Global Journal of ComputerandEngineering Technology (GJCET)

Vol. 1, No. 1, January, 2021 ISSN: 2767-1933

STOCK MARKET PRICEFORECASTINGUSING LSTM AND ANNWITH COMBINED APPROACH OF WAVELET BASED DATA PREPROCESSING

RichaHanda, D.P.Vipra College, India(proffhanda@gmail.com)
H.S. Hota, Atal Bihari Vajpayee University, India(proffhota@gmail.com)

ABSTRACT

Stock market has gone deeply in our economic life and affects the researchers and academiciansto deal with this issue from last few decades. This research work focus on prediction of two stock data NSE50 and BSE30 including period of COVID-19 pandemic in India. Due to fluctuating behaviour of stock data, we have used wavelet signal denoiser tool of MATLAB to remove unwanted signals called noise from data. The pre-processed data is given to ANN and LSTM model for next day ahead prediction and found that LSTM is outperforming.Theperformance of predictive model is measured with two measures MAPE=1.759 and RMSE=0.016 for NSE50 and MAPE=2.224 and RMSE=0.020 for BSE30.

Keywords: Artificial Neural Network(ANN), Long Short-Term Memory (LSTM), COVID-19, Stock market.

1. INTRODUCTION

Financial time series forecasting is a wide area which always attracts researcher's community from last few decades.To predict the financial time series data, it is very necessary to get the proper understanding of economic situation (Huang et al., 2020). Stock market is one of the most popular area for researchers due to its non-linear and non-stationary behaviour for forecasting time series data (Nevasalmi, 2020). The main objective of time series forecasting is to collect data based on past observations and develop forecasting model to predict the future using intelligent techniques by understanding the past data (Agrawal et al., 2013).

Many intelligent techniques found useful while forecasting stock market. Authors (Guresen et al., 2011)analysed multi-layer perceptron (MLP), dynamic artificial neural network (DAN2) and the hybrid neural networks which use generalized autoregressive conditional heteroscedasticity (GARCH) to extract new input variables. Authors (Khandelwal et al., 2015) proposed hybrid method and compared this with ARIMA, ANN, and Zhang's hybrid models and found best accuracy in proposed model.

Authors (Handa et al., 2019) proposed a hybrid approach of ANN and ANFIS with combined approach of Feature extraction techniques and wavelet for FX data prediction. Authors (Ardabili et al., 2020) has done comparative analysis of machine learning models for highly non-linear data of COVID-19 outbreak and found that machine learning worksas effective tool to handle this pandemic situation. Author (Selvin et al., 2017) uses different architectures of deep learning for comparative study of financial prediction of NSE listed companies. Authors (Huang et al., 2020)propose an improved VMD and LSTM based financial data forecasting model and results indicate that the model is accuracy-promising and superior to the baseline models in one-step-ahead forecasting of financial time series. Authors(Liang et al., 2019) presents wavelet threshold-denoising method, which has been widely applied in signal denoising, is adopted to pre-process the training data and found that multioptimal combination wavelet transform (MOCWT) is performing well in terms of model accuracy.

This study consider Artificial Neural Network (ANN) technique and Long Short-Term Memory (LSTM)(Chimmula& Zhang, 2020)todevelop forecasting model and done comparative analysis between them. In this research work, we have worked with two stock data National Stock Exchange (NSE50) and Bombay Stock Exchange (BSE30) collected from www.nseindia.com and www.bseindia.com. The nature of stock data is non-linear, and fluctuate mostly due to occurrence of any event. The event we consider in this proposed research work is

the time duration of COVID pandemic (From March 01, 2020). The prediction of stock data during this period is proposed by many authors(Ahmar& del Val, 2020; Contessi & Pierangelo, 2020; Takyi & Bentum-Ennin, 2020).It is difficult for researches to develop an accurate and robust predictive model with the data of non-linear behaviour. To overcome the issue of non-linearity from the data, it is necessary to remove noise from it. In this proposed work to remove non-linearity, we have used MATLAB tool i.e. wavelet signal denoiser(MathWorks, 2019), which remove the noise from the data and make it clean to get more accuracy in predictive result (Kozlowski, 2005). The performance of predictive model iscalculated by two error measures: Mean Absolute Percentage Error(MAPE) and Root Mean Square Error(RMSE). The rest of the paper is organized as follows: section 2 describes about dataset and methodologies we have used in this proposed work. We evaluate our model in section 3 and finally paper is concluded in section 4.

2. MATERIAL AND METHODOLOGY

2.1 Dataset

In order to develop a financial time series forecasting model, it is needed to collect data from various sources for observation based on historical data to predict the future value. In this research work we have focused on stock data. While using stock data, we have worked with two stock data: NSE50 and BSE30 for time series forecasting. The NSE data is collected from www.nseindia.com and BSE30 data is collected from www.bseindia.com. We have collected 3 years data from January 1, 2018 to December 31, 2020. This dataset also contain data during COVID pandemic in India. The both stock data has four index open, high, low and close.

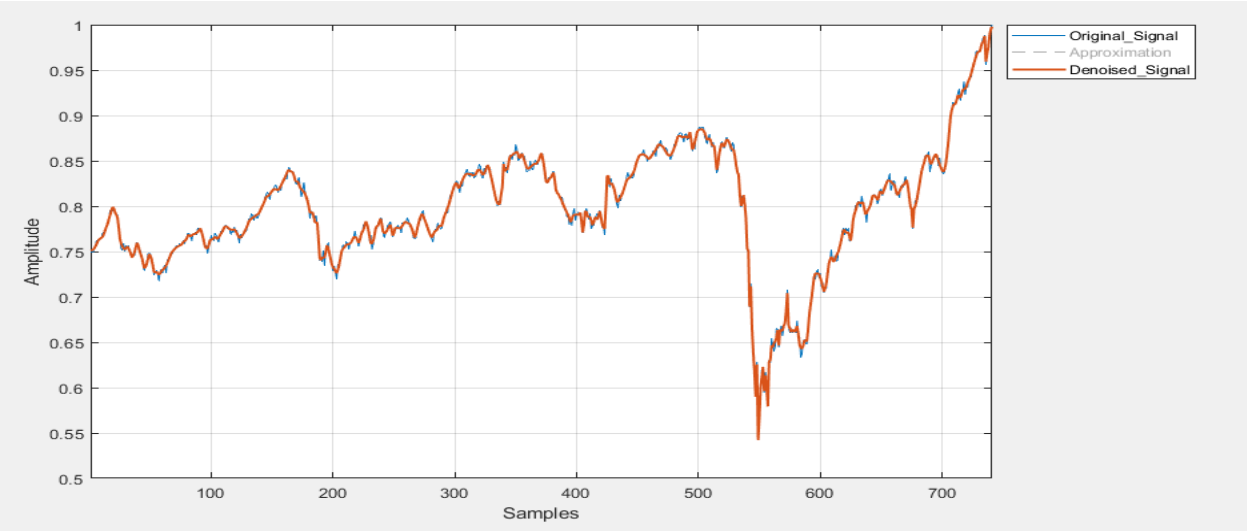
After data collection, the data is normalized to scale the data between [0 1](Sola & Sevilla, 1997). Due to nonlinear behaviour of stock data some pre-processing techniques can be applied to remove the noise and inconsistency from data. In this research work we have worked with wavelet signal denoising to clean the data for developing a robust forecasting model. After pre-processing, the data is partitioned into training, testing and validation dataset to check the robustness of model. Before COVID pandemic, the data is partitioned into training and testing dataset and the stock data during the COVID pandemic in India is considered as validation data. Table 1 shown the summary about data we have used in this research work.

Table 1: Summary of Dataset.

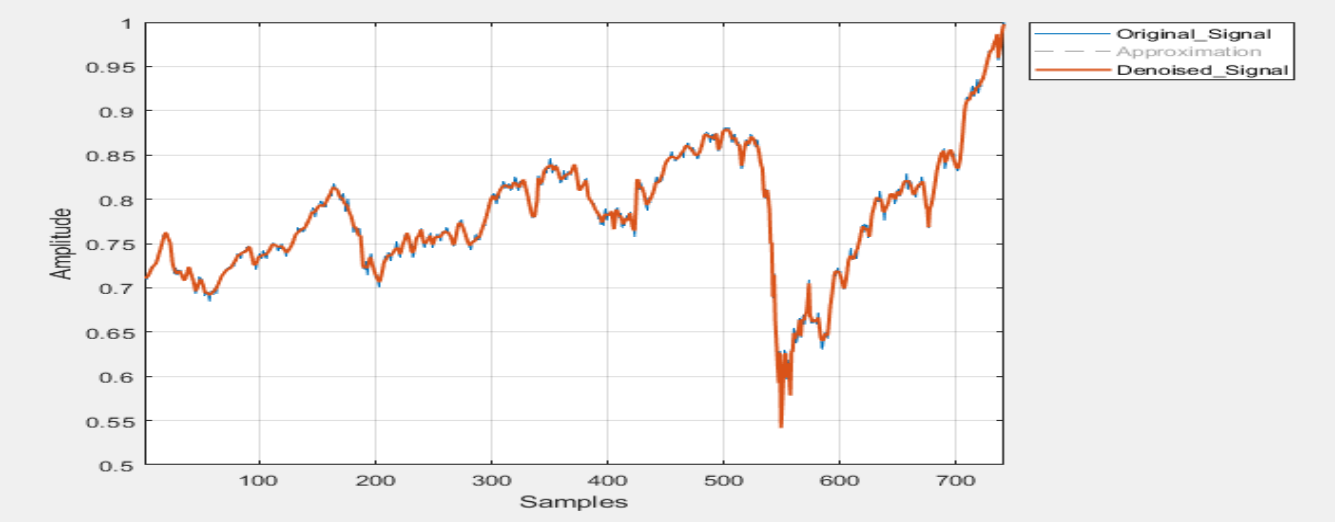
Particular	**Description**
Nature of Data	Stock Data (Daily)
Type of Stock Data	National Stock Exchange (NSE50) Bombay Stock Exchange (BSE30)
Periods	January 01, 2018 to December 31, 2020
Downloaded from	NSE50 (www.nseindia.com) BSE30 (www.bseindia.com)
#of Observations	741
Partition	Training (January 01, 2018 to September 25, 2019 Testing (September 26. 2019 to February 28, 2020) Validation (March 01, 2020 to December 31, 2020)

2.2 Wavelet Denoising

The fluctuating and non-linear behaviour of time series data can degrade the quality of forecasting result (Ustazhanov, 2014) so it is very important to remove this noise from signals while developing a robust predictive model. The wavelet signaldenoising method have been considered in this research work by using MATLAB tool called wavelet signal denoising, for reducing unwanted noise signal from data (MathWorks, 2019).Wavelet signal denoiser tool is an interactive tool by MATLAB to remove the noise and inconsistency from time series data to clean the stock data for accurate prediction. The graphical representation between original and denoised signal of NSE50 and BSE30 dataset is shown in figure 1.



(a)



(b)

Figure 1: Original and denoised signal of (a) NSE50 and (b) DOW30.

2.3 Artificial Neural Network

Artificial Neural Network(Zurada, 2003)is work in the way as brain perform.Neural networks can be used for both classification (to predict the class label of a given tuple) and numeric forecasting (to predict a continuous-valued output). The ANN is organized in layers and classified into feedforward and feedback neural network. Neural network has input, hidden and output layers as shown in figure 2. Each connection between the layers has some weights that affects the accuracy of model.

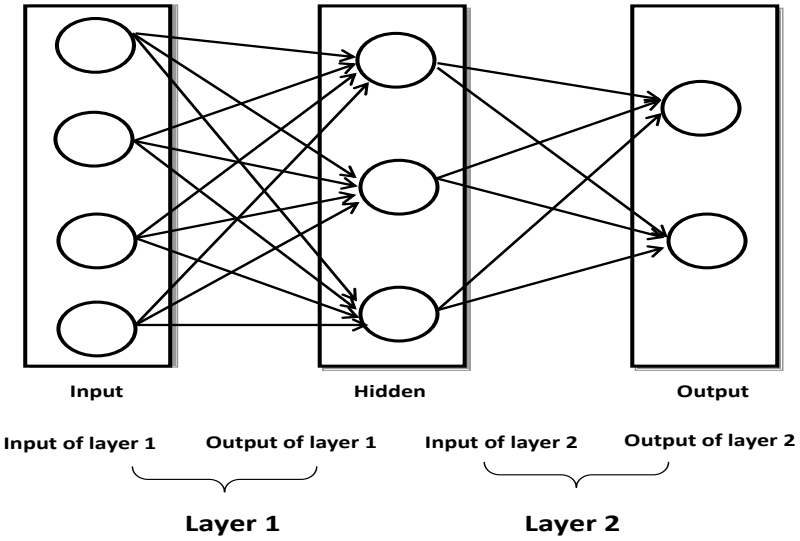


Figure 2: Artificial Neural Network architecture.

In this research work, the another method of artificial neural network (ANN) called Back Propagation Neural Network (BPN) is used to construct the forecasting model., where the data is given to the input layer and it is propagated forward to the network through another layers until it reaches to the output layer then output which we get from this layer is compared with actual output and then error value is calculated and propagated back to the network to updates the weight to get more accurate output and repeat until it comes closer to the desired output.

2.4 Long Short-Term Memory (LSTM)

LSTM is an architecture of deep learning with plenty of layers. LSTM is a combination of five special components (Ni et al., 2019) called gates which can model both long term as well as short term data they are: cell state, hidden state, input gate, forget gate and output gate. Cell state has the ability to add or remove information, hidden state is used to predict future data, Input state is used to decide the information flows from current input to cell state, Forget gate decided flow of information from previous state to current state and output gate describes the output acquired by LSTM. The architecture of LSTM is shown in figure 3.

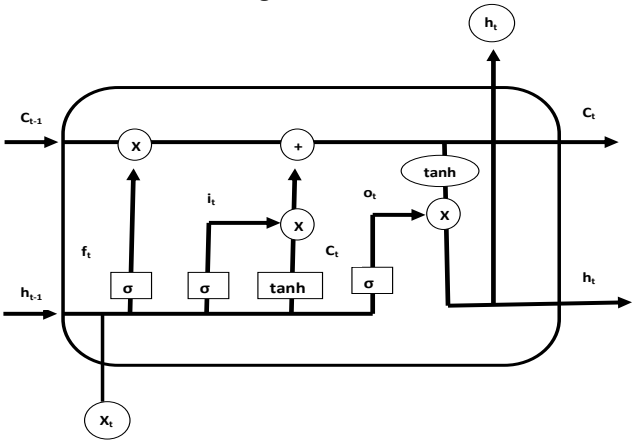


Figure 3: Architecture of LSTM.

3. RESULT AND ANALYSIS

In this proposed research work MATLAB tool is used to implement machine learning techniques for development of time series forecasting model. In this research work we have worked with two time series data: NSE50 and BSE30 stock data collected from www.nseindia.com and www.bseindia.com respectively. We have done static partitioned of data, where 80% data is used for training and 20% data is used for testing purpose. The rest of data from periods of COVID pandemic in India is considered as validation data. After partitioning of data wavelet denoisingis done by using wavelet signal denoisingtool of MATLAB to remove noise from nonlinear stock data. After datapre-processing the input data is given to ANN and LSTM model for next day ahead prediction. The accuracy of model is measured by two error measures: Mean Absolute Percentage Error (MAPE) and Root Mean Square Error(RMSE).The process flow diagram of proposed work is shown in figure 4. The comparative result between ANN and LSTM of both stock data is shown in table 2. From the table is has been found that LSTM is outperform ANN.

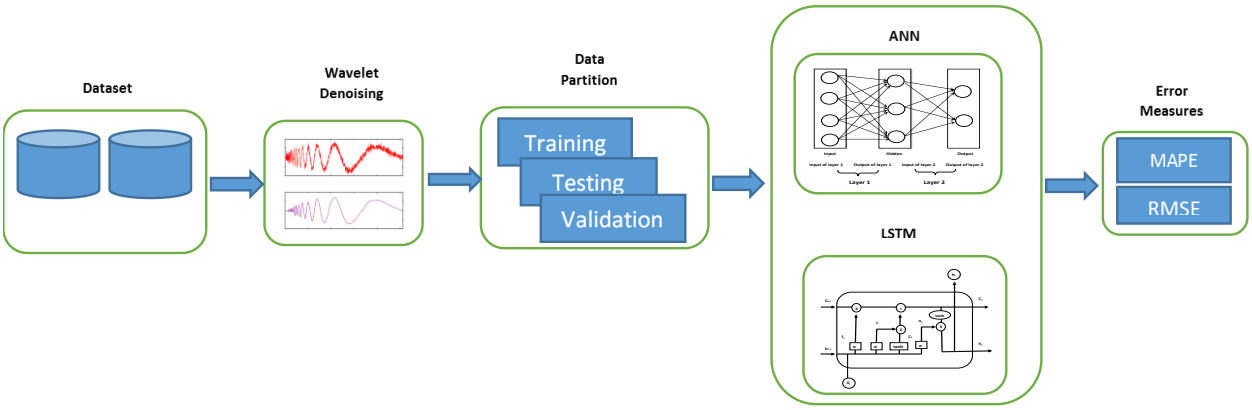


Figure 4: Process flow diagram of proposed work.

Table 2: Comparative analysis between ANN and LSTM on testing dataset.

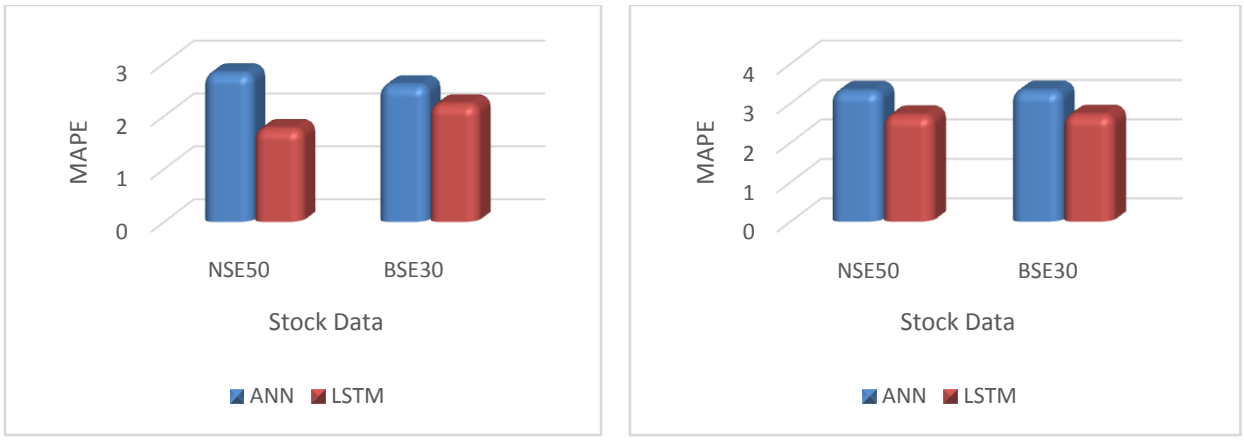
Dataset	ANN		LSTM	
	MAPE	RMSE	MAPE	RMSE
NSE50	2.811	0.030	1.759	0.016
BSE30	2.586	0.035	2.224	0.020

The comparative result on validation dataset in shown in table 3 and depict that error value in validation dataset is somehow higher than testing dataset as shown in table 2, which is somehow obvious as we have taken validation dataset during the period of COVID pandemic, at which time stock data shown a significant decline.

Table 3: Comparative analysis between ANN and LSTM on validation dataset.

Dataset	ANN		LSTM	
	MAPE	RMSE	MAPE	RMSE
NSE50	3.307	0.038	2.687	0.027
BSE30	3.316	0.046	2.703	0.027

Figure 5 depicts the graphical representation between ANN and LSTM for both NSE50 and BSE30 testing as well as validation dataset and found that LSTM is outperform in both dataset as well as in both partition.



(a) (b)

Figure 5: Graphical representation between ANN and LSTM of NSE50 and BSE30 (a) testing data and (b) validation data

4. CONCLUSION

In this research work we have worked with two stock data NSE50 and BSE30 and remove non-linearity using wavelet signal denoiser tool to develop a model with accurate prediction using combined approach of Wavelet and ANN/LSTM for next day ahead prediction.The data is partitioned into training(80%) and testing(20%) using static data partition. The data during COVID-19 is considered as validation data and given to ANN and LSTM to check the robustness of model. The result came from LSTM is performing better in both dataset with MAPE=1.759 and RMSE=0.016 for NSE50 and MAPE=2.224 and RMSE=0.020 for BSE30 in testing dataset and MAPE=2.687and RMSE=0.027 for NSE50 and MAPE=2.703 and RMSE=0.027 for BSE30 in validation dataset. However, the result came from validation dataset is higher because at that time stock data shown a significant decline.

REFERENCES

Agrawal, R. K. R. A. K., Adhikari, R., & Agrawal, R. K. R. A. K. (2013). An Introductory Study on Time Series Modeling and Forecasting Ratnadip Adhikari R. K. Agrawal. *ArXiv Preprint ArXiv:1302.6613*, *1302.6613*, 1–68.

Ahmar, A. S., & del Val, E. B. (2020). SutteARIMA: Short-term forecasting method, a case: Covid-19 and stock market in Spain. *Science of the Total Environment*, *729*, 138883. https://doi.org/10.1016/j.scitotenv.2020.138883

Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U., Rabczuk, T., & Atkinson, P. M. (2020). COVID-19 Outbreak Prediction with Machine Learning. *SSRN Electronic Journal*, *April*. https://doi.org/10.2139/ssrn.3580188

Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons and Fractals*, *135*. https://doi.org/10.1016/j.chaos.2020.109864

Contessi, S., & Pierangelo, D. P. (2020). The International Spread of COVID-19 Stock Market Collapses. *Finance Research Letters*, *November*, 101894. https://doi.org/10.1016/j.frl.2020.101894

Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, *38*(8), 10389–10397. https://doi.org/10.1016/j.eswa.2011.02.068

Handa, R., Shrivas, A. K., & Hota, H. S. (2019). Prediction of FX Data Using ANFIS and Ann with Combined Approach of Wavelet and Feature Extraction Technique. *International Journal of Computer Science and Engineering*, *7*(3), 15–18.

Huang, Y., Gao, Y., Gan, Y., & Ye, M. (2020). A new financial data forecasting model using genetic algorithm and long short-term memory network. *Neurocomputing*, *xxxx*. https://doi.org/10.1016/j.neucom.2020.04.086

Khandelwal, I., Adhikari, R., & Verma, G. (2015). Time series forecasting using HYBRID ARIMA and ANN Models based on DWT Decomposition. *Procedia Computer Science*, *48*, 173–179. https://doi.org/10.1016/j.procs.2015.04.167

Kozlowski, B. (2005). Time Series Denoising with Wavelet Transform. *Journal of Telecommunications and Information Technology*, *3*, 91–95.

Liang, X., Ge, Z., Sun, L., He, M., & Chen, H. (2019). LSTM with wavelet transform based data preprocessing for stock price prediction. *Mathematical Problems in Engineering*, *2019*. https://doi.org/10.1155/2019/1340174

MathWorks. (2019). *Denoise a Signal with the Wavelet Signal Denoiser*. https://www.mathworks.com/help/wavelet/examples/denoise-a-signal-with-the-wavelet-signal-denoiser.html

Nevasalmi, L. (2020). Forecasting multinomial stock returns using machine learning methods. *Journal of Finance and Data Science*, *6*, 86–106. https://doi.org/10.1016/j.jfds.2020.09.001

Ni, L., Li, Y., Wang, X., Zhang, J., Yu, J., & Qi, C. (2019). Forecasting of Forex Time Series Data Based on Deep Learning. *Procedia Computer Science*, *147*, 647–652. https://doi.org/10.1016/j.procs.2019.01.189

Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2017). Stock price prediction using LSTM, RNN and CNN-sliding window model. *2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017*, *2017*, 1643–1647. https://doi.org/10.1109/ICACCI.2017.8126078

Sola, J., & Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science*, *44*(3), 1464–1468. https://doi.org/10.1109/23.589532

Takyi, P. O., & Bentum-Ennin, I. (2020). The impact of COVID-19 on stock market performance in Africa: A Bayesian structural time series approach. *Journal of Economics and Business*, 105968. https://doi.org/10.1016/j.jeconbus.2020.105968

Ustazhanov, M. (2014). Digital Signal De-Noising using MATLAB-Simulink Wavelet. *Journal of Technical Science and Technologies*, *2*(2), 35–39.

Zurada, J. M. (2003). *Introduction to Artificial Neural Systems* (six). Jaico Publishing House, Mumbai.