STOCK MARKET PREDICTION SYSTEM USING ML

Dr. Suresh Subramanian[1], Sai Saketh Reddy[2], S.Rohith Reddy[3],S.Anil[4], T.Sai Teja[5]
[1]Professor, Dept of ECE Sri Indu Institute of Engineering & Technology, Hyderabad
[2-5] Student, Dept of ECE Sri Indu Institute of Engineering & Technology, Hyderabad.

Abstract: The stock market is highly volatile and complex in nature. Technical analysts often apply Technical Analysis (TA) to historical price data, which is an exhaustive task and might produce incorrect predictions. Machine learning coupled with fundamental and/or Technical Analysis also yields satisfactory results for stock market prediction. In this work, the effort is made to predict the price and price trend of stocks by applying optimal Long Short Term Memory (O-LSTM) deep learning and adaptive Stock Technical Indicators (STIs). We also evaluated the model for taking a buy-sell decision at the end of the day. To optimize the deep learning task we utilized the concept of Correlation-Tensor built with appropriate STIs. The tensor with adaptive indicators is passed to the model for better more and accurate prediction. The results are analyzed using popular metrics and compared with two benchmark ML classifiers and a recent classifier based on deep learning. The mean prediction accuracy achieved using the proposed model is 59.25%, over the number of stocks, which is much higher than benchmark approaches.

Keywords:- Long Short-term Memory, Support Vector Machine, Unified Model Labelling ,Jupyter Notebook,python, Scikit Learn.

I. INTRODUCTION

Prediction and analysis of the stock market are some of the most complicated tasks to do. There are several reasons for this, such as the market volatility and so many other dependent and independent factors for deciding the value of a particular stock in the market. These factors make it very difficult for any stock market analyst to predict the rise and fall with high accuracy degrees.

However, with the advent of Machine Learning and its robustalgorithms, the latest market analysis and Stock Market Prediction developments have started incorporating such techniques in understanding the stock market data. In short, Machine Learning Algorithms are being used widely by many organizations in analyzing and predicting stock values. This article shall go through a simple Implementation of analyzing and predicting a Popular Worldwide Online Retail. Storestock values using several Machine Learning Algorithms in Python. In short, Machine Learning Algorithms are being used widely by many organizations in analyzing and predicting stock values.

Stock Price Prediction using machine learning helps you discover the future value of company stock and other financial assets traded on an exchange. The entire idea of predicting stock prices is to gain significant profits. Predicting how the stock market will perform is a hard task to do. There are other factors involved in the prediction, such as physical and psychological factors, rational and irrational behavior, and so on. All these factors combine to make share prices dynamic and volatile. This makes it very difficult to predict stock prices with high accuracy.

The real commitments of this paper are:

- A pseudo comprehensive test technique to recognize all MSAFs in ROBDD based circuits which has immaterial test era exertion.
- All irredundant different flaws of the circuit under test are demonstrated testable.

II. LITERATURE REVIEW

STIs are statistical calculations based on the price, volume, or significance of a share, security, or contract. These do not depend on the fundamentals of a business, like earnings, revenue, or profit margins. Actives tock traders and technical analysts commonly use STIs to analyze short-term and long-term price movements and to identify entry and exit points.

Technical indicators can be useful while predicting the future prices of assets so they can be integrated into automated trading systems. There are two basic types of technical indicators: Overlays and Oscillators. In this work, we use SMA as it is one of the most widely used STIs. It filters out the noise which occurs due to random price variations and

ISSN : 0972-7175

helps to smooth out the price. It is said a trend followingindicator or simply lagging as it depends on past prices.

ALGORITHMS USED:

ARTIFICIAL NEURAL NETWORK(ANN):

Artificial Neural Network Neural networks are one the information processing system that is designed to mimic the way the human brain works in resolving a problemwith the learning process through its synaptic weight change. An artificial neural network is able to identify activities basedon past data. The data will be studied by an artificial neural network that has the capability for the ability to give a decision on the data that has not been studied. ANN is a parallel distributed processor that is very large with a tendency to store knowledge that is experienced and make it ready for use. ANN resembles a human brain in two ways, namely: a network of knowledge gained through the learning process; the strength of the relationship between nerve cells (neurons) known as synaptic weights are used to store knowledge.

Artificial neural networks (**ANNs**), usually simply called **neural networks** (**NNs**), are computing systems inspired by the biological neural networks that constitute animal brains. An ANN is based on a collection ofconnected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal to other neurons. An artificial neuron receives a signal then processes it and can signal neurons connected to it. The "signal" at a connection is a real number,and the output of each

output layer), possibly after traversing the layers multipletimes.

An artificial neural network is able to identify activities based on past data. The data will be studied by an artificial neural network that has the capability for the ability to give a decision on the data that has not been studied. ANN is a parallel distributed processor that is very large with a tendency to store knowledge that is experienced and make it ready for use. ANN resembles a human brain in two ways, namely: a network of knowledge gained through the learning process; the strength of the relationship between nerve cells (neurons) known as synaptic weights are used to store knowledge.

neuron is computed by some non-linear function of the sum of its inputs. The connections are called *edges*. Neurons and edges typically have a *weight* that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Neurons may have a threshold such that a signal is sent only if the aggregate signal crosses that threshold. Typically, neurons are aggregated intolayers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer) to the last layer (the output layer), possibly after traversing the layers multiple times. Neurons and edges typically have a *weight* that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Neurons may have a threshold such that a signal is sent only if the aggregate signal crosses that threshold. Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer) to the last layer (the

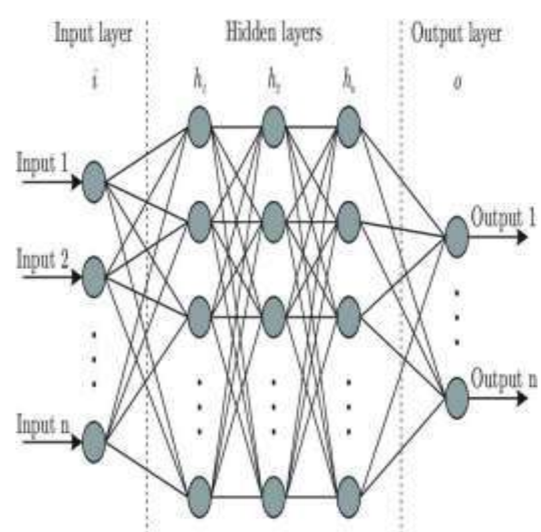

Figure2.1:Operation in Artificial Neural Network

Long Short-Term Memory

To develop a Machine Learning model to predict the stock prices of Microsoft Corporation, we will be using the technique of Long Short-Term Memory (LSTM). They are used to make small modifications to the information by multiplications and additions. By definition, long-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in deep learning.

Unlike standard feed-forward neural networks, LSTM has feedback connections. It can process single data points (such as images) and entire data sequences (such as speech or video).To understand the concept behind LSTM, let us take a simple example of an online customer review of a Mobile Phone.

Suppose we want to buy the Mobile Phone, we usually refer to the net reviews by certified users. Depending on their thinking and inputs, we decide whether the mobile is good or bad and then buy it. As we go on reading the reviews, we look for keywords such as “amazing”, “good camera”, “best battery backup”, and many other terms related to a mobile phone.

We tend to ignore the common words in English such as “it”, “gave”, “this”, etc. Thus, when we decide whether to buy the mobile phone or not, we only remember these keywords defined above. Most probably, we forget the other words.

This is the same way in which the Long short- term Memory Algorithm works. It only remembers the relevant information and uses it to make predictions ignoring the non-relevant data. In this way, we have to build an LSTM model that essentially recognises only the essential data about that stock and leaves out its outliers.

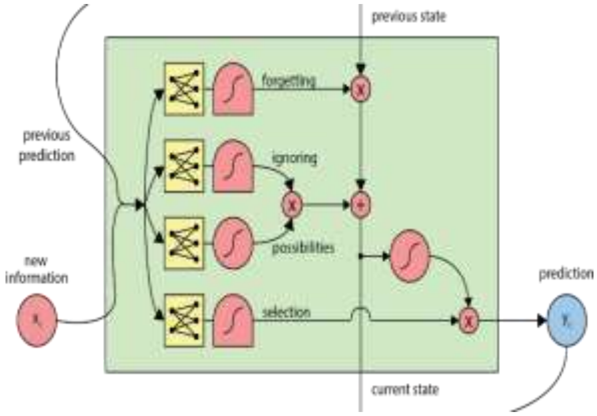

Figure 2.2: Long Short-Term Memory

III. EXISTING SYSTEM

To enhance the predictability of the daily stock price trends, Yuzheng Zhai et al. presented a system based on the SVM algorithm that combines the technical indicators and related news releases. For each trading day, seven technical indicators are computed from the prices in the past five days. Two groups of news releases are used. Two class categories, indicating the higher or equal price and lower price than the close price, are taken up for indicating the next day's price movement. The system achieved higher accuracy than achieved using single-source i.e. news or technical indicators. A Support Vector Machine (SVM) is a discriminative classifier that formally defined by the separating hyperplane. In other words, the given labeled training data (supervised learning), the algorithm outputs the optimal hyperplane which categorizes new examples. In the two-dimensional space, this hyperplane is a line dividing a plane into two parts wherein each class lay on either side. Support Vector Machine (SVM) is considered to be as one of the most suitable algorithms available for the time series prediction. The supervised algorithm can be used in both, regression and classification. The SVM involves in plotting of data as a point in the space of n dimensions.

IV. PROPOSED SYSTEM

The Deep Learning model characterizes numerical or categorical features as numbers. The indicator values are represented as simple vectors. The method has two primary drawbacks:

The vectors often acquire a size that is equivalent to the size and number of indicators overwhelming the size of memory. Such large dimension vectors make computation inconvenient.

1. Also, the representation makes processing and context semantic analysis cumbersome. A Deep Neural Network (DNN) operates on numbers where every neuron performs addition and multiplication operations on inputs and weights. The proposed model sequences are like stock data. The RNN has only a standard tanh layer in each repeating module while LSTM has four layers.

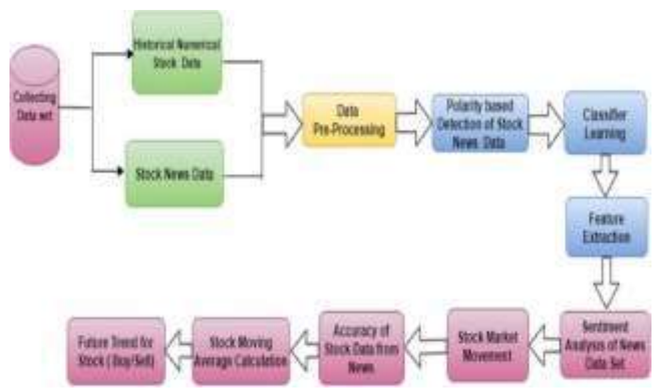

Figure 1.4: Proposed System for the future technology

V. IMPLEMENTATION

Collecting Dataset:- The foremost thing in order to implement the project is that we need to have the Dataset of the different stocks of a company. The data set includes Open price, Close price, Low Revenue, High Revenue, and the Volume of the Company's sales. It then divides the input data into two groups, one is called the Historical Numerical Stock data which provides us the past stock information, and the other is Stock News data which tend to give us the present stock values of the company. The stock market is a very dynamic market where nothing is as stable as a rock but as the technology is upgrading there are many ways and methods one can try to learn this dynamic change and be prepared accordingly.

Data Pre-Processing:-

In the Data Pre-Processing block, we import python libraries like NumPy, pandas, matplotlib, data line, and seaborn. NumPy library is used for the mathematical calculations of the data set, pandas library is responsible for data analysis and manipulation, matplotlib is the special library that is responsible for the data visualization, and the data line is used for representing labelling at the time of visualization, and seaborn is the library responsible for the output data representation.

Polarity-based detection of Stock News Data:-

As mentioned, the Stock News Data is the current stock of the company. The next step in the implementation involves detecting the stock prices of the company using present data. This is because in the furthermore steps we implement the detection of past data too and then we try to predict the futurestock price using both detections.

Classifier Learning:-

The Machine Learning techniques start here. This step involves the Supervised Learning method which involves the Classification of the data set according to their Open and Close prices, Low and High Stocks, etc.

Stock Market Movement:-

This block of operation includes the live movement of the stock in a company. This enhances our model to predict the day-wise stock price or weekly based stock price or a yearly based prediction. As our main motto is to predict the future

Accuracy of Stock Price Prediction:-

The results are analyzed using popular metrics and compared with two benchmark ML classifiers and a recent classifier based on deep learning. The mean prediction accuracy achieved using the proposed model is 59.25%, over the number of stocks, which is much higher than benchmark approaches.

Average Stock Volume Calculation:-

The Stock Volume represents the Average number of sales in a day or in a week or in a month or in a year. So, as per our requirement, for predicting the future stock trend we generally require the past stock data, present stock, and the volume of stock data. After all the requirements are satisfied then our desired machine learning model will be predicting the future stock trend.

VI. RESULTS

Fig(IV): Final Stock Prediction

VII. CONCLUSION

We inspected the efficiency of applying technical analysis to the stock prices. We analyzed whether investors could manage more profits than suggested by the recent research of Pang X. et al. (2018). We demonstrated the concept of deriving adaptive STIs and passing them as correlation tensors. The tensor is then supplied to the model. Finally, the model presents a well-organized approach that aids traders for long/short as well as daily trading and to earn profit. Optimal LSTM presents decision-based indicators (Price-rise (1) or Price-fall (0)) as well as trend-based analysis. The proposed Optimal Deep Learning Approach (Optimal LSTM) is a market independent approach as we are discovering the potential indicators existing in the data and applying LSTM dynamics of deep learning rather than fixing data or models. This work opens several research confronts to get more insights on stock trends forecasting. In the future, the research work can be extended by applying more STIs and can be evaluated against several ML and deep learning approaches. The proposed model can be further evaluated and optimized for stock indices. The proposed deep learning algorithm can also be further enhanced to optimize performance. To support investors, the proposed model can be further integrated into an automated system for trading specific stocks.

The proposed model can be further evaluated and optimized for stock indices. The proposed deep learning algorithm can also be further enhanced to optimize performance. We inspected the efficiency of applying technical analysis to the stock prices.

Future Trend Prediction:-

Finally, using the Regression model and scaler transformations we predict the future stock price of a company. As mentioned in the abstract, the Regression model in machine learning is responsible for future predictions, so in sk learn(Machine learning library in python) we have regression commands as regressor.predict(X_test). There are different commands existing in python and Jupyter notebook to implement our project and predict the future stock.

FUTURE SCOPE:

It is clearly depicted from the results that the application of Optimal LSTM along with correlation tensor of adaptive STIs improves prediction performance. The highest accuracy and mean accuracy achieved are 65.64% and 59.25% respectively, which is much higher than SVM, LR, and a deep learning approach (ELSTM) [21]. The proposed framework with Optimal LSTM is capable of exploring the correlative STIs. As an outcome, the model predicts the long and short-term market trends of any stock. The time period in which the predicted trends almost follow the actual trend is considered better for future investments.

However, the uncertain trends justify that there might be risk involved with particular stock for particular time periods. The outcomes of the model demonstrate that the investor will definitely fetch more profit by taking the efficient decision of buying and selling the stock. The model explicitly achieved uppermost prediction accuracy without over-fitting. The number of epochs is chosen to avoid over- fitting. The mean accuracy of the Classifier achieved is 59.25%, over a number of datasets. The MSE is least across all classifiers including Multi-Layer Perceptron. Further experiments reveal that STIs such as MA and MACD are highly correlated with close prices. The model built is market independent and is implemented in Python using powerful ML and NN libraries: Google - TensorFlow and Keras. Performance evaluation and comparison of has shown the pre-eminence of the proposed work.

VIII. REFERENCES

1. Sepp Hochreiter, Jürgen Schmidhuber, Long Short-Term Memory, Journal Neural Computation, Volume 9 Issue 8, November 15, 1997, Pages 1735-1780

2. Stelios A. Mitilineos; Panayiotis G. Artikis, “Forecasting of future stock prices using neural networks and genetic algorithms”, International Journal of Decision Sciences, Risk and Management, 2017 Vol.7 No.1/2, pp.2 – 25.

3. F. B. Oriani and G. P. Coelho, "Evaluating the impact of technical indicators on stock forecasting," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, 2016, pp. 1-8.

4. G. Zhang, X. Zhang, and H. Feng, “Forecasting financial time series using a methodology based on autoregressive integrated moving average and taylor expansion,” Expert Systems, vol. 33, no. 5, pp. 501–516, 2016.

5. M. Valipour, M. E. Banihabib, and S. M. R. Behbahani, “Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir,” Journal of Hydrology, vol. 476, pp. 433–441, jan 2013.

6. Sundaravelpandian Singaravel, Johan Suykens, Philipp Geyer, “Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction”, Advanced Engineering Informatics, Volume 38, 2018, Pages 81-90.

7. V.N. Gudivada, Chapter 1 - Cognitive Computing: Concepts, Architectures, Systems, and Applications, Editor(s): Venkat N. Gudivada, Vijay V. Raghavan, Venu Govindaraju, C.R. Rao, Handbook of Statistics, Elsevier, Volume 35, 2016, Pages 3-38.

8. Edward Gately, Neural Networks for Financial Forecasting, John Wiley & Sons, Inc., New York, NY, 1995

9. F. A. de Oliveira, C. N. Nobre, and L. E. Z´arate, "Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index – case study of PETR4, Petrobras, Brazil,"

10. Expert Systems with Applications, vol. 40, no. 18, pp. 7596– 7606, Dec 2013.

11. Zhai Y., Hsu A., Halgamuge S.K. (2007) Combining News and Technical Indicators in Daily Stock Price Trends Prediction. In: Liu D., Fei S., Hou Z., Zhang H., Sun C. (eds) Advances in Neural Networks – ISNN 2007. ISNN 2007. Lecture Notes in Computer Science, vol 4493. Springer, Berlin, Heidelberg

Zhang G-Y (2016) Application of improved convolution neural network in financial forecasting. Zhengzhou University, Henan.