

A hybrid approach to predicting daily stock market returns with deep learning

S. Vinoth^a ✉ | Satish Kumar^b | Manjula Jain^c

^aJAIN (Deemed-to-be University), Bangalore, India, Professor, Department of Finance.

^bIIMT University, Meerut, Uttar Pradesh, India, Professor, Department of Management.

^cTeerthanker Mahaveer University, Moradabad, Uttar Pradesh, India, Professor, Teerthanker Mahaveer Institute of Management and Technology.

Abstract It is quite difficult to correctly forecast returns on stock markets because of the economic stock industry's extreme volatility and complex nature. Programming methods of prediction have shown to be increasingly effective at forecasting stock values with the development of artificial intelligence and improved computational capability. Deep learning (DL) algorithms and big data analytics are becoming more and more crucial in a variety of application areas, including stock market investing. However, other research has focused on predicting daily stock market returns, particularly when employing DL approaches to carry out powerful analysis. The DL algorithm is used in this paper's big data analytics approach to forecast the SPDR S&P 500 ETF's daily stock market return direction. The complete dataset was then run through a DL algorithm, such as the MultiDepth NeuroNetwork (MD-NN) technique, to forecast path of the projected index for the stock market daily returns. The simulation results demonstrate that the MD-NN datasets provide much greater classification accuracy than those utilizing the existing approaches.

Keywords: BDA techniques, deep learning, daily stock market, MNED

1. Introduction

Market participants may safeguard their cash investments via futures contracts thanks to the superior hedging instruments provided by futures markets. This hedging strategy reduces some of the uncertainty brought on by erratic changes in market pricing. For instance, because to the unequal distribution of crude oil around the globe, manufacturers of products and services worldwide actively follow price changes, and there is a thriving futures marketplace for it (Ghosh et al 2022). Because good forecasting has the potential to benefit investors greatly, stock price forecasting is a crucial topic in the financial industry. Forecasting is of importance to financial professionals. Traders are also keen on these trends, which are behavioural models of stock prices (Cheng et al 2022). Insights may be gained through easy financial information for analysis, however in recent times, investment firms have increasingly turned to artificial intelligence (AI) systems to sift through vast volumes of real-time equities and economic data for trends. These systems assist with human decision-making, and since they have been in use for a while, it can be done to research and analyse their traits and performance to determine which ones outperform other methods in terms of prediction (Strader et al 2020).

Companies list their stocks on the stock market, a financial marketplace, to raise money for their activities. Investors invest in stocks in the hopes of making money through an annual dividend or a growth in the stock's value. The price change is erratic. The method that securities are exchanged on the stock market has evolved as a result of developments in communications and software technology (Tabar et al 2020). For anticipating the future market, there are several optimization technologies available, including genetic algorithms, particle swarm optimization, and artificial neural network (ANN) methodologies. For time series prediction, ANN is the finest optimization tool. They could anticipate hidden and undiscovered records (Chandar 2019). The basic goals of every investing in the share market are to generate a high return and limit losses since the stock market is the foundation of any economy. Therefore, because improving stock markets is linked to economic development, nations should work to do so. Making successful stock market forecasts is a possible path to economic self-reliance since the stock market can provide speedy returns on investment. It is more challenging to anticipate the stock prices of a certain company in a specific market since the stock market's prediction is not linear (Aldhyani and Alzahrani 2022).

Technical assessment and basic evaluation have historically been each of them most typical often used to evaluate the stock exchanges data. Since fundamental analysis is predicated on the idea of intrinsic worth, in order to establish the current price, data of both types are used. This strategy long-term embraces the Emit seems to imply there may be any inefficiency in the near run. However, technical assessment makes advantage of previous analytics to spot trends and predict eventual shifts in stock prices (Ayala et al 2021). It could be challenging to anticipate the cost of stocks. Many stock market ideas have



existed and evolved throughout the years. Whether they try to describe how markets work or ask whether it's feasible to exceed the market. According to the Efficient Market Hypothesis (EMH), at any given moment, every detail concerning a stock is included into its market price company, is one of the most popular and fiercely debated hypotheses put out by Fama (1970). The stock is suitably valued if everything happens priced (Shah et al 2019). Deep learning and machine learning methods may aid traders and investors in making judgments in programs that forecast the stock price. These methods seek to automatically identify and recognize connections amid enormous amounts of data. These methods of efficiently self-learning and can deal with forecasting price changes to enhance trading tactics. Numerous methods for predicting changes in stocks have been improved in recent years (Nabipou et al 2020).

Ghosh et al (2019) presented for predicting future returns in financial markets. The endeavour is particularly difficult because of how unpredictable, erratic, and nonlinear the financial market moves are. A three-stage strategy is suggested for completing this challenging undertaking. Kamalov et al (2020) investigated a hitherto understudied problem: utilizing machine learning algorithms to forecast substantial changes in stock price based on past changes. The efficiency of analysers using neural networks in the relevant context is of special interest to us. Singh et al (2019) focused on illustrating various machine learning techniques that have been used to create prediction predicts and valuation predictions for machines that utilize support vectors, among other stock exchanges, deep learning, random forests, boosted decision trees, ensemble methods, and a few hybrid techniques. By offering an updated systematic evaluation of the stock market forecasting methods, including their categorization, characterisation, and comparison, this work seeks to close this gap (Bustos and Pomares-Quimbaya 2020). Shahi et al (2020) evaluated the relevance applying emotional reactions to economic information when predicting the stock markets comparing the results of an averaged examination long short-term memory (LSTM) and gated recurrent unit (GRU) for stock market forecasting under the identical settings. The cooperative deep-learning architecture that we presented is the subject of this comparative research. Lu and Ma (2020) two fresh machine learning models built around hybrid decision trees are suggested to provide more precise predictions of short-term water quality. Extreme gradient boosting (XGBoost) and random forest (RF) are the two blended variables' base designs, and they both add an advanced data denoising approach called complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN). Chen et al (2021) offered a unique approach for predicting stock trends using a graph convolutional feature-based convolutional neural network (GC-CNN) model, which considers both stock market and company-specific information. The DL algorithm is used in this paper's big data analytics approach to forecast the SPDR S&P 500 ETF's daily stock market return direction. The complete dataset was then run through a DL algorithm, such as the MultiDepth NeuroNetwork (MD-NN) technique, to forecast the daily trajectory of the index of shares results in future years.

The following is the paper's key contributions:

1. A whole chain based on students' multi-source daily behaviour data; EPGO-ANN is suggested for the prediction of academic achievement. It can automatically extract characteristics without depending on specialized knowledge.
2. By combining ANN with an embedding layer, the time-series characteristics of each kind of behaviour data are effectively recovered.
3. ANN is used to find the correlation characteristics between different sorts of behaviours.
4. The trials are performed using a sizable actual data set, and the results demonstrate that our suggested technique works better than the conventional DL methods.

The remainder of the document is structured as follows: In section 2, the research methodology and techniques used to collect and evaluate the data are described along with recommendations for future research based on the findings. Before presenting the research results concisely and systematically, analysing and explaining them considering the study aims or objectives, we go through the Discussion and results in section 3 first. Section 4 provides an overview of the Study's main elements, as well as its relevance and contributions, potential ramifications for practice or policy, and potential future study areas.

2. Materials and Methods

2.1. Data collection and selection

The selection of relevant data for prediction is part of the data collecting process. The information used in the suggested technique was gathered from reliable sources. The historical S&P 500 ETF data used in the proposed approach was gathered from yahoo.finance.com. It is S&P 500 Index stock daily statistics. The result represents the stock's value on the specified time start of the day (open), its highest price throughout the day (high), its lowest price during the day (low), and its price at the end of the day (close) are the characteristics utilized as indicators. Results are the closing price of the stock the next day. The dataset spans the period from November 2008 and November 7, 2019. It has 2770 records in total. The neural network is trained with 1939 data, and it is tested using 831 records.

2.2. Data normalization

The most important problem for prediction is the quality of the data. Pre-processing the data is crucial because it may improve prediction accuracy even if the data contains missing values and mismatched samples. The data used in the suggested technique includes information on 2770 trade days. The process of dividing values from data across zero and one and one is known as normalization. The normalizing approach also has the benefit of uniformly scaling the data values, which equalizes their relative importance. The suggested technique employs min-max normalization. The approach known as min-max normalization transforms all values between the ranges of 0 and 1 linearly.

2.3. Prediction algorithm

2.3.1. MultiDepth NeuroNetwork (MD-NN)

The MultiDepth NeuroNetwork is the most widely used form of neural network of all the ones that have been created for data analysis uses recognition of patterns especially categorization. Fig. 1 depicts such a multidepth neural network. In Fig. 1, W_j , $j = 1, 2, \dots, j$, signifies the j th neuron in the hidden layer with j neurons; O_k , $k = 1, 2, \dots, K$, means the k th neuron in the output layer; and G_i , $i = 1, 2, \dots, l$, represents the i th output vector (layer) neural element comprising i components (neurons). Each neuron in the connections between the two neighbouring layers has weights that have been experimentally adjusted. For instance, the weight between the i th the j th neuronal in the concealed layer and the i th cell in the data input layer is shown by the symbol $[X]_{ij}$. Multilayer feed-forward neural networks with processes with a logarithmic threshold may accurately mimic any function if there are enough hidden neurons. Depending on how complicated the neural networks are, there may be any number of hidden layers. Typically, a boundary value of 10 is used to distinguish DNNs from shallow neural networks. In other words, feed-forward neural networks are DNNs if they include more than 10 hidden layers; otherwise, shallow neural networks are mentioned. Standard feed-forward ANNs frequently employ the back propagation learning algorithm, which is based on an iterative procedure wherein weights are placed of the connections exist various layers are repeatedly about the layer that is output via the layers that are hidden, and finally to the first buried layer, the adjustment is made reverse, in order to minimize the mean squared error (MSE), which measures the discrepancy between the predicted class and the true class. The conventional back propagation learning is still often employed to train recently generated DNNs, even though additional advanced learning there are currently built methods. throughout the years for certain purposes.

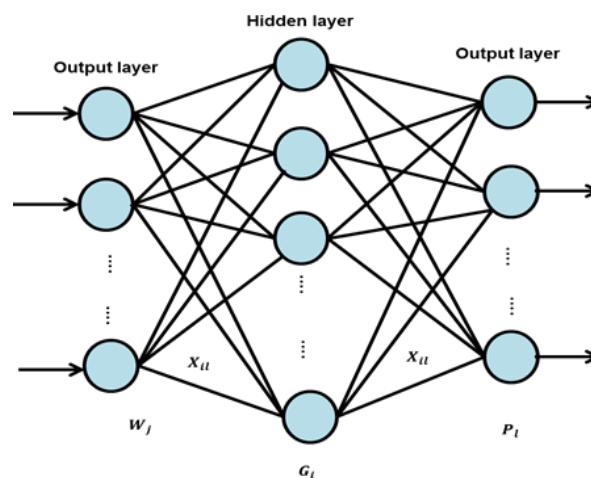


Figure 1 Topology of a multilayer feed-forward neural network used for classification.

3. Results

3.1. Error rate

In machine learning, the error rate is often used as a measure of the performance of a predictive model. It indicates how well the model generalizes to new, unseen data. Better performance is typically indicated by lower error rates because the model is operating more accurately. The cross-validation technique is used to calculate the error rate %. The number of errors is determined to be a proportion of 0.045. Three methods of machine learning are used in our design to forecast the upcoming price of stocks, including KNN, XG BOOST, RNN and MD-NN. The error rate % is used to determine the optimum methods. The best method for more precisely calculating a predictive model's test failure is cross-validation of the prediction error's deviation from the mean is known as RMSE. The distinction in the result actual and the desired output is measured by error in prediction. Figure 2 and table1 demonstrates the proposed and existing method. The modest result decreases the error rate.

Table 1 Numerical outcomes of error rate.

Methods	Error Rate (%)			
KNN	1.99	XG BOOST	RNN	MD-NN [Proposed]
XG BOOST	0.089	65.3	33.6	85.3
RNN	0.068	48.2	61.5	76.5

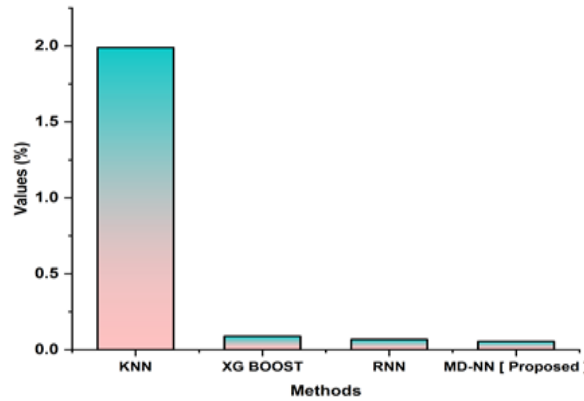


Figure 2 Error rate.

3.2. Actual and predicted price

The result graph showing the actual pricing as compared to what was anticipated. Within the graph, the blue line indicates the expected values, while the red line displays the stock's actual cost. L-fold verification is used to optimize this variable. technique to decrease the mistake and to calculate the mistake rate %. At last, using this method, a 1.98 percent error rate is calculated. Figure 3 and table 2 demonstrates the proposed and existing method. It is concluded that the proposed MD-NN features a 3% increase in efficiency for actual and predicted prices.

Table 2 Numerical outcomes of Actual and predicted price.

Methods	Actual and Predicted Price (%)			
KNN	86	XG BOOST	RNN	MD-NN [Proposed]
XG BOOST	75	55.5	66.2	98.2
RNN	69	38.7	43.4	85.6

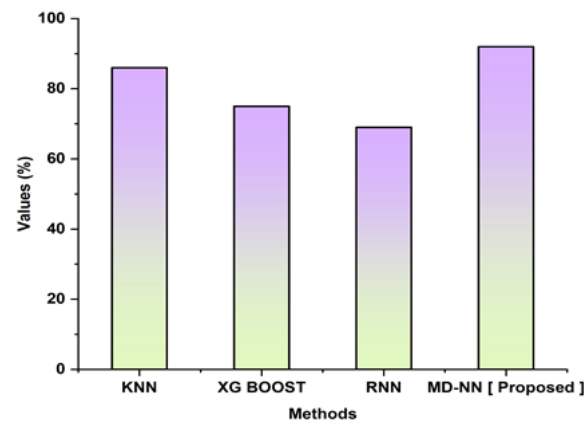


Figure 3 Actual and predicted price.

3.3. Accuracy and precision

Accuracy is a commonly used performance metric that measures the correctness of predictions or classifications made by a model. It is described as the proportion of appropriately predicted cases relative to all instances in a dataset. Accuracy is typically expressed as a percentage. Precision, or the percentage of true positive predictions among all positive predictions made by the model, is a measurement of how accurately a model or algorithm makes positive predictions. To put it another way, precision is the percentage of real positives out of all the occurrences the model correctly identified as positive.

$$Precision = \frac{|{\text{relevant documents}} \cap {\text{retrieved documents}}|}{|{\text{retrieved documents}}|} \quad (1)$$



Figure 4 and Table 3 demonstrates the proposed and existing method. The proposed MD-NN is better than existing method.

Table 3 Numerical outcomes of Accuracy and precision.

Methods	Accuracy Precision (%)			
	KNN	XG BOOST	RNN	MD-NN [Proposed]
1	26.7	55.5	66.2	98.2
2	38.5	38.7	43.4	85.6

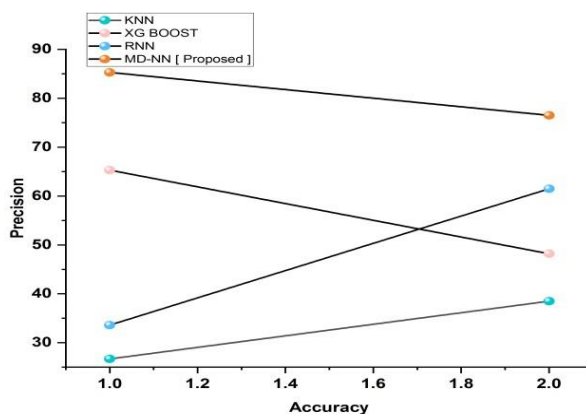


Figure 4 Accuracy and precision.

3.4. Recall and F1 Score

It is also referred to as sensitivity and represents the percentage of relevant documents that were found compared to all relevant instances. It is the percentage of pertinent instances that have been located out of all pertinent cases and documentation. 99.98% of the recall was successful.

$$Recall = \frac{|{\text{relevant documents}} \cap {\text{retrieved documents}}|}{|{\text{retrieved documents}}|} \tag{2}$$

It is a measurement of the test's accuracy and is also referred to as the F-measure. It includes both recall and precision. The estimated harmonic average of Precision and Recall is the F1 score.

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

Figure 5 and table 4 demonstrates the proposed and existing method. The suggested (MD-NN) technique outperforms the current one.

Table 4 Numerical outcomes of Recall and F1 Score.

Methods	Recall F1 Score (%)			
	KNN	XG BOOST	RNN	MD-NN [Proposed]
1	77.8	55.5	66.2	98.2
2	68.9	38.7	43.4	85.6

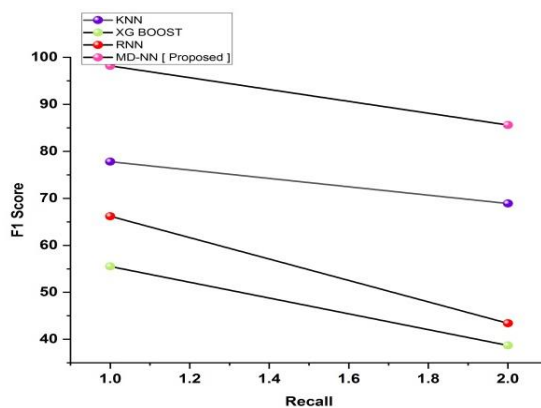


Figure 5 Recall and F1 Score.



4. Conclusions

The proposed system using MD-NN is discovered to be more reliable than the current SVR algorithm in forecasting daily returns on the stock market. The technique aided businesses in making more money while directing traders' market investments. To anticipate returns on stocks, a method like Principal Component Analysis, or PCA, or Deep Neural Network (DNN) has been further refined. The results highlight how important it is to compare the suggested approach's error rate, accuracy and precision, recall and f1score to cutting-edge methods. By combining several classifiers and feature selection techniques, we want to assess the classifier's selecting scheme in future work.

Ethical considerations

Not applicable.

Declaration of interest

The authors declare no conflicts of interest.

Funding

This research did not receive any financial support.

References

- Aldhyani TH, Alzahrani A (2022) Framework for predicting modelling stock market prices based on deep learning algorithms. *Electronics*.
- Ayala J, García-Torres M, Noguera JLV, Gómez-Vela F, Divina F (2021) Technical analysis strategy optimization using a machine learning approach in stock market indices *Knowledge-Based Systems*.
- Bustos O, Pomares-Quimbaya A (2020) Stock market movement forecast A systematic review. *Expert Systems with Applications*.
- Chandar SK (2019) Stock market prediction using subtractive clustering for a neuro fuzzy hybrid approach *Cluster Computing*.
- Chen W, Jiang M, Zhang WG, Chen Z (2021) A novel graph convolutional feature based convolutional neural network for stock trend prediction. *Information Sciences*.
- Cheng CH, Tsai MC, Chang C (2022) A Time Series Model Based on Deep Learning Integrated Indicator Selection Method for Forecasting Stock Prices Evaluating Trading Profits, *Systems*.
- Ghosh I, Chaudhuri TD, Alfaro-Cortés E, Gámez M, García N (2022) A hybrid approach to forecasting futures prices with simultaneous consideration of optimality in ensemble feature selection advanced artificial intelligence, *Technological Forecasting Social Change*,
- Ghosh I, Jana RK, Sanyal MK (2019) Analysis of temporal pattern, causal interaction predictive modelling of financial markets using nonlinear dynamics, econometric models machine learning algorithms *Applied Soft Computing*.
- Kamalov F (2020) Forecasting significant stock price changes using neural networks, *Neural Computing Applications*.
- Lu H, Ma X (2020) Hybrid decision tree-based machine learning models for short-term water quality prediction, *Chemosphere*.
- Shah D, Isah H, Zulkernine F (2019) Stock market analysis A review taxonomy of prediction techniques, *International Journal of Financial Studies*.
- Shahi TB, Shrestha A, Neupane A, Guo W (2020) Stock price forecasting with deep learning A comparative study, *Mathematics*.
- Singh S, Madan TK, Kumar J, Singh AK (2019) Stock market forecasting using machine learning Today a tomorrow. In (2019) 2nd International Conference on Intelligent Computing, Instrumentation Control Technologies.
- Strader TJ, Rozycki JJ, Root TH, Huang YHJ (2020) Machine learning stock market prediction studies review research directions *Journal of International Technology Information Management*.
- Tabar S, Sharma S, Volkman D, (2020). A new method for predicting stock market crashes using classification artificial neural networks, *International Journal of Business and Data Analytics*.