JOURNAL OF NORTHEASTERN UNIVERSITY
Volume 26 Issue 02, 2023 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

STOCK MARKET ANALYSIS WITH VARIOUS MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

Gummadavelly Sai kumar
M.Tech, Department of CSE, Vaagdevi College of engineering, Warangal

Dr.N.Satyavathi
Associate Professor, Head of Department of CSE, Vaagdevi college of Engineering, Warangal

ABSTRACT:
Due to a variety of influencing factors, the nature of stock market movement has long been unclear to investors. This work uses machine learning and deep learning techniques to significantly reduce the risk related to 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. Nine machine learning models (Decision Tree, Random Forest, Adaptive Boosting, eXtreme Gradient Boosting, Support Vector Classifier, Naive Bayes, K-Nearest Neighbors, Logistic Regression, and Artificial Neural Network) and two potent deep learning techniques (Recurrent Neural Network and Long Short-Term Memory) are compared in this study. 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, the results showed that those deep learning techniques are the best for evaluating binary data; however, the difference between them is decreasing as a result of the second method's performance clearly rising.
Keywords: LSTM, ANN, STOCK MARKET

INTRODUCTION:

Predicting the performance of a stock market has historically been an challenging topic for statisticians and financial analysts. This forecast is based on the idea that investors should put their money into stocks that are expected to rise in value and sell those that are expected to decrease in value. There are two main strategies for evaluating the trajectory of the stock market. One of these is called "fundamental analysis," and it uses a company's method and basic data like market position, expenses, and annual growth rates to draw conclusions. The second approach is technical analysis, which looks at stock prices and values in the past to predict future performance. In order to forecast future prices, this study looks at charts and patterns in the past.

In the past, financial professionals could reliably forecast the stock market. With the development of learning strategies, however, data scientists started resolving issues of prediction. Additionally, computer scientists have started employing machine learning

Submitted: 28/07/2023 **Accepted**: 18/08/2023

techniques to boost the effectiveness of prediction models. The next step in developing more accurate prediction models was to use deep learning [3, 4]. Data scientists often run into hurdles while building a stock market prediction model due to the inherent complexity of stock market prediction.

The unpredictable nature of the stock market and the connection between investor psychology and market behavior present two major challenges: complexity and nonlinearity.

It's clear that the movement of the stock market can be affected by a wide variety of unpredictable variables, such as public perception of businesses or the political environment in other countries. Therefore, it is possible to foretell the movement of stock prices and the index provided the data gleaned from stock prices is preprocessed well and appropriate algorithms are used. Machine learning and deep learning techniques can aid investors and traders in making decisions in the stock market.These strategies aim to learn and discover patterns in large data sets automatically. In order to enhance trading tactics, the algorithms can learn on their own and take on the responsibility of anticipating price swings. [6]. Many techniques for forecasting stock market movements have been refined in recent years. Hassan et al. [7] proposed using a model combination consisting of Genetic Algorithms (GA), Artificial Neural Networks, and Hidden Markov Model (HMM) to achieve the goal of translating the daily stock prices into separate sets of values for use as input to HMM. By analyzing the weekly trend of the NIKKEI 225 index, Huang et al. [8] looked into whether or not financial trends could be predicted using the SVM model. Their objective was to evaluate SVM, LDA, EBN, and QDA against one another. Based on the findings, SVM emerged as the superior classifier technique. Using an SVM ensemble, Sun et al. [9] introduced a new financial prediction technique. The suggested method for selecting base classifiers for an SVM ensemble takes into account both diversity analysis and individual performance. The final findings clearly demonstrated that SVM ensemble outperformed individual SVMs when it came to classification. Ou et al. [10] used ten data mining techniques to forecast the movement of the Hang index on the Hong Kong stock exchange. Tree-based, K-nearest neighbor, Bayesian, support vector machine, and neural network approaches were used. The SVM was shown to be more accurate than competing predictive models. Liu et al. [11] used historical stock price data to assume the positions and decisions of investors, then used a built Legendre neural network to predict future price fluctuations. For the forecasting model, they also looked at the impact of a random function (time strength). The morphological rank linear forecasting method was suggested by Arajo et al. [12] for evaluation against the time-delay added evolutionary forecasting method and the multilayer perceptron networks.

Based on the aforementioned literature, it's easy to see that all three of these algorithms offer viable solutions to the issue of stock prediction. It is essential to remember, however, that each has its own unique constraints. Prediction outcomes are sensitive to not only the input data representation but also the prediction technique used. Additionally, the accuracy of the prediction models can be significantly improved by employing only notable features and recognizing them as input data rather than all features.

Research has recently focused on using tree-based ensemble methods and deep learning algorithms to forecast stock and stock market trend. Tsai et al. [13], in light of their usage of bagging and majority vote approaches, employed two types of ensemble classifiers: heterogeneous and homogeneous. In order to evaluate the efficacy of the models, they take into account macroeconomic characteristics and financial ratios from the Taiwan stock market. When compared to individual classifiers, the results showed that ensemble classifiers provided greater investment returns and prediction accuracy. The results of multiple models involving SVM, KNN, Logistic Regression, and ANN were compared by Ballings et al. [14]. They make price forecasts for European corporations a full year in advance. According to the final results, Random Forest was the most effective model. Based on historical data, Basak et al. [15] used XGBoost and Random Forest algorithms for the classification problem to predict whether stock levels will rise or fall. According to the findings, some businesses' forecast performances have improved in contrast to the status quo. Weng et al. [16] enhanced four ensemble models—the boosting regressor, the bagging regressor, the neural network ensemble regressor, and the random forest regressor—to better analyze macroeconomic factors and predict the stock market a month in advance. Furthermore, a hybrid LSTM approach was used to demonstrate that macroeconomic variables are the best forecasters for the stock market.

Next, Long et al. [17] looked at a deep neural network model employing public market data and the transaction records to assess stock price movement using deep learning methods. In order to make sound financial decisions, the experimental results demonstrated that bidirectional LSTM could accurately predict stock prices. Rekha et al. [18] used convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to evaluate the accuracy of two algorithms against historical stock market data. Pang et al. [19] attempted to enhance a cutting-edge neural network technique in order to provide more accurate stock market forecasts. To analyze the stock market, they suggested using LSTM with an integrated layer and LSTM with an automatic encoder. For the Shanghai Composite Index, the findings showed that the LSTM with integrated layer performed better, with an accuracy of 57.2%, followed by the baseline model at 56.9%. The deep convolutional LSTM model was employed as a predictor by Kelotra and Pandey [20] to analyze stock market fluctuations. Minimal MSE and RMSE values of 7.2487 and 2.6923 were obtained after training the model using a Rider-based monarch butterfly optimization algorithm. Prediction LSTM and overfitting avoidance LSTM modules were proposed by Baek and Kim [21] for stock market index forecasting. The outcomes validated the high predictive quality of the suggested model.

in contrast to a model without an LSTM module for preventing overfitting. Chung and Shin [22] used an LSTM and GA hybrid to enhance a new stock market prediction model. The final results demonstrated that the LSTM network/GA hybrid model outperformed the gold standard model.

In sum, the previously mentioned literature shows that previous research generally neglected proper preprocessing approaches in favor of focusing on macroeconomic or technical aspects with modern machine learning techniques to detect stock index or value change.

Most state-owned firms are being privatized under the general policies of article 44 in the Iranian constitution, and people are allowed to buy the shares of newly privatized firms under the specific circumstances, which has led to a surge in interest in the Iranian stock market and a rise in the Tehran Price Index in recent years. A unique aspect of the Chinese stock market is a daily limit of 5% on trading activity for all indexes, which helps to smooth out the market by distributing the effects of shocks, politics, and other factors over a longer period of time. However, the impact of fundamental parameters is disproportionately significant, making it difficult to predict the market's direction in the future.
The method of predicting future patterns for stock market sectors, which are critical to investors, is the focus of this research. Despite the recent boom in Iran's stock market, not nearly sufficient attention has been paid to studying how stocks will perform in the future using the latest methods in machine learning.

LITERATURE SURVEY:

This analysis uses historical data covering the ten-year period from November 2009 to November 2019 for four stock market segments: petroleum, diversified financials, basic metals, and non-metallic minerals. All data were obtained from the www.tsetmc.com website. Figures 1-4 display the number of cases for each category that increased or decreased over the course of ten years. We chose ten technical indicators for this article based on prior research [24–26]. There are many technical indicators available for predicting stock market movement, and each of them has a unique capacity to forecast market patterns in the future. Technical indicators and their formulas are shown in Table 10 of the appendix section, and Table 11 provides summary statistics for the indicators of four stock groupings. The open, close, high, and low values of each trading day serve as the inputs for calculating indicators. Two ways are used in this work as input information. Binary data is presented with a preprocessing step to convert continuous data to binary one with respect to each indicator type. Continuous data is displayed as being based on actual time series.

A. Continuous data :

This method computes input values for prediction models for each technical indicator using formulas from Table 1. Prior to use, the indicators are normalized in the range of (0, +1) to avoid greater numbers overwhelming smaller ones. The procedure of stock trend prediction using ongoing data is shown in Figure 5.

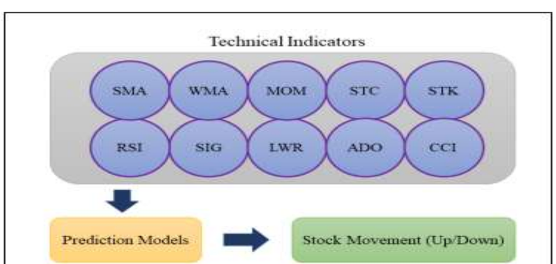


FIGURE 5. Predicting stock movement with continuous data.

Figure 1: Continous data

B. Binary data :

According to the nature and properties of each indicator, a new step is added in this method to convert continuous values of indicators to binary data. The method of stock trend prediction using binary data is shown in Figure 6. In this case, binary data is introduced via the signs +1 for an upward trend and -1 for a downward trend.

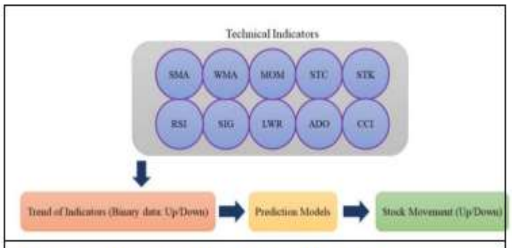


FIGURE 6. Predicting stock movement with binary data

Figure 2:Binary Data

METHODOLOGY:

Stock Price Prediction

stock Price Prediction uses machine learning to forecast the value of business shares and other financial assets traded on an exchange. To make significant profits, stock price predictions are made. It could prove challenging to forecast how the stock market will fare. The forecast is additionally affected by other variables, including biological and psychological aspects, rational and irrational conduct, etc. These all have a part in how dynamic and unpredictable share prices are. This makes it very challenging to predict stock values with any level of accuracy.

Artificial Neural Network

An ANN algorithm is shown in Figure 2 as having an input layer, a hidden layer, and an output layer. The procedure and the formulas linked to the ANN learning approach are described in the paragraphs that follow.

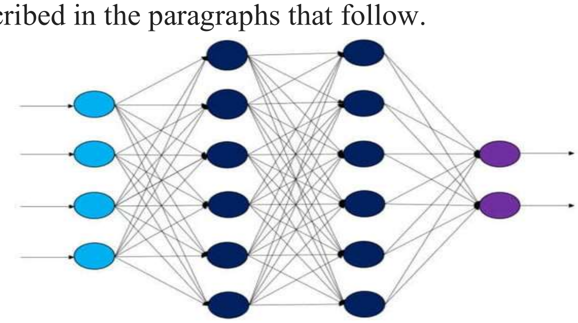


Figure 3. ANN algorithm structure.

Long Short-Term Memory

An method that makes up for RNN's weaknesses is called long short-term memory. In David Rumelhart's study from 1986, RNN was first used. It is a particular kind of ANN that is distinguished by having an internal circular data structure [21]. In order to prevent forgetting, it requires preserving earlier data and feeding it back while entering new data. RNN processes input data in its internal memory in order to make all input values related, in contrast to ANN, where all input variables are independent [22]. RNN may therefore be used to train time series data that have temporal correlation [23].

Furthermore, RNN employs "back-propagation through time" during training, which involves returning mistakes to the earliest time step for each time step [24]. Gradient vanishing, in which the learning rate is not updated and long-term patterns cannot be learned, can happen if the time step is too lengthy. Sepp Hochreiter and Jürgen Schmidhuberdm invented LSTM in 1997 [22] to address these RNN's drawbacks. Figure 3 depicts the LSTM's structural layout.

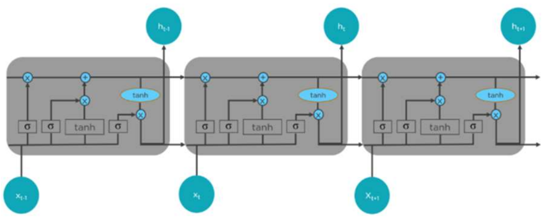


Figure 4. LSTM algorithm structure.

You can see that LSTMs have a chain-like structure in the top image. There is only one neural network layer in general RNNs. On the other hand, LSTMs have four interacting layers that communicate incredibly well.

LSTMs operate in three steps.

•The initial stage of LSTM is to choose which data should be left out of the cell at that specific time step. A sigmoid function is employed to make the decision. It computes the function while considering the current input xt, the previous state (ht-1), and both.

•The second layer has two tasks to do. The sigmoid function and the tanh function are the first and second, respectively. Which numbers to pass through are determined by the sigmoid function (0 or 1). The tanh function assigns the values passed weight, determining their relevance on a scale of -1 to 1.

•The third phase entails choosing the final product. Run a sigmoid layer first, which chooses which components of the cell state are sent to the output. The cell state must then be multiplied by the output of the sigmoid gate after being passed through the tanh function to push values between -1 and 1.

You can go on to the practical demonstration section of this course on stock price prediction using machine learning once you have a fundamental understanding of LSTM.

DATSET:

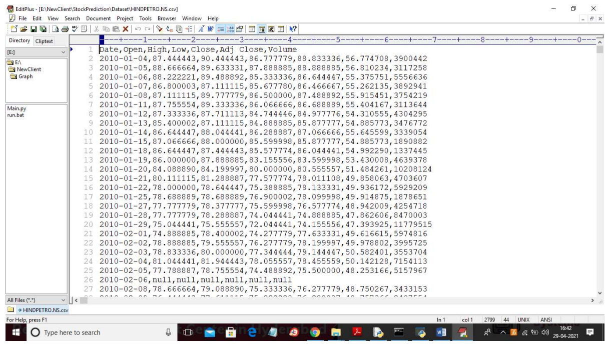


Figure 5:DATASET

In this work, the author compares the performance of four stock datasets that contain normal values (continuous) and binary data, which the author uses to convert stock values to binary data using indicators that verify the accuracy of the predictions. If the old stock price was lower than the current stock price, the dataset will be updated with 1; otherwise, -1).

Classification metrics

F1-Score, Accuracy and Receiver Operating Characteristics- Area Under the Curve (ROC-AUC) metrics are employed to evaluate the performance of our models. For Computing F1-score and Accuracy, Precision and Recall must be evaluated by Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). These values are indicated in Equations 7 and 8.

Precision = TP/TP + FP	(7)
Recall = TP/TP + FN	(8)

By calculation of above equations, F1-Score and Accuracy are defined in Equations 9 and 10.

Accuracy = TP + TN/TP + FP + TN + FN	(9)
F 1 − Score = 2 Precision Recall Precision + (10) Re call	(10)

Accuracy is a good measure, but it is not enough for all classification situations. To make sure the model is accurate, it is often necessary to look at some other measures. F1-Score may be a better measure to use if the results need to strike a balance between Recall and Precision, especially if the number of students in each class is not the same. The area under the ROC-AUC curve from prediction scores is used to calculate ROC-AUC, which is another powerful metric for classification issues.

SCREENSHOTS:

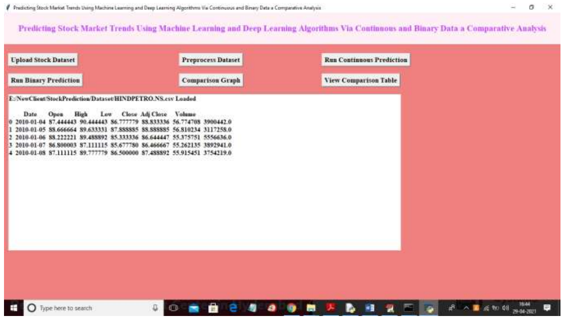
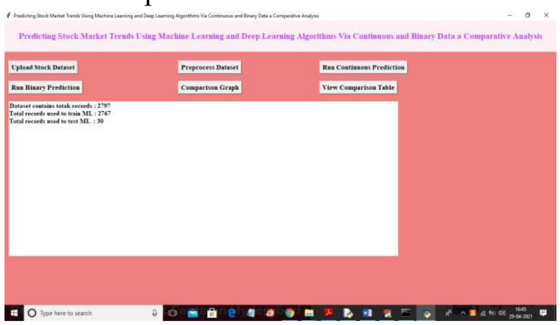


Figure 6:Preprocessing

In the screen above, the dataset was loaded, and some of the values were missing. To get clear of the missing values and split the dataset into train and test parts, click on the "Preprocess Dataset" button.



In above screen dataset contains total 2797 records and application using 2797 records for training and 30 records for testing and now train and test data is ready and now click on 'Run Continuous Prediction' button to train all algorithms with above dataset

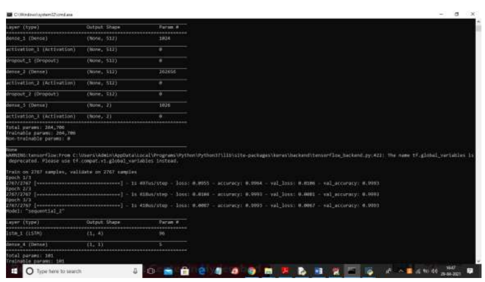


Figure 7:ANN

In above screen you can see we have created ANN and LSTM model and after building model will get predicted stock price for 30 test days

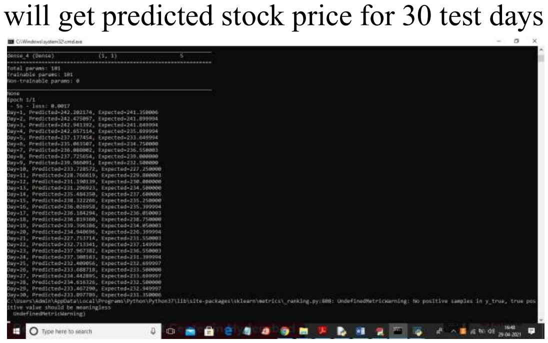


Figure 8:LSTM

Above, we can see actual and predicted values from day1 to day30, and we can verify that both prices are quite similar, indicating that LSTM accurately predicts stock prices. Below, we can see a graph comparing these values.

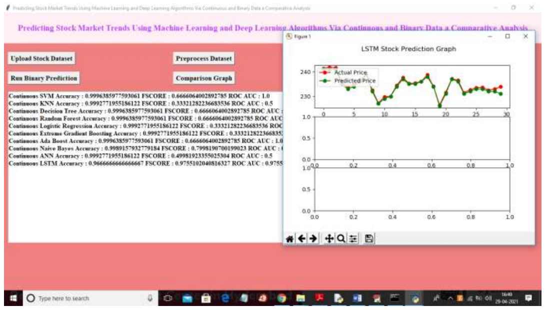


Figure 9: accuracy, FSCORE, and ROC_AUC

In the text area of the above screen, we can see the accuracy, FSCORE, and ROC_AUC values for every algorithm using continuous data. In the above graph, the x-axis shows the number of days and the y-axis shows the stock price. The red line shows the actual price and the green line shows the predicted price. We can see that the actual and predicted prices aren't too far apart, so LSTM's performance is good. Click on the "Run Binary Prediction

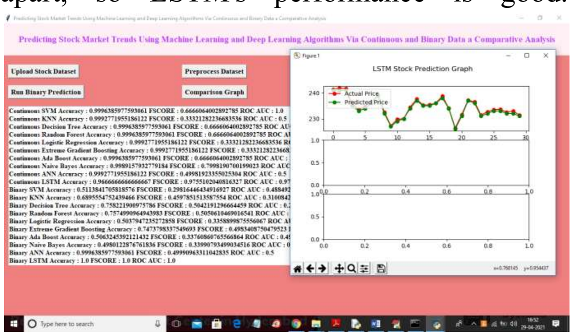


Figure 10: LSTM accuracy

In above screen binary prediction also giving best result and in text area we can see LSTM accuracy is 1.0 which means 100% accurate. Now click on 'Comparison Graph' button to get graph between all algorithms

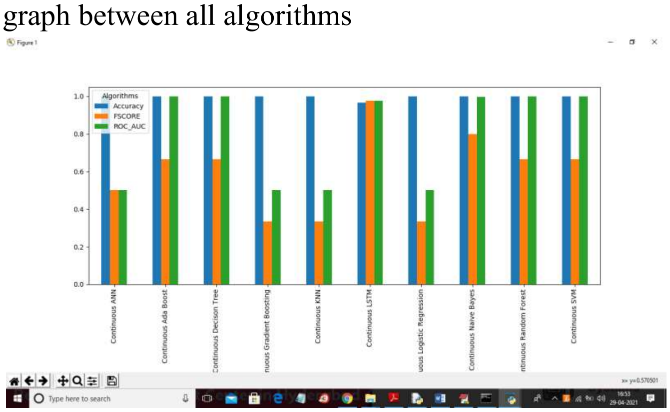
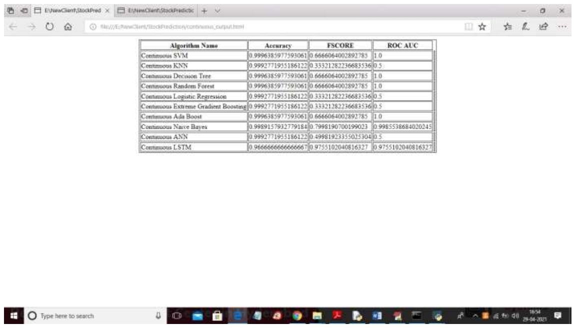


Figure 11: for continuous data ANN and LSTM Comparison Graph

In above graph for continuous data ANN and LSTM is giving better result and now click on 'View Comparison Table' button to get below screen



Algorithm Name	Accuracy	FSCORE	ROC AUC
Continuous SVM	0.9996385977593061	0.6666064002892785	1.0
Continuous KNN	0.9992771955186122	0.33321282236683536	0.5
Continuous Decision Tree	0.9996385977593061	0.6666064002892785	1.0
Continuous Random Forest	0.9996385977593061	0.6666064002892785	1.0
Continuous Logistic Regression	0.9992771955186122	0.33321282236683536	0.5
Continuous Extreme Gradient Boosting	0.9992771955186122	0.33321282236683536	0.5
Continuous Ada Boost	0.9996385977593061	0.6666064002892785	1.0
Continuous Naive Bayes	0.9989157932779184	0.7998190700199023	0.9985538684020245
Continuous ANN	0.9992771955186122	0.49981923355025304	0.5
Continuous LSTM	0.9666666666666667	0.9755102040816327	0.9755102040816327

In above screen for continuous data LSTM FSCORE is high and below we can see binary data result

Algorithm Name	Accuracy	FSCORE	ROC AUC
Binary SVM	0.5113841705818576	0.29816446434916927	0.4884925958898098
Binary KNN	0.6895554752439466	0.4597851513587554	0.31008420790858443
Binary Decision Tree	0.758221900975786	0.5042191296664459	0.24138381419298752
Binary Random Forest	0.7574990964943983	0.5050610469016541	0.24214449613180694
Binary Logistic Regression	0.5037947235272858	0.3358899875556067	0.496201553546575
Binary Extreme Gradient Boosting	0.7473798337549693	0.4983408750479523	0.2522574908362904
Binary Ada Boost	0.5063245392121432	0.33760860765566864	0.4933150372608763
Binary ANN	0.9996385977593061	0.49990963311042835	0.5
Binary LSTM	1.0	1.0	1.0

In above screen with binary data LSTM got 100% accuracy, FSCORE and ROC_AUC. Below is the binary data comparison graph between all algorithms

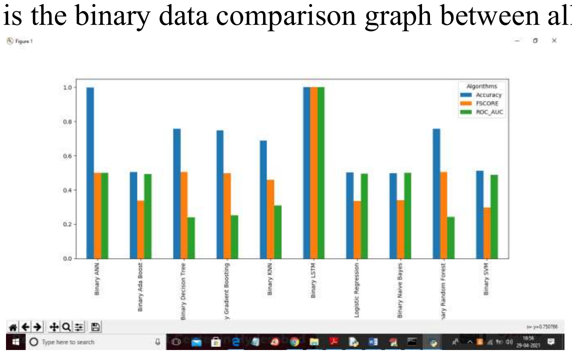


Figure 12: compare to all algorithms

In above graph LSTM is giving better output result compare to all algorithms

CONCLUSION:

The goal of this study is machine learning and deep learning systems were able to forecast how the stock market would move. Four stock market groups from the Tehran Stock Exchange were picked. These were diversified financials, petroleum, non-metallic minerals, and basic metals. The dataset was based on ten years of historical records and ten technical features. Also, as indicators, nine machine learning models (Decision Tree, Random Forest, Adaboost, XGBoost, SVC, Naive Bayes, KNN, Logistic Regression, and ANN) and two deep learning methods (RNN and LSTM) were used. We thought that models could take either continuous data or binary data as input, and we used three classification measures to evaluate them. The results of our experiments showed that models perform much better when they use binary data instead of continuous data. In fact, our best models for both methods were based on deep learning algorithms (RNN and LSTM).

REFERENCES:

[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] Maqsood, Haider, et al. "A local and global event sentiment based efficient stock exchange forecasting using deep learning." International Journal of Information Management 50 (2020): 432-451.

[4] Long, Wen, Zhichen Lu, and Lingxiao Cui. "Deep learning-based feature engineering for stock price movement prediction." Knowledge-Based Systems 164 (2019): 163-173.

[5] Duarte, Juan Benjamin Duarte, Leonardo Hernán Talero Sarmiento, and Katherine Julieth Sierra Juárez. "Evaluation of the effect of investor psychology on an artificial stock market through its degree of efficiency." Contaduría y Administración 62.4 (2017): 1361-1376.

[6] Lu, Ning. "A machine learning approach to automated trading." Boston, MA: Boston College Computer Science Senior Thesis (2016).

[7] Hassan, Md Rafiul, Baikunth Nath, and Michael Kirley. "A fusion model of HMM, ANN and GA for stock market forecasting." Expert systems with Applications 33.1 (2007): 171-180.

[8] Huang, Wei, Yoshiteru Nakamori, and Shou-Yang Wang. "Forecasting stock market movement direction with support vector machine." Computers & operations research 32.10 (2005): 2513-2522.gg

[9] Sun, Jie, and Hui Li. "Financial distress prediction using support vector machines: Ensemble vs. individual." Applied Soft Computing 12.8 (2012): 2254- 2265.

[10] Ou, Phichhang, and Hengshan Wang. "Prediction of stock market index movement by ten data mining techniques." Modern Applied Science 3.12 (2009): 28- 42.

[11] Liu, Fajiang, and Jun Wang. "Fluctuation prediction of stock market index by Legendre neural network with random time strength function." Neurocomputing 83 (2012): 12-21.

[12] Tsai, Chih-Fong, et al. "Predicting stock returns by classifier ensembles." Applied Soft Computing 11.2 (2011): 2452-2459.

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

[14] Ballings, Michel, et al. "Evaluating multiple classifiers for stock price direction prediction." Expert Systems with Applications 42.20 (2015): 7046-7056.

[15] 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.

[16] Weng, Bin, et al. "Macroeconomic indicators alone can predict the monthly closing price of major US indices: Insights from artificial intelligence, time-series analysis and hybrid models." Applied Soft Computing 71 (2018): 685-697.
[17] Long, Jiawei, et al. "An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market." Applied Soft Computing (2020): 106205.
[18] Rekha, G., et al. "Prediction of Stock Market Using Neural Network Strategies." Journal of Computational and Theoretical Nanoscience 16.5-6 (2019): 2333-2336.
[19] Pang, Xiongwen, et al. "An innovative neural network approach for stock market prediction." The Journal of Supercomputing (2018): 1-21.
[20] Kelotra, A. and P. Pandey, Stock Market Prediction Using Optimized Deep-ConvLSTM Model. Big Data, 2020. 8(1): p. 5-24.
[21] Baek, Yujin, and Ha Young Kim. "ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module." Expert Systems with Applications 113 (2018): 457-480.
[22] Chung, H. and K.-s. Shin, Genetic algorithm-optimized long short-term memory network for stock market prediction. Sustainability, 2018. 10(10): p. 3765.
[23] Kara, Yakup, Melek Acar Boyacioglu, and Ömer Kaan Baykan. "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange." Expert systems with Applications 38.5 (2011): 5311-5319.
[24] Patel, Jigar, et al. "Predicting stock market index using fusion of machine learning techniques." Expert Systems with Applications 42.4 (2015): 2162-2172.
[25] Patel, Jigar, et al. "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques." Expert systems with applications 42.1 (2015): 259-268