Financial Analysis: Stock Market Prediction Using Deep Learning Algorithms

Naman Arora, Parimala M*

VIT, Vellore-632014, India

ARTICLE INFO

Article history:
Received 16 January 19
Received in revised form 30 January 19
Accepted 24 February 19

Keywords:
Stock Market
Neural Network
Back Propagation
Long Short Term Memory
Support Vector Machine

ABSTRACT

Big Data Analytics plays a major role in collecting large amount of data, analyzing the information and Deep Learning comes into the mind as it analyses a massive amount of data (Big Data).It can be used to formulate the exact and valued information that is buried in a Big Data. The current share market is an associate example of these social networks. Stock market is a good means of generating income but when to buy or sell the stocks, has not been determined yet. Deep Learning proves beneficial in handling large amount of unstructured or unsupervised data. In this paper, we adopt Deep Learning concept in order to improve correct classification using various algorithms to predict different stock data values so as to exceed the current accuracy of stock price prediction.

Peer review under responsibility of International Conference on Sustainable Computing in Science, Technology and Management.

1. Introduction

Stock market prediction is fundamentally an important criteria in investing. Predicting short-term movement of any stock or the market in general is not easy as it can lead to major losses if prediction gone wrong .It seems to be almost impossible for all investors to correctly and consistently predict these things. Therefore in order to predict on the basis of huge amount of data in the financial market we use the term 'Big data'. Big data refers to managing and analyzing the digital data which is difficult to handle using traditional software tools and technologies. There are 4 main keys that define Big Data namely

- volume
- variety
- velocity
- veracity

Primarily by increasing the data volume the complexity of the data also increases as well. Most of the social media corporations together with Facebook, Twitter, LinkedIn have an oversized quantity of information, hence because the size of information increase, Deep Learning approach become additional necessity to supply massive information analysis.

Velocity is that feature of Big Data where information is being generated and to be processed quickly. Examples embrace E-Promotions that support your current location, your purchase history, what you wish to send promotions straight away for store next to you.

Variety is another feature of big data which involves different formats, types, and structures into which complex data is structured. Data can be text, numeral, picture image, audio for listening, video for visualising, sequential, social media data etc.

Last feature is veracity. Veracity is outlined as trustiness of the information in big data. On expanding the amount of information sources and sorts, data turns into a sensible challenge. Additionally, there are innumerable challenges like information cleaning, feature engineering, high-dimensionality, and information redundancy that massive information analytics face.

2. Related work

There are many algorithms to predict stock market value. Neural Network, Genetic Algorithm, Association, Decision Tree and Fuzzy systems, recurrent neural networks are commonly and widely used. An overview of researches on comparing algorithms suggests the applicable use of technical indications for stock price prediction over the ultimate two decades.

Kim and Han analyzed algorithm approach in neural systems at the expectation of stock cost esteem. This Genetic Algorithm isn't just used to enhance the preparation calculation, yet in addition to limit the complexity nature in highlight space. Genetic algorithmic rule optimizes the association weights between layers and thresholds for feature discretization at the same time. The weights that evolved in genetic form overcome the general limitations of the gradient descent algorithm. Hence genetic based model overcomes the conventional models.

Ming-Chang Lee, Chang To, proposed back propagation and support vector machine to predict the stock market value and there study showed support vector machine algorithm more accurate and correct.

Salim Lahmiri proposed the comparison of Probabilistic Neural Networks (PNN) and Support vector machines (SVM) algorithm for stock market analysis using economic and technical information which showed an extra ordinary performance of PNN in technical indications, while SVM perform in economic information.

Yang, Min and Lin proposed the use of fuzzy neural systems in stock exchange value estimating based on Genetic Algorithm finding fuzzy principles (Gupta, A, & Sharma, D. S. D, 2014). This paper gives a technique to enhance discovery display considering issues existed in its application. The change in information is accomplished primarily by applying Genetic Algorithm in fuzzy framework to find rules, dispense with errors caused by noisy data, and hence form valid set of principles. Likewise, fuzzy thinking approach is used based on the rule sets to anticipate value of stock market.

Alaa F.Sheta , Sara Elsir M.Ahmed, Hossam Faris showed a Comparative analysis between 3 algorithms as Regression, Artificial Neural Networks and Support Vector Machines for predicting Index of stock market and finally there study found that the designed SVM model with RBF kernel model provided better prediction capabilities with respect to the regression and ANN model. The results were validated using the number of criteria.

Sapkal et al., compared four Stock Market Prediction Algorithms such as Forecasting Algorithm, Moving Averages Algorithm, Regression Algorithm and Neural Network Algorithm and gave the best prediction decision based on these algorithm as artificial neural network (ANN).

Mahdi, Hamidreza and Homa examined stock exchange esteem forecast utilizing neural systems. In this paper, two sorts of neural systems, a feed forward multilayer perception (MLP) and an Elman recurrent network, are used to predict stock value dependent on its stock share value history. The utilization of MLP neural system is more significant in foreseeing stock value changes instead of Elman recurrent network and linear regression method. However, the direction of changes of stock value can be predicted better by Elman recurrent network and linear regression than MLP.

Shweta, Rekha and Vineet proposed foreseeing future trends in stock market by decision tree rough set based hybrid framework with Hierarchical Hidden Markov Model (HHMM). It introduces hybrid framework based on decision tree rough set, for foreseeing the trends in the Bombay Stock Exchange with the combination of Hierarchical Hidden Markov Model. It likewise exhibits future trends based on price earnings and dividend.

Vivek Rajput and Sarika Bobde tried stock market forecasting techniques using sentiment analysis from social media and various other data mining techniques so as to predict the stock market value but after the detailed study, got the conclusion that it is very complex task and various essential parameters should be considered for stock prediction.

3. Algorithms

Stock prediction in my paper is based on 3 main algorithm, on analysis of which best algorithm for the prediction of stock market can be evaluated.

- Back Propagation
- Long short term memory
- Support Vector Machine

3.1 Back Propagation

Back Propagation Algorithm (Figure 1a) is used for both Classification and Regression problem. Here, Stock Price Prediction is a Classification problem. We have Implemented Back Propagation algorithm for stock price prediction using Numpy and Pandas library. Back propagation (backward propagation of errors) is a general supervised learning method to train artificial neural networks Navale, G. S., Dudhwala, N., Jadhav, K., Gabda, P., & Vihangam, B. K. (2016). used with an optimization method such as gradient descent. Back propagation calculates the error (gradient) of network regarding the network's weights that can be modified.

Back propagation is a multilayer feed forward network. A Multilayer feed forward neural network is a network that consisting of an input layer, hidden layer may be one or more than one and one output layer. Neural network that has no hidden layer is called Perceptron. In Back Propagation, there is connection between input layer and hidden layer and hidden layer is connected to the output layer by interconnection of weights. As the number of layers increases, the computational complexity of the neural network also increases. This may results in the time taken for convergence and to minimize the error to be very high.

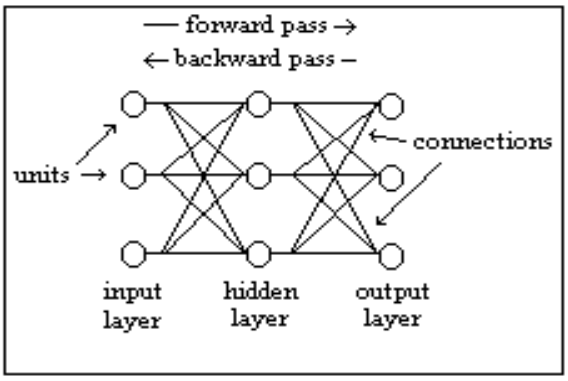

Fig.1 - (a) Back Propagation

3.2 Long Short Term Memory (LSTM)

Long Short Term Memory (LSTM) is structural block of neural network. Blocks of LSTM are used to build a recurrent neural network. Recurrent Neural Network (RNN) is a type of neural network (Ican & Celik, 2017) where the output of a block is fed as input to the next iteration. Main components of LSTM are:

- cell
- input gate
- output gate
- forget gate

The cell functions as remembering the values at constant or arbitrary time intervals. Among these 3 gates any of them can be called as conventional artificial neuron, as in multi-layer feed forward network. That means they compute an activation of weighted sum. There are connections between cell and gate some of which are recurrent and some are not.

As mentioned above, stock market prediction is a time series problem. LSTM (Figure 2a) can be used for time series prediction. LSTM do not have the vanishing gradient problem which a traditional RNN has.

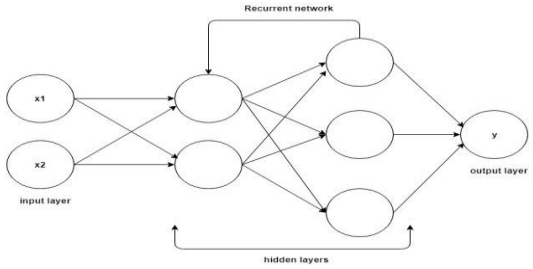

Fig.2. (a) Long Short Term Memory

3.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) is among the prevalent algorithms utilized for some characterization issues (Figure 3a). It is one among the regulated learning models with related learning algorithms that examine the information utilized for arrangement and relapse analysis. From the given arrangement of training algorithm (Hegazy, O., Soliman, O. S., & Salam, M. A., 2014), SVM training algorithm constructs a model that assigns out new models to either class, making it a non-probabilistic paired direct classifier. SVM model is also referred to as general representation of examples, mapped with the goal that the models of the different classifications are partitioned by clear gap i.e. as wide as could

be allowed. New examples are additionally mapped into a similar space and anticipated to have a place with the class dependent on which side of the gap they fall.

SVMs can also proficiently play out a non-straight arrangement utilizing the kernel trick, verifiably mapping their contributions to high dimensional component spaces. As for given problem our data was not linearly separable, so we implemented "RBF" kernel (Deepak, R. S., Uday, S. I., & Malathi, D., 2017) which gives better result for nonlinear kernel.

For our problem we implemented SVM using Scikit Learn (sklearn) Library. Using python code, we imported libraries, tried to run SVM on train dataset and then predicted values on that machine for test dataset.

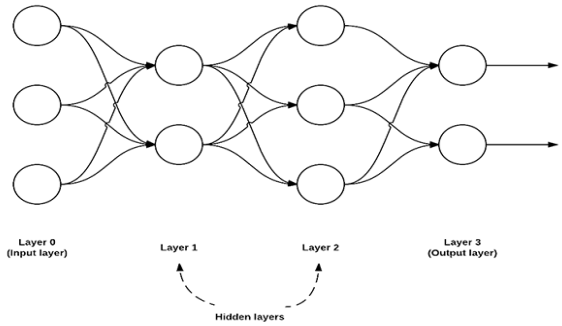

Fig.3- (a) Support Vector Machine

4. Comparison and Result

To make sure that one algorithm has a consistent performance and it performs better than other, we tried to run same algorithm multiple times. Each algorithm ran 30 times (Figure 4) and for each time, prediction accuracy is calculated on test data. For each run, different training and testing datasets are used.

```
Accuracy: 64.98450641876937
Accuracy: 65.30070195408841
Accuracy: 65.26275848985013
Accuracy: 65.28805413267564
Accuracy: 65.19951938278632
Accuracy: 65.21216720419908
Accuracy: 64.9718585973566
Accuracy: 65.31334977550117
Accuracy: 65.27540631126288
Accuracy: 65.28805413267564
Accuracy: 65.19951938278632
Accuracy: 65.04774552583318
Accuracy: 65.21216720419908
Accuracy: 65.32599759691394
Accuracy: 65.14892809713527
Accuracy: 65.18054765066718
Accuracy: 65.37658888256497
Accuracy: 64.78214127616518
Accuracy: 65.12363245430974
Accuracy: 65.16789982925441
Accuracy: 65.0414216151268
Accuracy: 65.28173022196927
Accuracy: 65.6295453108202
Accuracy: 64.99715424018213
Accuracy: 65.29437804338203
Accuracy: 65.20584329349269
Accuracy: 65.2690824005565
Accuracy: 65.31334977550117
Accuracy: 65.11730854360336
Accuracy: 65.19319547207994
n_epoch: 30
Mean_Accuracy: 65.20015177385694
Standard_Deviation: 0.15193494335888166
total time : 719.4723827838898
```

Fig.4- SVM

```
Accuracy:=68.026307
Accuracy:=68.114842
Accuracy:=60.412319
Accuracy:=64.579776
Accuracy:=67.185227
Accuracy:=64.877000
Accuracy:=68.114842
Accuracy:=68.114842
Accuracy:=61.752988
Accuracy:=66.805793
Accuracy:=67.754379
Accuracy:=67.337001
Accuracy:=62.050212
Accuracy:=68.121166
Accuracy:=62.227281
Accuracy:=66.382091
Accuracy:=66.065895
Accuracy:=67.836590
Accuracy:=68.089547
Accuracy:=66.110163
Accuracy:=67.994688
Accuracy:=63.599570
Accuracy:=68.114842
Accuracy:=68.076899
Accuracy:=68.114842
Accuracy:=64.301524
Accuracy:=64.428002
Accuracy:=62.764814
Accuracy:=66.628723
Accuracy:=68.114842
n_epoch: 30
learning rate: 0.5
Mean_Accuracy: 66.06990029300785
Standard_Deviation: 2.3292274542540126
total_time : 731.2970454692841
```

Fig.5- Back Propagation

```
Accuracy:63.81001548786783
Accuracy:67.95732231973842
Accuracy:67.75081741524694
Accuracy:67.47547754259163
Accuracy:65.90948201686456
Accuracy:66.02994321115126
Accuracy:67.95732231973842
Accuracy:65.87506453278266
Accuracy:64.54999139562898
Accuracy:64.60161762175186
Accuracy:67.63035622096025
Accuracy:67.87127860953365
Accuracy:63.99931165031836
Accuracy:68.1638272242299
Accuracy:65.84064704870075
Accuracy:67.38943383238686
Accuracy:67.33780760626398
Accuracy:67.50989502667355
Accuracy:67.8884873515746
Accuracy:67.66477370504217
Accuracy:67.80244364136982
Accuracy:64.75649630012046
Accuracy:65.8578557907417
Accuracy:66.27086559972466
Accuracy:65.47926346584065
Accuracy:66.54620547237997
Accuracy:64.9802099466529
n_epoch: 30
Mean_Accuracy: 66.55825159180864
Standard Deviation: 1.3068310850096652
```

Fig.6- LSTM

4.1 SVM Result

Figure 4 shows, for 30 runs of SVM algorithm we got around 65.20 mean accuracy with 0.15 standard deviation. This shows that SVM performance is consistent for 30 runs due to nurture of SVM algorithm. Algorithm will keep training until it can classify max testing data. This results into almost same network at the end resulting same close accuracy values on test data.

4.2 Back propagation Result

Figure 5 shows, for 30 runs of Back propagation algorithm we got around 66.0 mean accuracy with standard deviation of 2.32. As we can see Back propagation performs well as compared to SVM but it has huge fluctuation in accuracy so, it may cause an issue when we want steady accuracy.

4.3 LSTM Result

Figure 6 shows, for 30 runs of LSTM algorithm we got around 66.55 mean accuracy with standard deviation of 1.30. As we can see LSTM performs well as compared to SVM and Back propagation. Also there is no huge fluctuation in accuracy, so this has an overall performance improvement as compared to other algorithms.

	SVM	Back Propagation	LSTM
Mean Accuracy	65.20	66.0	66.55
Standard Deviation	0.15	2.32	1.30
Variation	0.023(approx.)	6.243(approx.)	1.851(approx.)

Table: 1 Comparison Results

4.4 T-Test

After getting the accuracy for each algorithm, we have to see which one performs better than the other, therefore we need to do T-Test.

T-0.95(58) = +-1.701

T test value for 95% accuracy for 30 epoch run was in range of +-1.701. If the value for each pair is under that range that means chances of getting improvement is 0.05% with other pair algorithms. As we can see in back propagation and SVM implementation, T-test result falls nearest to the range, therefore chances of getting better result from back propagation as compared to SVM is very less, as there is big variations for each run in back propagation as shown in Table 1 and Table 2.

Comparison	T-Test Result
Back propagation and SVM	1.885
Back Propagation and LSTM	1.133
LSTM and SVM	5.65

Table: 2 T-Test Results

5. Conclusion

In this project, we have demonstrated a machine learning approach (deep learning) to predict stock market trend using different neural networks. Results show how history data has been used to predict stock movement with reasonable accuracy. Also, with T test result analysis we can conclude that LSTM performs better in comparison to Back propagation and SVM. For this implementation, we can conclude that if we incorporate all the factors that affect performance of the stock and feed them to neural network with proper data preprocessing and filtering, after training the network we will be able to have a model which can predict stock momentum more accurately and precisely for the better idea of stock value so that firms may have increased profit ratio as compared to what is might be going currently at that time. This will also lead to more transparency regarding stock as it will be easier for firms to analyze loses and achieve great success.

6. Future work

In future new data set record can be used for pre-processing and training to meet the trend pattern that would be going on at that particular time. This can result into better stock forecasting/prediction and profit for financial firms. In future, new algorithms can be designed with deep learning algorithms to predict the stock market trend analysis more accurately and efficiently.

REFERENCES

Kim, K. J., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. Expert systems with Applications, 19(2), 125-132.

Kim, K. J. (2006). Artificial neural networks with evolutionary instance selection for financial forecasting. Expert Systems with Applications, 30(3), 519-526.

Lee, M. C., & To, C. (2010). Comparison of support vector machine and back propagation neural network in evaluating the enterprise financial distress. arXiv preprint arXiv:1007.5133.

Naeini, M. P., Taremian, H., & Hashemi, H. B. (2010, October). Stock market value prediction using neural networks. In Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on (pp. 132-136). IEEE.

Tiwari, S., Pandit, R., & Richhariya, V. (2010). Predicting future trends in stock market by decision tree rough-set based hybrid system with HHMM. International Journal of Electronics and Computer Science Engineering, 1(3).

Lahmiri, S. (2011). A comparison of PNN and SVM for stock market trend prediction using economic and technical information. International Journal of Computer Applications, 29(3), 24-30

Yang, K., Wu, M., & Lin, J. (2012, May). The application of fuzzy neural networks in stock price forecasting based on genetic algorithm discovering fuzzy rules. In Natural Computation (ICNC), 2012 Eighth International Conference on(pp. 470-474). IEEE.

Sheta, A. F., Ahmed, S. E. M., & Faris, H. (2015). A comparison between regression, artificial neural networks and support vector machines for predicting stock market index. Soft Computing, 7(8).

Sapkal, S., Barate, T., Kadbane, N., Pimple, M., & Kumbharkar, P. (2016). Comparative Study and Analysis of Stock Market Prediction Algorithms. International Journal of Innovative Science, Engineering & Technology, 3.

Rajput, V., & Bobde, S. (2016). Stock market forecasting techniques: Literature survey. International Journal of Computer Science and Mobile Computing, 5(6), 500-506.

Gupta, A., & Sharma, D. S. D. (2014). A survey on stock market prediction using various algorithms. International Journal of Computer Technology and Applications, 5, 530-533.

Hegazy, O., Soliman, O. S., & Salam, M. A. (2014). A machine learning model for stock market prediction. arXiv preprint arXiv:1402.7351.

Navale, G. S., Dudhwala, N., Jadhav, K., Gabda, P., & Vihangam, B. K. (2016). Prediction of Stock Market using Data Mining and Artificial Intelligence. International Journal of Engineering Science, 6539.

Ican, Ö., & Çelik, T. B. (2017). Stock Market Prediction Performance of Neural Networks: A Literature Review. International Journal of Economics and Finance, 9(11), 100.

Deepak, R. S., Uday, S. I., & Malathi, D. (2017). Machine Learning Approach in Stock Market prediction. International Journal of Pure and Applied Mathematics, 115(8), 71-77.