Stock Market Trend Prediction Using High-Order Information Of Time Series

Vignesh J[1], Dr.L.Josephine Mary[2]
[1]Dept of MCA
[2]Professor, Dept of MCA
[1, 2] Sri Muthukumaran Institute Of Technology, Chennai

Abstract- *The aim of this paper is to examine a number of different forecasting techniques to predict future stock returns based on past returns and numerical news indicators to construct a portfolio of multiple stocks in order to diversify the risk. We do this by applying supervised learning methods for stock price forecasting by interpreting the seemingly chaotic market data.*

Linear Regression is a machine learning technique used in to forecast stock prices. Stock market price forecasting is the process of determining the future value of a company's stock. This study uses daily closing prices for many technology stocks to calculate price volatility and momentum for individual stocks and for the overall sector. These are basically used as parameters to the linear model. This model attempts to predict whether a stock price sometime in the future will be higher or lower than it is on a given day. We find little predictive ability in the short-run, but definite predictive ability in the long-run. We can add more algorithms like random forest, Adabooster,Gradient Boosting Classifier and linear regression. Finally we predict which algorithm is provide a good result.

Keywords- Support vector machine (SVM), Long Short-Term Memory(LSTM),Machine Learning, Artificial Neural Network(ANN)

I. INTRODUCTION

Stock prices are considered to be chaotic and unpredictable. The Efficient Market Hypothesis deems stock prices to follow the Random Walk Model. This problem is challenged by Stock Market Prediction Models. Predicting the future stock prices of financial commodities or forecasting the upcoming stock market trends can enable the investors to garner profit from their trading by taking calculated risks based on reliable trading strategies.

There are various number of approaches based on which a prediction model can be developed and implemented. They can broadly be classified as Fundamental Analysis, Technical Analysis and Machine Learning Analysis. Fundamental Analysis is governed by the school of thought that the market may initially miscalculate the price of a tradable financial asset but the right price will be reached eventually

In recent times stock market predictions is gaining more attention, maybe due to the fact that if the trend of the market is successfully predicted the investors may be better guided. The profits are gained by investing and trading in the stock market greatly depends on the predictability. If there is a system that can consistently predict the direction of the dynamic stock market will enable the users of the system to make informed decisions. More over the predicted trends of the market will help the regulators of the market in taking corrective measures.

A stock market, equity market or share market is the aggregation of buyers and sellers (a loose network of economic transactions, not a physical facility or discrete entity) of stocks (also called shares), which represent ownership claims on businesses; these may include securities listed on a public stock exchange as well as those only traded privately. Examples of the latter include shares of private companies which are sold to investors through equity crowd funding platforms. Stock exchanges list shares of common equity as well as other security types, e.g. corporate bonds and convertible bonds.

Stock price prediction is one of the most widely studied problem, attracting researchers from many fields. The volatile nature of the stock market makes it really difficult to apply simple time-series or regression techniques. Financial institutions and active traders have created various proprietary models to beat the market for themselves or their clients, but rarely did anyone achieve consistently higher than the average returns on investment. The challenge of stock market price forecasting is so appealing because an improvement of just a few points of percentage can increase the profit by millions of dollars.

This paper discusses the application of Support Vector Machines and Linear Regression in detail along with

the pros and cons of the given methods. The paper introduces the parameters and variables which can be used to recognize the patterns in stock prices which can be helpful in future stock prediction and how boosting can be integrated with various other machine learning algorithms to improve the accuracy of our prediction systems.

MOTIVATION:

The aim of the paper is to examine a number of different forecasting techniques to predict future stock returns based on past returns and numerical news indicators to construct a portfolio of multiple stocks in order to diversify the risk. We do this by applying supervised learning methods for stock price forecasting by interpreting the seemingly chaotic market data. This paper is to analyse these daily news feeds by using deep learning approaches and investigate the correlation between the large-scale yearly news feeds and the stock market values over time.

The remainder of this paper is organized as follows. Section 2 describes Literature survey. Section 3 briefly introduces some related works and background knowledge about this work. Section 4 describes the methodology of this work.Section 5 is Conclusion and Future enhancement.

II. LITERATURE SURVEY

The stock market prediction has become an increasingly important issue in the present time. One of the methods employed is technical analysis, but such methods do not always yield accurate results. So it is important to develop methods for a more accurate prediction.

Generally, investments are made using predictions that are obtained from the stock price after considering all the factors that might affect it. The technique that was employed in this instance was a regression. Since financial stock marks generate enormous amounts of data at any given time a great volume of data needs to undergo analysis before a prediction can be made[1].

The use of machine learning and artificial intelligence techniques to predict the prices of the stock is an increasing trend. More and more researchers invest their time every day in coming up with ways to arrive at techniques that can further improve the accuracy of the stock prediction model. Due to the vast number of options available, there can be n number of ways on how to predict the price of the stock, but all methods don't work the same way. The output varies for each technique even if the same data set is being applied. In the cited paper the stock price prediction has been carried out by using the random forest algorithm is being used to predict the price of the stock using financial ratios form the previous quarter. By using the financial ratio along with a model that can effectively analyze sentiments the accuracy of the stock price prediction model can be increased[2].

Accurately predicting the stock market is a challenging task, but the modern web has proved to be a very useful tool in making this task easier. Due to the interconnected format of data, it is easy to extract certain sentiments thus making it easier to establish relationships between various variable and roughly scope out a pattern of investment. Investment pattern from various firms show sign of similarity, and the key to successfully predicting the stock market is to exploit these same consistencies between the data sets[3].

The stock market prediction process is filled with uncertainty and can be influenced by multiple factors. Therefore, the stock market plays an important role in business and finance. The technical and fundamental analysis is done by sentimental analysis process. The method usually involves gathering various social media data, news to extract sentiments expressed by individuals. Other data like previous year stock prices are also considered. The relationship between various data points is considered, and a prediction is made on these data points. The model was able to make predictions about future stock values[4].

The recent studies provide a well-grounded proof that most of the predictive regression models are inefficient in out of sample predictability test. The reason for this inefficiency was parameter instability and model uncertainty. The studies also concluded the traditional strategies that promise to solve this problem. Support vector machine commonly known as SVM provides with the kernel, decision function, and sparsity of the solution. It is used to learn polynomial radial basis function and the multi-layer perceptron classifier. It is a training algorithm for classification and regression, which works on a larger dataset. There are many algorithms in the market but SVM provides with better efficiency and accuracy. The correlation analysis between SVM and stock market indicates strong interconnection between the stock prices and the market index[5].

In statistics, linear regression is a linear approach to modelling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression.

III. RELATED WORK

Many efforts to financial time series analysis problem have focused on price or trend prediction, and they roughly include traditional time series analysis methods, Machine Learning methods, Deep Learning methods and their combination.

Traditional approaches to stock market analysis and stock price prediction include fundamental analysis, which looks at a stock post performance and the general credibility of the company itself, and statistical analysis, which is solely concerned with number crunching and identifying patterns in stock price variation.Then predictions were achieved with the help of Genetic Algorithms (GA) or Artificial Neural Networks(ANN's), but these fail to capture coorelation between stock prediction is the phenomenon of exploding / vanishing gradient, where the weights of a large network either become too large or too small, drasticallyslowing their convergence to the optimal value. This is typically caused by two factors: weights are initialized randomly, and the weight closer to the end of the network also trend to change a lot more than those at the beginning.

IV. THE IMPLEMENTATION OFSTOCK MARKET TRENDPREDICTION USINGHIGH-ORDER INFORMATION OF TIME SERIES

The System Architecture of Stock Market Prediction using Long Short-Term Memory**(**LSTM) and Explainable AI (XAI) is shown in Fig 1. It shows that how the data is initially processed, divided into train and test set to build further predictions and evaluate the biases using XAI.

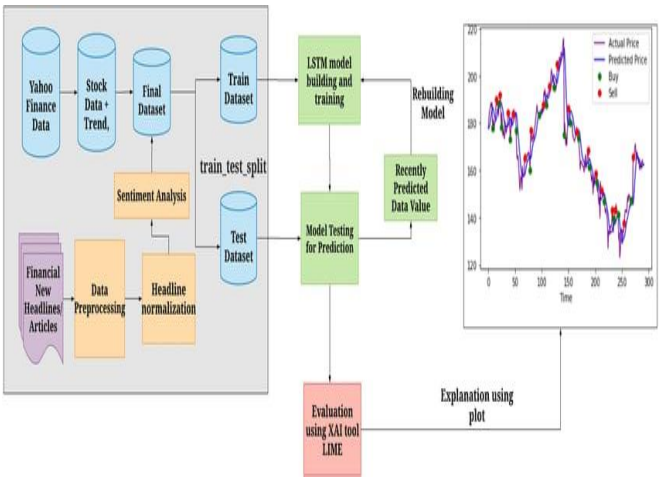

Fig.1. System Architecture of Stock Market Prediction

The Network Architecture of LSTM Network is shown in Fig 2. The network Architecture of the LSTM network depicts how the processed headlines are embedded and passed through a CNN Layer and further through an LSTM cell to output the prediction.

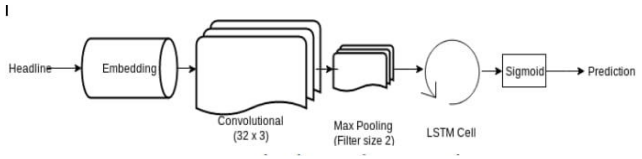

Fig 2. Network Architecture of LSTM Network

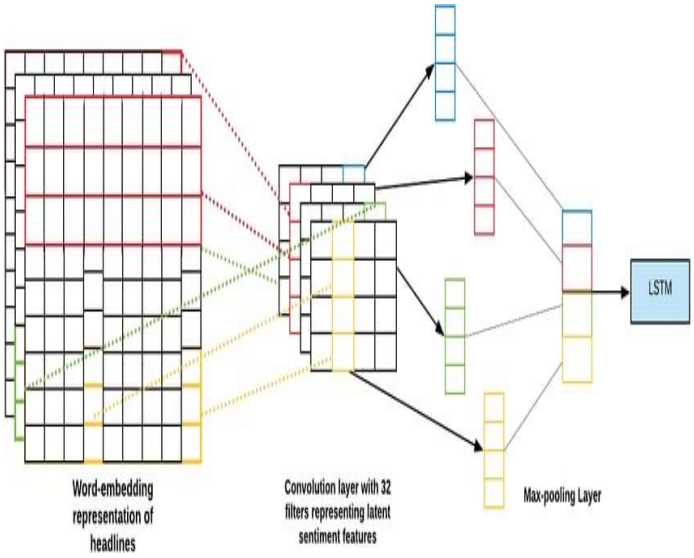

Fig. 3. CNN MaxPooling1D

The CNN MaxPooling1D layer is shown in Fig 3. The MaxPooling1D layer minimized the spatial oriented size. The layer works separately on each feature.

The Hyperparameters, loss function and optimizer and their values used along with detailed comments is shown in table 1.

Title	Value	Comments
Loss function	Binary cross-entropy	Since the type of classification performed is binary
Optimizer	Adam optimization	Used to enhance the network
Learning rate	0.001	Set after trials
Decay	0.1	Validation loss does not ameliorate significantly any more for 5 epochs, there will be decay.

Table 1.Hyperparameters,loss function and optimizer and their values

Set of predictions made:-

The final set of predictions along with comparisons with actual openings prices, calculations for Rise/Fall, Error Percentage, and the chi-squared test is shown in Table 2.

Test case	Date	Previous close	Predicted open	Actual open	Rise/fall	Error(%)	Chi-squared test
1	07/07/2020	36487.28	36558.12	36660.35	170.84	0.27	0.28
2	08/07/2020	36674.52	36594.59	36738.38	−79.93	0.39	0.56
3	09/07/2020	36329.01	36512.42	36450.69	183.41	0.16	1.52
4	10/07/2020	36737.69	36678.35	36555.13	−59.34	0.33	0.41
5	13/07/2020	36594.33	36610.27	36,880.66	15.94	0.73	1.98
6	14/07/2020	36693.69	36649.78	36,517.28	−43.91	0.21	0.48
7	15/07/2020	36033.06	36257.20	36,314.76	224.14	0.15	0.01

Table 2. The final set of predictionwith comparisons with actual openings prices.

Interpolating or recovering the missing data and removing the redundant data:-

This step also involves creating any useful feature from the existing ones. This step involves searching in the space of possible feature subsets. We then pick the subset that is optimal or near-optimal with respect to some objective function. This is done so as to avoid problems of overfitting/underfitting the dataset.Data is needed to be normalized for better accuracy by ensuring that all features are not given excessive/low weightage.

Analysis of various supervised learning methods Classification Methods:

This phase would involve supervised classification methods like linear regression, Ensemble classifiers (like Adaboost, Random Forest Classifiers), etc.

The implementation of this paper is divided into following steps

1.Data Pre-processing:

The entries are present in the dataset. The null values are removed using df = df. dropna() where df is the data frame. The categorical attributes (Date,High,Low,Close,Adj value) are converted into numeric using Label Encoder. The date attribute is splitted into new attributes like total which can be used as feature for the model.

2. Feature selection:

Features selection is done which can be used to build the model. The attributes used for feature selection are Date,Price,Adj close, Forecast Xcoordinate, Y coordinate, Latitude, Longitude, Hour and month,

3.Building and Training Model:

After feature selection location and month attribute are used for training. The dataset is divided into pair of xtrain,ytrain and xtest, y test. The algorithms model is imported form sklearn. Building model is done using model. Fit (xtrain, ytrain). This phase would involve supervised classification methods like linear regression, Ensemble classifiers (like Adaboost, Random Forest Classifiers), etc.

4.Visualization:

Using mathplotlib library from sklearn. Analysis of the Stock dataset is done by plotting various graphs.

V. CONCLUSION AND FUTURE ENHANCEMENT

Machine learning methods were then tested on a wide range of data sources. The result of some models looked hopeful, but ultimately failed when they were put through realistic trading simulations. This highlights that the stock market is prone to differences between theory and practice.

Future Enhancement:

We apply neural network to predict the stock market prices day by day and compare previous year and predict future year.

REFERENCES

[1] Survey of Stock Market Prediction Using Machine Learning Approach, Ashish Sharma ; Dinesh Bhuriya ; Upendra Singh, 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)

[2] Impact of Financial Ratios and Technical Analysis on Stock Price Prediction Using Random Forests:, K. S. Loke, 2017 International Conference on Computer and Drone Applications (IConDA)

[3] Stock Market Prediction via Multi-Source Multiple Instance Learning,Jieyun Huang ; Binxing Fang ; Philip Yu, 2018 IEEE Access

[4] Stock Market Prediction: Using Historical Data Analysis,Saloni Mohan ; SahityaMullapudi ; Sudheer Sammeta Parag Vijayvergia ; David C. Anastasiu, 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)

[5] A Survey on Stock Market Prediction Using SVM, Yelin Li ; Junjie Wu ; Hui Bu, 2016 13th International Conference on Service Systems and Service Management (ICSSSM)