STOCK MARKET PRICE PREDICTION USING MACHINE LEARNING

M.Monikasri[1], S. Varshini[2]

[1]*Student, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India*
[2]*Student, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India*

Abstract – *Analyzing the stock market is one of the difficult one. Investors investing in stock must know the daily opening and closing prices of their investment. Values keep changing day to day . Therefore we built a model using machine learning to predict the future price of the stocks. The model is built in a way that the investors can see the daily prices and make the most profit out of it. Deep learning approach plays a vital role in prediction of financial time series data. The method used in our project is LSTM(long short term memory).By using the above model we can predict the closing prices of a particular share at the end of the day. Prediction of closing price is done at march1 to march12. Dataset is taken from Kaggle website and we follow all cleaning activities and then apply the algorithm and finally visualize the results in the form of graph. There are three evaluation metric namely RMSE(root mean squared error),MSE(mean squared error),MAE(mean absolute error) .*

Keywords : **Stocks , Forecasting , Prediction , LSTM , Deep Learning , Accuracy.**

1.INTRODUCTION

Stock Market's Future Forecasting is the project on predicting the closing price at the end of the day using data provided by Maruti.The project is developed as a dashboard which is user friendly and the stock price can be viewed by the investors at any time. We are Using pandas to get stock information, next we perform all the required preprocessing steps, apply algorithm and can see the future prediction and also we can find out the risks in stock using the old data .We predicted future stock prices through a LSTM model. In this proposed work, Long Shot Term Memory (LSTM) have been utilized for predicting the intraday closing price for Maruti company belonging to different sectors of operation. LSTM will not process a single data point therefore Lstm needs large sequence of data to process and store the values that are previously evaluated. LSTM is the best model built to make predictions with high accuracy.

Stock market is like a ocean which has many old dataset that contains data with different dates and closing prices. LSTM model is trained in such a way that we yield a high accuracy . Investors can invest without fear of loss, because they can see the closing of a particular day so that they can come to a conclusion either to sell or buy a stock. LSTM Can work well even when the dataset is large.

2.METHODOLOGY

The modules in our research work includes,

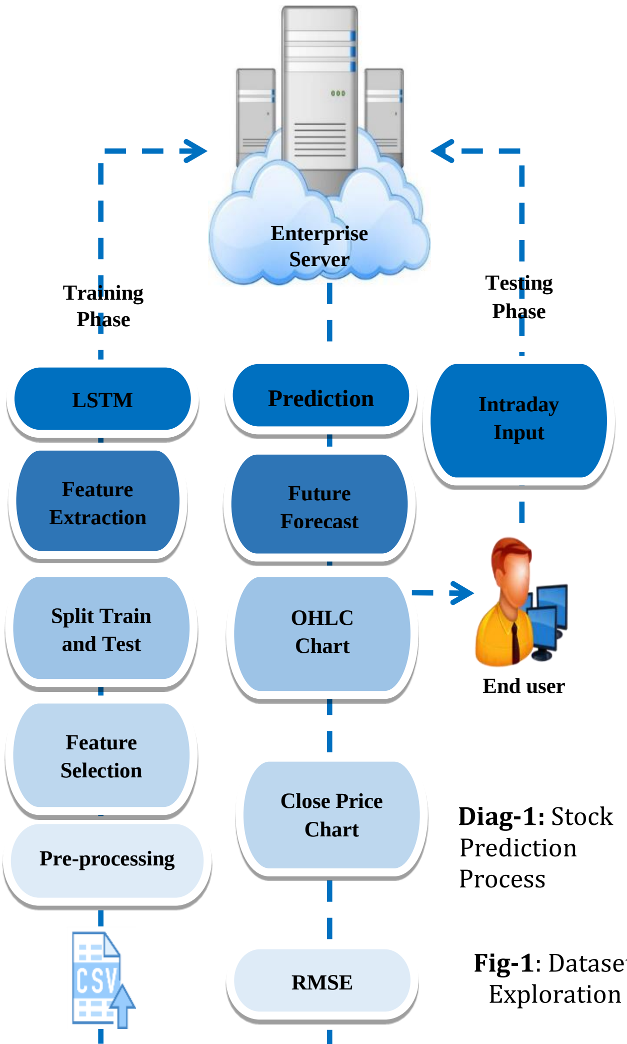
- Data preparation and exploration
- Preprocessing
- Feature Selection
- Feature Extraction
- LSTM prediction model
- Evaluation measures

Hardware specification

- Processors: Intel Core i5
- Disk space: 320 GB
- Operating systems: Windows10, macOS, and Linux

Software specification

- Server Side : Python 3.7.4(64-bit) or (32-bit)
- Client Side : HTML, CSS, Bootstrap
- IDE : Flask 1.1.1
- Back end : MySQL 5.
- Server : WampServer 2i
- OS : Windows 10 64 –bit

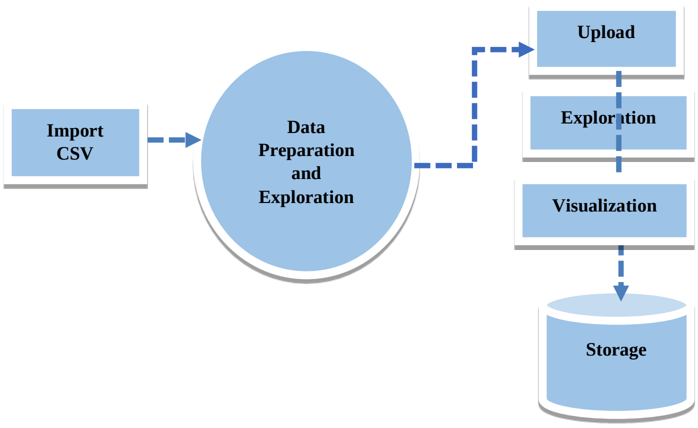


Diag-1: Stock Prediction Process

Fig-1: Dataset Exploration

2.1 Data Preparation and Exploration

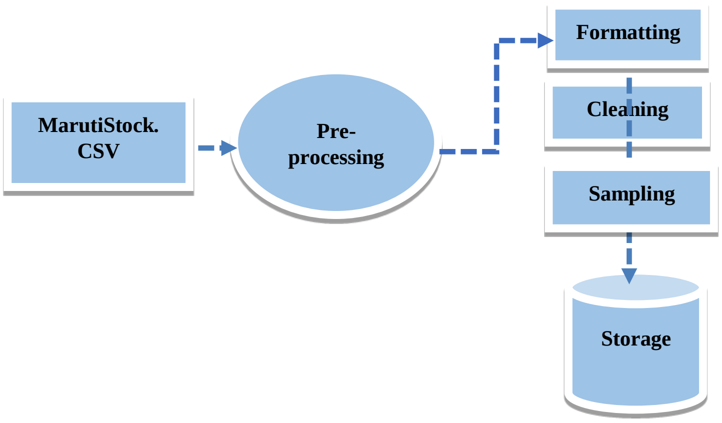
The stock price data set of Maruti was taken from Kaggle web page which contains the stock prices from 2003–01-01 to 2021–02–12 with comma-separated value(.csv) format also it has a opening ,closing ,high, low, date and etc. By obtaining a data set, then come up with finalized characteristics and behaviour of the stock prices. Seven features are obtained.



Date	Open	High	Low	Close	Adj Close	Volume
2009-01-02	12.268572	13.005714	12.165714	12.964286	11.253528	186503800
2009-01-05	13.310000	13.740000	13.244286	13.511429	11.728474	295402100
2009-01-06	13.707143	13.881429	13.198571	13.288571	11.535025	322327600
2009-01-07	13.115714	13.214286	12.894286	13.001429	11.285772	188262200
2009-01-08	12.918571	13.307143	12.862857	13.242857	11.495339	168375200

Fig-1: Data Preparation and Exploration

2.2 Pre-processing



Diag-2: Data Pre-processing

Every dataset consists of various types of anomalies such as missing values, redundancy or any other problem for removing this problem there is need of certain step called as processing data. Pre-processing step is needed to overcome from such problem. There are three pre-processing steps:

2.2.1 Formatting: The data set is used for implementation is taken from Kaggle; it may contain certain attributes whose names are not clear in the (dataset name) also contain certain unrelated attribute which is not useful for the greater performance of proposed work.

2.2.2 Cleaning: Pre-processing or cleaning means is to remove or fixing of missing out entry in the data frame. Row containing these incomplete columned to be removed also for removing certain redundant entries in data frame this step is recommend

2.2.3 Sampling: Sampling is also done on the dataset to enhance the performance of the algorithm on sample data set may lead algorithm to take longer time.

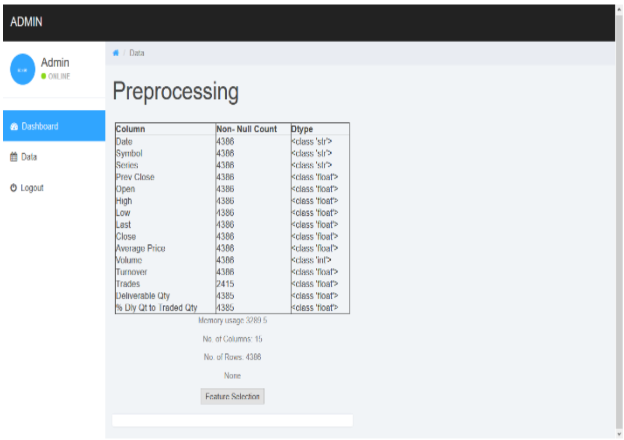


Fig-1: Data Pre-processing

2.3.Feature Selection

In this step, data attributes are selected that are going to be given to the next layer that is neural network. In this study Date, Open, High, Low and Last VWAP Close Price are chosen as selected features.

Diag-3: Feature Selection

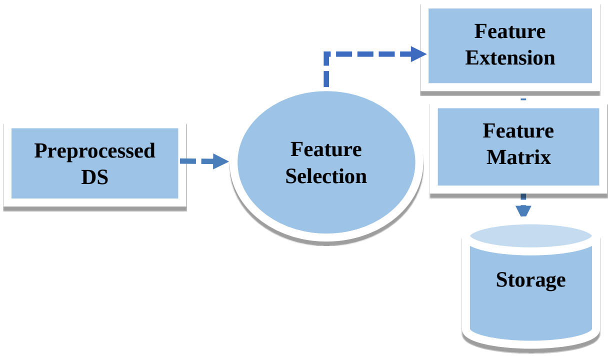


Fig-2: Feature Selection

Applying feature extension

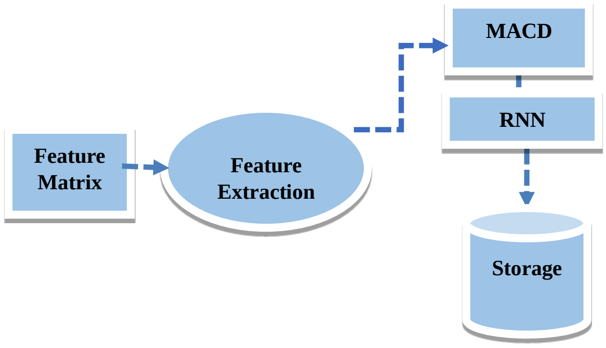
The first and foremost step after feature selection is applying feature extension. Here,we give the input data to three main feature extension methods like min max scaling, polarizing and calculation fluctuation percentage but the technical indices will not be suitable for all the methods. So we chose meaningful extension methods based on how the calculations of indices are made. After this, the extended features are combined with mostly used technical indices.

Applying recursive feature elimination

Followed by feature extension we apply recursive feature elimination. In this step the unnecessary features in the dataset are removed using recursive feature elimination algorithm. The next step is to give input to PCA but before that preprocessing of features is done. The output is in the form of matrix.

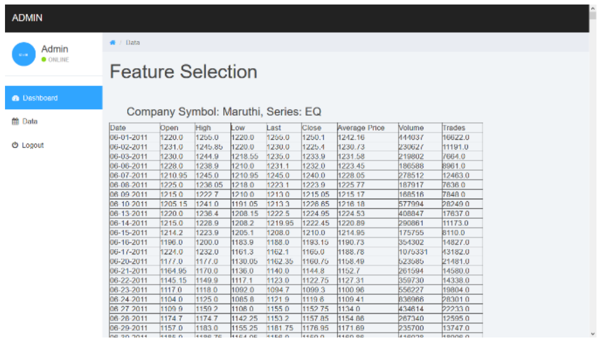
2.4 Feature Extraction

Closing price is being predicted in our work. So, the features that contributes much to the prediction of closing price is extracted in this step. Then, the features are given as input to MACD (Moving Average Convergence/Divergence Oscillator). Followed by this the features are passed on to recurrent neural network algorithm.



Diag-4: Feature Extraction

Before the LSTM Model split the dataset into training dataset 80% and testing dataset 20%.



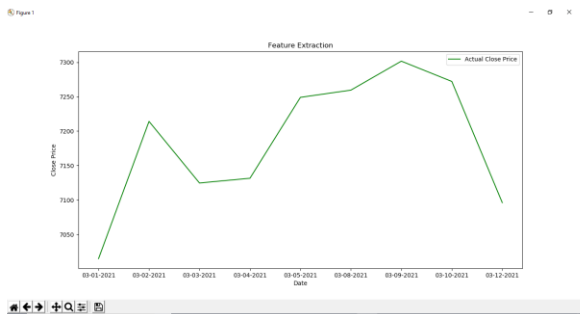


Chart-1: Actual Stock Price Graph

2.5 LSTM Prediction model

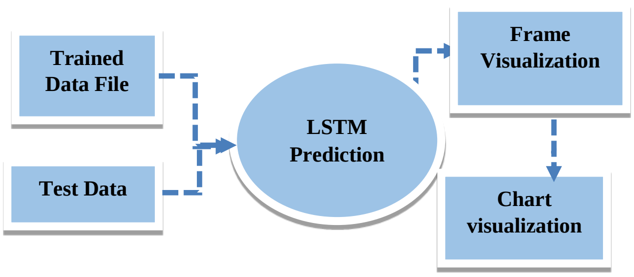
After going through some prediction model we chose LSTM model for prediction because it best suits for the stock market analysis. It uses the aforementioned dataset, with S&P 500 datapoints over the time period of 2003-01-01 to 2021-02-12 and splits it up in 80%

training data and 20% testing data. With the selected feature(s) of the current and previous number of days (equal to time step) as input (X), the closing price of the following day is predicted as output (y).
The model is a Sequential model using two LSTM layers and one Dense layer. The LSTM layers use 50 units each and hyperbolic tangent as activation function. The model is currently not using any Dropout and thus runs greater risk of overfitting, but does perform better than with Dropout on our test data.

In the selection process, different settings narrowing the spread between high and low values where tried, not too far from how many root finding methods work in mathematics.

When we tested the samples were accurate in performance and efficiency. As epoch sizes above 20 generated little to no improvements, 20 was chosen as the number of times to iterate over the entire dataset. For loss function, the mean squared error (MSE) is used. Out of the different optimizers available, Adam generated the best results and was thus chosen in favour of stochastic gradient descent (SGD).

The Adam optimization algorithm is a modern alternative to the SGD algorithm which updates network weights iteratively in training data. The SGD algorithm on the other hand maintains a single learning rate for all weight updates. Then the results have demonstrated that Adam works well in practice, even comparing favourably to other stochastic optimization methods, hence its application on this model. Lastly, different number of time steps are tested to see how they affect the model's performance. The time steps chosen to be tested are 5, 10, 25, 50 and 100. For each of these time steps, the model is trained and tested five times to generate enough data to evaluate their relative performance.



Diag-5: LSTM Prediction

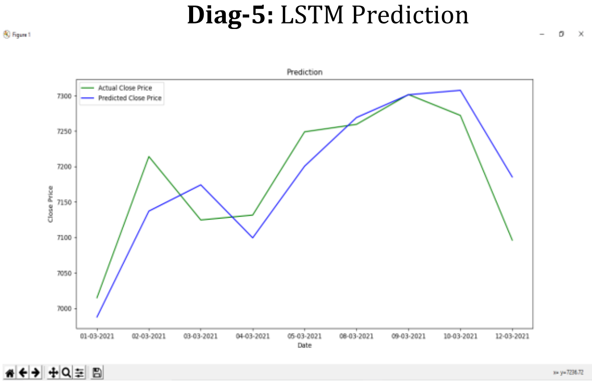


Chart-2: Actual Vs Predicted Graph

Here is the final graph that is plotted between actual stock price and the predicted stock price. From March 1,2021 to March 12,2021.

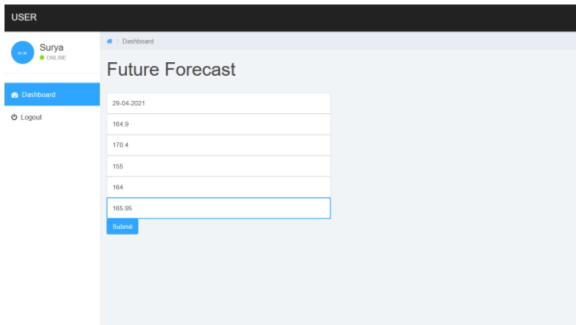


Fig-3: Future Forecasting

Future Forecasting has also been done in our work. A particular day's open, high,low is given as input to predict the closing price of that particular day.

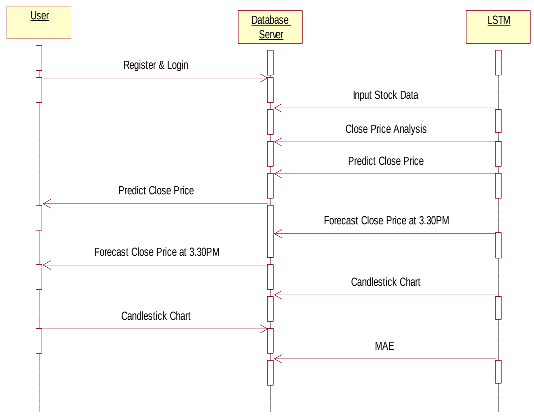
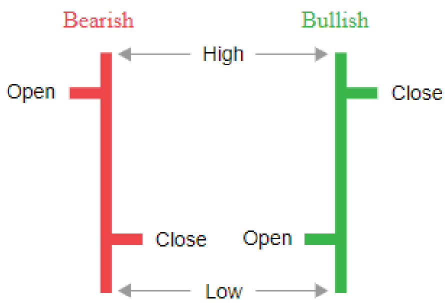


Fig-4: Sequence diagram

Bar Chart OHLC

The OHLC chart is used to see the open, close. high and low stock prices. Here there are two main elements namely bullish and bearish. When there is a bullish graph, the investor has profit where the close price is greater than opening price of the stock. Similarly, When there is a bearish graph, the investor faces loss where the close price is lower than the opening price. Bullish graph is represented in green colour whereas the bearish graph is represented in red colour.



So here in our work,the OHLC chart is depicted from March1,2021 to March 12,2021.

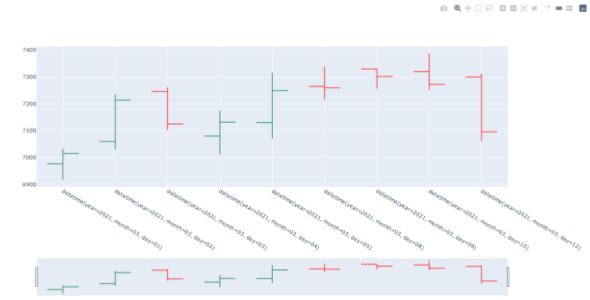


Chart-3: OHLC Chart

This OHLC chart particularly represents every single day's stock price and helps us to know the profit/loss for that particular day.

2.6 Evaluation Measures

In this research work, we included four evaluation metrics to find out the accuracy of the prediction model. These metrics are used in the result analysis. The actual and predicted close price is used to calculate these measures.

2.6.1 Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) is gives us the accuracy of the prediction model in terms of percentage. In our work, we achieved an accuracy of nearly 86% with MAPE being 14.466.The formula used is,

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100,$$

2.6.2 Mean Absolute Error

The mean absolute error from our model is 0.144666.It is found using the below formula,

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t|,$$

2.6.3 Relative Root Mean Square Error

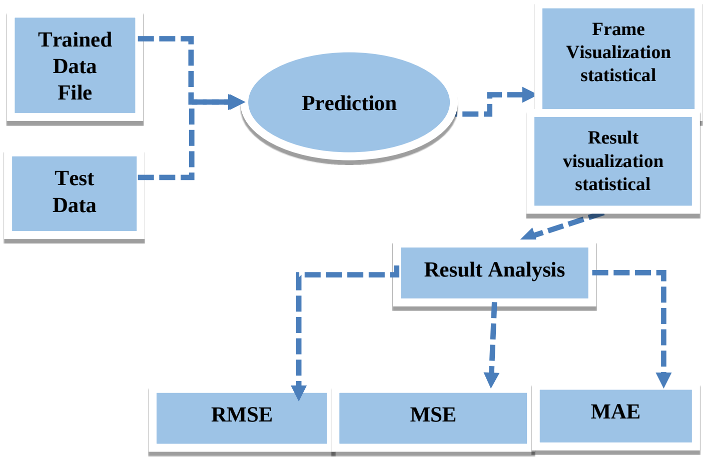
The RRMSE of our model is 0.38003.The formula used is,

$$\text{RRMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{A_t - F_t}{A_t}\right)^2},$$

2.6.4 Mean Squared Error

The Mean Squared Error (MSE) of our model is 0.18835.

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2,$$



Diag-6: Prediction Process

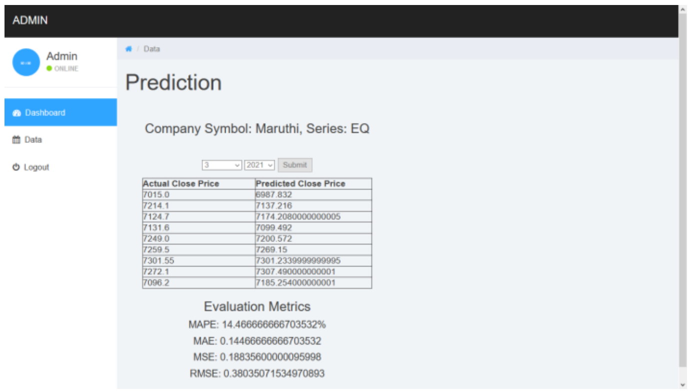


Fig-5: Evaluation

In our work, we achieved an accuracy of nearly 86% with MAPE being 14.466.

3. CONCLUSION

This paper presented an implementation of stock market price prediction which will be very helpful for the investors to invest in stock markets. After experimenting many models which had shortcomings, we concluded that LSTM model best suits for stock market price prediction.

REFERENCES

[1] A. Altunkaynak and T. A. Nigussie, ``Monthly water consumptionprediction using season algorithm and wavelet transform_based models,''*J. Water Resour. Planning Manage.*, vol. 143, no. 6, Jun. 2017,Art. no. 04017011.
[2] P. Liang, H.-D. Yang, W.-S. Chen, S.-Y. Xiao, and Z.-Z. Lan, ``Transferlearning for aluminium extrusion electricity consumption anomaly detectionvia deep neural networks,'' *Int. J. Comput. Integr. Manuf.*, vol. 31,nos. 4_5, pp. 396_405, Apr. 2018.
[3] D. G. Gloubos, ``Estimating corporate failure as an auditor's going concernevaluation factor,'' M.S. thesis,

School Bus. Admin., Univ. Macedonia,Thessaloniki, Greece, 2016.

[4] K. A. Althelaya, E.-S.-M. El-Alfy, and S. Mohammed, ``Evaluation ofbidirectional LSTM for short-and long-term stock market prediction,'' in*Proc. 9th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2018, pp. 151_156.

[5] E. Cambria, ``Affective computing and sentiment analysis,'' *IEEE Intell.Syst.*, vol. 31, no. 2, pp. 102_107, Mar. 2016.

[6] J. Grif_th, M. Najand, and J. Shen, ``Emotions in the stock market,''*J. Behav. Finance*, vol. 21, no. 1, pp. 42_56, 2020.

prediction: Methodology, data representations, andcase studies,'' *Expert Syst. Appl.*, vol. 83, pp. 187_205, Oct. 2017.

[7] H. Yan and H. Ouyang, ``Financial time series prediction based ondeep learning,'' *Wireless Pers. Commun.*, vol. 102, no. 2, pp. 683_700,Sep. 2018.

[8] W. Bao, J. Yue, and Y. Rao, ``A deep learning framework for financial timeseries using stacked autoencoders and long-short term memory,'' *PLoSONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0180944.

[9] M. Kim, ``Cost-sensitive estimation of ARMA models for financial assetreturn data,'' Math. Problems Eng., vol. 2015, pp. 1-8, Jan. 2015.

[10] D. Marcek, ``Forecasting high frequency data: An ARMA-soft RBF networkmodel for time series,'' Appl. Mech. Mater., vol. 596, pp. 160-163,Jul. 2014.