SJIF Impact Factor: 6.260| ISI I.F.Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 5 | Issue: 4 | April 2020 - Peer Reviewed Journal

APPLYING MACHINE LEARNING MODELS IN STOCK MARKET PREDICTION

Vignesh CK
MCA Scholar, Department of MCA, School of CS & IT,Jain (Deemed-to-be) University, Bengaluru

Article DOI: https://doi.org/10.36713/epra4361

ABSTRACT

This paper deals with the techniques of attempting to calculate the future value of a company stock or any other financial instrument which is being traded in a stock exchange. This prediction plays a great role in many financing and investing decisions. This calculation can be done by Machine learning by training a model to identify the trend from past data in order to predict the future. The main topic of study here will be the comparative analysis of the SVM and LTSM algorithms.

KEYWORDS: *Machine learning, Stock price, Stock market, Support vector machine, neural network, long short term memory.*

I. INTRODUCTION

Stock markets have been operating on the digital paradigm after the advent of Information Technology. Artificial Neural Networks, which serve as mathematical function approximators, make the crux of this application. The popularly implemented ANN

in use for this is the feed forward network. Apart from that there are Back propagation networks. They utilize the backward propagation of errors algorithm to adjust weights in the model. SVM (Support vector machine) Algorithm along with Random Forest has had its implementation for stock prediction, which is included. A modified type of Recurring Neural Network called LSTM is also implemented. It memorizes historical or past data for prediction. The result of this project includes a brief conclusion for how the algorithm performs vis-a-vis the real world figures thereof.

SVM and Back propagation have shown reasonable accuracy in the previous studies [1] [2] [3] [4] [5] [6] [7] [8]. We cannot afford to tweak the model to improve the accuracy as the market operates in very volatile circumstances. This involves time series problems for which we need LSTM to analyse past data and come up with predictions.

II. DESCRIPTION

Problem statement was to predict increase or decrease in price for any given day in future. I addressed this as classification problem. The main goal is to compare performance of SVM and Back propagation algorithm's results.

III. DATAFLOW DIAGRAM

The flow diagram can be briefly represented by this diagram.

Figure 1: Data Flow Diagram

First, the past data is fetched from the dataset. Then, it's organised and plotted according to our project's requirement. There are 2 sets created – the Training set and the Testing set.

SJIF Impact Factor: 6.260| ISI I.F.Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 5 | Issue: 4 | April 2020 - Peer Reviewed Journal

After training a model, it's tested using the test model. If the accuracy is reasonable, we can assume that the model is reasonably trained.

Once the predicted results are out, the model is saved after assessment and analysis. The accuracy of the model depends upon how the model is trained.

IV. ARCHITECTURE

The architecture of the model which we create briefly appears like this.

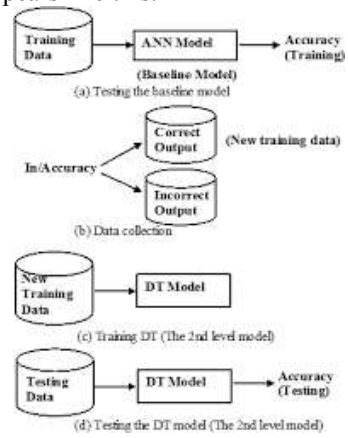

Figure 2: Architecture

V. APPROACH

The approach for this project consists of:

1. Creating dataset.
2. Implementing the algorithm.
3. Comparing result.
4. Analysis of the result.

1. Dataset creation :

Yahoo Finance contains stock prices for various companies. Hence that is Dataset here. The collected data is from January 2011 to December 2015. The 2008 subprime financial crisis created an unexpected change in the trends and hence that has been avoided. If not, it would cause unusual and unexplained behaviour by the model. [3]

The data set contains stock data of the following companies:

- Yahoo
- Microsoft

The stock dataset which we got from yahoo finance contains the following parameters:

1. Date
2. Open
3. High
4. Low
5. Close

The closing value of a day is assumed as the stock price of that day.

Parameters calculated for input dataset

The below mentioned are some of the other parameters. [9]

- **Momentum:** If price of stock is more than yesterday then the momentum for given day is +1 as there is an increase in price. It's -1 if vice versa.
- **Volatility:** Represents how big or small the changes in values are. Volatility is the difference between values of today and yesterday, divided by the closing value of the previous day.
- **Index Momentum:** Calculated based on market performance for last 3 days. It's an average of 3 days index momentum.
- **Index Volatility:** Calculated as the average index Volatility over the last 3 days.
- **Stock Momentum:** Calculated as the last 3 days' average momentum for the given momentum.
- **Stock Price Volatility:** Calculated as the average of last 3 days of the given stock.

Output: *If closing stock price for a stock today day is more than yesterday's closing stock price for the same stock, then the corresponding output is denoted by 1 else it is denoted as 0.*

2. Implementation of Algorithms

I) Support vector machine

SVM is the algorithm used for classification problems. It is a supervised learning model with associated learning algorithms that analyzes data used for classification and also regression analysis. A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving SVM model sets of labelled training data for each category, they're able to categorize new text. [10] [11]

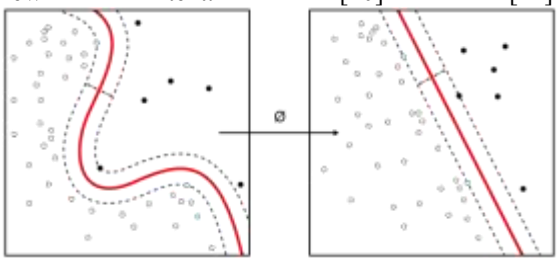

Figure 3: SVM Algorithm

SVM using Scikit Learn Library [12] has been implemented in this study. Using python codes, import the library, try SVM on training dataset and later apply them on the test dataset.

SJIF Impact Factor: 6.260| ISI I.F.Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 5 | Issue: 4 | April 2020 - Peer Reviewed Journal

II) Long Short Term Memory

LSTM [13] stands for Long Short Term memory. It is building block of a neural network (like perceptron which is use for supervised learning of binary classifiers). LSTM is an algorithm that consists of many blocks which are used to build a Recurring Neural Network. An LSTM block is typically composed of four parts. They are:

1. Cell
2. Input gate
3. Output gate
4. Forget gate

The cell remembers values over arbitrary time intervals therefore involving the concept of memory in the LSTM model. This is part of the cell's primary duties.

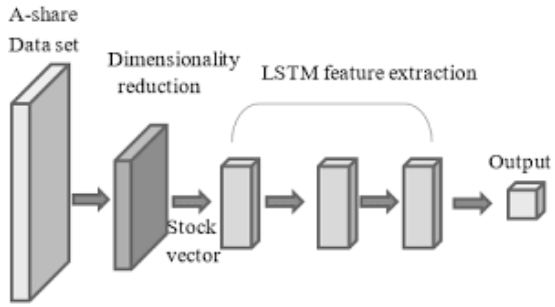

Figure 4: LTSM Model

3. Comparison of result and analysis

The purpose of this paper was to make sure that one among the many algorithms used, performs consistently and even better than others against which it has been run and tested numerous times. For each run of the algorithm, the prediction accuracy is calculated for the test data. Each algorithm mentioned above, was run and checked for more than 10 times. The same training dataset and testing dataset is never used for the same run. The accuracy results for each algorithm are mentioned below.

SVM Result

The SVM algorithm was run 30 times, as shown in Figure 6. The mean accuracy of these results is 65.20 while the standard deviation was 0.15. This shows the performance consistency of the SVM. It can be trained further with more datasets to improve the accuracy.

```
Accuracy: 65.26275848985813
Accuracy: 65.2800541326756
Accuracy: 65.19951938278632
Accuracy: 65.21216720419008
Accuracy: 64.9718585973566
Accuracy: 65.31334077550117
Accuracy: 65.27540631126208
Accuracy: 65.2800541326756
Accuracy: 65.19951938278632
Accuracy: 65.04774552583318
Accuracy: 65.21216720419008
Accuracy: 65.32599759691394
Accuracy: 65.14892809713527
Accuracy: 65.18054765066718
Accuracy: 65.37650808256497
Accuracy: 64.78214127616518
Accuracy: 65.12363245430974
Accuracy: 65.16789902925441
Accuracy: 65.0414216151268
Accuracy: 65.28173022196927
Accuracy: 65.6295453100202
Accuracy: 64.99715424018213
Accuracy: 65.29437804338203
Accuracy: 65.20584329349269
Accuracy: 65.2690824005565
Accuracy: 65.31334977550117
Accuracy: 65.11738854360336
Accuracy: 65.19319547207994
n_epoch: 30
Mean_Accuracy: 65.20015177385694
Standard_Deviation: 0.15193494335888166
total_time : 181.6788330078125

Process finished with exit code 0
```

Figure 5: SVM Result

LSTM Result

The LSTM algorithm is also run 30 times, to get a better perspective while comparing. 66.83 was the mean accuracy for this algorithm. The standard deviation was 1.36 in this case. This performs well compared to SVM. Also there in no significant fluctuation in accuracy compared to other algorithms.

```
Accuracy:66.27886559972466
Accuracy:68.06057477198418
Accuracy:68.07770351402513
Accuracy:68.2498709344347
Accuracy:65.44484598175873
Accuracy:66.98758985524085
Accuracy:66.64945792462571
Accuracy:68.28428841851661
Accuracy:67.01804150748581
Accuracy:67.83686112545173
Accuracy:64.56720013766993
Accuracy:66.51178798829805
Accuracy:63.20770951643435
Accuracy:66.8903803131991
Accuracy:67.44186005850973
Accuracy:68.19824470831182
Accuracy:67.76002615728789
Accuracy:65.28996730339011
Accuracy:68.47350458096713
Accuracy:65.42763723971777
Accuracy:68.2498709344347
Accuracy:66.32249182584754
Accuracy:67.47547754259163
Accuracy:64.3434864911375
Accuracy:67.58909502667355
Accuracy:66.63224918258476
Accuracy:67.05406906749268
Accuracy:68.09499225686608
Accuracy:67.2861813881411
Accuracy:64.68766133195663
Accuracy:66.27886559972466
n_epoch: 30
Mean_Accuracy: 66.83416508919863
Standard_Deviation: 1.3644332040103173

Process finished with exit code 0
```

Figure 6: LTSM Result

VI. FUTURE SCOPE

This model can be further trained and developed to carry out advanced tasks like volume deduction i.e. volume of the stock prices which can be sold/purchased in a way which is beneficial to

SJIF Impact Factor: 6.260| ISI I.F.Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 5 | Issue: 4 | April 2020 - Peer Reviewed Journal

the investor. Past datasets can be used to train the model to gain more accuracy and get a better prediction which has a 70% or more accuracy. With inclusion of a variety of other factors that affect the stock prices, it can be used to provide accurate financial advice.

CONCLUSION

This project is a demonstration of the application of machine learning to solve the problems in stock prediction. The past data of the stocks was considered to train the model in a way where it could find out trends and patterns and thereby predict the data in future. This project also proved that LSTM worked better compared to back propagation and SVM algorithms. For this implementation, it can be summed up that incorporation of all the factors that affect stock performance being fed into neural network with proper data processing and filtering, a model which can predict stock market prices very accurately can be developed.

REFERENCES

1. *Iacomin, Radu, "Stock Market Prediction," 2015 19th International Conference on System Theory, Control and Computing (ICSTCC), October 14-16, Cheile Gradistei, Romania, 2015.*
2. *Parmar, Ishita, "2018 First International Conference on Secure Cyber Computing and Communication(ICSCCC)," IEEE, 2018.*
3. *Divit Karmiani, Ruman Kazi, Ameya Nambisan, Aastha Shah, Vijaya Kamble, "Comparison of Predictive Algorithms: Backpropagation, SVM,LSTM and Kalman Filter for Stock Market,"*
4. *A. Zheng, J. Jin, "Using AI to make predictions on stock,"*
5. *Y. Xia, Y. Liu, Z. Chen, "Support vector regression for prediction of stock trend,"*
6. *Madge, S., "Predicting stock price direction using support vector,"*
7. *Olah, C., "Understanding LSTM networks," github.com, http://colah.github.io/posts/2015-08-Understanding-.*
8. *G. Bonde, R. Khaled, "Extracting the best features for predicting stock prices using machine learning,"*
9. *Heinz, Sebastian, "A simple deep learning model for stock price prediction using TensorFlow," Medium.com, https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877.*
10. *Stecanella, Bruno, "An Introduction to Support Vector Machines (SVM)," Monkey Learn, https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/.*
11. *Judith Hurwitz, Daniel Kirsch, Machine Learning for Dummies, s.l. : John Wiley and Sons.*
12. *Developers, Scikit-learn, "Support Vector Machines," SciKit, https://scikit-learn.org/stable/modules/svm.html.*
13. *Brownlee, Jason, "Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras," Machine Learning Mastery, https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/.*
14. *Pratik, "CloudXLab," NumPy and Pandas Tutorial – Data Analysis with Python, https://cloudxlab.com/blog/numpy-pandas-introduction/.*
15. *Ng, Yibin, "Machine Learning Techniques applied to Stock Price Prediction," Towards Data Science, https://towardsdatascience.com/machine-learning-techniques-applied-to-stock-price-prediction-6c1994da8001.*
16. *Hiransha M, Gopalakrishnan EA, Vijay Krishna Menon,Soman KP, "NSE Stock Market Prediction Using Deep-Learning Models,"*