Analysis of Stock Price Prediction using CNN model based on the Historical Stock Prices

Vedansh Sanjay Kokra

ABSTRACT

Forecasting stock price is an essential financial subject issue that has concerned researchers' concentration for numerous years. Demanding of stock price prediction is rapidly increased as it plays major role for company or individual to do prediction for better investment and profit. Deep convolutional neural network model is prepared with four convolutional 1D layers and two fully connected layers. This model is applied to yahoo finance datasets which are collected through python API. Experiment is performed with yahoo finance stock price datasets of last 5 years. Stock price prediction is based on closing price of yahoo datasets. Testing datasets are analysed with proposed 1D CNN model for last one year of datasets. Proposed method of stock price prediction is resulted with prediction error is 1.21 and model accuracy is 95%.

Keywords*: Convolutional neural network, stock price prediction, Deep Learning*

1. INTRODUCTION

Stock price prediction is very crucial and challenging key factor for investors in the stock market (Choudhry & Garg, 2008)(Sharma, Bhuriya, & Singh, 2017). The precise forecasting of stock is very demanding task and helps to investors for making the profit (Tsang et al., 2007). Demand of stock price forecasting is rapidly increased day to day (Ismail, Md Noorani, Ismail, Abdul Razak, & Alias, 2020). Deep learning methods give better performance for many applications. Deep learning approaches handle large datasets applications which are in numeric, image or video form. SVM classification method is applied for stock price prediction by many researchers'(Patil, Patidar, & Jain, 2016)(Sirimevan, Mamalgaha, Jayasekara, Mayuran, & Jayawardena, 2019). Deep learning methodologies are widely used for prediction of stock data in today's ara (Khan et al., 2020)(Yu & Yan, 2020). Stock market is affected by many highly inter-related economic, political, and sentimental aspects, which often interrelate with one another in a very complex

way. As such, it has been always very much difficult to predict the movements of the stock price and stock market index (Ismail et al., 2020). It is very crucial for every researcher's to find out best model which fit for perfect stock price prediction (Ismail et al., 2020).

The methods that had been used to forecast stock market prices mainly characterize into three categories namely fundamental analysis, technical analysis and traditional time series forecasting. Fundamental analysis is a good approach only for long term basis. It is also complicated to formalize fundamental analysis for automated decision support because the interpretation of financial analysis is often extremely subjective. Technical analysis refers to the variety of approaches that aim to predict future price movements using past stock prices and amount of information. It is based on the hypothesis that the past repeats itself and that future market guidelines can be resolute by timid historical price data. Most of the methods used in technical analysis are exceedingly subjective in nature and have been shown not to be statistically valid (Rosenzweig, 2015) . Most challenging task of using data is to produce useful rules from raw data in a dataset/database for users to make decision, and these rules may be hidden extremely in the raw data. The problem with forecasting stock market price is that the quantity of data is too large and huge.

The aim of stock price prediction is based on accurate prediction for next day stock value. Machine learning approaches find out best features from stock price datasets. Various deep learning models are used for prediction of stock price. CNN model is very appropriate as it find out data loss which balance through increasing epochs and achieve high training and testing data accuracy(Dangarwala & Hiran, 2020). This paper focuses on gathering of yahoo finance dataset description, Deep convolutional neural network approach and outcome of methods which apply to stock datasets.

2. REVIEW OF LITERATURE

Machine learning approach which combine particle swarm optimization and LS –SVM methods is used to predict stock price by applying different technical indicators. This integrated approach performance gives better compared to other models. This approach is applied to many stock companies for check out their performance (Kanade, 2020). Combine approach gives less error rate compared to single methodologies of LS-SVM & PSO (Kanade, 2020). Psychological thinking of people approach is applied for developing stock

price indication. Twitter users' psychological data are used to do analysis of people basic emotions. SVM and neural network methodologies are applied for stock price predictions with lexicon based approach (Porshnev, Redkin, & Shevchenko, 2013).

Support vector machine is applied for the stock market prediction. SVM algorithm works on the large dataset value which assembled from various global financial markets. Correlation is the important factor to understand the relationship between the market stock index and global markets (Porshnev et al., 2013). System which forecast the stock market movements are based on the past stock prices and market sentiment analysis. Researcher used data of standard and poor's 500 (S&P 500) from Yahoo finance. Naïve Bayes classification is applied for sentiment analysis and stock movement were predicted using support vector machine, logistic and neural network methods (Deshmukh, Jain, Patwardhan, & Kulkarni, 2016).

Clustering and multiple regression techniqeues is used to forecaste the stock price. Clustering is performed on stock data obtained from NSE, which makes the name of the finest companies as output. Then comparison between partitioning based, hierarchical, model based and density based methods are carry out with the help of validation index such as c-index, Jaccard index, rand index and silhouette index (Bini & Mathew, 2016) . Data mining methods are referred which shown huge potentials in financial applications and will continue to flourish in the new knowledge-based economy. Clustering analysis is used to segment a huge set of data into subsets or clusters. Each cluster is a collection of data objects that are similar to one another within the same cluster but dissimilar to objects in other clusters. Sequential pattern and time-series mining looks for patterns where one event leads to another later event (Zhou & Zhang, 2004).

3. YAHOO FINANCE DATASET

Stock price prediction is applied for yahoo finance stock data. API for Yahoo finance provides Stock data, reports of finance, new related to finance and many more. The datasets provided by Yahoo is totally free so it is very easy for every researcher to gather datasets for their research work. Python API is used to collect data of yahoo finance daily price detail online. First day of January 2015 to last day of June 2020 stock price data are considered for this proposed model. Almost last one year of this dataset are considered as testing datasets.

This dataset has features like date, high, low, close, volume and adjusted close price of yahoo finance. Stock price prediction is major goal for these researches so close prices are considered for it. Features of stock data description is shown in following Table 3.1. Sample Yahoo Finance dataset is shown in following Table 3.2.

Table 3.1 Stock Price Data Features

Features	**Description**
Date	Current date of stock price data
Open	Current day open price of the stock
Low	Current day minimum price of the stock
High	Current day maximum price of the stock
Close	Current day close price of the stock
Volume	Total turnover of shares
Adj. Close	Closing price at starting point

Table 3.2 Sample Yahoo Finance Dataset

Date	High	Low	Open	Close	Volume	Adj Close
02-01-2015	111.44	107.35	111.39	109.33	53204600	99.76601
05-01-2015	108.65	105.41	108.29	106.25	64285500	96.95543
06-01-2015	107.43	104.63	106.54	106.26	65797100	96.96458
07-01-2015	108.2	106.7	107.2	107.75	40105900	98.32424
08-01-2015	112.15	108.7	109.23	111.89	59364500	102.1021
09-01-2015	113.25	110.21	112.67	112.01	53699500	102.2116
..	..	..	..	..	..	..
24-06-2020	368.79	358.52	365	360.06	48155800	359.412
25-06-2020	365	357.57	360.7	364.84	34380600	364.1834
26-06-2020	365.32	353.02	364.41	353.63	51314200	352.9936
29-06-2020	362.17	351.28	353.25	361.78	32661500	361.1289
30-06-2020	365.98	360	360.08	364.8	35055800	364.1434

4. STOCK PRICE PREDICTION CNN MODEL

Convolutional neural network is applied to prediction of stock price for yahoo finance datasets for 2015 to 2020 duration. Input for CNN is ‘close’ price of datasets which is in numeric form. Yahoo finance datasets are distributed into training and testing data. Here last one year data are considered as testing datasets. Four year data from 2015 to 2019 are considered as training datasets. CNN has basic structure of layers in which convolutional layer, max pooling layers and fully connected layers(Dangarwala & Hiran, 2020). The

proposed CNN 1D model is prepared for prediction of stock price. CNN 1D model contains four convolutional layer and two fully connected layers. The 1D convolutional layer has three parameters namely number of filters, filter size and input data size. The CNN 1D model summary is shown in table 4.1. Model summary indicates four CNN 1D layer with 50 filters of size 2. Convolutional operations perform convolved function which applies to input data with filter. These filters functionality is to create feature map. Flatten layer is used to generate single dimensional feature matrix. After this layer, two fully connected layers are applied for prediction. First FC layer contains 25 units which is followed by second FC layer of 1 unit.

Figure 4.1 'Close' Price History of datasets in USD ($)

Table 4.1 Model Summary(Dangarwala & Hiran, 2020)

Model Layers	Output	No. of Features
CNN 1D layer 1	(None, 59, 50)	150
CNN 1D layer 2	(None, 58, 50)	5050
CNN 1D layer 3	(None, 57, 50)	5050
CNN 1D layer 4	(None, 56, 50)	5050
Flatten layer	(None,2800)	-
FC layer 1	(None, 25)	70025
FC layer 2	(None,1)	26

Steps for stock price prediction with CNN 1D model (Dangarwala & Hiran, 2020)

1. Distribution occur with Yahoo finance datasets of 1384 into training and testing sets. Training sets contain 1108 data and testing set contains 276.

2. Deep learning CNN of four 1D layers applied to yahoo datasets . CNN 1D layer of python keras library contains following parameters.

   a) total number of filters
   b) size of filter
   c) size of input

   Each CNN 1D layer generates learnable parameters as shown in table 4.1. These learnable parameters

calculation done with mathematical formula described as follow:

CNN 1D layer learnable paramters = total number of filters in current layer (n) * (preceding layers of total number of filters(n-1) * size of filter)+1)

3. Two fully connected layer is applied after CNN 1D fourth layer which has also learnable value as seen from model summary table 4.1 . These learnable values are calculated with following formula.

   FC learnable values = total number of units by FC layer (n)* (preceding layers of number of filters +1)

4. Apply testing datasets to find out stock price prediction. Here batch size =1 & epoch size are vary with 10,20, 40,60,80 & 100 to find out prediction accuracy. Prediction error is calculated using difference between actual 'close' price and prediction 'close' price .

4. OUTCOME OF STOCK PRICE CNN 1D PREDICITON MODEL

Stock price prediction CNN 1D model generates model accuracy which is almost 95%. June 2019 to July 2020 data of yahoo finance are considered as testing datasets. CNN 1D model testing datasets prediction is based on 'close' price of stock. Model prediction is shown in following figure 4.1 which will indicate that testing actual 'close' price and prediction 'close' price is almost very near to each other after epochs 100. Prediction error is calculated with epoch size 10,20,40,60, 80 & 100 which listed on following table 4.1. Root mean square error in calculated with following formula (Patel, Shah, Thakkar, & Kotecha, 2015). RMSE values are decreased with increasing epochs of training. Prediction error is almost 1.21 after 100 epochs .

RMSE = Sqrt (mean((predictions_ price-acutal_price)**2)))

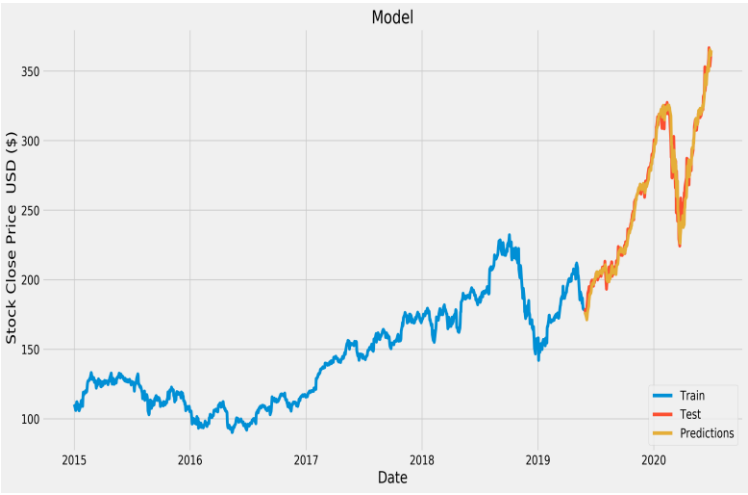

Figure 4.1 Testing dataset Prediction of 'close' price

Table 4.1 RMSE of CNN 1D model with various epochs

EPOCHS	ERROR RATE FOR STOCK PRICE PREDICTION (RMSE) (%)
10	9.76
20	6.48
40	4.25
60	2.10
100	1.21

5. CONCLUSION AND FUTURE WORK

This paper presents a proposal to about for predicting the stock price of yahoo finance stock in the market. In this paper, CNN 1D model is proposed for past data values of the stock price to predict the stock price which gives indication to the investor about whether he or she buy or sell in the stock market. Such proposed model can be supportive for company or investors to take the right decision regarding their stocks in order to derive and predictive information from the past data. The result for this proposed model is shown that prediction of 'close' price is very near to actual 'close' price. Prediction error is 1.21 which is very good approximation for investors. After 100 epochs , rmse values are decreased rapidly which is shown our model perfection. The results for the proposed model were not idyllic because so many other factors such as political events, investors' expectations influence and general economic conditions not covered. As for the future, this model will apply for prediction of any stock data.

6. REFERENCES

Bini, B. S., & Mathew, T. (2016). Clustering and Regression Techniques for Stock Prediction. *Procedia Technology*, *24*, 1248–1255. https://doi.org/10.1016/j.protcy.2016.05.104

Choudhry, R., & Garg, K. (2008). A Hybrid Machine Learning System for Stock Market Forecasting, (July), 315–318.

Dangarwala, K. J., & Hiran, D. (2020). Alphabet classification of indian sign language with deep learning. In *Lecture Notes on Data Engineering and Communications Technologies* (Vol. 46, pp. 569–576). Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-38040-3_64

Deshmukh, B. G., Jain, P. S., Patwardhan, M. S., & Kulkarni, V. (2016). Spin-offs in Indian stock market owing to twitter sentiments, commodity prices and analyst recommendations. *ACM International Conference Proceeding Series*, *12–13–Augu*. https://doi.org/10.1145/2979779.2979856

Ismail, M. S., Md Noorani, M. S., Ismail, M., Abdul Razak, F., & Alias, M. A. (2020). Predicting next day direction of stock price movement using machine learning methods with persistent homology: Evidence from Kuala Lumpur Stock Exchange. *Applied Soft Computing Journal*, *93*, 106422. https://doi.org/10.1016/j.asoc.2020.106422

Kanade, P. A. (2020). Machine Learning Model for Stock Market Prediction. *International Journal for Research in Applied Science and Engineering Technology*, *8*(6), 209–216. https://doi.org/10.22214/ijraset.2020.6030

Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2020). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, (0123456789). https://doi.org/10.1007/s12652-020-01839-w

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, *42*(4), 2162–2172. https://doi.org/10.1016/j.eswa.2014.10.031

Patil, S. S., Patidar, K., & Jain, M. (2016). Stock market trend prediction using support vector machine. *International Journal of Current Trends in Engineering & Technology*, *2*(1), 18–25. Retrieved from http://casopisi.junis.ni.ac.rs/index.php/FUAutContRob/article/view/585

Porshnev, A., Redkin, I., & Shevchenko, A. (2013). Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis. *Proceedings - IEEE 13th International Conference on Data Mining Workshops, ICDMW 2013*, 440–444. https://doi.org/10.1109/ICDMW.2013.111

Rosenzweig, E. (2015). Successful user experience: Strategies and roadmaps. *Successful User Experience: Strategies and Roadmaps*, 1–344. https://doi.org/10.1016/c2013-0-19353-1

Sharma, A., Bhuriya, D., & Singh, U. (2017). Survey of stock market prediction using machine learning approach. *Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017*, *2017–Janua*, 506–509. https://doi.org/10.1109/ICECA.2017.8212715

Sirimevan, N., Mamalgaha, I. G. U. H., Jayasekara, C., Mayuran, Y. S., & Jayawardena, C. (2019). Stock Market Prediction Using Machine Learning Techniques. *2019 International Conference on Advancements in Computing, ICAC 2019*, 192–197. https://doi.org/10.1109/ICAC49085.2019.9103381

Tsang, P. M., Kwok, P., Choy, S. O., Kwan, R., Ng, S. C., Mak, J., … Wong, T. L. (2007). Design and implementation of NN5 for Hong Kong stock price forecasting. *Engineering Applications of Artificial Intelligence*, *20*(4), 453–461. https://doi.org/10.1016/j.engappai.2006.10.002

Yu, P., & Yan, X. (2020). Stock price prediction based on deep neural networks. *Neural Computing and Applications*, *32*(6), 1609–1628. https://doi.org/10.1007/s00521-019-04212-x

Zhou, D., & Zhang, L. (2004). Discovering golden nuggets: data mining in financial application. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *34*(4), 513–522.