A Comparative Study on Useful Learning Algorithms Used for the Stock Market Analysis or Prediction

Ahsan Masroor
Sir Syed University of Engineering and Technology, Karachi Pakistan.
Ahsanmath@yahoo.com
Sallar Khan
Sir Syed University of Engineering and Technology, Karachi Pakistan.
Sallarkhan_92@yahoo.com
Jawaria Hafeez Siddiqui
Sir Syed University of Engineering and Technology, Karachi Pakistan.
h.jawaria@yahoo.com
Abdul Khaliq
Sir Syed University of Engineering and Technology, Karachi Pakistan.
akhaliq.ciit@gmail.com
Shariq Ahmed
Sir Syed University of Engineering and Technology, Karachi Pakistan.
shariqahmedssuet@gmail.com
Parshan Kumar
Sir Syed University of Engineering and Technology, Karachi Pakistan.
parshankumarssuet@gmail.com

Abstract
Predicting, analyzing, and forecasting financial markets always remain a challenging task for investors and researchers around the globe. Many authors implemented and discussed various learning models for their studies and conclude many hypothesizes, the fundamental objective of our study is to discuss all the possible learning algorithms or models used by the authors in this area of financial marketing also to highlight the optimal or state of art learning algorithm which produced the effective results among the many studies. This study will help out future researchers to choose the best possible learning technique for their desire financial market problem, also it can help many researchers to choose the optimal technique for the creation of any of their future products for the field of the stock market.
Keywords: Learning Algorithms, Stock Market Analysis, Stock Market Prediction.

1 Introduction and Related work

Regarding the reward-based training pipeline, few studies on expanding ensembling steps where most of the work depend exclusively on one single agent that receives one or several sources of data as input then a single agent completed the output that is trained on a modified number of iterations. In a three-layer stock trader, this methodology is just doing the opposite by increasing ensembling steps, which accepts loaded inputs from other classifiers and combines its set of classes via different training iterations, fusing these steps on a Reinforcement Learning trader agent. Pre-processing layer which is layer 1 that performs on data of stock market at various time resolutions and transforming these time-series data to images and using within the day trading signals generated by thousand (1000) convolutional neural networks as meta-features to be utilized by the upcoming layer. The reinforcement meta-learner is layer 2 where a reward-based classifier is considered and the preceding layer handles the output, this ensemble process is commonly called preprocessing or stacking. Deep Double Q-Learning is the technique in which meta-learner performs preprocessing. In Ensembling Layer, which is layer 3, to get the last decision, the layer combines different signals of different training iterations of the meta-learner[1].This approach is a combination of multiresolution analysis for data decomposition and deep learning architectures for data modeling and data training. Multi-scale wavelet analysis is decomposing financial time-series data. A set of deep learning neural networks is trained for each scale to operate the forecasting per each stage. An Empirical multiresolution wavelet analysis method is applied, and its results are compared with stationary wavelet analysis. The training process includes 2 stages. The predicting and learning of each sub-stage of the time series is completed in separation in the first stage. The earlier stage is combined to predict the last target cost. Though the proposed methodology has been evaluated by conducting experiments for both short-term and long-term predicting using 2 targeted datasets, namely the S&P500 dataset and Mackey-glass time series[2].The research recommends padding-based Fourier transform denoising that removes the noise waveform in the frequency domain of financial time series data and the problem of data divergence was solved at both sides when reinstating to the initial time series. Experiments were conducted to calculate the prices of S&P500, SSE, and KOSPI by utilizing information, from which padding-based Fourier transform denoising removed noise, to different deep learning models founded on time series. Results show that the combination of the proposed denoising technique and the deep learning models not only outperforms the essential models in terms of predictive performance but also reduces the time break problem[3].An ensembling technique for Reinforcement Learning trader agents involves the usage of distinct models trained at multiple epochs (iterations), with additional assessment of the performance of these ensembles over distinct agreement levels. The Deep Double Q-learning algorithm in Reinforcement Learning is used which was initially offered by Mnih et al. (2015), useful to stock market intraday trading. Even when combined with a very primary network structure, the result of the ensemble has a sustainable presentation against the Buy-and-Hold target in some future marketplaces. This is the initial work that discovers combining Deep Double Q-learning Reinforcement Learning strategies in ensemble methods reflecting different training epochs to predict different future sales, various combinations of actions to be completed by the trader. Multi-resolution, where various time resolutions of data are chosen and join up in a more effective way and hence gives the appropriate information that the agent must learn to achieve an improved trading performance[4].The fundamental goal of this paper is to study the performance of improving portfolio optimization models with ML (Machine Learning) and DL (Deep Learning) model return prediction. In this regard, this study has 2 parts to satisfy the gaps in existing researches. First, this paper researches the performance of 2 ML models (i.e., SVR and RF) and 3 deep learning models (i.e., DMLP, LSTM, CNN) in the stock pre-selection technique before portfolio formation. These models have a great performance than usual time series models, which guarantees high-quality stocks are selected before creating portfolio optimization models. Also, these models need a few amounts of hypotheses that are more appropriate for practical applications. Second, this study combines the predictive output of these models in advancing classic MV and omega portfolio optimization models for the 1st time. These enhanced portfolio optimization models not only have the benefits of ML and DL models in return prediction but also keep the essence of classical MV and omega models in portfolio optimization. Thus, these models can more enhance the out-of-sample performance of present models[5].A sentiment information extraction method based on deep learning is proposed in this study and sentiment features are applied to stock movement prediction. Sentiment analysis based on deep learning techniques is achievable for stock movement prediction and sentiment features have a positive effect on the prediction. Although deep neural networks, such as CNN and GRU attained the best polarity classification outcomes in the sentiment feature extraction process, with the LR method, they have almost the same contributions to the last prediction. Thus, ML-based feature extraction can attain similar performance to DL-based methods. PCA is an efficient and robust feature decomposition technique for both trade only and hybrid features. Among the feature selection methods, it attained the finest performance. More sentiment data was filtered after feature engineering via feature visualization. Sentiment analysis is beneficial for stocks with lower beta risk values and higher P/B ratios. This behavior proves that investors make plans based on the overall sentiment of the market and focus more likely on stocks with a development space[6].In this research, LSTM DL (Deep Learning) network is utilized to predict the daily stock final price series. The proposed hybrid framework is merged with 4 parts, namely, EWT (Empirical Wavelet Transform), LSTM (Long Short-Term Memory), PSO (Particle Swarm Optimization), and ORELM. The EWT algorithm is used to divide the raw stock closing price into serval sub-stage used for data pre-processing making training data more stable and regular. The dropout strategy is employed for enhancing the training process of the LSTM deep network and improving the generality capability of the model. The PSO algorithm is used to choose appropriate hyperparameters of the LSTM connection. The ORELM is used for the correction of errors founded on earlier forecasting output for each sub-stage[7].In this research, daily data of a particular bank over a specific period is considered and the effect of different variables on it is explored. No. of shares, WAP (Weighted Average Price), No. of trades, Total Turnover in currency, percent deliverable quantity to traded quantity, deliverable quantity, spread open and close, spread high and low, and the high stock price of the organization for this research was stated. The high stock price was measured as output whereas other parameters were utilized as input. For the analysis, Biovia software's Pipeline Pilot module was utilized. For the development of an ML (Machine Learning) model and prediction, Biovia software offers various built-in modules. As a result, the relationship between Weighted Average Price (WAP), No. of trades, No. of shares, percent deliverable quantities to traded quantity, deliverable quantity including spread open and close have a substantial effect on a high stock price of the bank[8].In this study, an efficient ARIMA (Auto-Regressive Integrated Moving Average) model is built to predict the Indian stock market volatility. An ARIMA model is a vibrant univariate forecasting technique to estimate the future values of a time series. For this research, the publicly available time-series data of the Indian stock market has been utilized. The predicted time series and actual time series are compared, which reveals approximately a variation of 5% mean percentage error on average for both Sensex and Nifty. ADF test and the L-Jung box tests are utilized for the purpose of validation. For handling time-series data, an ARIMA methodology is good enough and as such can be very useful in numerous real-world challenges as the education sector, health, finance, and other effective areas for forecasting[9].For stock predictions, deep functional link neural networks as FLANN (Functional Link Artificial Neural Networks) are single-layer neural networks that can solve complicated problems by producing non-linear decision boundaries. To handle complex problems generated from several hidden layers, it is assumed that deep FLANN would take advantage of the power of deep learning and interconnection of neurons as well as from the inherent capability of single layer FLANN. Primary findings indicate that the average performance of Deep FLANN is better than the single-layer FLANN as well as Convolutional Neural Network (CNN) for the prediction of stock prices of a set of companies listed on the Korean Stock market. Using various architecture, the model can be trained more and assessed on different companies for prediction[10]. The predictive abilities on the DJIA index of a simple solution based on deep recurrent neural network is analyzed, which in several research areas have shown superior abilities to detect long-term dependencies in sequences of data, such as in speech recognition and in text understanding. The objective of this work was to focus more on the time-series data of stock rather than the latest complex trends based on the use of non-structured data. From this perspective, the work follows the idea of the ARIMA methodology offered in 1970 by Box and Jenkins. The contribution of this work is that the set of effective choices that have directed to the implementation and training of a deep RNN for trading with the DJIA (Dow Jones Industrial Average). As a result, it executes trading actions like buy/sell/none, the closing price movements of DJIA, which appear irrational conflicting with the stock market instability[11].A comparison of the US and Japanese stock market to achieve a greater awareness of long-term stock market movements, using monthly data over the last 40 years has been performed. Evidence of a single cointegration vector between stock prices, inflation, IP, and the long-term interest rate is found by utilizing data from the US stock market. From the cointegrating vector, the coefficients are stabilized on the stock price, proposed United States stock prices were affected, as anticipated, negatively by inflation and the long-term interest rate and positively by IP. It is found that the money supply had an insignificant impact on the stock price. Two cointegrating vectors were found in Japan. Stock prices are negatively associated to the money supply but positively associated to IP. For the second vector when normalized on IP, IP was negatively associated to the rate of inflation and interest rate. A clarification of the variance in performance between the two stock markets may lie in Japan's slump after 1990 and its resultant liquidity trap of the late 1990s and early 21st century[12].

Comparative Analysis

Financial exchange expectation utilizing AI procedure targets creating viable and proficient models that can give a superior and higher pace of forecast precision.Throughout the most recent couple of many years, an enormous number of studies have proposed and created various techniques to break down and figure securities exchange actions like as shown in *fig.1*,deep learning and ANNmodel are applied to securities exchange examination and expectationbecause they are in high frequency now a days as compared to other models. While Autoregressive and Multiresolution Wavelet models place second in terms of usage in prediction of financial markets. Lastly, Deep hybrid framework, Random Forest, and Logistic Regression found least used classifiers for predicting or forecasting in the paradigm financial market.

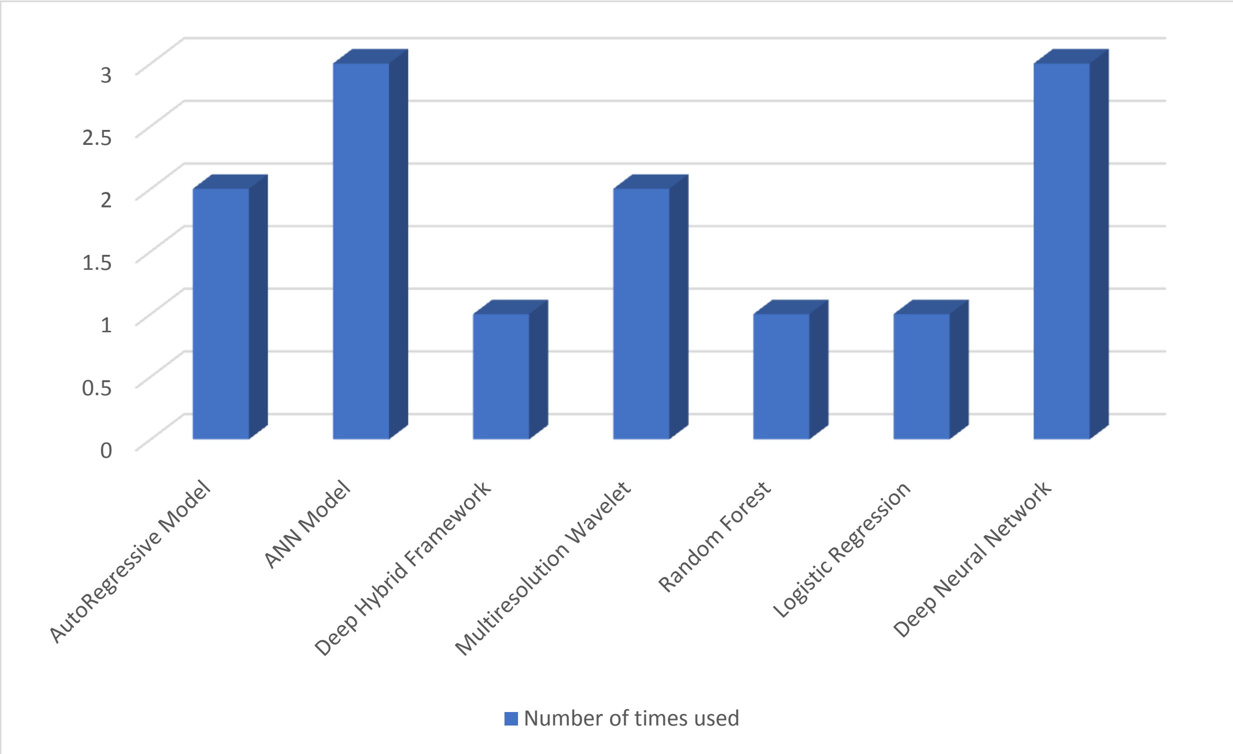

Figure 1. Comparative Analysis of Different Learning Classifiers

As the analysis shown below in *fig 2.* the highest usage is found by Deep Learning models these dayswith the percentage of 50, and its data profound learning procedures have progressivelybeen analyzed to evaluate whether they can further develop market estimating whencontrasted and customary methodologies.While supervised models came second with the 31.25% usage and reinforcement learning models are comparatively less used in the financial market analysis or prediction with the 18.75%.

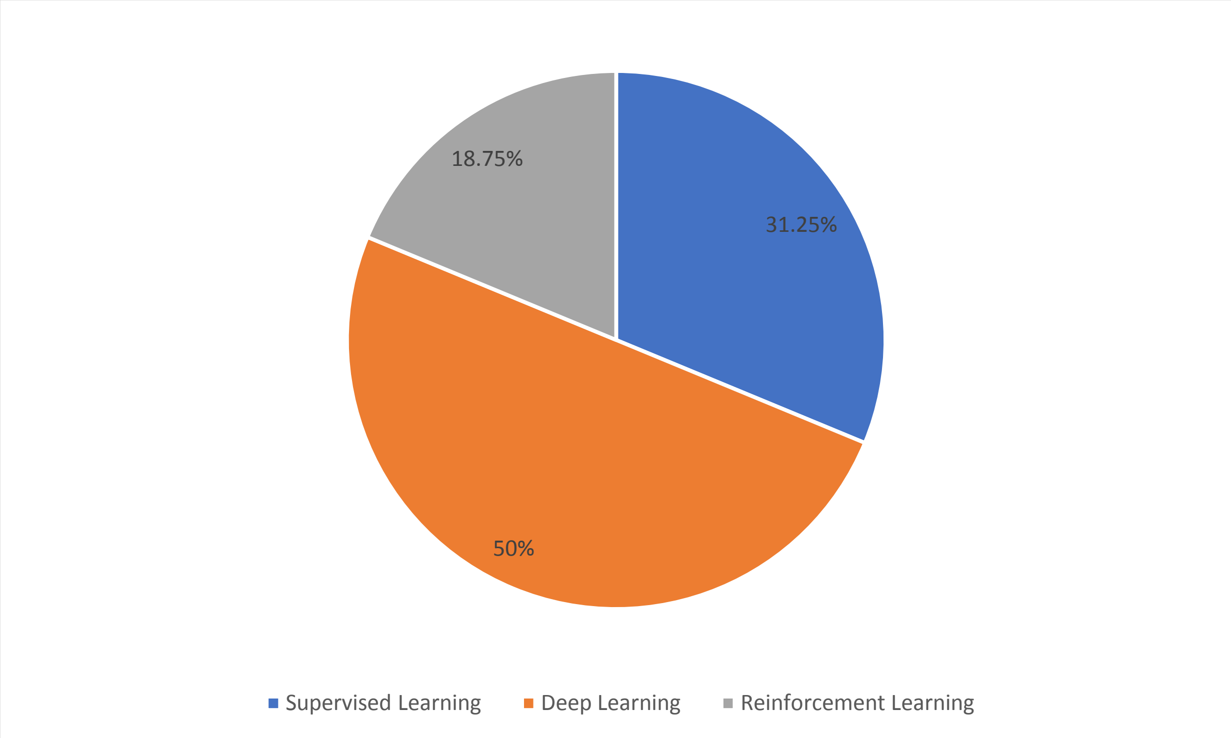

Figure 2. Usage of different Learning techniques.

Conclusion

One of the most challenging application of Machine Learning is the stock market forecasting, as it contains naturally unstable and noisy historical data. As seen from studies, Supervised techniques considered most of the successful approaches, training data is labelled as being negative or positive moments of the market. However, overfitting occurs when trying to train the machine learning classifiers, since various external factor like political events, market trends, etc make impact on market behavior. In this research, we tried to conclude a comparative study of different learning classifiers and their usage in the financial market forecasting and prediction. As per recent advancements, Deep learning modelsare taking high bridge by making impact in the accuracy as well as dealing high amount of data, and can be recommended for further studies or in the designing of any framework in the field of Financial market forecasting.

References

[1] S. Carta, A. Corriga, A. Ferreira, A. S. Podda, and D. R. Recupero, “A multi-layer and multi-ensemble stock trader using deep learning and deep reinforcement learning,” *Appl. Intell.*, vol. 51, no. 2, pp. 889–905, 2021, doi: 10.1007/s10489-020-01839-5.

[2] K. A. Althelaya, S. A. Mohammed, and E. S. M. El-Alfy, “Combining deep learning and multiresolution analysis for stock market forecasting,” *IEEE Access*, vol. 9, pp. 13099–13111, 2021, doi: 10.1109/ACCESS.2021.3051872.

[3] D. Song, A. M. Baek, and N. Kim, “Forecasting Stock Market Indices Using Padding-based Fourier Transform Denoising and Time Series Deep Learning Models,” *IEEE Access*, vol. 9, pp. 83786–83796, 2021, doi: 10.1109/ACCESS.2021.3086537.

[4] S. Carta, A. Ferreira, A. S. Podda, D. Reforgiato Recupero, and A. Sanna, “Multi-DQN: An

ensemble of Deep Q-learning agents for stock market forecasting," *Expert Syst. Appl.*, vol. 164, no. July 2020, p. 113820, 2021, doi: 10.1016/j.eswa.2020.113820.

[5] Y. Ma, R. Han, and W. Wang, "Portfolio optimization with return prediction using deep learning and machine learning," *Expert Syst. Appl.*, vol. 165, no. September 2020, p. 113973, 2021, doi: 10.1016/j.eswa.2020.113973.

[6] Y. Shi, Y. Zheng, K. Guo, and X. Ren, "Stock movement prediction with sentiment analysis based on deep learning networks," *Concurr. Comput.* , vol. 33, no. 6, pp. 1–16, 2021, doi: 10.1002/cpe.6076.

[7] H. Liu and Z. Long, "An improved deep learning model for predicting stock market price time series," *Digit. Signal Process. A Rev. J.*, vol. 102, p. 102741, 2020, doi: 10.1016/j.dsp.2020.102741.

[8] T. Samantara, K. P. Rath, and M. Siddique, "CONSEQUENCE OF DIFFERENT PARAMETERS ON THE HIGH STOCK INDEX OF ALLAHABAD BANK USING DEEP LEARNING TECHNIQUES," vol. 9, no. 4, pp. 864–874, 2020.

[9] S. M. Idrees, M. A. Alam, and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," *IEEE Access*, vol. 7, pp. 17287–17298, 2019, doi: 10.1109/ACCESS.2019.2895252.

[10] P. Gaurav, A. Singhal, and A. Mani, "Towards A Deep FLANN For Prediction Of Stock Market Returns," *2019 3rd Int. Conf. Recent Dev. Control. Autom. Power Eng. RDCAPE 2019*, vol. 2010, pp. 508–513, 2019, doi: 10.1109/RDCAPE47089.2019.8979088.

[11] M. Fabbri and G. Moro, "Dow Jones Trading with Deep Learning: The Unreasonable Effectiveness of Recurrent Neural Networks," *DATA 2018 - Proc. 7th Int. Conf. Data Sci. Technol. Appl.*, no. Data, pp. 142–153, 2018, doi: 10.5220/0006922101420153.

[12] A. Humpe and P. Macmillan, "Can macroeconomic variables explain long-term stock market movements? A comparison of the US and Japan," *Appl. Financ. Econ.*, vol. 19, no. 2, pp. 111–119, 2009, doi: 10.1080/09603100701748956.