Survey on combined supervised learning for optimized daily Stock price

D.SHALINIGAYATHRI[1] and A.S.ARUNACHALAM[2]

[1]Research Scholar, [2]Associate Professor

[1&2]Department of Computer Science, School of Computing Sciences,

Vels Institute of Science, Technology and Advanced Studies (VISTAS), Tamilnadu, India.

[1]san.thiru@gmail.com, [2]arunachalam1976@gmail.com

Abstract

Equity Market is the place of shares and stock trading system which collection of buyers and sellers. In this domain, diverse public or private shares are traded by the listed companies. The Stock market places an important role in predicting the growth of a country's economy. The prediction of the stock market becomes a difficult task due to its non-linear scenario. The Machine Learning algorithm handles previously logged data as training samples and makes supports for forecasting the stock price for future trends. Supervised Learning is a powerful Machine Learning task. The basic idea of Supervised Learning is to classify and process data using Machine Learning. This paper denotes the survey and usage of Supervised Learning in a stock market application. This survey expresses the various aspect of the stock market then describes Supervised Learning and its algorithm for future prediction of the stock market. This paper shows a detailed idea about the supervised learning techniques and future scope of developing new features

Keyword: Prediction, Classification, Regression, Optimization, Propagation, Volatile

I.INTRODUCTION

Many traders trade their equities by many indices. The Financial System of any country depends on Stocks. Most of people like to get a good return in the stock market. The Stock market holds important pillars of policymakers', buyers, Investors, sellers, markets. Investors investing money in the stock market. They are watching stock prices regularly. But in nature, the stock price is messy and unstable. To forecast the daily stock price is a more complicated task. Researchers are giving more effort to find Stock market prediction. Even though several ranges of methods are existing for forecasting stock price, but anyone has touched particular correctness of price movements. Fundamental and Technical Analysis is used by market researchers to get approximate results. The fundamental analysis uses future growth, revenues, return on equities, profit margin to forecast the stock price. Technical analysis uses some indicators such as Exponential Moving Average (EMA), Moving average (MA), Relative Strength Index (RSI) to forecast the stock price. Historical data are supported by price prediction for both analysis techniques.

II. Literature Review

In [5] the author compares the performance of Bidirectional and stacked LSTM and results show BiLSTM networks show a better output for both short and long-term predictions.

In [1] the author has pointed out that in earlier days forecasting financial analysis is handled by various statistical tools. But nowadays, we have many numbers of Neural Network models to find the exact solution of financial data. ANN is combined with a statistical model is given a better result.

In [12] the author has tried to compare the Machine Learning algorithms such as SVM, Perceptron, Logistic regression for good prediction. Here, the author expressed that by using historical data of the stocks Machine Learning Algorithms are trained to predict future data.

In [3] the author has evaluated a model for short term stock prediction. Here the author used the LSTM model. This algorithm is evaluated by root-mean-square and mean absolute percentage. Recurrent neural network along with Long Short-Term Memory competent of good prediction with lower error rate approximately.

In [10] the author has tried to predict stock price by candlestick pattern direction and regression method. Based on the candlestick graph, the author predicts market movement with the best level of accuracy. Market trend prediction using the Knn regression algorithm shows greater accuracy over other supervised learning algorithms.

In [6] the author has proposed combining two methodologies EMD-SVM and implemented them in the stock market to predict future trends. Here the author split the financial data into several regions with the help of the EMD Model. Then the author used various SVMs functions to get the desired predictions.

In [2] the author has tried to compare LSTM with BiLSTM and says that the BiLSTM model gives accuracy in forecasting stock price than LSTM & ARIMA. Then the author says that the BiLSTM model support one or more training by traversing the input data twice so that accuracy is getting more.

In [4] the author has promoted economic news and ending value of stock scrutinize as input. In the usage of the text mining approach, the author processed the economic news. By using some of the indicators find closing values of the stock price. By applying the feature extraction method, and classifier algorithms, the author finds a correlation between economics news and stock market values.

In [14] the author has to determine stock prediction by different deep learning models. He suggests that some limitations in every model. He declared Convolutional Neural Network takes a long learning time, Support Vector Machine becomes slower when we give more input. Training time needs to increase if the size of the machine is increased in the Deep Belief Network.

III. SUPERVISED LEARNING

Machine learning is split into two categories, namely Supervised, Unsupervised learning. Here the Supervised learning existing data is given for training samples for an algorithm. This type of learning is a kind of predictive model. The predictive model is using past data to predict the feature output. Here the input and output are well known. Input data is called features(v), respective output (w) which are compared by its

real prediction among their output value for seeking out errors, finally it modifies the model according to it. Different types of Learning methods are illustrated in the figure.1


Fig. Machine Learning
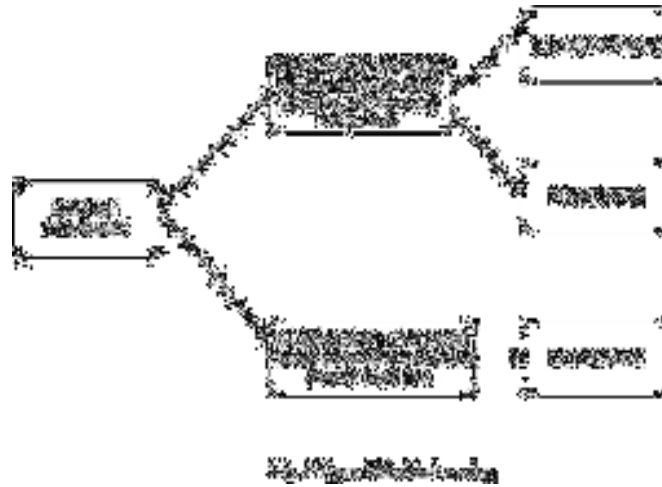
A. Applications
- In statistical data analysis, to recognize patterns and regularities in data.
- In computer speech recognition, to recognize and translate spoken language into text by the computers
- In airports, OCR is used for passport recognition and information extraction.
- In Database Marketing, to generate personalized communications to promote a product.
- In Stock Market, to predict the stock price of the next day.

B. Supervised Learning Algorithms
  1. Support vector machine

   The optimal methodology in machine learning is the Support Vector Machine. It is mainly supported for classification and regression analysis. It is very easy to learn because of its simple structure. An SVM algorithm trained from samples and allocate label to the object. It may differentiate the samples into categories. In n-dimensional space, every data element is plotted. In the fig. 2 SVM shows A binary classification.

Fig 2 Support Vector Machine
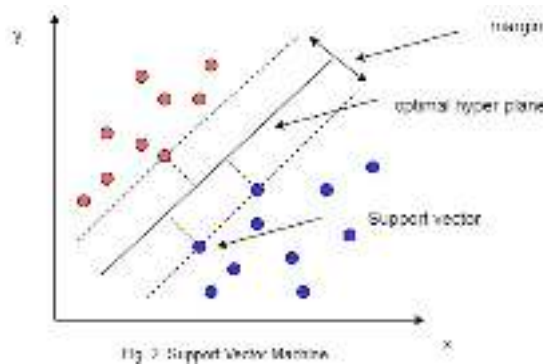
The algorithm includes the following steps:
  1. Find hyperplane. It can be defined mathematically as:
     $$f(y)=b+w^T y$$
     here y is the training sample closest to the hyperplane, b is called bias, w is the weight vector.
  2. Find distance (D). It is calculated between training samples y & hyperplane

D=b+w^{T}y

3. To make training samples yi correctly, Dirichlet L-function L(w) is used as:

$$\min_{w,b} L(w) = \frac{1}{2}\|w\|^2$$

In this way, we can understand that forecast stock prices with SVM.

2. K nearest neighbors regression (K-NN regression)

It is the modest learning Algorithm under the category of supervised. It is mainly focused on classification. In[10] Knn classifies its data with its neighbors' orientation. Based on similarity measure Knn stores all available data set and classifies new cases. The Distance formula is used for finding the weight of the data set. At last nearest distance is identified and c classified [17].

The algorithm includes the following steps:
1. Loading and Initializing the training and test data
2. From the training, data choose the opt k values
3. Testing data for every point is calculated as follows:
   - Euclidean distance is calculated to all training data points by

$$D(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

   - keep follow the list of Euclidean distances and make them to sort it.
   - select the first k points
   - Final selected points of the testing data are measured by the majority of the classes of the data set.
4. Stop.

Knn-Regression can be used in the financial sector and also used in time series prediction to forecast the financial market.

3.Random Forest

It is one of the algorithms which is mainly used for classification and regression problem of machine learning. This model formed with the decision tree along with separating the model data according to the features. This model is depending on a number of features [16]. The number of branch nodes will be increased if the feature increases.

Guidelines for building a decision tree as follows:
1. Fixing root by collecting all training data and then derive all features, and optimal features are filtered out. Finally, existing samples are split into a number of the subset.
2. Assigning an optimal dataset as a leaf node.
3. To create a new subset, divide the existing subset of data by traversing all features, and choose the best one.

Creation of decision tree model by continuing step b & c up to their condition stops the split.

until the conditions for stopping the split are met. A structure of Random Forest is represented in the figure.3

Fig. 3 Random Forest

Predictive analysis in Radom forest having many applications for an especially optimal way of stock prediction.

4. Multiple Linear Regression

The Regression technique is one of the most analytical tools for stock markets in the present scenario. It is widely used past values to predict the future [8]. Multiple Linear Regression (MLR) is the process of connecting a single to a more experimental variable which is calculated by the subsequent equation.

$$y_i = \beta_0 + \beta_1 x_i 1 + \beta 2 x_i 2 + \dots \beta_p x_i p + \text{\euro}$$

y_i=dependent variable
x_i = experimental variable
β_0=y-intercept
β_P=slope coefficient for each independent variable
Є=the models error term

5. Long Short-Term Memory

LSTM Model is a semi-supervised learning algorithm that makes memory cells with an information loop. Time series problems are solved efficiently with a large storage capacity of memory and permit feedback of stock information at various levels which accumulates knowledge of forecasting tasks [3]. The main idea of the financial series prediction problem can store long term data for their various levels of real-time data. Long term availability outlines the LSTM neural Network with graphic analysis for stock and financial securities are justified by their tracking record [2] supports for predicting financial behavior. Dramatize the data by arithmetic operations of multiplications and additions flows through its cell states by an active selection or not [15]. The cell state can derive a piece of information by its dependencies of stock prices such as the previous cell state contains the information of the previous timestamp, hidden State consists of output for the previous cell, current Time Stamp having information new updates of data for predicting factors and trend of stocks.

LSTM Network is represented in Figure. 4
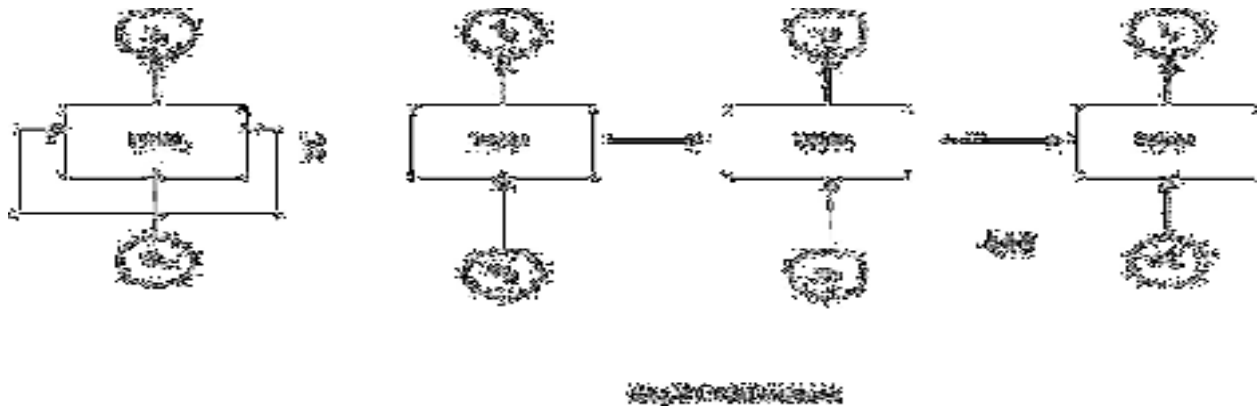
IV. COMPARATIVE ANALYSIS AND FUTURE TRENDS

TABLE: 1.   COMPARATIVE ANALYSIS

Name of the Supervised Learning Models	Compared with	Dataset used	Advantages	Future Trends
K Nearest Neighbor regression[10]	Linear Regression SVM	NSE	Better Accuracy	This Model can be expanded by combining sentiment analysis to make better action in the market prediction
Random Forest[7]	SVM ANN	China stock market	Highly predictive model for long term stock price	The feature selection algorithm needs to expand to achieve new predictable features.
Support Vector Machine [12].	Perceptron model with Logistic Regression	NSE's NIFTY50 Index	Higher accuracy & Lower error rate	This model can be enhanced and applied by former data with technical indicators.
Multiple Linear Regression [8]	Neural Network	BSE	Higher Accuracy	This model can be implemented Neural Network with backpropagation to get a good result.
Long Short-Term Memory [3]	Neural Network	Euro–Dollar assets, Period between September 2018 to July 2019	Greatest potential for short term investment	This model can be expanded by planning to add other tactics and finding financial activities in the short term with numerous periods.

V. Conclusion

This paper shows the elementary awareness and ideologies of Supervised Machine Learning. Also, shows the basic algorithm of supervised machine learning including RF, Knn, SVM, MLR, etc. From this survey, we conclude that various supervised machine learning algorithm produces more optimized results. Finally, we describe a comparative study of current Supervised Machine Learning methods along with future patterns.

References

[1]. D. A. Kumar and S. Murugan, "Performance analysis of Indian stock market index using neural network time series model," in *2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*, Salem, Feb. 2013, pp. 72–78, doi: 10.1109/ICPRIME.2013.6496450.

[2]. S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM," *ArXiv191109512 Cs Stat*, Nov. 2019, Accessed: Jun. 12, 2020. [Online]. Available: http://arxiv.org/abs/1911.09512.

[3]. S. M. Thampi et al., Eds., Applied Soft Computing and Communication Networks: Proceedings of ACN 2019, vol. 125. Singapore: Springer Singapore, 2020.

[4]. S. E. Seker, C. Mert, K. Al-Naami, U. Ayan, and N. Ozalp, "Ensemble classification over stock market time series and economy news," in 2013 IEEE International Conference on Intelligence and Security Informatics, Seattle, WA, USA, Jun. 2013, pp. 272–273, doi: 10.1109/ISI.2013.6578840.

[5]. K. A. Althelaya, E.-S. M. El-Alfy, and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," in *2018 9th International Conference on Information and Communication Systems (ICICS)*, Irbid, Apr. 2018, pp. 151–156, doi: 10.1109/IACS.2018.8355458.

[6]. H. Yu and H. Liu, "Improved Stock Market Prediction by Combining Support Vector Machine and Empirical Mode Decomposition," in *2012 Fifth International Symposium on Computational Intelligence and Design*, Hangzhou, China, Oct. 2012, pp. 531–534, doi: 10.1109/ISCID.2012.138.

[7]. X. Yuan, J. Yuan, T. Jiang, and Q. U. Ain, "Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market," *IEEE Access*, vol. 8, pp. 22672–22685, 2020, doi: 10.1109/ACCESS.2020.2969293.

[8]. H. R. Patel, S. M. Parikh, and D. N. Darji, "Prediction model for stock market using news based different Classification, Regression and Statistical Techniques: (PMSMN)," in *2016 International Conference on ICT in Business Industry & Government (ICTBIG)*, Indore, India, 2016, pp. 1–5, doi: 10.1109/ICTBIG.2016.7892636.

[9]. B. Jeevan, E. Naresh, B. P. V. kumar, and P. Kambli, "Share Price Prediction using Machine Learning Technique," in *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, Bangalore, India, Oct. 2018, pp. 1–4, doi: 10.1109/CIMCA.2018.8739647.

[10]. M. Ananthi and K. Vijayakumar, "Stock market analysis using candlestick regression and market trend prediction (CKRM)," *J. Ambient Intell. Humaniz. Comput.*, Apr. 2020, doi: 10.1007/s12652-020-01892-5.

[11]. K. A. Althelaya, E.-S. M. El-Alfy, and S. Mohammed, "Stock Market Forecast Using Multivariate Analysis with Bidirectional and Stacked (LSTM, GRU)," in *2018 21st Saudi Computer Society National Computer Conference (NCC)*, Riyadh, Apr. 2018, pp. 1–7, doi: 10.1109/NCG.2018.8593076.

[12]. I. R. Parray, S. S. Khurana, M. Kumar, and A. A. Altalbe, "Time series data analysis of stock price movement using machine learning techniques," Soft Comput., Apr. 2020, doi: 10.1007/s00500-020-04957-x.

[13]. L. Chen, Z. Qiao, M. Wang, C. Wang, R. Du, and H. E. Stanley, "Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market?," IEEE Access, vol. 6, pp. 48625–48633, 2018, doi: 10.1109/ACCESS.2018.2859809.

[14]. V. Kumar and M. L. Garg, "Deep learning in predictive analytics: A survey," in 2017 International Conference on Emerging Trends in Computing and Communication Technologies (ICETCCT), Dehradun, Nov. 2017, pp. 1–6, doi: 10.1109/ICETCCT.2017.8280331.

[15]. A.S. Arunachalam, T.Velmurugan. "A Survey on Educational Data Mining Techniques." International Journal of Data Mining Techniques and Applications 5.2 (2016): 167-171

[16]. S. Perumal,T.Velmurugan, "Lung cancer detection and classification on CT scan images using enhanced artificial bee colony optimization",International Journal of Engineering & Technology,7 (2.26) (2018) 74-79.

[17]. Arunarani, S., Gobinath, R. "A relative analysis of multimodal biometric fusing face, ear and fingerprint"International Journal of Scientific and Technology Research, 2020, 9(4), pp. 1996–2002.

[18]. Arunachalam, A. S., & Velmurugan, T. (2018). Measures for predicting success factors of ELearning in Educational institutions. International Journal of Pure and Applied Mathematics, 118(18), 3673-3679.