Stock Price Prediction using SVM and LSTM

Harit Ahuja[1], Sonakshi Mehra[2], Dr.G.Manju[3], Aishwarya D[4], Yashaswi Bhavaraju[5]

[1]*Student, CSE, SRM Institute of Science and Technology, Kattankulathur, Chennai, India.*
[2]*Student, CSE, SRM Institute of Science and Technology, Kattankulathur, Chennai, India.*
[3]*Associate Professor, CSE, SRM Institute of Science and Technology, Kattankulathur, Chennai, India.*
[4]*Research Scholar, CSE, SRM Institute of Science and Technology, Kattankulathur, Chennai, India.*
[5]*Student, ICT, Manipal Institute of Technology, Manipal, Karnataka, India.*

Abstract*—Stock market prediction is the process of determin-ing the future value of a stock of a company on an exchange. Predicting a stock market price is a huge challenge due to its dynamic environment. Most of the stockbrokers use fundamental, technical or time series analysis to make the prediction about the prices. The main motive of this paper is to predict the stock trend. This paper implements and compares the two artificial neural network (ANN). The first algorithm is the SVM (Support Vector Machine) and the second algorithm is the LSTM (Long Short Term Memory) which memorizes the history data for prediction. The paper focuses on the US market that include companies like Yahoo, Microsoft, Apple, Intel etc. and uses stock parameters like the date, open, high, low, close and volume. Each algorithm was run 30 times to calculate the prediction accuracy. SVM resulted in 65% mean accuracy and LSTM achieved 66% accuracy. In conclusion, LSTM performed better than SVM as there was no major fluctuation in the accuracy.*

Key words—Stock market, Machine learning, SVM, LSTM

1. INTRODUCTION

Predicting the stock price is difficult as there are many uncertainties involved. There are two types of analysis which are used by investors before investing in a stock. First is the fundamental analysis which focuses on the intrinsic value of the stocks, it's performance in the market and industry. Second is the technical analysis which involves the study of the historical data of the stocks like the price and volume. The main aim of this paper is to develop stock price predictor which will predict the future prices with the help of the dataset which contains the historical stock prices of several companies. The companies chosen belong to the sector of technology. The data collected spans from 2010 till 2019. The prices and other values of the stocks were extracted from Yahoo Finance. The stock momentum depends on its past performance and the price of a given company will depend on how market is doing and how given sector is doing. So, keeping these things in mind, the index momentum, index volatility, sector momentum and stock momentum is also extracted to increase the accuracy for prediction. For the first algorithm, SVM is used with the Scikit Learn (sklearn) Library. The SVM achieves 65% accuracy with around 0.15 standard deviation. The second algorithm is LSTM as it is a time series predictor. It achieved 66% accuracy with a standard deviation of 1.3.

2. LITERATURE SURVEY

The paper written by Jigar Patel [1] predicts the price movement of the stock of the Indian stock market. It compares four different models, Na¨ıve-Bayes, random forest, Support Vector Machine and Artificial Neural Network. It implements two different methods to compute the movement. The first involves the manipulation of technical parameters like open, close, high and low prices. The second method is based on the trend formed by analysing the following parameters. Finally, the accuracy is evaluated of both the methods. The dataset contains the prices of ten years of NIFTY and SP index. After evaluating all the methods, the random forest algorithm performs better than other three models. In [2], David M. Q.

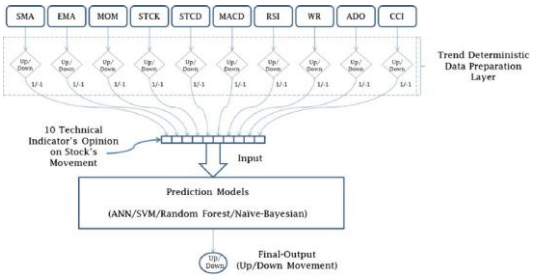

Fig. 1. Architecture diagram to predict the trend

Nelson, uses LSTM to predict the stock trend which depends on several factors like the historical prices, technical analysis etc. The Long Short term memory (LSTM) was used as it can differentiate between the past and present data by assigning different weights to them. So, it is able to memorize long sequence of data as compared to other models. The dataset was extracted from the Brazilian stock exchange. The result achieved was compared with the basic technical strategies. The random forest achieved highest accuracy of 86% while the Na¨ıve Bayes had the lowest performance with accuracy of 73%.

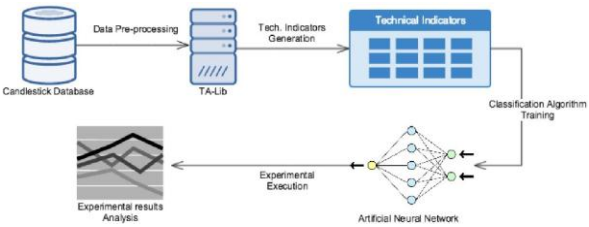

Fig. 2. Methodology

The paper by Jae Won Lee [3] uses the technique of re- inforcement learning to predict the price of the stock. The TD(0) reinforcement learning algorithm is used which learns from experience. The dataset is derived from the Korean stock market. The elements used to compute the learning were taken like state, action, reward and policy. Each price of the stock is mapped to the state of the reinforcement learning. Then, a neural network is used to retrieve the approximate value for the corresponding state. The results represent that this method can be used as an important indicator tool for stock trading. In [4],

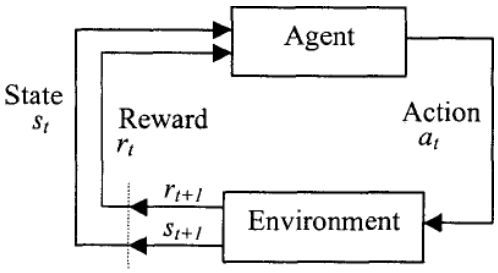

Fig. 3. Agent-environment interaction

Mehak Usmani uses machine learning techniques like Multi Layer Perceptron (MLP), Single Layer Perceptron (SLP), Radial Basis Function (RBF) and Support Vector Machine (SVM) to predict the market performance. The dataset was used from the historical prices of the Karachi Stock Exchange (KSE). The factors that are taken into consideration were the interest rate, social media, news, oil rates etc. Out of all algorithms, Multi Layer perceptron achieved the highest accuracy of 77%. It performed the best result for petrol price.

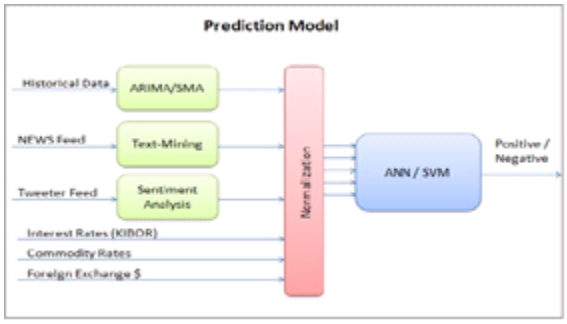

Fig. 4. Prediction model

The paper by Vatsal H. Shah [5] uses different types of algorithms like Linear Regression, SVM and Expert Weighting to determine the price pattern. The factors to predict the prices were Moving Average (MA), Exponential Moving Average (EMA) and the Rate of Change (ROC). The Support Vector Machine along with Boosting gave the best results. Linear Regression tend to give errors in the case of EMA. Technol-ogy stocks were giving better results compared to the other

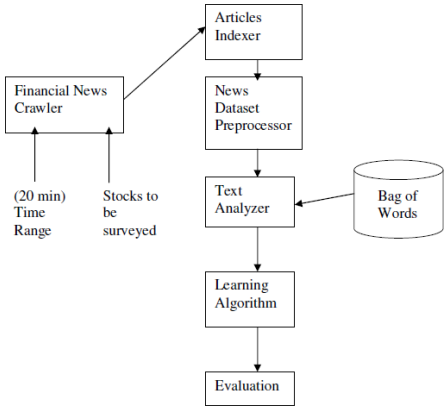

Fig. 5. Architecture Diagram

stocks like Energy. The accuracy was not high due to other uncontrollable factors like the election results and rumours. Figure 5 shows the architecture diagram.

3. IMPLEMENTATION

A. Creating the Dataset

Yahoo finance is a type of database which contains the prices of the stock. The stock prices were scraped from the site using python request library. The data that was extracted from the companies belonging to the technology sector. The data is ranges from January 2010 till December 2019. The stocks mentioned in the database are Oracle, Amazon, Intel, Lonovo, Google etc. The most important factors were extracted like the date, volume, close, high, low. Few other parameters taken into considerations are stock momentum, index volatility, index momentum and sector momentum.

Date	Open	High	Low	Close	Adj Close	Volume
########	30.49	30.64286	30.34	30.57286	27.40653	1.23E+08
########	30.65714	30.79857	30.46428	30.62571	27.45392	1.5E+08
########	30.62571	30.74714	30.10714	30.13857	27.01722	1.38E+08
########	30.25	30.28572	29.86429	30.08286	26.96728	1.19E+08
########	30.04286	30.28572	29.86572	30.28286	27.14657	1.12E+08
########	30.4	30.42857	29.77857	30.01572	26.90709	1.16E+08
########	29.88429	29.96714	29.48857	29.67429	26.60102	1.49E+08
########	29.69572	30.13286	29.15714	30.09286	26.97624	1.51E+08
########	30.01572	30.06571	29.86	29.91857	26.82001	1.08E+08
########	30.13286	30.22857	29.41	29.41857	26.37179	1.49E+08
########	29.76143	30.74143	29.60572	30.72	27.53843	1.83E+08
########	30.70143	30.79286	29.92857	30.24714	27.11456	1.53E+08
########	30.29714	30.47286	29.60143	29.72429	26.64584	1.52E+08
########	29.54	29.64286	28.16571	28.25	25.32424	2.2E+08
########	28.93	29.24286	28.59857	29.01	26.00554	2.66E+08
########	29.42143	30.53	28.94	29.42	26.37307	4.67E+08
########	29.55	30.08286	28.50429	29.69714	26.62152	4.31E+08
########	29.27572	29.35714	28.38571	28.47	25.52146	2.93E+08
########	28.72571	28.88571	27.17857	27.43714	24.59557	3.11E+08
########	27.48143	28	27.32857	27.81857	24.9375	1.87E+08
########	27.98714	28.04571	27.62571	27.98	25.08221	1.75E+08

Fig. 6. Dataset of Apple

B. Algorithms

1) Support Vector Machine (SVM): SVM is generally con- sidered to implement a classification approach but I can be used for regression problems too. A hyperplane is generated in multidimensional space to differentiate the classes. To minimize the error, the optimal hyperplane is constructed in an iterative manner. The main feature of SVM is to neatly segregate the corresponding data and the hyperplane with the maximum margin is opted between the support vectors. Basically, the algorithm is generated using a kernel. The kernel trick is a technique where the low dimensional input space is taken and it is converted into a higher dimensional space. The data was not linear separable, so RBF (Radial Basis Function) Kernel was used which to give better results for non-linear kernel. So, firstly SVM is applied to the train dataset and the values are predicted for the test dataset.

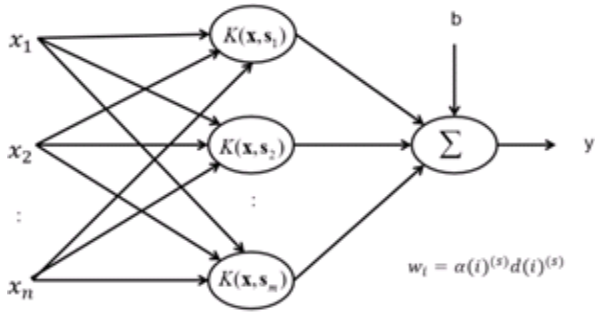

Fig. 7. SVM architecture

2) Long Short-Term Memory: The Long Short Term mem- ory is used to build a recurrent neural network. It contains four main parts : cell, input gate, output gate and forget gate. The cells play the role of memorizing the values. An activation is computed with the weighted sum by the three gates. The forgot gate decides which data should be remembered or thrown away. The value ranges between 0 and 1. The input gate plays the role of updating the cell state. The values which fall between 0 and 1 are not important and the values that has the value is considered important. The output gate makes the decision about the updated hidden state and this state is used for prediction.

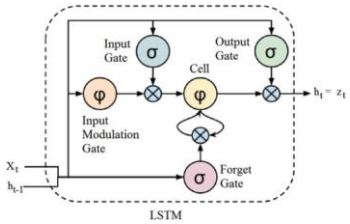

Fig. 8. LSTM architecture

3) Computing the accuracy: To get a better result, both the algorithms are implemented 30 times and the accuracy is generated for each time. Different training and testing data is used for each independent run of the algorithm. After the run time ends, the final mean accuracy is calculated along with the standard deviation and the total time taken for the algorithm to execute.

Algorithm	Mean accuracy	SD	Epoch
SVM	65.20	0.15	30
LSTM	66.83	1.3	30

Fig. 9. Comparison Table

4. RESULT

In the case for SVM, the mean accuracy for the 30 runs is 65.20% and the standard deviation is 0.15. It is noticed that the performance is consistent due to its nature.

```
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: 65.04774552833318
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: 65.62954531082O2
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 : 181.6788330078125
```

Fig. 10. Result of SVM

For the 30 runs for LSTM, it achieved a mean accuracy of 66.83% and a standard deviation of 1.36. Figure 11 displays the result of accuracy performed by LSTM algorithm.

5. CONCLUSION

Different machine learning techniques like Support Vector Machine and Long Short Term Memory to predict the prices of the stock. The results display that it is possible to predict the movement of the stock using the historical data with decent accuracy. By comparing the accuracy, it is evident that the LSTM algorithm performed better than SVM. To conclude, if all the factors which predict the price movement are taken into consideration and fed to neural network with proper filtering, the accuracy can be improved and can predict better movement of the prices of stocks.

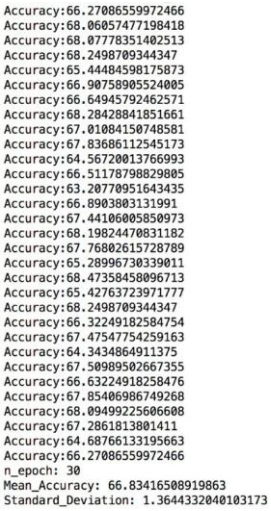

Fig. 11. Result of LSTM

REFERENCES

[1] Patel, J., Shah, S., Thakkar, P., Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert systems with applications, 42(1), 259-268.

[2] Nelson, D. M., Pereira, A. C., de Oliveira, R. A. (2017, May). Stock market's price movement prediction with LSTM neural networks. In 2017 International joint conference on neural networks (IJCNN) (pp. 1419-1426). IEEE.

[3] Lee, J. W. (2001, June). Stock price prediction using reinforcement learning. In ISIE 2001. 2001 IEEE International Symposium on Indus- trial Electronics Proceedings (Cat. No. 01TH8570) (Vol. 1, pp. 690-695). IEEE.

[4] Usmani, M., Adil, S. H., Raza, K., Ali, S. S. A. (2016, August). Stock market prediction using machine learning techniques. In 2016 3rd international conference on computer and information sciences (ICCOINS) (pp. 322-327). IEEE.

[5] Shah, V. H. (2007). Machine learning techniques for stock prediction. Foundations of Machine Learning— Spring, 1(1), 6-12.