Published Online First: July, 2024
https://doi.org/10.21123/bsj.2024.10017
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

An Analytical Comparison of the Behavior of Machine Learning and Deep Learning in Stock Market Prediction

Hasanen S. Abdullah*[1], *Nada Hussain Ali* *[1], *Ammar Hussein Jassim*[2], *Syed Hamid Hussain*[3]

[1]Department of Computer Science, University of Technology, Baghdad, Iraq.
[2]Department of Computer Science, Collage of Science for Women, University of Baghdad, Baghdad, Iraq.
[3]Department of Computer Science, Faculty of Computing, University Technology Malaysia, Johor, Malaysia.
*Corresponding Author.

Received 24/10/2023, Revised 23/03/2024, Accepted 25/03/2024, Published Online First 20/07/2024

Abstract

Machine learning is considered a powerful technique in many applications such as classification, clustering, recognition and prediction. Deep learning is a modern, vital and superior machine learning that gives stunning performance, especially with huge data. Stock market price prediction is the process of determining the future value of a prospect of a financial instrument traded in the market, to gain a great profit a successful prediction must be conducted, in order to achieve that machine learning is used, in this article, two approaches are proposed to predict the stock market prices and movement using two datasets, the first approach employs two machine learning models (J48 & logistic regression) while the second approach based on recurrent neural network (proposed long short term memory(LSTM) model). The proposed LSTM architecture is designed and trained with inefficient optimizer, tuned hyperparameters and a good choice dropout ratio to avoid overfitting. The aim of this article is to conduct an experimental comparison between the classical machine learning approach (J48 & logistic regression) and deep learning represented by LSTM. The experimental results show that the proposed approach of LSTM outperforms other approaches with the two datasets in predicting the price and movement of the stock market.

Keywords: Stock Market Prediction, Machine learning, Deep learning, Recurrent neural network, LSTM, J48, logistic regression.

Introduction

A growing request for prediction technique has made it more helpful and powerful tool. Therefore, when it has a correct format for all data and practically true values, a high-accuracy prediction process will be obtained. An important role in the prediction process is data preprocessing and preparation. It provides potential values of data and also manipulates unnecessary, missing and gabs in data in order to get useful data. The price prediction depends on a variety of characteristics and features. Accordingly, those characteristics and features subscribe to their property for predicting a unique output named price [1].

Machine learning (ML) is a significant field of AI, that is, algorithms are searching in large data sets for targets, learning more, and achieving tasks without

Published Online First: July, 2024
https://doi.org/10.21123/bsj.2024.10017
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

telling how to formulate the problem and fix its solution [2]. The assumption application concentrates on improving prediction based on ML which can have big impacts on policy. In the problem-solving approach of prediction techniques, there needs to be a description of how ML embeds value over approaches of traditional regression [3].

At present, two main approaches are considered proffered ways in predicting the price movement of the stock market as using ML ways or deep learning (DL) ways. ML is a powerful tool to accomplish some tasks in financial transactions[4].The stock market comprises many worldwide stock exchanges. It is one of the major platforms for raising money [5]. The stock market is considered as non-linear time series data and has a big fluctuation. The time series data is a dataset that is measured during the time for acquiring some activity status [6].

The stock price movements are non-stationary, non-linear, and noisy, the price of stock is based on a variety of factors. Several factors can move the price of stock down or up including "demand-and-supply" for stock, factors of legal or political, competition from competitors, management of company, the policy of a government, the central bank policy for a country, and the price of a stock is influenced with news that linked to company [7,8].For supporting the real economy, due to the financial service system's important part, the stock market can also become the part of competitiveness for the country's core [9].Stock market is divided into two main components, "primary market" and "secondary market", primary market means that new issues are presented to market during "Initial Public Offerings". A secondary market means that the securities are traded by investors that they already possess. Linear models are used for stock market prediction, the associated problem with such models is that they are working only for a certain time series data; i.e., these models identified for a certain company will not achieve well for another[6].

The problem of predicting the stock market movement and prices efficiently relay mainly on two factors, the dataset and the model used for this task, as the success of this task highly depends on these factors, the proposed models in this work overcomes the problem of these two factors, as the proposed LSTM model can work efficiently with normal size data and huge data.

In this article, a promising approach for stock market prediction is presented based on ML and proposed model of LSTM by employing two different datasets, the main contributions of this work are:

1. Building three models that are able to efficiently predict the stock market movement and prices.

2. Conducting experimental results to show the power of deep learning in comparison with traditional machine learning models in prediction task.

3. The advantage of deep learning (based on LSTM) in dealing with huge data compared to ordinary machine learning.

In literature, several works are presented, in[10] the authors presented the ability of RNN to achieve time series forecasting. The non-linearity auto-regressive with exogenous inputs (NARX) model forecasted time series such as "EURO/ALL" and "USD/ALL "exchange rates, "Consumer Price Index (CPI)" and Interest Rate about credits in Euro. The RNN performed high-accuracy forecasting of time series. For this time series, the mean absolute error (MAE) is about 0.25 and the regression value is 0.8462. The authors in[11]proposedamachine learning model that is trained with available stock data and then used acquired knowledge to predict accurately stock prices based on a "support vector machine". In their proposal, they used data collection from various global financial markets for predicting index movements of stock. The model produced a high profit in comparison with selected benchmarks. Experimental results showed that there is typically a convergence between the original and predicted value for the stock. In[12], the authors proposed hybrid DNN architecture (CNN and LSTM) as a multi-source information fusion framework to predict stock prices. The approach evaluation was tested with stock data (from Jan. 2017 to Jan. 2020) about the Ghana Stock Exchange. The results gave a 98.31% accuracy prediction. The authors in[13]produced a method of time series prediction-based stock price and investment portfolio model. The regression scheme is executed on LSTM-basedDNN.NIFTY-50 large stock dataset experiments have been implemented. Experimental results showed that the proposed method outperforms several standard prediction methods. The results of the LSTM model according to root mean square error (RMSE), as an average of the best five tests, gave 1.715. In[14], an ANN or Feed-forward DNN and CNN are the two approaches which have been used for predicting the prices of the stock market. The ANN approach performs 97.66%, while 98.92% for the CNN approach as accuracy. The proposal was tested during the pandemic of COVID-19, which caused the stock market downfall suddenly; the study's experimental results produced 91% accuracy. The authors in[15] proposed structured framework to perform high accuracy predicted stock price that combines CNN and LSTM. As a comparison with previous works, the proposal performance of prediction results outperforms compared works. Prediction accuracy for SSACNN, SSALSTM and SACLSTM datasets starts from 71% and reaches its highest at 95.1% from 1 to 7 days. In[16], the authors presented a machine-learning approach to predict the stock market based on statistical data. The overall accuracy of the proposed prediction technique is 80.3% that is the SVM method is about 78.7% while the random forest method performs 80.8%. In[17], the authors used an LSTM, a particular ANN architecture, for predicting the closing price of the next day of the "S&P 500 index's" closing price. By using chosen input variables, LSTM Single-layer and multilayer architectures were developed. The test results showed that single-layer LSTM produced more RMSE fit prediction in comparison to multilayer LSTM, that is single-layer LSTM produced an average RMSE of about 5.411 while multilayer LSTM produced an average RMSE of about 8.637.

The rest of this work is organized as follows, methodology where the proposed models and the techniques used in the proposed models are explained. The results where a comprehensive discussion of the results is conducted and finally the conclusion, in this section the conclusions gained during this work are illustrated.

Methodology

In this section, a detailed description of the datasets, theoretical background and ML and DL models will be discussed.

Datasets Description

In the past few years, it has been an obvious trend to use deep learning in untraditional and various fields such as wisdom model building [18], stock market prediction [19], micro-content recognition [20]. Two datasets are used in this article to test the feasibility of the developed two prediction approaches, the two datasets are downloaded from the "Kaggle" repository, the first dataset namely as "Daily News for Stock Market Prediction" (DNSMP) and the second namely as "stock market predication" (SMP). For the first dataset "DNSMP", this dataset consists of two parts, the first part includes data from historical news headlines from the Reddit world news channel, the data is ranked by users' votes and the top 25 headlines according to users' votes are considered, the dataset also contains the second part which is the stock data from Dow Jones Industrial Average (DJIA) from the range (2008 to 2016). The combined file of the two parts contains 25 columns and 1990 records[21]. For the second dataset "SMP", it contains several rows grouped by the name of the company each row contains several columns including the age, the date, the price, the volume and some statistics of the price and volume, and the target. The target is defined as 1 if the close price is increased by 15% in 20 days (market days) without any loss higher than 10% from the starting price in every day in that period, 0 otherwise. The dataset includes 77 columns and 100045 rows (records)[22]. Table 1 illustrates the details of the two datasets.

Table 1. datasets description

Dataset name	Number of records	Number of columns
DNSMP	1990	25
SMP	100045	77

The Proposed Prediction System Design

In this research article, the proposed prediction system based on time series data is presented over two separate approaches to reduce the preference of

Published Online First: July, 2024
https://doi.org/10.21123/bsj.2024.10017
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

the approach that has been proposed as a contribution of this research. The proposed prediction system aims to predict the financial economy movements of counties, states or large organizations by working on understanding, analysis and anticipation for companies and sectors in business. The first approach is build based on machine learning techniques, while the second approach, which achieves the major purpose of the article, is developed to improve the quality, efficiency and other criteria in stock market prediction based on modern age approach represented by deep learning. The proposed system design presents the collective intelligence description of two main approaches.

Several scientific theories are employed in the proposed system including preprocessing, features reduction, essentials of logistic regression and J48 machine learning and a descriptive construction of DNN for prediction task.

The Stock Market Prediction Approach based on Machine Learning

In this approach, the prediction way is developed and improved according to the sequence of stages including preprocessing, features extraction and prediction machine learning to achieve the stock market (companies/sectors) financial movements and price prediction.

1. Preprocessing Stage: The preprocessing stage for the employment datasets includes several steps commensurate with their contents as well as the partitioning of the datasets. The objective of this stage is to organize the data in datasets and manipulate it to transfer raw data into the structured formula to enhance performance. The preprocessing stage goes through several steps, including normalization, this step is necessary when datasets have multiple dimensions as well as the standardization step of features causes the values to relate to each other.

The preprocessing stage includes Min-Max Scaler normalization. The implementation importance is determined by the need to reduce the sensitivity of the model for feature values in the dataset to increase the adequacy of the tested model.

$$X_{scaled}=\frac{X-\min(x)}{\max(x)-\min(x)} \ldots\ldots \text{ Eq. } 1$$ [23].

Where x_{scaled} represents the normalized attribute, X represents the current feature of a dataset, min(X) represents the minimum value for attribute X, while max(X) represents the maximum value for attribute X [23].

Also dataset splitting is done, in this step each dataset is split according to the training phase and testing phase.

2. Features Reduction Stage: the Generalized Discriminant Analysis (GDA) method is used as features extraction and reduction process. GDA is an efficient method that is involved for reducing dimensions. In the general formation, GDA pursues a nonlinear projection that simultaneously maximizes the dissimilarity between classes and minimizes or reduces the within-class non-similar to increase separation class ability.

GDA uses kernel functions to map datasets to high dimensional space of feature that leads to "nonlinear discriminant analysis" [24]. GDA has offered a very powerful approach for extracting features that is non-linear with low use of the system's resources associated with high efficiency [25]. The methods of statistical learning have been considered as the backbone for developing machine intelligence. Machine learning algorithms aims to design models which will make these models learn and enable them to achieve various tasks, such as prediction [26].

3. Prediction Machine Learning for Stock Market: in this stage, two powerful machine learning are employed as prediction models, each model runs independently from the other, which are the Logistic Regression (LR) and the J48 techniques.

The LR method is considered the famous classification model. LR is cherished and widely used in statistics, machine learning, and classification applications. The benefit of the LR model includes a powerful statistical issue and probabilistic method that helps in data analysis. LR is considered an effective classifier and strong prediction method. It was built for handling only the certainty data which is a single value for each feature. Recently, in many applications, uncertain data appeared. In an uncertainty dataset, information cannot be represented ideally as a single value thus, when dealing with uncertain data, it's difficult to achieve satisfying results unless managing these data for classification task [27].

Published Online First: July, 2024
https://doi.org/10.21123/bsj.2024.10017
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

The J48 has features such as rule derivation, continuous range value feature, and pruning of decision tree, among others. When possible, pruning of overfitting can be employed as a precision tool. J48 is used for rules creation to lead the formation of the data as a unique identity. The aim of employing this model is to stream the decision tree gradually until a balance is performed between accuracy and versatility [28].

The main property for these two techniques to work as stock market prediction models is considered as an appropriate machine learning that deals with time series data, the purpose of that is to ensure and getting well performance and acceptable results according to the ability of each model so to make sure a potential preference and comparison with the second approach which is considered the article major aim, which is described in next section.

The Proposed Stock Market Prediction Approach Based on Recurrent Neural Network

In this section, the proposed approach for the stock market prediction requires a comprehensive understanding process of the data stock price, analysis, movement and company business expectation, this comprehensive data treatment means that each dataset is handled as an overall dataset that is the DNSMP contains 1990 records which is considered a suitable dataset for the first and second approach, while the SMP dataset which is considered a huge dataset contains 100045 records, this dataset is considered as more compatible for the RNN approach but override the ability of first approach.

In recent years, deep learning has proven to be more accurate on many tasks that outperform humans. This was obvious in the gained results recently from deep learning algorithms that surpass human ability and performance [29]. The architectures of deep learning are essentially built as multi-layered structures via high-level characteristics that are computed through nonlinear functions about low-level properties [30]. Many architectures of deep learning must be proposed corresponding to prediction tasks associated with tuned hyperparameters that decide the NN stability such as in RNN [29]. RNN is a substation of NN which analyzes a stream of data by using hidden units. The real output of any application, such as regression, recognition or prediction, depends on the last computations [31]. A simple RNN architecture has three layers in its design: input, recurrent hidden and output layers, in the input layer there are N input units. The inputs are a sequence of vectors sequenced through time. These units in the input layer are connected to the hidden units in the recurrent hidden layers in a fully connected RNN. These connections between the input layer and hidden layer units are defined with a weight matrix. The units in the hidden layer are connected with recurrent connections which work as a memory in a RNN model [32]. RNN architecture determines the way information moves between different neurons and choosing the suitable design is important to have a robust learning system. RNN architecture consists of three nodes (gates), input nodes which do not have any incoming connections, output nodes that do not have outgoing connections and hidden nodes that have both ingoing and outgoing connections [33]. LSTM is a special architecture of RNN that had been originally produced by Hochreiter and Schmidhuber for overcoming the limitations of ANN sand deals with the problem of vanishing gradient which gets from a small gradient of vanishing that prevents the change of weight values effectively [34]. The behavior of LSTM depends on using the network gates which decide what will be kept and what must be ignored from series [35]. The LSTM hidden units have three gates: input gate, output gate and forget gate. The input gate decides what the unit should contain, this gate is similar to the forget gate but with different weights, the input gate has the capability of adding a candidate to the cell state which works as a saving mechanism that controls the amount of input data to be saved in the unit. The forget gate controls the amount of the raw input, the unit remembers for both the weighted raw input and weights from the previous hidden state. The final gate is the output gate which decides what part of the unit state to be returned as input [36]. Fig. 1 shows the architecture of the LSTM layer.

Published Online First: July, 2024
https://doi.org/10.21123/bsj.2024.10017
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

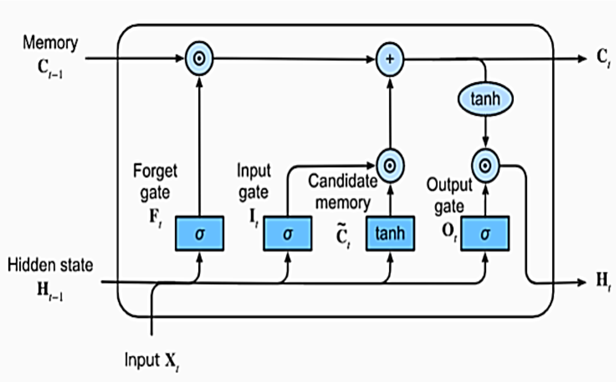


Figure 1. LSTM Architecture [37].

A proposed Recurrent Neural Network (RNN) is suggested to handle the dataset (DNSMP or SMP) in order to deduce the prediction results. This approach includes three stages, the preprocessing stage, the features reduction stage and the prediction stage. The suggested RNN model handles the problem of the stock market prediction with any dataset size since the suggested RNN deep learning is modelled for this purpose.

1. Preprocessing Stage: in this approach, the preprocessing stage contains normalization (min-max scaler) and standardization steps, since there is no dataset splitting because the RNN need a huge dataset to work under dataset availability conditions and the RNN takes upon itself the task of dataset splitting according to only the ratio of each phase (training and testing phases) that is previously determined under RNN environment condition.
2. Features Reduction Stage: the same stage of feature reduction (by employing the GDA method) that is used in the first approach will be used in the second approach.
3. RNN Prediction for Stock Market: in this stage, the proposed prediction approach is build based on the proposed RNN (LSTM layer), this stage is divided into two phases, the training phase to train the model and then save it after reaching learning, and the testing phase to check the feasibility of the proposed LSTM. RNN is a type of ANN which can process series data and recognize patterns to predict the final output (results). The logic behind RNN is storing the output of a current layer to feed it back to the input layer in order to predict the output. The RNN has fixed weight values inside each layer while the feed-forward (like CNN) tends to have different weights in each neuron. The proposed LSTM is built to train a model that is used to predict time series stock market data based on historical data which produces a prediction for future results. The stock market dataset is fed to the proposed LSTM model to build a machine learning model that can predict future stock price and movement in the business market based on the knowledge the model leans it during the training phase depend on historical data, this process helps the investors to make decision for their investments. Fig. 2 represents the proposed architecture of RNN model-based LSTM layers for solving the stock market prediction problem to implement the prediction task of the second approach. The dataset is divided into 75% training data and 25% testing data, Min-Max data normalization is used to scale the data between the 0-1 range, in which a linear transform is used on the original data, this technique keeps the relation among data unattached but decreases the standard deviation value which restrain the outlier effect. The proposed LSTM model architecture is as follows:

- One input layer is used that takes the features of the stock market dataset as input.
- Five LTSM layers are used with one step (step is a parameter that determines the number of previous steps that the model saves to use for predicting the next step).
- Sigmoid activation function after each LSTM layer.
- One dens layer (fully connected layer) with softmax activation function.

-Adam optimizer with a learning rate of 0.001, decay of 0.9, decay2 of 0.99.

The model is trained for 100 epochs and a batch size of 1. The mean square error is used as an evaluation metric.

Published Online First: July, 2024
https://doi.org/10.21123/bsj.2024.10017
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

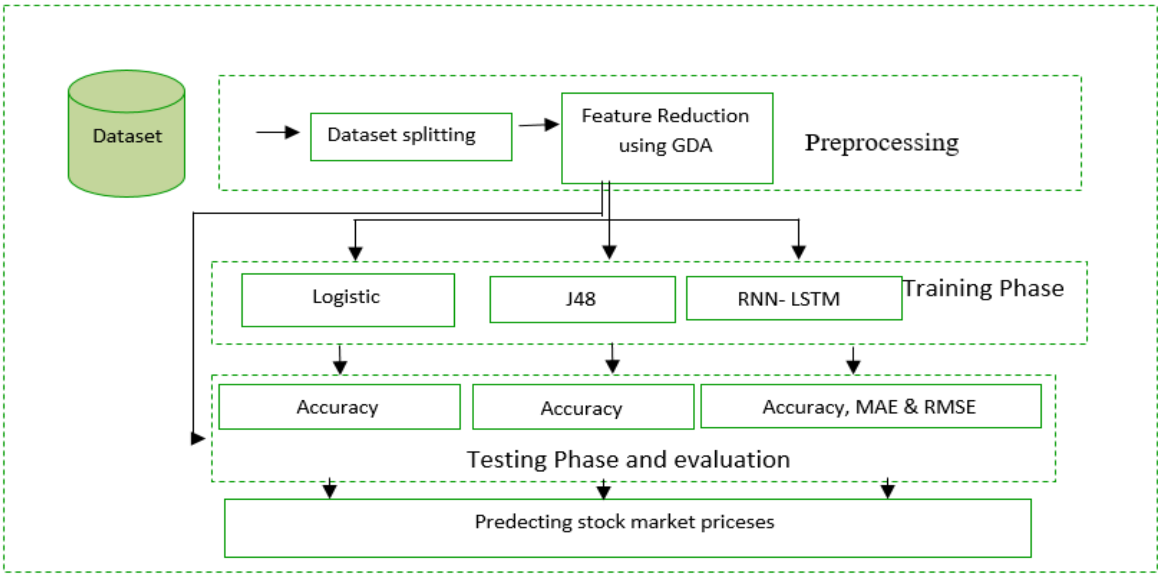


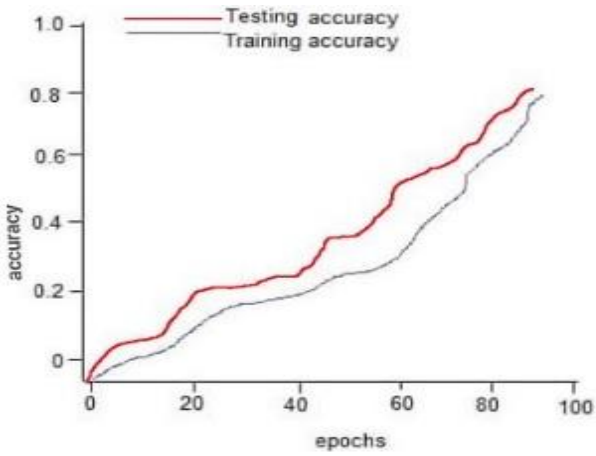
Figure 2. The architecture of RNN model based LSTM

Results and Discussion

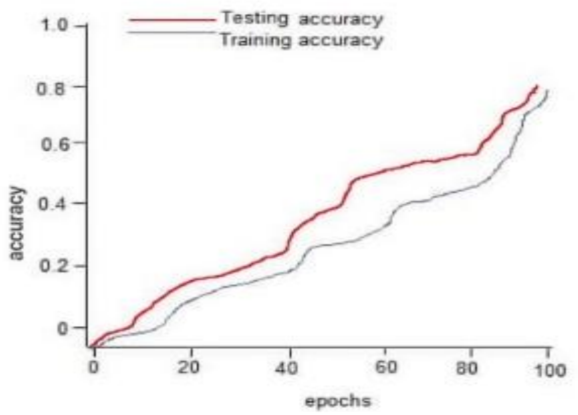
The implementation of the proposed system includes two approach results, which are the prediction stock market results of the machine learning (LR and J48) approach and the proposed approach of deep learning represented by LSTM which are achieved using two datasets, DNSMP and SMP. For the ML approach, the results are subjected to the accuracy measure (scale) to determine the efficiency of each model while for the DL approach; the results are subjected to the accuracy measure (scale) and error rate (MAE, and RMSE) to determine the efficiency of the RNN (proposed LSTM).

Results of DNSMP Dataset

The DNSMP Dataset is considered an acceptable dataset size to work with machine learning and deep learning; this dataset has three sectors to be predicted for the prices and movement of the stock market that represents it. The sectors of DNSMP are DN headlines, DJIA and Combined DN & DJIA. As a machine learning, the prediction of the stock market prices and movement is accomplished based on an approach that uses two ML models (LR and J48), the experimental results of this approach are shown in Table 1 bellow which are obtained with the use of the DNSMP dataset. While the deep learning approach that represented by a proposed LSTM (RNN) which is the fundamental objective of this article, the experimental results of this approach are obtained with the use of the DNSMP dataset as shown in Table 2 below. Fig. 3 shows the training and testing accuracy for the LR model

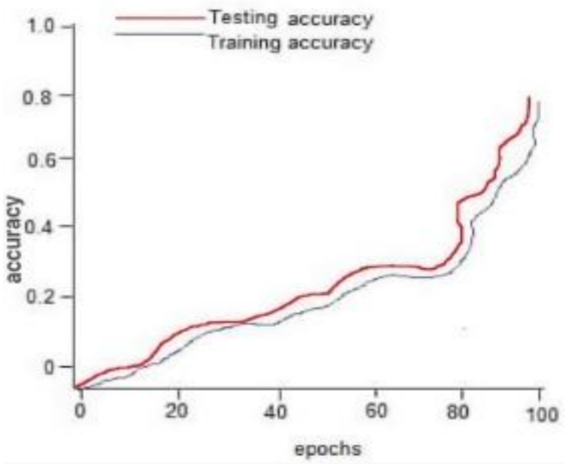


(a)



(b)

Published Online First: July, 2024
https://doi.org/10.21123/bsj.2024.10017
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

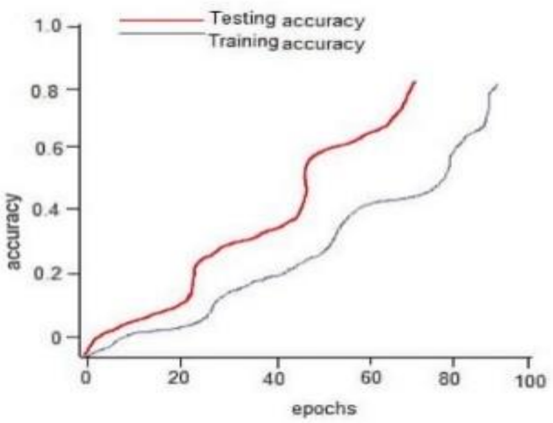


(c)

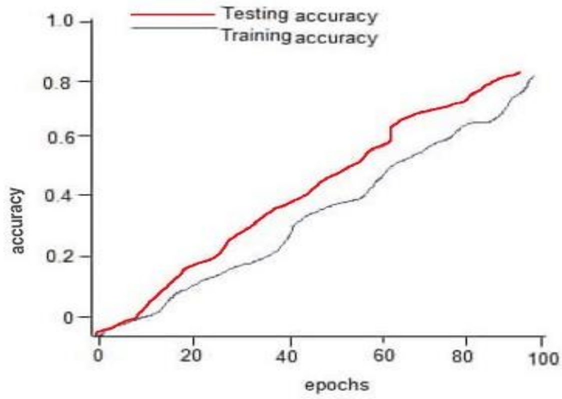
Figure 3.LR prediction accuracy, (a) prediction accuracy of the DN headlines, (b) the prediction accuracy of DJIA, (c) the accuracy of the combined DN & DJIA.

From Fig. 3, it can be noticed that the learning and testing accuracy are over passing 80% with 100 epochs and the accuracy boosted after the epoch 60 when using the LR model, as it seen in Fig. 3-a and Fig. 3-b the training performance is smoother than the testing performance gradually reaching the peak of training accuracy while the testing performance suffered from few ups and downs reaching the testing accuracy peak this is due to the nature of LR model learning behavior, this issue has slightly vanished when the two dataset are combined as it can be seen in Fig. 3-c.

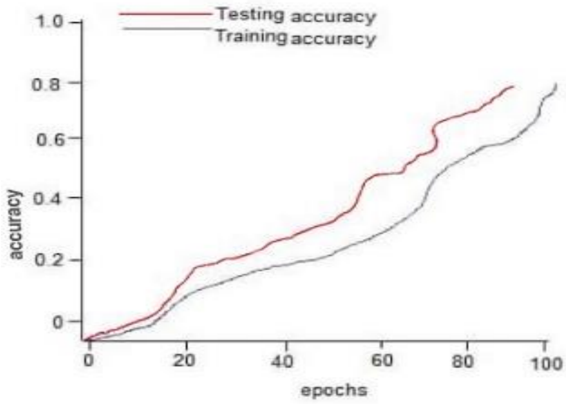
While Fig. 4 shows the training and testing accuracy for theJ48 model.



(a)



(b)

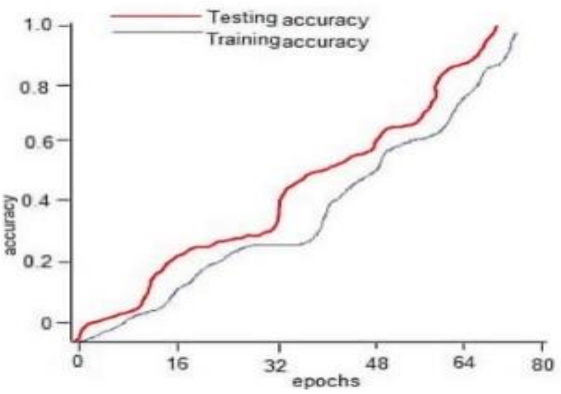


(c)

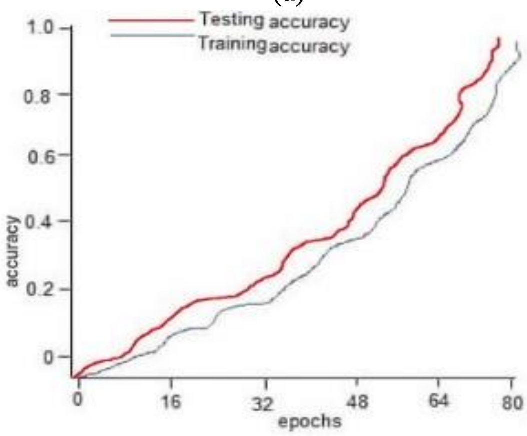
Figure 4. J48 prediction accuracy, (a) prediction accuracy of the DN headlines, (b) the prediction accuracy of DJIA, (c) the accuracy of the combined DN & DJIA.

In Fig. 4 the accuracy stably rises along the epochs reaching an accuracy of over 80% during 100 epochs when using theJ48 model nevertheless the model suffered with the DN dataset as the testing accuracy outperforms the training accuracy reaching the peak sooner than expected which cause the testing accuracy to not improve further making this model to achieve the least accuracy among the three proposed models.
As for Fig. 5, it shows the training and testing accuracy for the LSTM model.

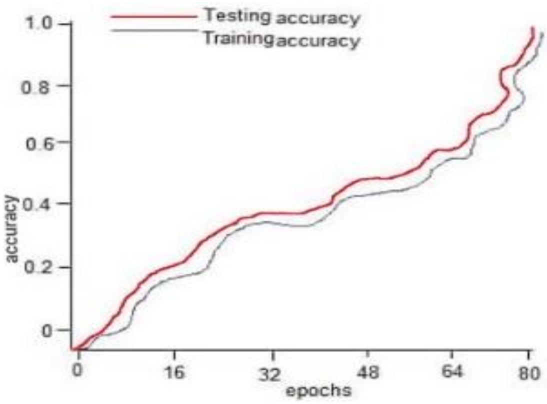
Published Online First: July, 2024
https://doi.org/10.21123/bsj.2024.10017
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal



(a)



(b)



(c)

Figure 5. LSTM prediction accuracy, (a) prediction accuracy of the DN headlines, (b) the prediction accuracy of DJIA, (c) the accuracy of the combined DN & DJIA.

In Fig 5, it is can be seen easily that the accuracy of the LSTM model is stable reaching almost 99% with only 80 epochs outperforming the two models of machine learning, although the dataset is not huge which might raise the problem of overfitting when using it with deep learning but the proposed model overcomes this problem by achieving a remarkable result compared with LR and J48 considering that the size of this dataset is suitable for J48 and LR models. Tables 2 and 3 illustrate the difference in results between the machine learning approach and the LSTM approach for the DNSMP dataset.

Table 2. Result of DNSMP dataset for ML mode

Sector	Prediction Accuracy of LR	Prediction Accuracy of J48
DN headlines	86.67%	84.04
DJIA	86.34	85.21
Combined DN & DJIA	84.68	82.88

Table 3. Result of DNSMP dataset for LSTM model

Company	Prediction Accuracy	Error rate	
		MAE	RMSE
DN headlines	97.88%	0,0129	0.0403
DJIA	98.56	0,0129	0.0353
Combined DN & DJIA	98.92	0,0123	0.0347

The combined file of the two parts of the DNSMP Dataset contains 25 columns and 1990 records. From the results in the above tables, although the dataset is not considered a huge dataset, which means that it is theoretically closest to the ML approach, but practically, the results show the opposite to that, this is due to the efficient optimizer of the proposed LSTM approach and the good choice dropout ratio to avoid overfitting. Thus, it can be seen that the deep learning approach outperforms than ML approach, in addition, the LSTM approach is closer to the perfect accuracy ratio. It can also be noticed that in the ML approach, the LR is achieved better than J48 in terms of accuracy. One can see also that the proposed LSTM approach performs an excellent error rate in terms of MAE and RMSE.

Results of the SMP Dataset

The SMP dataset is considered a huge dataset, which is theoretically closest to the DL approach, this dataset contains 77 columns and 100045 rows (records). Now practically, results show that the

Published Online First: July, 2024
https://doi.org/10.21123/bsj.2024.10017
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

proposed LSTM approach clearly outperforms than ML approach for both the individual sector as well as an overall dataset. Tables 4 and 5 present the experimental result for the ML approach and proposed LSTM approach respectively, while Figs. 6,7 and 8 show the training and testing accuracy.

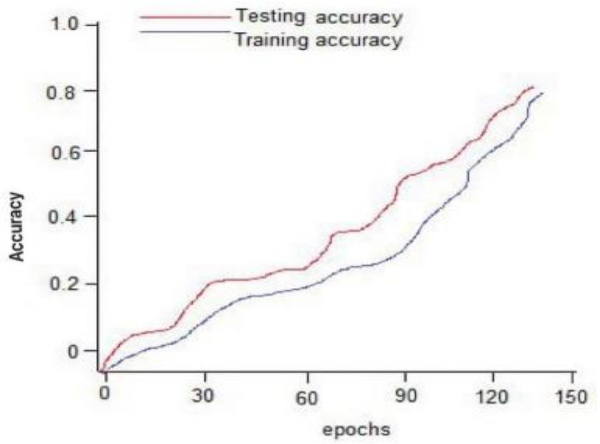


Figure 6. Accuracy of LR model.

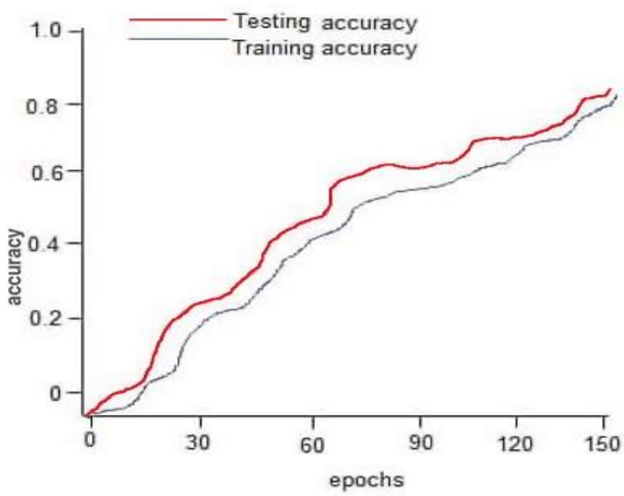


Figure 7. Accuracy of J48 model Figure

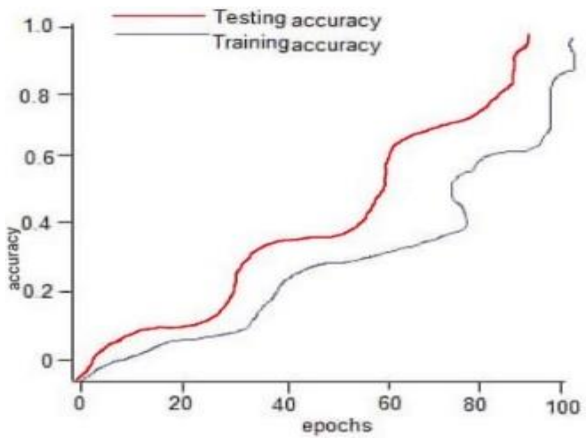


Figure 8. Accuracy of the LSTM model

From the Figs. 6 and 7, when training the LR and J48 models with the SMP dataset the accuracy is increasing in a slower manner which is considered That the models' accuracy is not improving anymore and the models reached the learning level that is possible for them because after reaching epoch 80 the model's accuracy did not improve more than 20 percent of the total accuracy that the models achieve even though that each model is trained for 150 epochs. This is due to the huge size of SMP dataset which exceeds the ability of these models, while the accuracy of LSTM model is boosting after the epoch 80 reaching almost 99% with 100 epochs which proofs the ability of deep learning to handle huge datasets and achieve an outstanding result compared with other machine learning techniques, as can be seen in Fig. 8.

Table 4. Result of samples of SMP dataset for LSTM approach

Company (sector)	Prediction Accuracy of LR	Prediction Accuracy of J48
ABEO	82.45%	82.17%
ABIO	85.32%	82.38%
ABTX	81.60%	79.82%
ABUS	85.62%	86.23%
BATRK	82.09%	73.08%
BBCP	84.22%	81.71%
BBGI	77.53%	82.44%
CALT	79.72%	80.11%
CARG	81.44%	83.07%
CBNK	88.76%	84.73%
CERE	84.64%	82.62%
CLFD	85.31%	81.85%
CNCE	87.30%	83.51%
CLRO	83.22%	82.37%
BWFG	84.65%	84.52%
BYFC	89.34%	82.42%
CAAS	84.58%	82.47%

Table 5. Result of samples of SMP dataset for ML approach

Company (sector)	Prediction Accuracy	Error Rate MAE RMSE
ABEO	99.12%	0,0122 0.0347
ABIO	99.36%	0,0110 0.0345
ABTX	99.32%	0,0111 0.0345
ABUS	98.73%	0,0127 0.0348

Published Online First: July, 2024
https://doi.org/10.21123/bsj.2024.10017
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

BATRK	99.37%	0,0110 0.0345
BBCP	99.57%	0,0108 0.0343
BBGI	99.28%	0,0115 0.0345
CALT	99.43%	0,0109 0.0344
CARG	97.18%	0,0136 0.0413
CBNK	99.49%	0,0109 0.0344
CERE	99.81%	0,0101 0.0341
CLFD	99.52%	0,0108 0.0343
CNCE	99.81%	0,0101 0.0341
CLRO	98.64%	0,0127 0.0348
BWFG	99.23%	0,0116 0.0346
BYFC	98.73%	0,0127 0.0348
CAAS	99.47%	0,0109 0.0344

From Table 4 and 5, the ML approach gives an acceptable accuracy rate of price prediction for each company as well as the overall dataset accuracy. It is clear that the proposed LSTM approach with a suitable optimizer deduces powerful, potential, near to perfect and robust results, that is, powerful is for the power of the proposed approach to predict results during the acceptable number of epochs, potential due to very high accurate results for each sector, near to perfect refers to the average of the accuracy for all integrated dataset sectors and the robust property is a metaphor for proposed LSTM stability for giving very high ranking prediction accuracy for multiple environments represented by two different size datasets. The proposed LSTM achieves the prediction task of stock market prices with a low error rate in terms of MAE and RMSE. Table 6 illustrates a comparison between the machine learning approach and deep learning approach to show the difference in accuracy for the whole datasets as well as a comparison with related works according to accuracy or error rate measurements.

Table 6. comparative among the proposed models and related work

Reference	**Used model**	**Used dataset**	**Outcome measure**	**Result ratio**
Proposed model	LSTM	DNSMP	accuracy MAE	98.453% 0,0129
		SMP	accuracy MAE	98.548% 0.0114
Proposed model	LR	DNSMP	accuracy	85.896%
		SMP	accuracy	83.239%
Proposed model	J48	DNSMP	accuracy	84.043%
		SMP	accuracy	80.692%
[10]Dezdemona G, et. al.	RNN	EURO/ALL" and "USD/ALL" exchange	MAE	0.25
[11] Kranthi S. R., et. al.	SVM	data collection of various global financial markets	accuracy	convergence between original and predicted value
[12] Nti IK, et. al.	hybrid DNN (CNN and LSTM)	Ghana Stock Exchange	accuracy	98.31%

[13] Rather AM.	LSTM based DNN	NIFTY-50	RMSE	1.715
[14] Mukherjee S, et. al.	ANN or Feed-forward DNN and CNN	NSE (NIFTY price index)	accuracy	ANN 97.66%, CNN 98.92%
[15] Wu JM-T, et. al.	Combined CNN and LSTM	SSACNN, SSALSTM and SACLSTM	accuracy	start from 71% reach its highest at 95.1%
[16] Akhtar MdM, et. al.	SVM random forest	Collection dataset	accuracy	SVM about 78.7% Random Forest 80.8%
[17] Bhandari HN, et. al.	single layer LSTM multilayer LSTM	S&P 500 index, (a popular US stock market)	RMSE	single layer LSTM 5.411 multilayer LSTM 8.637

As it's shown form Table 6, the model that achieved approximate results to the proposed LSTM model is a hybrid model between CNN and RNN, even though the proposed model is just LSTM (not hybrid) and still achieved very high results without the complication of using a hybrid model with another technique. Even for the traditional machine learning models J48, LR, SVM and random forest the proposed approach (J48, LR) outperforms other models due to the employment of GDA to extract and select the most relevant and important features.

The limitation

The limitation of this work can be summarized with two significant points:

1. When dealing with a huge dataset the ability of the traditional machine learning approaches is modest which leads to inadequate results.

2. When dealing with normal size dataset, the deep learning suffers from overfitting problem which leads to use additional techniques to overcome that, therefore the system will suffer from extra complications.

Conclusion

Stock market movement and price prediction have an impressive role in gaining good profit in financial market exchange; the trading process requires a high-level prediction system to avoid losses. As it is known, the trading volume in the stock market involves huge data and many details which requires powerful tools for managing and configuring such data.

Traditional prediction systems, like statistical or manual systems, lack to precise outcomes in addition to the time factor. Also, ordinary ML usually falls at work with huge data. Therefore, intelligent systems especially those that are based on machine learning have contributed to overcome many limitations that were facing stock market prediction systems.

The proposed approach took advantage of the capabilities of machine learning in two ways, ML approach and proposed deep learning-based LSTM approach. The LSTM approach has demonstrated that deep learning has a great ability to deal with prediction systems with huge stock market datasets and multiple features with overwhelming evaluation results. The use of proposed LSTM achieved superior performance in the stock market prediction due to the memory nature of LSTM cells that avail the historical data for reliable performance. The efficient optimizer, hyperparameter tuning and dropout ratio give the proposed LSTM a positive effect in upgrading the performance. The suitable design of the proposed LSTM architecture plays an

Published Online First: July, 2024
https://doi.org/10.21123/bsj.2024.10017
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

important role in the stability of the model with an acceptable epoch number.

The proposed study contributes in some useful points; a qualitative utilization has been achieved by automating the prediction of the stock market; both approaches produced better results when conducting the DNSMP dataset (normal dataset size), while the results retreat when the models are conducted with a huge SMP dataset. The second approach (proposed LSTM) exceeded the first approach with both datasets with high prediction accuracy and low error rates (in terms of MAE & RMSE).

For future suggestions, it is important to potentially improve the insight of the stock market prediction system. It can be said that the suggested approach may study the effect of other factors on stock market movement (such as oil, gold or wars), also it can be included the CNN as a features extraction stage to be fed to RNN, finally, an integration between the stock market and cryptocurrency technology can be taken in consideration for more comprehensive financial prediction system.

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- No animal studies are present in the manuscript.
- No human studies are present in the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee at University of Baghdad.

Authors' Contribution Statement

The authors contribute to this paper and as follows: All the authors contributed in conceptualization, methodology, validation and analysis. They also contributed in writing the original draft and the revision of the final draft.

References

1. Ramprasath J, Murugesan B, Muthanantha M, Arockia C. Prediction Of Data Analysis Using Machine Learning Techniques. Int J Aquat Sci. 2021; 12(3): 2755-2762.
2. Alnedawe SM, Aljobouri HK. A new model design for combating covid -19 pandemic based on SVM and CNN approaches. Baghdad Sci J. 2023. https://doi.org/10.21123/bsj.2023.7403
3. Kleinberg J, Ludwig J, Mullainathan S, Obermeyer Z. Prediction policy problems. Am Econ Rev. 2015; 105(5): 491–5. https://doi.org/10.1257/aer.p20151023
4. Sheng C, Hongxiang H. Stock prediction using Convolutional Neural Network. IOP Conf Ser.: Mater Sci Eng. 2018; 435: 012026. https://doi.org/10.1088/1757-899x/435/1/012026
5. Chhajer P, Shah M, Kshirsagar A. The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction. Decis. 2022; 2: 100015. https://doi.org/10.1016/j.dajour.2021.100015
6. Hiransha M, Gopalakrishnan E, Vijay K M, Soman K. NSE stock market prediction using deep-learning models. Procedia Comput Sci. 2018; 132: 1351–62. https://doi.org/10.1016/j.procs.2018.05.050
7. Jagdish C, Manish K. Convolutional Neural Network-based a novel Deep Trend Following Strategy for Stock Market Trading, CEUR Workshop Proceedings. 2020; 3053. https://doi.org/10.1111/exsy.12514
8. Wen Y, Lin P, Nie X. Research of stock price prediction based on PCA-LSTM model. IOP Conf Ser.: Mater Sci Eng. 2020; 790(1): 012109. https://doi.org/10.1088/1757-899x/790/1/012109
9. Li J. Research on Market Stock index prediction based on Network Security and Deep Learning. Secur Commun Netw. 2021; 2021: 1–8. https://doi.org/10.1155/2021/5522375
10. Dezdemona G, Eljona P, Alketa H. Recurrent Neural Networks in Time Series Prediction. J Multidiscip Eng Sci Technol. October 2018; 5(10). Available from: https://www.researchgate.net/publication/320269918

11. Kranthi S R. Stock Market Prediction Using Machine Learning. IRJET. 2018 October; 5(10): 2395-0072. https://doi.org/10.13140/RG.2.2.12300.77448
12. Nti IK, Adekoya AF, Weyori BA. A novel multi-source Information-Fusion Predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. J Big Data. 2021; 8(1). https://doi.org/10.1186/s40537-020-00400-y
13. Rather AM. LSTM-based deep learning model for stock prediction and predictive optimization model. EURO J Decis. Process. 2021; 9: 100001. https://doi.org/10.1016/j.ejdp.2021.100001
14. Mukherjee S, Sadhukhan B, Sarkar N, Roy D, De S. Stock market prediction using Deep Learning Algorithms. CAAI Trans Intell Technol. 2021; 8(1): 82–94. https://doi.org/10.1049/cit2.12059
15. Wu JM-T, Li Z, Herencsar N, Vo B, Lin JC-W. A graph-based CNN-LSTM stock price prediction algorithm with leading indicators. Multimed Syst. 2021; 29(3): 1751–70. https://doi.org/10.1007/s00530-021-00758-w
16. Akhtar MdM, Zamani AS, Khan S, Shatat AS, Dilshad S, Samdani F. Stock market prediction based on statistical data using machine learning algorithms. J King Saud Univ Sci. 2022; 34(4): 101940. https://doi.org/10.1016/j.jksus.2022.101940
17. Bhandari HN, Rimal B, Pokhrel NR, Rimal R, Dahal KR, Khatri RKC. Predicting stock market index using LSTM. MLWA. 2022; 9: 100320. https://doi.org/10.1016/j.mlwa.2022.100320
18. 18.Mahmood I, Abdullah H. Wisdommodel: Convert Data Into Wisdom. Appl Comput Inform. 2021. https://doi.org/10.1108/aci-06-2021-0155
19. Vijh M, Chandola D, Tikkiwal VA, Kumar A. Stock closing price prediction using Machine Learning Techniques. Procedia Comput Sci. 2020; 167: 599–606. https://doi.org/10.1016/j.procs.2020.03.326
20. Ali NH, Abdulmunim ME, Ali AE. Intelligent system for multi-layer lip reading based Microlearning. Iraqi J Sci. 2022; 3977–93. https://doi.org/10.24996/ijs.2022.63.9.29
21. Daily News for Stock Market Prediction Dataset. 2019.
22. Luis A G, Carlos T, Stock market prediction .2021.
23. Izonin I, Tkachenko R, Shakhovska N, Ilchyshyn B, Singh KK. A two-step data normalization approach for improving classification accuracy in the medical diagnosis domain. Mathematics. 2022; 10(11): 1942. https://doi.org/10.3390/math10111942
24. Bahmaninezhad F, Hansen JHL. Generalized discriminant analysis (GDA) for improved I-vector based speaker recognition. Interspeech 2016. https://doi.org/10.21437/interspeech.2016-1523
25. Shailendra S, Sanjay S, Generalized Discriminant Analysis algorithm for feature reduction in Cyber Attack Detection System. IJCSIS. 2009; 6(1): 173-180.
26. Ali NH, Abdulmunem ME, Ali AE. Learning evolution: A survey. Iraqi J Sci. 2021; 4978–87. https://doi.org/10.24996/ijs.2021.62.12.34
27. Abdallah B M, Logistic Regression Classification for Uncertain Data. J Math Stat Sci. 2014; 2(2): 1-6.Available from :
28. Ali A T, Abdullah HS, Fadhil MN. Withdrawn: Voice recognition system using machine learning techniques. Mater Today. 2021; https://doi.org/10.1016/j.matpr.2021.04.075
29. Ali NH, Abdulmunem ME, Ali AE. Constructed model for micro-content recognition in lip reading based Deep Learning. Bull Electr Eng Inform. 2021; 10(5): 2557–65. https://doi.org10.11591/eei.v10i5.2927
30. Alsaedi EM, Farhan A kadhim. Retrieving encrypted images using convolution neural network and fully homomorphic encryption. Baghdad Sci J. 2023; 20(1): 0206. https://doi.org/10.21123/bsj.2022.6550
31. Ashour MA. Optimized artificial neural network models to time series. Baghdad Sci J. 2022; 19(4): 0 899. https://doi.org/10.21123/bsj.2022.19.4.0899
32. Salehinejad H, Sankar S, Barfett J, Colak E, Valaee S. Recent advances in recurrent neural networks . 2018 [cited 2023 Nov 26]. arXiv. https://doi.org/10.48550/arXiv.1801.01078
33. Bianchi FM, Maiorino E, Kampffmeyer MC, Rizzi A, Jenssen R. Recurrent neural networks for short-term load forecasting. Springer Br. 2017; https://doi.org/10.1007/978-3-319-70338-1
34. Hasan AM, Qasim AF, Jalab HA, Ibrahim RW. Breast cancer MRI classification based on fractional entropy image enhancement and deep feature extraction. Baghdad Sic J. 2023; 20(1): 0221. https://doi.org/10.21123/bsj.2022.6782
35. Wotaifi TA, Dhannoon BN. An effective hybrid deep neural network for Arabic fake news detection. Baghdad Sci J. 2023. https://doi.org/10.21123/bsj.2023.7427
36. Skansi S. Introduction to deep learning: From logical calculus to artificial intelligence. Cham Switzerland: Springer. 2018. https://doi.org/10.1007/978-3-319-73004-2
37. Bin Shah SR, Chadha GS, Schwung A, Ding SX. A Sequence-to-Sequence Approach for Remaining Useful Lifetime Estimation Using Attention-augmented Bidirectional LSTM. Intelligent Systems with Applications J. 2021; DOI: 10.1016/j.iswa.2021.200049

Published Online First: July, 2024
https://doi.org/10.21123/bsj.2024.10017
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

مقارنة تحليلية لسلوك تعلم الالة والتعلم العميق في التنبؤ بسوق الاوراق المالية

حسنين سمير عبدالله[1]، ندى حسين علي [1]، عمار حسين جاسم[2]، سيد حامد حسين[3]

[1]قسم علوم الحاسوب، الجامعة التكنولوجية، بغداد، العراق.
[2]قسم علوم الحاسوب، كلية العلوم للبنات، جامعة بغداد، بغداد، العراق.
[3]قسم علوم الحاسوب، الجامعة التكنولوجية الماليزية، جوهور، ماليزيا.

الخلاصة

يعتبر تعلم الماكنة تقنية قوية في كثير من التطبيقات مثل التصنيف، العنقدة، التمييز والتنبؤ. التعلم العميق هو تقنية تعلم ماكنة حديثة وحيوية ومتفوقة والتي تعطي اداء مبهر خصوصا مع البيانات ضخمة. التنبؤ بأسعار سوق الاوراق المالية هي عملية تحديد القيمة المستقبلية لأدوات مالية متعامل بها في السوق، لتحقيق مكاسب كبرى يجب توظيف عملية تنبؤ ناجحة ومن اجل تحقيق هذا الغرض تم استخدام تعلم الماكنة. في هذا البحث، تم اقتراح نهجين للتنبؤ بأسعار وتحركات سوق الاوراق المالية باستخدام مجموعتين بيانات، النهج الاول يوظف نموذجين تعلم ماكنة& J48) الانحدار اللوجستي (بينما النهج الثاني يعتمد على الشبكات العصبية المتكررة (LSTM المقترحة .(معمارية LSTM المقترحة صممت وتم تدريبها باستخدام محسنات كفوئة، ضبط المعلمات الفائقة واختبار معدل اسقاط مناسب لتجنب مشكلة التجهيز الزائد. الهدف من هذا البحث هو اجراء مقارنة تجريبية بين مناهج تعلم الماكنة التقليدية& J48) الانحدار اللوجستي(وتعلم الماكنة العميقLSTM) المقترحة (النتائج التجريبية اظهرت ان نظام التعلم العميق المقترح LSTM تفوق على النهج الآخر (لكلا لنموذجين) ومن خلال مجموعتي البيانات بتنبؤ اسعار وحركة سوق الاوراق المالية.

الكلمات المفتاحية: التنبؤ بسوق الأوراق المالية، التعلم الآلي، التعلم العميق، الشبكة العصبية المتكررة، LSTM، J48، الانحدار اللوجستي.