COMPUTATIONAL INTELLIGENCE AND MACHINE LEARNING
e-ISSN: 2582-7464

Computational Intelligence and Machine Learning
Vol-2, Issue-1, April 2021

# Comparative Analysis of Stock Price Prediction by ANN and RF Model

**Lopamudra Hota[1], Prasant Kumar Dash[2]***

[1]Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India.
[2]Department of Computer Science and Engineering, C. V. Raman Global University, Bhubaneswar, India.
*Corresponding Author Email: [2] prasant.oitburla@gmail.com

***Abstract***

*The elementary goal of this paper is to predict the best model for estimation of stock market. Machine Learning is a blooming field in computer science that has contributed to many predictions and analysis-based algorithm in Financial and economical field. Some of the algorithms used for predictions are Random Forest (RF), Support vector machine (SVM), Long-Short Term Memory (LSTM), Artificial Neural Networks (ANN). Random Forest is an ensemble supervised learning algorithm for classification problems with high accuracy factor. ANN has matured to a great extend over the past years. With the advent of high-performance computing ANN has assumed tremendous significance and huge application potentials in recent years. The innovation of ANN technology mimics the large interconnections and networking that exists between the nerve cells to process complex task. The paper has presented ANN and RF model for stock price estimation based on historical data and computed the future price, with comparative result analysis of their performance. Further, a candlestick model is designed of the stock to show the variation in price of stock over a stipulated period of time.*

***Keywords***

*ANN, Candle-stick, CNN, Deep Learning, Random Forest, RNN, Support Vector Machine.*

## INTRODUCTION

The unpredictable dynamic nature of stock market has given rise to challenges for future analysis of the stock based on present value. The Efficient-Market Hypothesis (EMH) states that prices of asset reflect all the information available. Implicating that it is not possible to "beat the market" steadily based on risk adjustment, as stocks prices in market are sensitive to updated information in regular basis. Furthermore, the fluctuation in stock prices rely on economic, political, psychology and expectations of investors, price of commodities and many other factors. The arbitrariness and turbulences in market on every day basis led off challenges in future prediction of stocks. The accuracy in stock price and asset's future trend prediction aids in minimizing risk and maximizing profit. Various parameters are considered for stock price prediction based on statistical data computation and collection [1]. There are two aspect of predictions of asset based on past price and trends. The future prediction of stock gives a clear scenario of estimating economic status of the country. The growth of economy of a country is proportional to stock capitalization [2]. The dynamic, varied parameters, and non-linear nature of stock values generally weaken the performance analysis of statistical models for accurate prediction of future value [3].

One of the definitions of Machine Learning that states "Algorithms that parse data, learn from that data, and then apply what they've learned to make informed decisions". Machine learning models basically train machines to learn the training dataset and test the dataset trained for future prediction. Machine learning algorithms are significantly used for performance analysis based on case studies. It identifies the data pattern and validate information to compute it as reference for future prediction and analysis. Mostly stock market prediction is done by implementation of stochastic or random walk methodologies by estimation of future price based on successive previous price inspection and examination. Some of the common machine learning algorithms for prediction are Random Forest, Boosting and Bagging. The modernized trend in ML algorithms incorporated the Deep Learning (DL) concepts, based on non-linear topology is used in financial series of services for future trends and estimations [4].

Random Forest (RF) is one of the best models for prediction on tabular data. It takes the capabilities of multiple Decision Trees (DT) model (one of the base methods for prediction) in which the machine learns the tree having maximum influence and sets the weight values on feature points. A decision is made by branch traversal depending on various parameters. There is a problem of overfitting that arises in DT which increases the depth of the tree making it a complex structure. To overcome this problem Radom Forest is designed that is an ensemble supervised ML algorithm. The DT is created based on random features of datasets, and finally the RF is created based on the outcome of DT that is fetched maximum time.

DL is nothing but a subset of ML with variance in capabilities. DL uses a layered logical structure for estimation designed in accordance to human nervous system called Artificial Neural Network (ANN). As per the scenario of stock market, the future estimation of market is not only based on present or latest datasets but also have to deal with historical dataset for prediction accuracy of trained model. Artificial Neural Network (ANN), Recurrent Neural Network (RNN), Long-Short Term Memory (LSTM) are some of efficient DL models for economic and financial estimation

and statistical analysis [5]. Ensemble Learning model are reliable models used in today's scenario for performance and statistical prediction with minimization of overfitting issues (causes failure in accuracy of future prediction).

The tree-based models such as Decision Tree, Random Forest, Bagging, AdaBoost, XGBoost, Gradient Boosting are commonly used supervised learning ensemble estimator models for stock market prediction. ANN provides a flexible framework for computation including a high range of time series with good accuracy in future prediction. In view of this advantage of ANN we have modelled our proposal for stock market prediction based on ANN. Provided ANN has a less processing power consumption and efficient computational resource utilization, whereas Random Forest is considered best model for tabular data unlike ANN that can be implemented on audio, video and images as well.

## REVIEW OF LITERATURE

Isaac et. Al in [6] have proposed a novel "homogeneous" ensemble classifier implemented with Genetic Algorithm (GA) for SVM parameter optimization and feature extraction for prediction stock price of Ghana stock exchange (GSE) for ten days. Authors use Decision Tree, Random Forest and Neural Network model for prediction analysis and compared the accuracy of these models. It was proved that the proposal of thew author based on Genetic Algorithm concept provides better accuracy of nearly 93.7 percent compared to other algorithms. Similarly, in [7] authors have compared the performance analysis of Neural Network models based on multiple linear regression (MLR), Elman, Jordan, Radial Basis Function (RBF), and Multilayer Perceptron (MLP). This was tested on six most traded stocks of Brazilian Stock Exchange based on RSME, number of inputs and hidden layer. As per the result MLR being the simplest model of all showed good accuracy in prediction with less computational overhead.

The researchers have proposed many models and algorithm with Machine Learning capabilities for prediction of stock prices and financial domain in past as well as today. ML is found to be one of the most powerful algorithms for information validation and future data pattern prediction. The ensemble methods in ML outperforms many of the traditional methods in time series prediction [8]. Two of the popular algorithms for prediction problem are Boosting and Bagging. Financial data analysis is one of the most blooming topics. Computing techniques such as Neural Networks are designed for analysis of buy and sell on daily basis with less time consumption and accuracy. Many proposals have been proposed for substantial predictions [9-12].

Improved Bacterial Chemotaxis Optimization (IBCO) with ANN proposed by Zhang et.al [13] satisfies the prediction of stock price for short time of one day as well as long-time of fifteen days. Asadi et. Al [14] have proposed a Feed Forward Network along with Genetic Algorithm and Levenberg–Marquardt (LM) for learning model, basically their model is reliable dealing with the fluctuations of stock market with the capability of pre-processing data and selection of input variables. Jigar et al. [15] laid down efforts by implementing ML techniques incorporating hybrid combination of models for specifically Indian stock market index analysis and prediction. Similarly, Linear Regression model has also played an important role stock market prediction and analysing market behaviour [16].

Emioma et. al [17] proposed a model based on least square regression for prediction of future stock price handling the random changes. Some of the researchers have also used Support Vector Machine (SVM) which is a classifier that discriminate the datasets for stock market prediction [18]. Prediction of time-series has been done by methodologies based on Convolutional Neural Networks (CNN) [19], Similarly, in [20] stock market forecasting is done by using Artificial Neural Networks and Wavelet Transformations. In [21] authors have incorporated an Artificial Fish Swarm Optimization technique in neural network to achieve a more accurate model for stock prediction, taking the dataset of Shanghai Stock Exchange. Random Forest with LSTM has also been implemented to get one of the best prediction results for stock market in [22].

In [23], authors have proposed an LSTM model for stock market prediction with computation of Linear Regression. They have used K-NN classifier for classification of dataset, computed the moving average of the stock TITAN and NIFTY50. They have used K-Means clustering concept to provide a data-frame of the stock. Bollinger bands tool is used to measure volatility of stock to help brokers and investors to predict their value of stock in future.

## PROBLEM STATEMENT

The prediction of stock market is an essential factor to determine future value and movement of stock in financial sector. The more the accuracy in the share price prediction the more it will lead to profit making for investors as they get to do more accurate market analysis and study stock price pattern. This is basically done by utilizing and analysis of datasets by implementing on various models based on ML, DL, and other prediction and detection techniques. The fluctuations in the market mostly depends on the sell, buy and opinions of share-holders. The fluctuations are very rapid as the market and financial sector is fully dependent on political and social news and occasions which plays the trick for changes in stock price. As recently there has been swing of prices on day-to-day basis due to outburst COVID-19, vaccines, closing and opening of financial, economic and other sectors which drastically impact the country's economy has also a direct impact on stock market.

With a brief analysis of stock market, we have tried to predict the future stock market by implementing ANN and RF mechanism with pre-defined python libraries taking the datasets of 5-6 years from Yahoo Finance of TELA stock. The proposal has demonstrated the use of ANN and RF in prediction along with the implementation details and outcome that we got after the implementation in the

remaining section of the paper. Finally, the accuracy of algorithm for predicting the stock price is compared.

## METHODOLOGY

### ANN Model

Our proposal is based on designing multilayered ANN model for an accuracy efficient stock market prediction scheme. ANNs are basically single or multi-layer network structure in a fully connected fashion with input, output and hidden layers nodes. In ANN model the input in each layer nodes depends on the output of previous layer nodes i.e N depends on N-1.

As we say Artificial Neural Network, it strikes the mind that what is natural neural network? And that is the human brain which is a highly complex, nonlinear and does parallel computations with organised structural constituents "Neuron". The neuron is interconnected in a complex way between each other and one to another that gives visualization of network structure that is complex, non-linear and massively parallel. The working of ANN follows a node with weighted sum of inputs, then summed to a bias value, and passed through an activation function (non-linear function). The result that is output of this node becomes input of node in the next layer. It continues in a chain fashion and then the final output is retrieved after the result travelled the intermediate hidden layers. The increase in number of hidden layers tends to deepen the network structure [5]. Associated weights and biases train the network model.

Some of the capabilities of ANN are; exploitation of non-linearity to deal more efficiently with real-world problem which is distributed in nature, Input/Output Mapping (feed the input and wait for the desired output). For example, there may be a difference between the actual output and desired output, in that scenario we can take some free parameters (weights or input strength in ANN and Synaptic connection in biological term) to minimize the margin between the actual and desired output by finding a closest value. ANN has a facility of learning in which the system can learn about the desired output to get the output closest to the desired output value, this feature makes it different from traditional computational unit, that is called the Learning ability with adaptability to certain specific characteristics due to changes in the environment. Ability for fault tolerance is another most important capability of ANN, fault being directly proportional to performance degradation.

### *Perceptron (Computer Neuron)*

A perceptron algorithm is based on binary classifiers, that takes an input signal and provides an output signal after processing or computation, basically a single-layered neural network. The input layer resembles the dendrites of the neuron and output signal the axon. The input signals are assigned weights which are multiplied by input values, and the weighted sum of all inputs are stored in the neuron. The computation of the weights is done by gradient decent and back-propagation algorithms, these adjust the free parameters to minimize the loss/cost function. Back-propagation model are generally used in training of Feed-Forward Neural networks.

The ANN consists of hidden layers in multi-layer perceptron scenario. For the hidden or output nodes, nodes take the weighted sum of inputs, add to a bias, and then passes it through an activation function non-linear function. The result fetched at this node becomes input for another node in next layer. The process is iterated for all the nodes for determining the final output. And the network is trained by learning process of associated weights and biases of nodes.

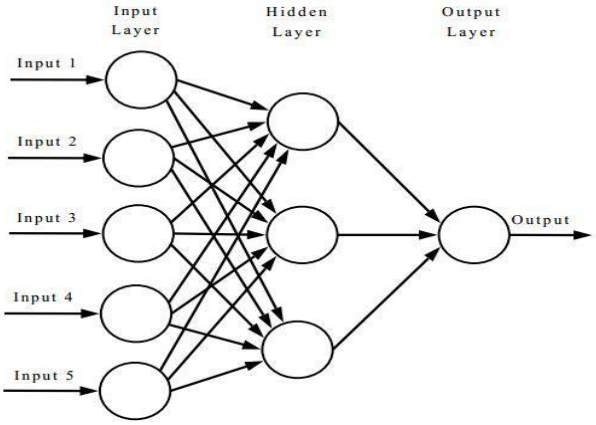


**Fig 1:** Schematic Structure of ANN

The equation below [6] show the relation between weights, biases and activation function for computation of the learning model.

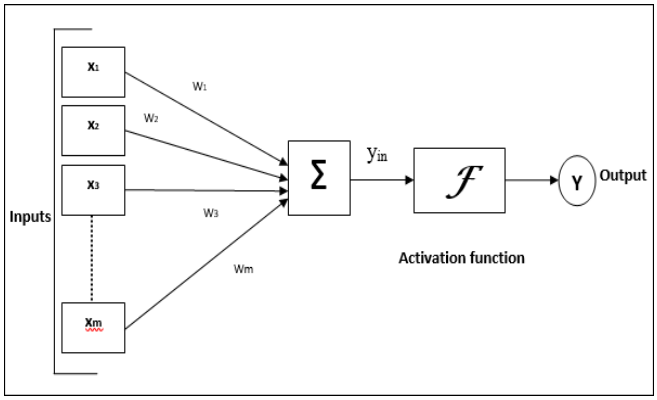


**Fig 2:** Working of ANN Model

It basically works same as the principle of human brain, take input values and iterate it over multiple processing steps to fetch desired result, the inputs in case of human brain can be smell, seeing, hearing, touch, and taste. The feelings and emotions are the two hidden layers through which processing is carried out that make us to take decisions (output). This gives a brief understanding although we know that computation in brain is much more complex and have many more hidden layers for processing tasks.

### Random Forest Model

It is basically a ML based supervised ensemble learning mechanism used for classification as well as Regression

problems. The procedure merges various classifiers to solve complex tree overfitting problem thereby improvising model performance by enhancing the prediction accuracy of datasets. Rather than considering the outcome of a single DT, it combines the outcome of all DTs and takes the outcome with majority votes to predict the end result. The accuracy of the RF is directly proportional to the number of tree structures in the forest. RF has higher accuracy, less time to train as compared to other algorithms even on larger datasets.

Here, the selection of feature is done by Bootstrap or Bagging mechanism. The training sets are created from the feature set of datasets by selecting random ones and a feature may be repeated in other training sets. This random selection procedure ensures less correlation and minimizes overfitting.

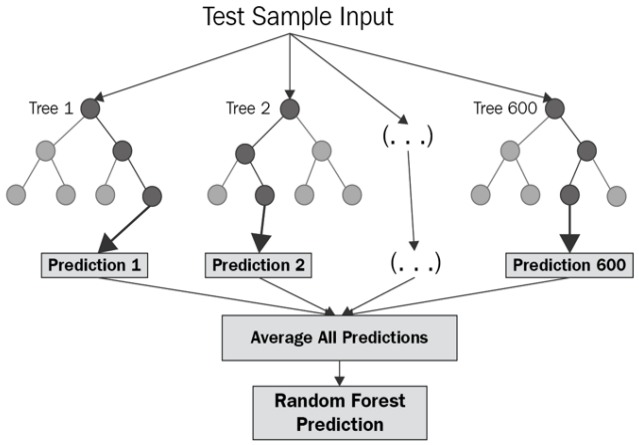


**Fig 3:** Schematic Structure of Random Forest

***Basic Structure of Random Forest***

The basic RF model contains the following working procedure to be implemented on the dataset.

1. Selection of K datapoints from training dataset.
2. Decision tree construction with the selected datapoints
3. Select N Decision Trees
4. Again process 1 and 2
5. For computation of new data-points, evaluate the predictions of DTs and assign datapoints to the ones that has maximum votes.

## PROPOSED WORK

Here, we have used the discussed ANN model as well RF model for prediction of price in stock market. In ANN, as per the diagram structure there are the hidden layer which depicts open/close price and their difference, volume. The model is trained based on weight applied to an activation function for getting a specified output. The final output is calculated by sum of output compute for each neuron.

Initially the ANN train itself for the given data set specifically the Open, High, Low, Close and Volume (OHLCV) consisting of variables; timestamp (epoch time), open price, close price, high price, low price and volume of stock (quantity of assets sold or purchased). The flat file data field consists of all historical data in .csv format. We have taken the input values as the open, close price to train the model and then compute the actual output Y, which is somewhat different from the predicted output Y'.

The weights of the variables are modified for each neuron by cost function minimization, that defines the cost of prediction making, that is gap between the actual output value and predicted output. Basically, cost function measures "how acceptable" a neural network is with respect to the training sample given and the output expected.

We have computed the cost function as the sum squared deviation between that of computed actual output and predicted output values. Initially the cost function is computed for the given datasets with a specified set of weights for each neuron. Neural networks learn according to the weights associated with the neuron which is updated after every forward passage of data through the neural network structure. The weights are adjusted to aid in reconciliation of the differences between actual value computed and predicted output for consecutive forward propagation. Question arises, how can we adjust these random weights assignment? As we have already discussed about the differences in actual and predicted outcomes, the error factor becomes an important consideration for computation of weights. At each neuron errors are computed and the reverted back to the neurons through the network by Back-Propagation for update of weight facilitated by forward pass. The procedure is repeated till we achieve a minimized cost factor.

In RF model, we fetch the dataset then create the input variables based on Open-Close and High-Low values and output variables by setting it to 1 if next day closing price is greater than today's and -1 otherwise, done randomly. The decision is made by dividing data into sets that are heterogeneous with others and homogeneous among them. This division is based on some criteria like Mean Squared Error (MSE) termed as Information Gain that predicts how our model accuracy improves as we split further.

The parameters taken for RF implementation are estimators, depth value of DT, bootstrap, and number of samples. The estimator signifies the number of DT incorporated. Bootstrap pick the random values from samples for maximized accuracy and visualization of result by building multiple DT models. RMSE is computed to estimate the best DT computed from the dataset. For each DT, fetch a subset of training dataset randomly and fit the DT. Repeat for all DTs created. As the subset are randomly chosen computer the error for each model and then by taking average of these model the information is merged to get the predicted or computed dataset.

**Design of Trading Model:**

We have implemented our model by the use of Neural Network and Random Forest methodologies in python for stock market future estimations and accurate prediction for trader's welfare. Python, use powerful libraries for building robust, efficient and reliable Trading Model.

We have used predefined python libraries like numpy and pandas for dataset computations. Talib for computation of Relative Strength Index (RSI) (value between 0 and 100) and

William %R Oscillator (value between 0 and -100) in ANN. Williams %R Oscillator computation has a significance in measuring closing price for a specified time period within the trading range. RSI computes the consistency of variation in prices over time, so high value of RSI signifies frequent increase in price than the declined rate over a time duration. RandomForestRegressor and bootstrap with quantrautil and RandomForestClassifier module for Random forest computation.

The data set has been imported from Yahoo Finance, and we have implemented our model for TESLA datasets to compute the stock return and planning return of a stock invested by the investor. These are computed by initially computing the return for the next day, for 'true' predicted value long leaps are taken whereas for 'false' value shorter leaps are taken. The rate of return is computed, if long leaps are present at the end of a day, squared at the end of the next day. After computation of return the cumulative return values are calculated, based on this value a graph is plotted to demonstrate how our planning return performs against the stock return. We have also implemented a candlestick model that estimate the possible price movement based on past values for a specified time period of one year form MAR 2020 to MAR 2021.

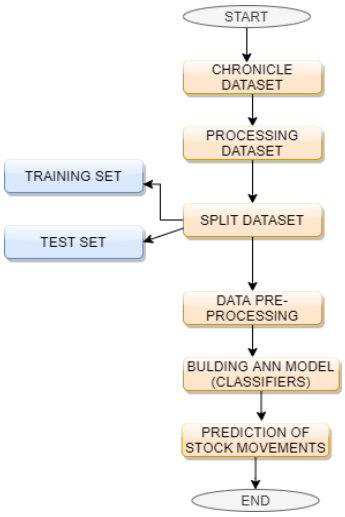


**Fig 4:** Flow Chart for ANN

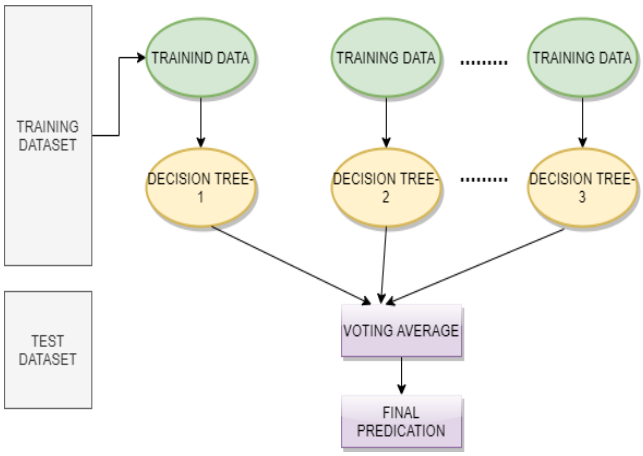


**Fig 5:** Flow Chart for RF

## EXPERIMENTAL RESULTS

The data has been extracted from Yahoo Finance of TSLA stock with seven attributes stating the high-low, close-open values of stock. The dataset was implemented for ANN as well as RF for prediction of stock price and analysis of accuracy based on actual and predicted value. The adjusted close price of the specified duration is computed and graph is plotted. A candlestick model of the stock is also implemented to record the fluctuation in the stock price in a specified duration (MAR 2020 to MAR 2021).

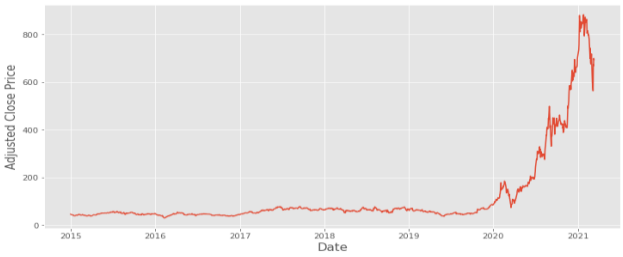


**Fig 6:** Graph for Adjusted Close Price (RF model)


**Fig 7:** Candle-Stick Chart

In the RF model, the confusion matrix was built to examine various possible outcome of prediction to reach at an accurate prediction. It predicts the correctness and in-correctness of the classifier's prediction giving the accuracy of the model. Graph is designed for real and predicted price over time vs price.

Along with the same dataset, the ANN is modelled and stock prediction is made with computation of buy/sell profit, cumulative profit and predict the future price after 15days with accuracy of 82.8% and MSE of 5.89. Graph is designed for actual and predicted over year vs price.

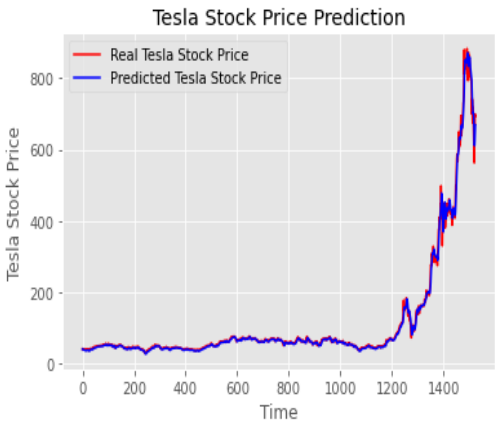


**Fig 8:** Graph Stock Price vs Time (RF)

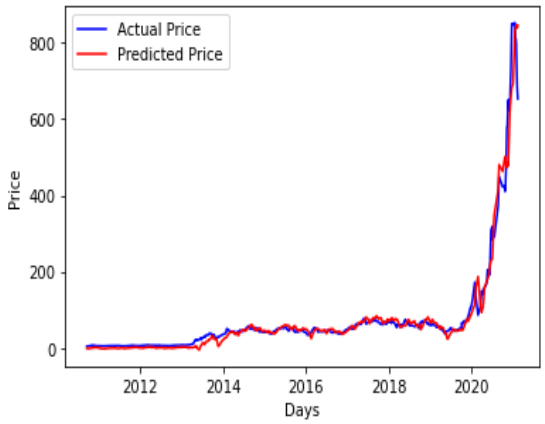


**Fig 9:** Graph Price vs Days (ANN)

By measure of various parameter and accuracy computation it is found that although Random Forest is a better model for financial and stock prediction based on historical dataset still, the advanced Artificial Neural Network model shows better result with RSME between 4 to 7. Although every ML algorithm has its own way and perform as they are designed; still Neural Network models are marginally better in performance that traditional ML models whether it is textual data for prediction or audio/visual data for recognition.

**Table 1:** Comparison of Algorithm

| Accuracy of Algorithm | |
|---|---|
| Random Forest (RF) | 79.6 |
| Artificial Neural Network (ANN) | 82.8 |

## CONCLUSION AND FUTURE SCOPE

In the dilemma of choosing the best model for stock market or any financial prediction, the decision should be made based on data and parameters for ease of computation. Although Neural Network performs better in all type of data values it can be specifically used in audio, video and pictorial dataset. Contrast to this model like Random Forest, Decision Tree, Linear Regression and Support Vector Machine can do a good job with tabular data due to their simplicity in computation and implementation thereby minimizing the overhead to work with complex structure of Neural Network. Still there is always an opportunity to switch to neural network to achieving better performance and accurate prediction.

For future work, we would like to implement our model on larger datasets and also implement the datasets using Recurrent Neural Network (RNN) and Convolution Neural Network (CNN) with LSTM and compare with result of ANN. The use of Deep Reinforcement Learning (DRL) with Q-Learning to get an optimized result and able to predict financial and stock market with better accuracy can be done in a simpler way with lesser historical data. The DRL models an intelligent system to handle fluctuations and test updated actions and approaches; keeping a check on failure and success rates in a continuous basis.

## REFERENCES

[1] J. Lehoczky, M. Schervish, Overview and History of Statistics for Equity Markets. Annu. Rev. Stat. Its Appl. 2018, 5, 265–288.
[2] A. Aali-Bujari, F. Venegas-Martínez, G. Pérez-Lechuga, Impact of the stock market capitalization and the banking spread in growth and development in Latin American: A panel data estimation with System GMM. Contaduría y Administración 2017, 62, 1427–1441.
[3] M.P. Naeini, H. Taremian, H.B. Hashemi, Stock market value prediction using neural networks. In Proceedings of the 2010 international conference on computer information systems and industrial management applications (CISIM), Krakow, Poland, 8–10 October 2010; pp. 132–136.
[4] R.C. Cavalcante, R.C. Brasileiro, V.L. Souza, J.P. Nobrega, A.L. Oliveira, Computational intelligence and financial markets: A survey and future directions. Expert Syst. Appl. 2016, 55, 194–211.
[5] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, Deep learning for Stock Market Prediction. arXiv 2020, arXiv:2004.01497.
[6] Isaac Kofi Nti, Adebayo Felix Adekoya, and Benjamin Asubam Weyori, Efficient Stock-Market Prediction Using Ensemble Support Vector Machine, Open Computer Science, 2020; 10:153–163, https://doi.org/10.1515/comp-2020-0199.
[7] S. Teixeira, Z. Pauli, M. Kleina, and W. H. Bonat, Comparing Artifcial Neural Network Architectures for Brazilian Stock Market Prediction, Annals of Data Science, 2020, 7(4):613–628 https://doi.org/10.1007/s40745-020-00305-w.
[8] S. Amari, The Handbook of Brain Theory and Neural Networks; MIT press: Cambridge, MA, USA, 2003.
[9] M. Ballings, et al., Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications, 2015. 42(20): p. 7046-7056.
[10] E. Hadavandi, H. Shavandi, A. Ghanbari, Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. Knowledge-Based Systems, 2010. 23(8): p. 800-808. 20.
[11] Y.S. Lee, and L.I. Tong, Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. Knowledge-Based Systems, 2011. 24(1): p. 66-72.21.
[12] M.F. Zarandi, E. Hadavandi, and I. Turksen, A hybrid fuzzy intelligent agent-based system for stock price prediction. International Journal of Intelligent Systems, 2012. 27(11): p. 947-969.
[13] Y. Zhang and L. Wu, Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. Expert systems with applications, 2009. 36(5): p. 8849-8854.
[14] S. Asadi, et al., Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. Knowledge-Based Systems, 2012. 35: p. 245-258.
[15] J. Patel, et al., Predicting stock market index using fusion of machine learning techniques. Expert Systems with Applications, 2015. 42(4): p. 2162-2172.
[16] S. Abdulsalam, K.S. Adewole, and R. Jimoh, Stock trend prediction using regression analysis–a data mining approach. 2011.

[17] C. C. Emioma and S. O. Edeki, Stock price prediction using machine learning on least-squares linear regression basis, International Conference on Recent Trends in Applied Research (ICoRTAR) 2020, doi:10.1088/1742-6596/1734/1/012058.

[18] Z. Guo, H. Wang, Q. Liu, and J. Yang, A feature fusion based forecasting model for financial time series. PloS one, 9(6):e101113, 2014.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Image-net classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[20] W. Bao, J. Yue, and Y. Rao, A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PloS one, 12(7):e0180944, 2017.

[21] W. Shen, et al., Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm. Knowledge-Based Systems, 2011. 24(3): p. 378-385.

[22] S. Han, Stock Prediction with Random Forests and Long Short-term Memory, 2019, https://lib.dr.iastate.edu/creativecomponents.

[23] A. Singh , P. Gupta, and N. Thakur, An Empirical Research and Comprehensive Analysis of Stock Market Prediction using Machine Learning and Deep Learning techniques, 2021, IOP Conf. Series: Materials Science and Engineering 1022 (2021) 012098, doi:10.1088/1757-899X/1022/1/012098.