data). They also reveal the fact that multivariate inputs are less accurate than univarient inputs.Financial news is considered in this study and GRU are not explored.

Can Yang and Zhai [14] proposed a framework combining a Convolutional neural network for feature extraction and a LSTM network for prediction. They used a three dimensional CNN for data input in the framework, including the information on time series, technical indicators, and the correlation between stock indices. And in the three-dimensional input tensor, the technical indicators were converted into deterministic trend signals and the stock indices were ranked by Pearson product-moment correlation coefficient (PPMCC). The study includes a fully connected network which was used to drive the CNN to learn a feature vector, which acted as the input of concatenated LSTM. After both the CNN and the LSTM were trained well, they were finally used for prediction in the testing set. This study indicated that the deterministic trend signals and the ranked stock indices in the three-dimensional input tensor play a significant role in improving the prediction performance.

Xuesong Yan [11] treated the financial product price data as a one-dimensional series generated by the projection of a chaotic system composed of multiple factors into the time dimension, and the price series was reconstructed using the time series phase-space reconstruction (PSR) method.In this paper, the PSR method for time series analysis was combined with a DNN-based LSTM network model which was then used to predict stock prices. They selected an optimum activation function and optimization method to optimize DNN model.This paper compared the results with different numbers of hidden layers and number of nodes in the hidden layers of the model and different activation functions(tanh and ReLU). This paper also compared ARIMA, SVR, deep MLP model, deep LSTM model with no PSR process and LSTM combined with PSR.
In [15] Dutta, Kumar and Basu propose a method where sets of exogenous and endogenous parameters are focused on. GRUs train faster and have not been used rarely for cryptocurrency predictions.

Model is compared with an LSTM network. Time series data is obtained from bitcoincharts.com. Features considered are price, daily lag, returns, volatility, transaction volume, fees hash rate, money supply, block size. Variance Inflation factor is considered for feature selection to address multicollinearity. TanH was used for learning and ReLU for activation as it gave best results. Look back periods of 15, 30, 45 and 60 were calculated to achieve best results at 30. Aftermarket trading(weekend fluctuations) and sentiment analysis are not taken into account.

Ghosh and Bose [10] proposed a framework using the LSTM Model to analyze which is the best time span to predict the future share price of a company from a particular sector. Firstly they collected the data from BSE official website. Then they preprocessed the data mainly data discretization, Data transformation, Data cleaning and Data integration. They predicted the future closing price of 5 different companies with the help of LSTM.Future prediction was done for 3 month, 6 month, 1 year and 3 years. LSTM model consists of a sequential input layer followed by 3 LSTM layers and then a dense layer with activation. They generated output using RNN and compared with target values and calculated the error difference. In this study the prediction was visualised using Keras. This study indicated that companies from a certain sector have the same dependencies as well as the same growth rate.

However, all the above mentioned studies did not take sentiment analysis into consideration.

Xu Jiawei and Tomohiro [8] used LSTM as their base model and had taken the Chinese stock market . Author preprossed the data like feature selection and dimensionality reduction and took different inputs in LSTM like stock data, technical index and macro index. Here, stock news sentiment is used as another input . This study mainly concludes that stock news emotions can have an impact on the stock market. Author has not considered comparing many models to predict the stock data.

Sidra and Jaydip [12] proposed a hybrid approach for stock price movement prediction using ML, DL, NLP. In this study NIFTY50 daily data was collected for a period of 4 years which consists of Date,Open, High value and Low value of the Index, Close, Volume of the stock traded on a given date. Nine variables were derived and used in their forecasting models. In this study 8 approaches for classification and 8 approaches of regression were implemented. Then they used the Twitter sentiment analysis of NIFTY50 related tweets during the training and test period. And then these two inputs were fed into the fuzzy neural network based SOFNN algorithm. This study proved that public sentiments in social media serve a very significant input in predictive model building. The prediction results were the best among all the papers. However, this study did not consider Neural networks which might have given better results compared to used Machine Learning techniques.

Satish and Girivarman [13] used a system combining the LSTM for technical analysis and sentiment analysis for fundamental analysis for stock price prediction. Firstly they collected the NSE historical data then removed the null values then removed the ‘,’ in close value. After that the dataset was split into train and test and implemented LSTM model to predict the close price for next ‘n’ days. In the sentiment analysis they extracted Google news data using feed parser and RSS feed[Basically NLP] then used the ‘newspaper’ package to access articles. Later they got the sentiment value by comparing with lexicons and then repeated this process for every

keyword.Finally the output from the technical and the sentiment analysis was compared and processed the individual output and then displayed the result. This paper showed that the proposed system will be useful for the user who is unaware of the stock market.

Tej Bahadur Shahi , Ashish Shrestha , Arjun Neupane and William Guo [8] have used LSTM and GRU to predict the stock price. Data set was taken from Nepal Stock Exchange . Stock news and stock data was scraped from the sharesansar webpage. The Authors had tried to integrate the stock news sentiment with both LSTM and GRU. By this study author concluded that Sentiment analysis is recommended for stock prediction and can get accurate results.

IV .PROPOSED METHODOLOGY TO SPECIFIC GAPS AND CHALLENGES

A major gap in the literature **[10][11][14][7][5]** we observed was the absence of sentiment analysis along with technical analysis in the training data. Market sentiments provide valuable information about the trends that can be included into the model. We will add sentiment scores for each day as a separate parameter to the dataset. This new column of data however, adds a step to the training pipeline accounting for more computational time. It also adds a risk of inaccuracy if sentiment scores are erroneously calculated. The literature that does include market sentiment fails to experiment with the lookback periods (Window Size). For example using 2 previous day's sentiment instead of only one day.
Also, a combination of stock news and multivariate time series will be tried in this paper. For example high, low, open ,close prices can all be used instead of only one in combination with sentiment analysis. Another finding is that a validation set is not used during training and testing. Our method will include this. In addition to this, the error analysis in the papers are confined to generally only one error metric. We wish to incorporate more such metrics for better analysis.

VI. END-SEM DELIVERABLES

Our typical end-sem deliverables include fully optimized code for prediction of stock prices using python language and Libraries in python to visualize the results such as Keras ,Tensorflow etc.We will be developing code for LSTM based prediction and GRU for technical analysis.We will also develop code for Scraping the news data and use Natural Language Processing tools such as Text Blob library or vader for determining the subjective and polarity of the particular text and derive the sentiment scores from that for sentiment analysis.Then compare it with various Machine Learning Models and Deep Learning Models such as ARIMA,SARIMA etc. We will experiment with the proposed model to increase accuracy and fill in the gaps in the literature.

VII. REFERENCES

[1] A Machine Learning Model for Stock Market Prediction

[2] Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE

[3] Fast, Compact, and High Quality LSTM-RNN Based Statistical Parametric Speech Synthesizers for Mobile Devices

[4]LongShortTermMemoryRecurrentNeuralNetwork(LSTM-RNN)BasedWorkloadForecastingModelForCloud Datacenters

[5] Stock Market Prediction Using Machine Learning Kranthi Sai Reddy Vanukuru

[6] NSE Stock Market Prediction Using Deep-Learning Models Hiransha Ma Gopalakrishnan E.Ab , Vijay Krishna Menonab, Soman K.P*

[7]Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models Sidra Mehtab1, Jaydip Sen2 and Abhishek Dutta3

[8]Stock Market Trend Prediction with Sentiment Analysis based on LSTM Neural Network Xu Jiawei, Tomohiro Murata

[9]Stock Price Forecasting with Deep Learning: A Comparative Stud Tej Bahadur Shahi , Ashish Shrestha 2 , Arjun Neupane 1 and William Guo 1,*

[10]Stock Price Prediction Using LSTM on Indian Share Market Achyut Ghosh , Soumik Bose , Giridhar Maji , Narayan C. Debnath , Soumya Sen

[11]Stock price prediction based on deep neural networks Pengfei Yu Xuesong Yan

[12]A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing Sidra Mehtab Jaydip Sen

[13]Stock Price Prediction Using Deep Learning and Sentimental Analysis R. Sathishkumar1 R. Girivarman2 S. Parameswaran2 V. Sriram2

[14]Deep Learning for Price Movement Prediction Using Convolutional Neural Network and Long Short-Term Memory Can Yang, Junjie Zhai , and Guihua Tao

[15] GRU approach to Bitcoin Prediction.