ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data: A Comparative Analysis

[1] A NAGARAJU, [2]M. DURGABHAVANI

[1](Assistant Professor), MCA, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

[2]MCA, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

ABSTRACT

The nature of stock market movement has always been ambiguous for investors because of various influential factors. This study aims to significantly reduce the risk of trend prediction with machine learning and deep learning algorithms. Four stock market groups, namely diversified financials, petroleum, non-metallic minerals and basic metals from Tehran stock exchange, are chosen for experimental evaluations. This study compares nine machine learning models (Decision Tree, Random Forest, Adaptive Boosting (Adaboost), eXtreme Gradient Boosting (XGBoost), Support Vector Classifier (SVC), Naïve Bayes, K-Nearest Neighbors (KNN), Logistic Regression and Artificial Neural Network (ANN)) and two powerful deep learning methods (Recurrent Neural Network (RNN) and Long short-term memory (LSTM).

Ten technical indicators from ten years of historical data are our input values, and two ways are supposed for employing them. Firstly, calculating the indicators by stock trading values as continuous data, and secondly converting indicators to binary data before using. Each prediction model is evaluated by three metrics based on the input ways. The evaluation results indicate that for the continuous data, RNN and LSTM outperform other prediction models with a considerable difference. Also, results show that in the binary data evaluation, those deep learning methods are the best; however, the difference becomes less because of the noticeable improvement of models' performance in the second way.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

1.INTRODUCTION

The task of stock prediction has always been a challenging problem for statistics experts and finance. The main reason behind this prediction is buying stocks that are likely to increase in price and then selling stocks that are probably to fall. Generally, there are two ways for stock market prediction. Fundamental analysis is one of them and relies on a company's technique and fundamental information like mar-ket position, expenses and annual growth rates. The second one is the technical analysis method, which concentrates on previous stock prices and values. This analysis uses historical charts and patterns to predict future prices [1], [2]. Stock markets were normally predicted by financial experts in the past time. However, data scientists have started solving prediction problems with the progress of learning techniques.

Also, computer scientists have begun using machine learning methods to improve the performance of prediction models and enhance the accuracy of predictions. Employing deep learning was the next phase in improving prediction models with better performance [3], [4]. Stock market prediction is full of challenges, and data scientists usually confront some problems when they try to develop a predictive model. Complexity and nonlinearity are two main challenges caused by the instability of stock market and the correlation between investment psychology and market behavior [5]. It is clear that there are always unpredictable factors such as the public image of companies or political situation of countries, which affect stock markets trend.

Therefore, if the data gained from stock values are efficiently preprocessed and suitable algorithms are employed, the trend of stock values and index can be predicted. In stock market prediction systems, machine learning and deep learning approaches can help investors and traders through their decisions. These methods intend to automatically recognize and learn patterns among big amounts of information. The algorithms can be effectively self-learning, and can tackle the predicting task of price puctuations in order to improve trading strategies [6].

Since recent years, many methods have been improved to predict stock market trends. The implementation of a model combination with Genetic Algorithms (GA), Artificial Neural Networks and Hidden Markov Model (HMM) was proposed by Hassan et

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

al. [7]; the purpose was transforming the daily stock values to independent groups of prices as inputs to HMM. The predictability of financial trend with SVM model by evaluating the weekly trend of NIKKEI 225 index was investigated by Huang et al. [8]. A comparison between SVM, Linear Discriminant method, Elman Back propagation Neural Networks and Quadratic Discriminant method was their goal. The results indicated that SVM was the best classifier method. New financial prediction algorithm based on SVM ensemble was proposed by Sun et al. [9].

The method for choosing SVM ensemble' s base classifiers was proposed by deeming both diversity analysis and individual prediction. Final results showed that SVM ensemble was importantly better than individual SVM for classification. Ten data mining methods were employed by Ou et al. [10] to predict value trends of Hang index from Hong Kong market. The methods involved Tree based classification, K-nearest neighbor, Bayesian classification, SVM and neural network. Results indicated that the SVM outperformed other predictive models. The value uctuations by a developed Legendre neural network was forecasted by Liu et al. [11] by assuming investors' positions and their decisions by analyzing the prior data on the values. Indeed, they examined a random function (time strength) in the prediction model. Araújo et al. [12] proposed the morphological rank linear forecasting approach to compare its results with time-delay added evolutionary forecasting approach and multilayer perceptron networks. From the above research background, it is clear that each of the algorithms can effectively solve stock prediction problems. However, it is vital to notice that there are specific limitations for each of them. The prediction results not only are affected by the representation of the input data but also depend on the prediction method. Moreover, using only prominent features and identifying them as input data instead of all features can noticeably develop the accuracy of the prediction models. Employing tree-based ensemble methods and deep learning algorithms for predicting the stock and stock market trend is a new area of research activities. In light of employing bagging and majority vote methods, Tsai et al. [13] used two different kinds of ensemble classifiers, such as heterogeneous and homogeneous methods. They also consider macroeconomic features

and financial ratios from Taiwan stock market to examine the performance of models. The results demonstrated that with respect to the investment returns and prediction accuracy, ensemble classifiers were superior to single classifiers. Ballings et al. [14] compared the performance of Ada Boost, Random Forest and kernel factory versus single models involving SVM, KNN, Logistic Regression and ANN. They predict European company's prices for oneyear ahead. The nal results showed that Random Forest outperformed among all models. Basak et al. [15] employed XGBoost and Random Forest methods for the classification problem to forecast the stock increase or decrease based on previous values. Results showed that the prediction performances have advanced for several companies in comparison with the existing ones. For examining macroeconomic indicators to accurately predict stock market for one-month ahead, Weng et al. [16] improved four ensemble models, boosting regressor, bagging regressor, neural network ensemble regressor and random forest regressor. Indeed, another aim was employing a hybrid way of LSTM to prove that the macroeconomic features are the most successful predictors for stock market. Moving on using deep learning algorithms, a deep neural network algorithm with the transaction records and public market data was investigated by Long et al. [17] to assess stock price trends. Their nal results indicated that bidirectional LSTM could forecast the future of market for investors, and the technique attained the greatest performance. The employment of RNN and CNN algorithms was examined by Rekha et al. [18] to compare the accuracy of those with real values from stock markets. LSTM with an automatic encoder and LSTM with an embedded layer were utilized by Pang et al. [19] to acquire better stock market estimations. The result of experimental works indicated that LSTM with an embedded layer outperformed for the Shanghai composite index with 57.2% accuracy. The deep convolutional LSTM algorithm was employed by Kelotra and Pandey [20] to ef-

ciently calculate stock market movements. They used a model with Rider-based monarch buttery optimization method and gained the RMSE and MSE of 2.6923 and 7.2487. A forecasting LSTM model and an overtting prevention LSTM module were suggested by Baek and Kim [21] to predict stock market. They showed that using the

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

overtting prevention module make results more accurate. Using a hybrid method of LSTM and GA was presented by Chung and Shin [22] to develop a new stock market prediction method. Theirresults indicated that the method outperformed the benchmark model. Overall, regarding the above literature, prior studies often concentrated on macroeconomic or technical features with recent machine learning methods to detect stock index or values movement without considering appropriate preprocessing methods. Tehran's stock market has been greatly popular lately due to the remarkable growth of the main index in the last decade. The important reason behind that is privatizing most of the state-owned in the Iranian constitution rms under the general policies of article 44. The shares of lately privatized rms can be bought by ordinary people under particular conditions.

The market has some special features compared to other country's stock markets; for example, dealing price limitation that is 5% of opening price for every index in each trading day. This matter hampers scatter market shocks and irregular market uctuations, political matters, etc. over a particular time and could form the market smoother. However, the effect of fundamental parameters on the market is considerable and the prediction task of future movements is not easy [23]. This study employed stock market groups (that are important for traders) to investigate the task of predicting future trends. In spite of remarkable progress in Tehran stock market in the recent decade, there has been not adequate papers on the stock price predictions and trends via novel machine learning algorithms. However, a paper has been published recently by Nabipour et al. [23] where they employed tree based models and deep learning algorithms to estimate future stock prices from 1 day ahead to 30 days ahead as a regression problem. The experimental results indicated that LSTM (as the superior model) could successfully predict values (from Tehran Stock Exchange) with the lowest error. In this research, we concentrate on comparing prediction performance of nine machine learning models (Decision Tree, Random Forest, Adaboost, XGBoost, SVC, Naïve Bayes, KNN, Logistic Regression and ANN) and two deep learning methods (RNN and LSTM) to predict stock market movement. Ten technical indicators are utilized as inputs to our models. Our study includes two different approaches for inputs,

continuous data and binary data, to investigate the effect of preprocessing; the former uses stock trading data (open, close, high and low values) while the latter employs preprocessing step to convert continuous data to binary one. Each technical indicator has its specic possibility of up or down movement based on market inherent properties. The performance of the mentioned models is compared for the both approaches with three classication metrics, and the best tuning parameter for each model (except Naïve Bayes and Logistic Regression) is reported. All experimental tests are done with ten years of historical data of four stock market groups (petroleum, diversified financials, basic metals and non-metallic minerals), that are totally crucial for investors, from Tehran stock exchange. We believe that this study is a new research paper that incorporates multiple machine learning and deep learning methods to improve the prediction task of stock groups' trend and movement.

2.LITERATURE SURVEY

A local and global event sentiment based efficient stock exchange forecasting using deep learning Stock exchange forecasting is an important aspect of business investment plans. The customers prefer to invest in stocks rather than traditional investments due to high profitability. The high profit is often linked with high risk due to the nonlinear nature of data and complex economic rules. The stock markets are often volatile and change abruptly due to the economic conditions, political situation and major events for the country. Therefore, to investigate the effect of some major events more specifically global and local events for different top stock companies (country-wise) remains an open research area. In this study, consider four countries- US, Hong Kong, Turkey, and Pakistan from developed, emerging and underdeveloped economies' list. We have explored the effect of different major events occurred during 2012–2016 on stock markets

Predicting stock returns by classifier ensembles

The problem of predicting stock returns has been an important issue for many years. Advancement in computer technology has allowed many recent studies to utilize machine learning techniques such as neural networks and decision trees to predict stock returns. In the area of machine learning, classifier ensembles (i.e. combining multiple classifiers) have proven to be a method

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

superior to single classifiers. In order to build a better model for predicting stock returns effectively and efficiently, this study aims at investigating the prediction performance that utilizes the classifier ensembles method to analyze stock returns. In particular, the hybrid methods of majority voting and bagging are considered. Moreover, performance using two types of classifier ensembles is compared with those using single baseline classifiers (i.e. neural networks, decision trees, and logistic regression). These two types of ensembles are 'homogeneous' classifier ensembles (e.g. an ensemble of neural networks) and 'heterogeneous' classifier ensembles (e.g. an ensemble of neural networks, decision trees and logistic regression). Average prediction accuracy, Type I and II errors, and return on investment of these models are also examined **A Morphological-Rank-Linear evolutionary method for stock market prediction This work presents an evolutionary morphologicalrank-linear approach in order to overcome the random walk dilemma for financial time series forecasting.** The proposed Evolutionary Morphological-Rank-Linear Forecasting (EMRLF) method consists of an intelligent hybrid model composed of a Morphological-Rank-Linear (MRL) filter combined with a Modified Genetic Algorithm (MGA), which performs an evolutionary search for the minimum number of relevant time lags capable of a fine tuned characterization of the time series, as well as for the initial (sub-optimal) parameters of the MRL filter. Then, each individual of the MGA

3. EXISTING SYSTEM

Many models of prediction have been proposed till date to forecast the stock prices and stock market trends. Some of the machine learning techniques have been discussed in this paper. All the techniques have been classified into various subcategories like classification techniques, regression techniques, ensemble algorithms, evolutionary techniques, deep learning, hybrid models and some other additional techniques. A. Classification Techniques 1) Support Vector Machine (SVM): One of machine learning algorithms that possesses the desired features such as the decision function, usage of kernel methods and also the sparsity of the solution is known as the Support Vector Machine (SVM) technique. Random Forest Classifier Random forest classifier is a type of ensemble classifier and

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

also a supervised algorithm. It basically creates a set of decision trees, that yields some result. The basic approach of random class classifier is to take the decision aggregate of random subset decision trees and yield a final class or result based on the votes of the random subset of decision trees. B. Random Forest Algorithm Random forest algorithm is being used for the stock market prediction. Since it has been termed as one of the easiest to use and flexible machine learning algorithms, it gives good accuracy in the prediction. This is usually used in the classification tasks. Because of the high volatility in the stock market, the task of predicting is quite challenging. In stock market prediction we are using a random forest classifier which has the same hyperparameters as the decision tree. The decision tool has a model similar to that of a tree. It takes the decision based on possible consequences, which includes variables like event outcome, resource cost, and utility. The random forest algorithm represents an algorithm where it randomly selects different observations and features to build several decision trees and then takes the aggregate of the several decision trees outcomes. The data is split into partitions based on the questions on a label or an attribute. The data set we used was from the previous year's stock markets collected from the public database available online, 80 % of data was used to train the machine and the rest 20 % to test the data. The basic approach of the supervised learning model is to learn the patterns and relationships in the data from the training set and then reproduce them for the test data. Proposed Model

3.1. PROPOSED SYSTEM

In the proposed system, the system concentrates on comparing prediction performance of nine machine learning models (Decision Tree, Random Forest, Adaboost, XGBoost, SVC, Naïve Bayes, KNN, Logistic Regression and ANN) and two deep learning methods (RNN and LSTM) to predict stock market movement. Ten technical indicators are utilized as inputs to our models. The proposed study includes two different approaches for inputs, continuous data and binary data, to investigate the effect of preprocessing; the former uses stock trading data (open, close, high and low values) while the latter employs preprocessing step to convert continuous data to binary one. Each technical indicator has its specific possibility

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

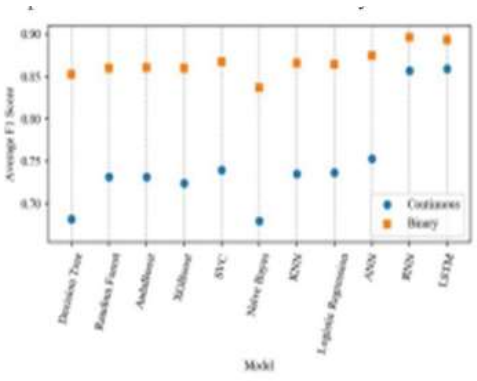
ISSN2454-9940

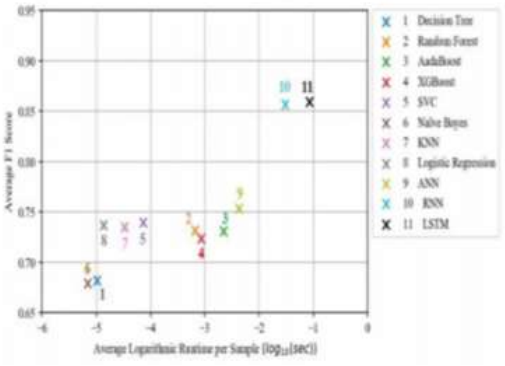
www.ijasem.org

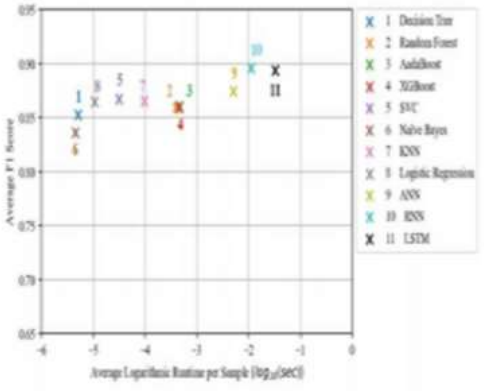
Vol 18, Issue 3, 2024

of up or down movement based on market inherent properties. The performance of the mentioned models is compared for the both approaches with three classification metrics, and the best tuning parameter for each model (except Naïve Bayes and Logistic Regression) is reported. All experimental tests are done with ten years of historical data of four stock market groups (petroleum, diversified financials, basic metals and non-metallic minerals), that are totally crucial for investors, from Tehran stock exchange. We believe that this study is a new research paper that incorporates multiple machine learning and deep learning methods to improve the prediction task of stock groups' trend and movement. Advantages • In the proposed system, each of the algorithms can effectively solve stock prediction problems.

4. OUTPUTSCREENS







5. CONCLUSION

The purpose of this study was the prediction task of stock market movement by machine learning and deep learning algorithms. Four stock market groups, namely diversied financials, petroleum, non-metallic minerals and basic metals, from Tehran stock exchange were chosen, and the dataset was based on ten years of historical records with ten technical features. Also, nine machine learning models (Decision Tree, Random Forest, Ada boost, XG Boost, SVC, Naïve Bayes, KNN, Logistic Regression and

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

ANN) and two deep learning methods (RNN and LSTM) were employed as predictors. We supposed two approaches for input values to models, continuous data and binary data, and we employed three classication metrics for evaluations. Our experimental works showed that there was a signicant improvement in the performance of models when they use binary data instead of continuous one. Indeed, deep learning algorithms (RNN and LSTM) were our superior models in both approaches..

6. REFERENCE

[1] Murphy, John J. Technical analysis of the financial markets: A comprehensive guide to trading methods and applications. Penguin, 1999

[2] Turner, Toni. A Beginner's Guide to Day Trading Online 2nd Edition. Simon and Schuster, 2007.

[3] Chen, Yingjun, and Yongtao Hao. "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction." Expert Systems with Applications 80 (2017): 340-355.

[4] Basak, Suryoday, et al. "Predicting the direction of stock market prices using tree-based classifiers." The North American Journal of Economics and Finance 47 (2019): 552-567

[5] Araújo, Ricardo De A., and Tiago AE Ferreira. "A morphological-rank-linear evolutionary method for stock market prediction." Information Sciences 237 (2013): 3-17