Stock Market Prediction Using Machine Learning Techniques

Pooja Iyer [1], Adarsh Badyal [2], Anusree Mohanan[3], Parag Iyyani[4], Krishnendu Nair[5]
[1, 2, 3, 4] Dept of Computer Science
[5]Professor, Dept of Computer Science
[1, 2, 3, 4, 5] Pillai College of Engineering Panvel, Navi Mumbai

Abstract- *In the finance world stock trading is one of the most important activities. Stock market is a place where shares of public listed companies are traded. Stock exchange facilitates stock brokers to trade company stocks and other securities. Though the stock market fluctuates randomly from day to day, experienced traders know that much of the stock market fluctuations are not random. In recent years, increasing prominence of machine learning in various industries have enlightened many traders to apply machine learning techniques to the field, and some of them have produced quite promising results. The techniques of machine learning and artificial intelligence offer significant benefits to financial decision makers in terms of new approaches to modeling and forecasting from data. Ours is an effort to use this technology to predict the current trends in the financial market that helps in better investment and usage of money.*

Keywords- machine learning, artificial intelligence, stock market

I. INTRODUCTION

The stock market is basically an aggregation of various buyers and sellers of stock. A stock in general represents ownership claims on business by a particular individual or a group of people. The attempt to determine the future value of the stock market is known as a stock market prediction. The prediction is expected to be robust, accurate and efficient. Machine learning, an application of Artificial Intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed can be used to tackle these problems head on.

Traditional methods of prediction in machine learning use algorithms like Backward Propagation, also known as Backpropagation errors. Lately, many researchers are using more of ensemble learning techniques. The datasets of the stock market prediction model include details like the closing price opening price, the data and various other variables that are needed to predict the object variable which is the price in a given day. The previous model used traditional methods of prediction like regression and support vector machines (SVM). Stock market prediction outperforms when it is treated as a regression problem but performs well when treated as a classification. The aim is to design a model that gains from the market information utilizing machine learning strategies and gauge the future patterns in stock value development. The LSTM can be used for classification. It has been observed that LSTMs are more used in prediction based problem like ours.

II. PROBLEM DEFINITION

The trading process has evolved massively, to a state where traders employ sophisticated parameters and combinations of factors to come up with a decision. From social sentiment scores, through technical indicators, to fundamental information — investing today is more complicated than ever. Machine learning has the potential to ease the whole process by analyzing large chunks of data, spotting significant patterns and generating a single output that navigates traders towards a particular decision based on predicted asset prices [11]. Supervised machine learning algorithms are used to build the models. As part of the daily prediction model, historical prices are combined with sentiments. Up to 70% of accuracy is observed using supervised machine learning algorithms on daily prediction model. Monthly prediction model tries to evaluate whether there is any similarity between any two months trend. Evaluation proves that trend of one month is least correlated with the trend of another month. The various algorithms used for forecasting can be categorized into linear (AR, MA, ARIMA, ARMA) and non-linear models (ARCH, GARCH, Neural Network).In this paper, we are using four types of deep learning architectures i.e Multilayer Perceptron (MLP), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Convolutional Neural Network(CNN) for predicting the stock price of a company. It has been observed that CNN is outperforming the other models.

III. LITERATURE SURVEY

During Literature survey we collected some of the information about stock market prediction mechanisms currently being in use.

1. Machine learning in finance: A topic modeling approach

In this applied Probabilistic Technique on the corpus dataset is used to find the suitable data and processing. Latent Dirichlet Allocation is used for topic modelling. Genetic algorithm is used on the dataset to obtain sustainable results [1].

2. A Machine Learning Framework for Stock Selection

A hybrid approach for AI domain technique for dataset construction purposes. It explores a new hybrid approach for solving the problem of finding the solutions based on user inputs. The proposed technique provides improvements in addressing two major challenges of this domain system. The evaluation of the system shows superiority of the solution compared to stand-alone regression techniques. It shows that the hybrid approach improves the various problems [3].

3. Stock Market Prediction using Machine Learning

This paper explains the prediction of a stock using Machine Learning. The technical and fundamental or the time series analysis is used by most of the stockbrokers while making the stock predictions. The programming language is used to predict the stock market using machine learning is Python. In this paper author propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context, this study uses a machine learning technique called Support Vector Machine (SVM) to predict stock prices for the large and small capitalizations and in the three different markets, employing prices with both daily and up-to-the-minute frequencies [4].

4. NSE Stock Market Prediction Using Deep Learning Models

LSTM,CNN,RNN is used for identification of patterns in the stock markets purpose. It presents the state of the Art and discusses the main issues related to this domain problem. It discusses various techniques starting from simple techniques for representing data according to the day-wise closing price of each stock ,the output of CNN is denormalized for acquiring original predicted values and output of testing is also denormalized for accuracy [2].

5. Stock Market Prediction Using ANN

It introduces the flow of the domain system and the specific realization procedures of data input, data preprocessing, and other related steps. Results show that this technique in this domain has shown great performance. It summarizes important techniques in machine learning which are relevant to stock prediction [7]. Various features such as stochastic indicator, moving averages, RSI are extracted. It highlights the use of prediction technique: ANN with back propagation. The training is done by phase forward propagation in neural network and Feed forward propagation is used to obtain results.

IV. EXISTING SYSTEM

Methodologies:

1. *SVM Algorithm*

SVM classifier is a type of discriminative classifier. The SVM uses supervised learning i.e. a labeled training data. The output are hyperplanes which categorizes the new dataset. They are supervised learning models that use associated learning algorithm for classification as well as regression. When it comes to trading, if you are using daily frequency data, then it is very likely that your data set is extremely limited, probably a few thousand data points. Let us say that you have created a trading strategy using a decision tree to extract a high probability rule from the past data. Now you want to understand how this rule would behave on unseen or test data. Creating an SVC to predict when the rule would be a success or failure would help you eliminate the bad trades.

2. *Logistic Regression*

A logistic regression model helps us to fit a model using binary behavior and forecast market direction. Logistic regression is one of the probabilistic models which assigns a probability to each event. We are going to use the quantmod package.

Characteristics used as input for this method include:

(1) Feature Index Variables are easy to both understand for the private investor and obtain from daily stock trading information.
(2) the prediction procedure includes unique and crucial operation of selecting optimizing prediction parameters. (3) Significant time-effectiveness and strong purposefulness

enable users predict stock price trend of next month just through considering current monthly financial data instead of needing a long term procedure of analyzing and collecting financial data.

3. Random Forest

Random forests are based on ensemble learning techniques. Ensemble, simply means a group or a collection, which in this case, is a collection of decision trees, together called as random forest. The accuracy of ensemble models is better than the accuracy of individual models due to the fact that it compiles the results from the individual models and provides a final outcome. Features are selected randomly using a method known as bootstrap aggregating or bagging. From the set of features available in the dataset, a number of training subsets are created by choosing random features with replacement. Once the features are selected, the trees are constructed based on the best split. Each tree gives an output which is considered as a 'vote' from that tree to the given output. The output which receives the maximum 'votes' is chosen by the random forest as the final output/result or in case of continuous variables, the average of all the outputs is considered as the final output.

V. PROPOSED SYSTEM

A. Methodology

LSTM (Long Short Term Memory):

LSTM is a special type of RNN. These networks are proficient in learning about long-term dependencies. These networks are clearly designed to evade the long- term dependency problem, but remembering information for a long time period back is their normal behavior. LSTM have a different structure compared to other neural networks. Conventional RNN has a very simple neural network with a feedback loop but LSTM consists of a memory block or cells instead of a single neural network layer. Each cell or block has 3 gates and a cell state tend to regulate the flow of data information through the cells.

The LSTM model is organized in the form of a chain structure. However, the repeating module has a different structure. Instead of a single neural network like a standard RNN, it has four interacting layers with a unique method of communication.

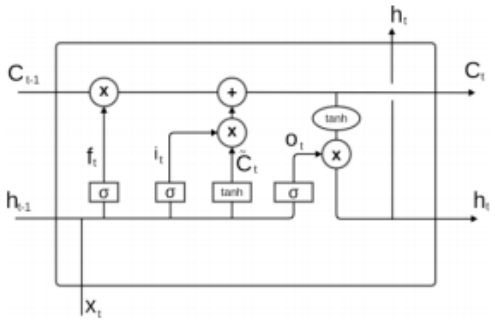

Fig. 1.1: LSTM Cell

The symbols used here have the following meaning:

a) X : Scaling of information
b)+ : Adding information
c) σ : Sigmoid layer
d) tanh: tanh layer
e) h(t-1) : Output of last LSTM unit
f) c(t-1) : Memory from last LSTM unit
g) X(t) : Current input
h) c(t) : New updated memory
i) h(t) : Current output

To overcome the vanishing gradient problem, we need a function whose second derivative can sustain for a long range before going to zero. *tanh* is a suitable function with the above property. As Sigmoid can output 0 or 1, it can be used to forget or remember the information.

Information passes through many such LSTM units.There are three main components of an LSTM unit which are labeled in the diagram:

1. LSTM has a special architecture which enables it to forget the unnecessary information .The sigmoid layer takes the input X(t) and h(t-1) and decides which parts from old output should be removed (by outputting a 0).
2. The next step is to decide and store information from the new input X(t) in the cell state. A Sigmoid layer decides which of the new information should be updated or ignored. A *tanh* layer creates a vector of all the possible values from the new input. These two are multiplied to update the new cell sate. This new memory is then added to old memory c(t-1) to give c(t).
3. Finally, we need to decide what we're going to output. A sigmoid layer decides which parts of the cell state we are going to output. Then, we put the cell state through a *tanh* generating all the possible values and multiply it by the output of the sigmoid gate, so that we only output the parts we decided to [19].

B. System Architecture

This system helps the user to find information by providing them with personalized suggestions for prediction. Our proposed system will allow investors to predict the next-day movement for a particular stock or index, which is important for daily traders. The machine learning models, Long short term memory, neural networks serve as the basis for our "inference engine". To build these models, a list of potential predictors from these data-sources have been generated. The list includes both variables (i.e. data in raw form) and features (e.g. summary statistics and predictors combining multiple variables) selected from these variables.

Working:

1. Current Data is split into training, testing and validation data
2. We make use of an LSTM model to input the training data
3. The model is then saved and trained[4]. The testing data is sent as an input to model for evaluating the accuracy
3. The model is then saved and trained[4]. The testing data is sent as an input to model for evaluating the accuracy

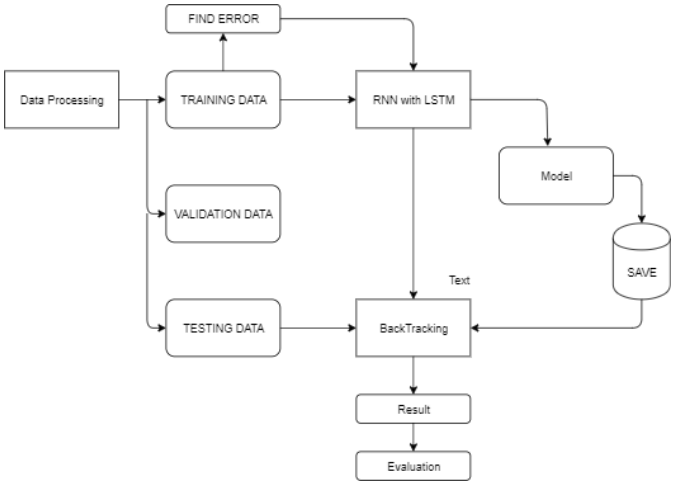

Fig.1.2: System Architecture

Description: Every required data will be available in the datasets provided online.With the change of market and technology regular update of system is required. Beside this the prediction result of stock exchange and their actual price will be compared regularly.

Prediction result will be handled and generated by program. The system will be built, through which the result of prediction and system performance will be analyzed. Accuracy and errors will be monitored to improve the prediction

Graphs can be used for better representation and easier analysis of the data and will help to display the growth and market value over a period of time. Indication will help to constantly compare the predicted and actual values[16].

VI. METRICS

Precision: A measure of exactness, determines the fraction of relevant items retrieved out of all items retrieved. Precision (P) It is given in Equation 2.1. It is the prediction of recommended investments that are actually good.

$$P = \frac{TP}{TP + FP} \qquad (2.1)$$

Recall: A measure of completeness, determines the fraction of relevant items retrieved out of all relevant items. Recall (R) is given in Equation 2.2. It is the proportion of all good stock predictions i.e. actual and predicted values

$$R = \frac{TP}{TP + FN} \qquad (2.2)$$

Accuracy: It is the ratio of the number of correct predictions to the total number of input samples. It works well only if there are an equal number of samples belonging to each class [14].

$$Accuracy = \frac{Number\ of\ Correct\ predictions}{Total\ number\ of\ predictions\ made} \qquad (2.3)$$

VII. RESULTS

Our results suggest that:

(a) the knowledge base of financial expert systems can benefit from data captured from nontraditional "experts" like Google and Wikipedia.
(b) diversifying the knowledge base by combining data from disparate sources can help improve the performance of financial expert systems.
(c) the use of simple machine learning models for inference and rule generation is appropriate with our rich knowledge database. Finally, an intelligent decision making tool is provided to assist investors in making trading decisions on any stock, commodity or index. A financial expert system to predict movements in the one-day ahead stock price/volume.
To construct our "knowledge base", we scrapped the data sets:
(i) historical stock market data,

(ii) commonly used technical indicators,
(iii) Statistics pertaining to the company's pages (i.e. general company profile, stock page, and pages pertaining to the company's main products). These data sources used in combination in an expert system, generate features from the two online sources to further improve our knowledge base.

Our AI framework consisted of two major phases:

(1) Variable/ feature selection, which helps to improve the performance of our AI algorithms by reducing the dimensions of the data without the loss of information.
(2) The incorporation of CNN, LSTM and RNN for prediction.

REFERENCES

[1] Helmi Hammami,Michael Dowling,Saqib Aziz,Anke Piepenbrink, Machine learning in finance: A topic modeling approach , 2019
[2] Hiransha Ma , Gopalakrishnan E.Ab , Vijay Krishna Menonab, Soman K.PNSE Stock Market Prediction Using Deep-Learning Models, November, 2018
[3] XingYu Fu,JinHong Du,YiFeng Guo,MingWen Liu,Tao Dong,XiuWen Duan, A Machine Learning Framework for Stock Selection, August , 2018
[4] Kranthi Sai Reddy Vanukuru , Stock Market Prediction using Machine Learning, November 2018
[5] Carol Hargreaves,Machine Learning Application in the Financial Markets Industry Research Paper: Stock Market Prediction Using Machine Learning,2017
[6] Raut Sushrut Deepak, Shinde Isha Uday, Dr. D. Malathi, Machine Learning Approach In Stock Market prediction, 2017
[7] Mruga Gurjar, Parth Naik ,Gururaj Mujumdar, Prof. Tejaswita Vaidya,STOCK MARKET PREDICTION USING ANN, 2018
[8] Nirbhey Singh Pahwa, Neeha Khalfay, Vidhi Soni, Deepali Vora, Stock Prediction usingMachine Learning a Review Paper, 2017
[9] Aparna Nayak, M. M. Manohara Pai, Radhika M. Pai, Prediction Models for Indian Stock Market, 2016
[10] Osman Hegazy,Omar Soliman, Mustafa Abdul Salam, A Machine Learning Model for Stock Market Prediction, 2013
[11] https://medium.com/datadriveninvestor/machine-learning-for-stock-market-investing-f90ad3478b64
[12] https://www.mdpi.com/2073-4441/11/7/1387/pdf
[13] https://skymind.ai/wiki/accuracy-precision-recall-f1
[14] https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234
[15] https://www.sciencedirect.com/science/article/pii/S0957417404001484
[16] https://towardsdatascience.com/machine-learning-techniques-applied-to-stock-price-prediction-6c1994da8001
[17] https://scholar.google.co.in/scholar?q=stock+market+prediction+using+machine+learning+synopsis&hl=en&as_sdt=0&as_vis=1&oi=scholart
[18] https://medium.com/neuronio/an-introduction-to-convlstm-55c9025563a7
[19] https://towardsdatascience.com/understanding-neural-networks-from-neuron-to-rnn-cnn-and-deep-learning-cd88e90e0a90
[20] https://skymind.ai/wiki/lstm#long
[21] https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/
[22] https://medium.com/mlreview/understanding-lstm-and-its-diagrams-37e2f46f1714