International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 12th Dec 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=Issue 12

DOI: 10.48047/IJIEMR/V11/ISSUE 12/18

Title STOCK MARKET PREDICTION USING DEEP LEARNING

Volume 11, ISSUE 12, Pages: 118-126

Paper Authors
Jadav Arunkumar, Dr. G Venkata Rami Reddy

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

STOCK MARKET PREDICTION USING DEEP LEARNING

1 Jadav Arunkumar, M.tech in software engineering (SE) SIT JNTUH

2 Dr. G Venkata Rami Reddy, Professor M. Tech , Ph. D

ABSTRACT: Due to a variety of deciding factors, the nature of stock market movement has long been unclear to investors. This work uses machine learning and deep learning techniques to drastically lower the risk associated with trend prediction. Four stock market groups from the Tehran Stock Exchange are selected for experimental evaluations: diversified financials, petroleum, non-metallic minerals, and basic metals. In this study, two potent deep learning techniques (Recurrent Neural Network (RNN) and Long Short-Term Memory) are compared to nine machine learning models (Decision Tree, Random Forest, Adaptive Boosting (Adaboost), eXtreme Gradient Boosting (XGBoost), Support Vector Classifier (SVC), Naive Bayes, K-Nearest Neighbors (KNN), Logistic Regression, and Artificial Neural Network (ANN)) (LSTM). Our input values are 10 technical indicators from ten years of historical data, and two methods are intended for using them. First, stock trading values are used to calculate the indicators, and then, before use, the indicators are converted to binary data. Based on the input methods, each prediction model is assessed using three metrics. The evaluation findings show that RNN and LSTM perform significantly better than other prediction models for continuous data. Additionally, results demonstrate that those deep learning techniques are the best for evaluating binary data; nevertheless, the difference between them is diminishing as a result of the second method's clearly improved model performance. As an extension, LSTM, Linear Regression, Lasso Regression, Ridge Regressor, Xgboost, Voting Regression, Decision Tree, Random Forest, SVM, Stacking Regression, Adaboost, SGDregressor, Adaboost, Catboost, LightBoost, Voting Regression-[catboost, lightboost], Stacking Regressor - [catboost, lightboost], are used.

Keywords *–Machine learning algorithms*

1. INTRODUCTION

Experts in finance and statistics have long found stock prediction to be a difficult task. The key strategy used to make this prediction is buying stocks with a high probability of price growth and selling stocks with a high probability of price decline. Generally, there are two approaches for stock market prediction. One of them is fundamental analysis, which depends on a company's strategy and fundamental data including market position, costs, and annual growth rates. The second method focuses on historical stock prices and values and is known as technical analysis. This technique forecasts future prices using past charts and trends [1], [2]. Financial analysts used to regularly forecast stock markets in the past. However, as learning techniques have advanced, data scientists have begun to address prediction issues. Additionally, computer scientists are now utilising machine learning techniques to raise the efficiency and precision of prediction models. The next step in developing prediction models with greater performance was to use deep learning [3, 4]. The task of predicting the stock market is difficult,

and data scientists frequently run into issues while attempting to create a predictive model. The instability of the stock market and the link between investing psychology and market behaviour are two major problems that are brought on by these factors [5]. It is obvious that unforeseen factors, such as a company's reputation or a nation's political climate, can alter the direction of the stock market. Therefore, the trend of stock values and the index can be anticipated if the data obtained from stock values is effectively preprocessed and appropriate algorithms are used. Machine learning and deep learning techniques can aid traders and investors in making decisions in stock market prediction systems. These techniques aim to detect and understand patterns among massive volumes of data automatically. The algorithms are capable of effective self-learning and can deal with forecasting price changes to enhance trading tactics. [6].

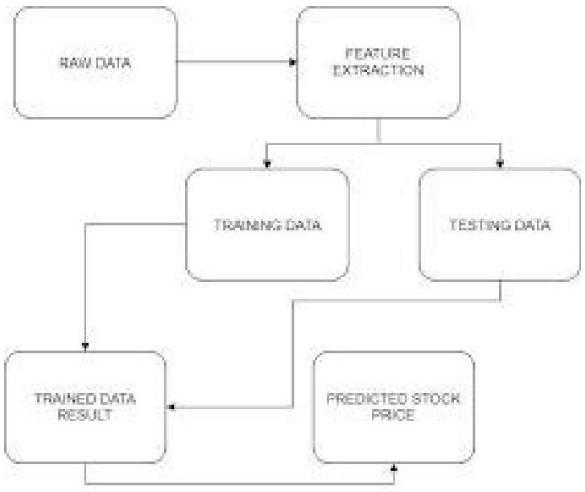

Fig.1: Example figure

Since the stock market is unpredictable, complex, and constantly changing, making it impossible to make accurate predictions, it has long piqued the curiosity of investors and researchers. The act of attempting to anticipate the future value of a company's shares or another financial instrument traded on an exchange is known as stock market prediction. A stock price forecast that is accurate could result in a sizable profit. According to the efficient-market hypothesis, stock prices accurately reflect all information that is currently known, and any price fluctuations that are not based on recently disclosed information are therefore intrinsically unpredictable. Others disagree, and individuals who hold this opinion claim to have access to a number of technologies and methods that enable them to learn about future prices. One of the most challenging tasks is predicting how the stock market will behave. Prediction is difficult due to the stock markets' inherent volatility on a global scale. There are so many variables at play in the prediction, including technological versus physical variables, rational versus irrational behaviour, etc. These factors all work together to make share values unpredictable and highly challenging to forecast accurately. Machine learning techniques have the potential to uncover patterns and insights we hadn't noticed previously using features like the most recent announcements about a firm, their quarterly sales statistics, etc. These can be utilised to produce incredibly accurate forecasts.

2. LITERATURE REVIEW

M. R. Hassan, B. Nath, and M. Kirley et al., claimed that In order to forecast the behaviour of the financial markets, we develop and put into practise a fusion model in this research that combines the Hidden Markov Model (HMM), Artificial Neural Networks (ANN), and Genetic Algorithms (GA). The

developed technology can be utilised for in depth analysis of the stock market. The daily stock prices are converted using ANN into separate sets of values that are then input to HMM. To optimise the HMM's starting settings, we look to GA. To find and locate related patterns in the historical data, the trained HMM is used. Calculated are the price variations between the matching days and the corresponding next day. To make a forecast for the necessary next step, a weighted average of the price differences of identical patterns is finally obtained. For a variety of securities in the IT sector, forecasts are obtained, and they are contrasted with a traditional forecasting method.

W. Huang, Y. Nakamori, and S.-Y. Wang et al., said that stock market investor sentiment plays a significant effect. In addition to stock market data, user-generated writing on the Internet serves as a valuable source for reflecting investor psychology and making stock price predictions. This study incorporates sentiment analysis into an SVM-based machine learning technique. Furthermore, we develop more accurate and realistic sentiment indexes by taking the day-of-week influence into account. Empirical findings show that, after introducing sentiment factors, the accuracy of predicting the SSE 50 Index's movement direction can increase by 18.6%, reaching a maximum of 89.93%. In the meanwhile, our model aids in the decision-making of investors. These findings also suggest that mood can be used as a leading indicator of the stock market and likely contains valuable information about the asset's basic values.

J. Sun and H. Li, et al., noted that Financial distress prediction (FDP) is crucial for both internal and external aspects of businesses. Although there is a wealth of literature that provides in-depth examination of the single classifier FDP approach, the ensemble method for FDP has only recently appeared and requires more research. In FDP, the Support Vector Machine (SVM) performs better than other single classifier approaches. This study makes a contribution by introducing a novel FDP approach based on SVM ensemble, where candidate single classifiers are trained using SVM algorithms with various kernel functions on various feature subsets of a single initial dataset. The filter feature selection/extraction techniques of stepwise multi discriminant analysis (MDA), stepwise logistic regression (logit), and principal component analysis (PCA) are used. SVM kernels such as linear, polynomial, RBF, and sigmoid are also utilised. The algorithm for picking SVM ensemble's base classifiers from candidate ones is built by considering both individual performance and diversity analysis. The combination process uses weighted majority voting based on the cross validation accuracy of the base classifiers on the training dataset. When the number of base classifiers in the SVM ensemble is properly selected, experimental results show that SVM ensemble is much better than individual SVM classifier. Additionally, when an individual SVM classifier is used, it demonstrates that RBF SVM based on features chosen by stepwise MDA is a good option for FDP.

P. Ou and H. Wang, et al., For market dealers or investors to maximise their earnings, the ability to predict the direction of the stock or index price with

accuracy is essential. It has been successfully demonstrated that data mining techniques can produce highly accurate stock price movement forecasts. Today, traders must employ a variety of forecasting approaches rather than just one to obtain different indications and more information about the direction of the markets. Ten distinct data mining approaches are presented and used in this research to forecast price movement of the Hang Seng index of the Hong Kong stock market. Linear discriminant analysis (LDA), Quadratic discriminant analysis (QDA), K-nearest neighbour classification, Nave Bayes based on kernel estimation, Logit model, Tree-based classification, neural network, Bayesian classification with Gaussian process, Support vector machine (SVM), and Least Squares support vector machine are some of the techniques (LS-SVM). According on experimental findings, the SVM and LS-SVM produce the best predictive performances of all the models. In terms of hit rate and error rate criteria, SVM is superior to LS-SVM for in-sample prediction, whereas LS-SVM is superior to SVM for out-of-sample forecasts.

F. Liu and J. Wang, et al., Financial time series prediction has long focused on stock market forecasting. In this research, using an enhanced Legendre neural network, we examine and forecast price fluctuation. A random time strength function is added to the forecasting model to provide a weight for each historical data point since we assume that investors decide their investing positions by analysing historical stock market data, which has the potential to affect the volatility of the current stock market. A random process is used to develop the impact strength of historical data on the market. To describe the behaviour of the time strength, a tendency function and a random Brownian volatility function are applied; in this case, Brownian motion gives the model the appearance of random movement while preserving the original fluctuation. Additionally, empirical research is conducted to examine the established model's ability to anticipate changes in SAI, SBI, DJI, and IXIC, and the related statistical comparisons of the aforementioned market indexes are also displayed.

3. METHODOLOGY

Deep learning-based efficient stock exchange forecasting in a local and global event is sentiment-based. They take into account four nations from the list of developed, emerging, and undeveloped economies: the United States, Hong Kong, Turkey, and Pakistan. We looked at how many significant occurrences between 2012 and 2016 affected stock markets. To determine the sentiment analysis for each of these occurrences, we use the Twitter dataset. The dataset used to calculate the event sentiment is made up of 11.42 million tweets. For stock exchange forecasting, we have utilised support vector regression, deep learning, and linear regression.

Disadvantages:

• Since this procedure takes time, deep learning model training is sluggish.

Recurrent Neural Network (RNN) and Long Short-Term Memory are two of the potent deep learning techniques we offer, out of the nine machine learning techniques (LSTM). Our input values are 10 technical indicators from ten years of historical data, and two

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

methods are intended for using them. First, stock trading values are used to calculate the indicators, and then, before use, the indicators are converted to binary data. Based on the input methods, three metrics are used to assess each prediction model.

Advantages:

1. The finest deep learning techniques are those for binary data assessment. It can accelerate and streamline this procedure.

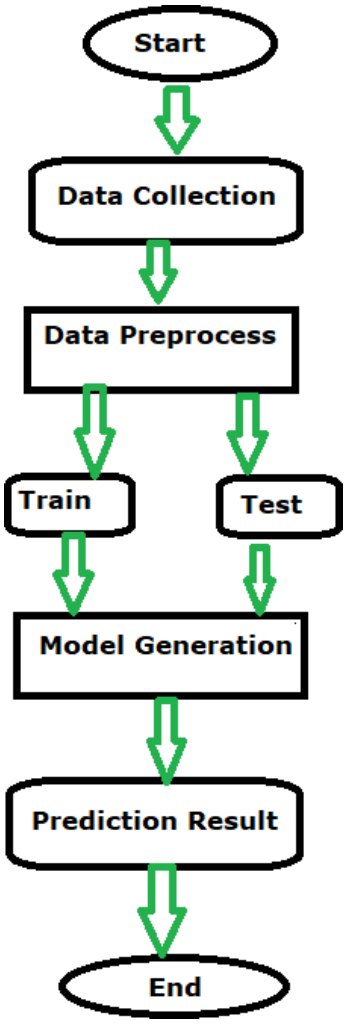

Fig.2: System architecture

MODULES:

To implement aforementioned project we have designed following modules

- Upload stock dataset: Using this module we will upload dataset into the application
- Preprocess dataset: Using this module we will read the data from the dataset for further processing.
- Feature extraction: using this module we will extract important features from the dataset
- Generate algorithms: Using this module algorithms will trained on loaded dataset
- Predict result: using this module predict the final outcome

4. IMPLEMENTATION

SVM Algorithm: Using different machine learning methods depending on the dataset, we forecast and classify data using machine learning. Support Vector Machine, sometimes known as SVM, is a linear model used to solve classification and regression issues. It works well for many real-world issues and can solve both linear and non-linear problems. The SVM concept is straightforward: The method divides the data into classes by generating a line or hyperplane. The radial basis function kernel, or RBF kernel, is a well-liked kernel function in machine learning that is utilised in a variety of kernelized learning techniques. It is frequently employed in support vector machine classification in particular. For a straightforward illustration, consider a hyper plane as a line that linearly separates and classifies a

set of data for a classification problem with only two features (such as the figure above).

Random Forest Algorithm: Because it is an ensemble algorithm, it will internally combine different classifier algorithms to create a precise classifier model. Internally, this method will create its training model for classification using the decision tree technique.

Decision Tree Algorithm: This algorithm will create a training model by grouping together all data that are comparable in a single branch of the tree and continuing until all records are grouped together. We'll refer to the entire tree as the classification train model.

Gradient Boosting Algorithm: A class of machine learning techniques known as gradient boosting classifiers combines a number of weak learning models to produce a powerful predicting model. Gradient boosting frequently makes use of decision trees. Gradient boosting models are gaining popular due of their ability at categorising difficult datasets, and have lately been used to win multiple Kaggle data science contests.

Deep Learning ANN Algorithm: A computational model based on the structure and operations of biological neural networks is known as an artificial neuron network (ANN). Because a neural network adapts or learns based on input and output, information that goes through the network impacts how the ANN is structured.

Long short-term memory (LSTM): in the study of deep learning is an artificial recurrent neural network (RNN) architecture. LSTM features feedback connections as opposed to typical feedforward neural networks. It can analyse whole data sequences in addition to single data points (like photos) (such as speech or video). For instance, LSTM can be used for tasks like linked, unsegmented handwriting identification, speech recognition, and network traffic anomaly detection, or IDSs (intrusion detection systems).

CNN algorithm: A CNN is a particular type of network design for deep learning algorithms that is utilised for tasks like image recognition and pixel data processing. Although there are different kinds of neural networks in deep learning, CNNs are the preferred network architecture for identifying and recognising objects. Convolutional neural networks (CNN or ConvNet) are a subclass of neural networks that are mostly employed in voice and image recognition applications. Its built-in convolutional layer minimises the high dimensionality of images without losing its information. CNNs are therefore very well suited for this use case.

5. EXPERIMENTAL RESULTS

6. CONCLUSION

The goal of this study was to use machine learning and deep learning algorithms to anticipate stock market behaviour. The dataset was based on ten years of historical records with ten technical features, and it consisted of four stock market groups from the Tehran Stock Exchange, including diversified financials, petroleum, non-metallic minerals, and

basic metals. Additionally, two deep learning techniques (RNN and LSTM) and nine machine learning models (Decision Tree, Random Forest, Adaboost, XGBoost, SVC, Naive Bayes, KNN, Logistic Regression, and ANN) were used as predictors. We considered two approaches for continuous data and binary data as model input values, and we used three classification metrics for assessments. Our experimental work demonstrated that using binary data instead of continuous data significantly improves the performance of models. Indeed, our best models for both approaches were deep learning algorithms (RNN and LSTM). As an extension, LSTM, Linear Regression, Lasso Regression, Ridge Regressor, Xgboost, Voting Regression, Decision Tree, Random Forest, SVM, Stacking Regression, Adaboost, SGDregressor, Adaboost, Catboost, LightBoost, Voting Regression-[catboost, lightboost], Stacking Regressor - [catboost, lightboost], are used.

7. FUTURE WORK

Future study can concentrate on merging data from sentiment analysis of stocks with numerical values relating to historical stock values to forecast stock prices. By utilising both information, more efficient stock recommendation algorithms may also be created. Further use of deep learning-based methodologies can lead to more effective feature extraction methods. It is possible to develop stock prediction engines using graph knowledge techniques, but research needs focus on the complexity and gradient of graphs with a lot of nodes. Our survey provides information on potential future research areas in this field.

REFERENCES

[1] M. R. Hassan, B. Nath, and M. Kirley, ‘‘A fusion model of HMM, ANN and GA for stock market forecasting,’’ Expert Syst. Appl., vol. 33, no. 1, pp. 171–180, Jul. 2007.

[2] W. Huang, Y. Nakamori, and S.-Y. Wang, ‘‘Forecasting stock market movement direction with support vector machine,’’ Comput. Oper. Res., vol. 32, no. 10, pp. 2513–2522, Oct. 2005.

[3] J. Sun and H. Li, ‘‘Financial distress prediction using support vector machines: Ensemble vs. Individual,’’ Appl. Soft Comput., vol. 12, no. 8, pp. 2254–2265, Aug. 2012.

[4] P. Ou and H. Wang, ‘‘Prediction of stock market index movement by ten data mining techniques,’’ Modern Appl. Sci., vol. 3, no. 12, pp. 28–42, Nov. 2009.

[5] F. Liu and J. Wang, ‘‘Fluctuation prediction of stock market index by legendre neural network with random time strength function,’’ Neurocomputing, vol. 83, pp. 12–21, Apr. 2012

[6] Yadav, A. and Vishwakarma, D.K., 2020. Sentiment analysis using deep learning architectures: a review. Artificial Intelligence Review, 53(6), pp.4335-4385.

[7] Sulandari, W., Suhartono, Subanar and Rodrigues, P.C., 2021. Exponential Smoothing on Modeling and Forecasting Multiple Seasonal Time Series: An Overview. Fluctuation and Noise Letters, p.2130003.

[8] Kumar, I., Dogra, K., Utreja, C. and Yadav, P., 2018, April. A comparative study of supervised machine learning algorithms for stock market trend prediction. In 2018 Second International Conference

on Inventive Communication and Computational Technologies (ICICCT) (pp. 1003- 1007). IEEE.

[9] Ingle, V. and Deshmukh, S., 2016, August. Hidden Markov model implementation for prediction of stock prices with TF-IDF features. In Proceedings of the International Conference on Advances in Information Communication Technology & Computing (pp. 1-6).

[10] Singh, Sukhman, Tarun Kumar Madan, J. Kumar and A. Singh. "Stock Market Forecasting using Machine Learning: Today and Tomorrow." 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) 1 (2019): 738-745.

[11] Pahwa, K. and Agarwal, N., 2019, February. Stock market analysis using supervised machine learning. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) (pp. 197-200). IEEE.

[12] Misra, Meghna, Ajay Prakash Yadav and Harkiran Kaur. "Stock Market Prediction using Machine Learning Algorithms: A Classification Study." 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE) (2018): 2475-2478.

[13] Vats, P. and Samdani, K., 2019, March. Study on Machine Learning Techniques In Financial Markets. In 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1-5). IEEE.

[14] Song, Y. and Lee, J., 2019, December. Design of stock price prediction model with various configurations of input features. In Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing (pp. 1-5).

[15] Werawithayaset, P. and Tritilanunt, S., 2019, November. Stock Closing Price Prediction Using Machine Learning. In 2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE) (pp. 1-8). IEEE.

[16] Xingzhou, L., Hong, R. and Yujun, Z., 2019, July. Predictive Modeling of Stock Indexes Using Machine Learning and Information Theory. In Proceedings of the 2019 10th International Conference on E-business, Management and Economics (pp. 175-179).

[17] S. Sarode, H. G. Tolani, P. Kak and C. S. Life, "Stock Price Prediction Using Machine Learning Techniques," 2019 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 2019, pp. 177-181. doi: 10.1109/ISS1.2019.8907958

[18] Jeevan, B., Naresh, E. and Kambli, P., 2018, October. Share Price Prediction using Machine Learning Technique. In 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C) (pp. 1-4). IEEE.

[19] Sharma, V., Khemnar, R., Kumari, R. and Mohan, B.R., 2019, September. Time series with sentiment analysis for stock price prediction. In 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT) (pp. 178-181). IEEE.

[20] Patil, P., Wu, C.S.M., Potika, K. and Orang, M., 2020, January. Stock market prediction using an ensemble of graph theory, machine learning and deep learning models. In Proceedings of the 3rd International Conference on Software Engineering and Information Management (pp. 85-92).

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org