International Journal of Computer Engineering and Information Technology

VOL. 12, NO. 8, August 2020, 48–51
Available online at: www.ijceit.org
E-ISSN 2412-8856 (Online)

A Hybrid Arima-Lstm Model for Stock Price Prediction

Umar Farouk Ibn Abdulrahman[1], Najim Ussiph[2] and Benjamin Hayfron-Acquah3

[1, 2] School of Technology, Christ Apostolic University College, Kwaadaso, Ashanti, Ghana

[1, 3] Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ashanti Ghana

umarfarouk@cauc.edu.gh, nussiph@yahoo.com, jbha@yahoo.com

ABSTRACT

The stock market offers investors the opportunity to trade in shares and equities. Making profit in the stock market depends on the ability to accurately predict the future stock prices. This is leading to the development of stock price prediction models using various methods such as ARIMA, LSTM, and Neural Network. Even though these models have proven to predict future stock prices, recent time is witnessing the development of better models using hybrid approaches. This study therefore propose a hybrid model by combining ARIMA and LSTM based on data decomposition with low-pass filter of the discrete Fourier Transform. Experiment carried out with data from the Ghana stock exchange reveals that, the hybrid model performs better than the individual ARIMA and LSTM models. This is confirmed with their R.M.S.E values.

Keywords: *Stock Market, stock prediction, Hybrid stock predictions, ARIMA, LSTM.*

1. INTRODUCTION

Advances in the computing field have led to the application of Artificial intelligence (AI) in many domains such as healthcare, agriculture, business and energy. One particular area that has seen significant application of AI in modern times is the stock price prediction. Prior to the use of AI in predicting stock prices, a lot more was done using statistical and econometrics approaches to predict the future prices of stock for decision making purposes.

Machine learning (ML), a branch of AI has seen the usage of various algorithms configured in building stock price predictive models. Amongst the ML used are such as ANN (Hamed et al, 2012), Support vector regression (Rustam and Kintandani, 2019), Long Short Term Memory

(M et al, 2019). Ultimately, the uses of different methods are to achieve near accurate prediction results. However results have varied from time to time. .Typically, an accurate prediction results can lead to better decisions with eventual higher profit yields for investors.

Recent time has seen the combination of two or methods as hybrid models to better the performance of prediction models (Yang et al, 2019). However stock data is a time series which contains both linear and non-linear components (Babu and Reddy, 2014). Modeling such time series therefore requires the segregation of the time series in their respective components before the development of predictive model

2. LITERATURE REVIEW

Stock prediction has received tremendous attention from both academicians and practitioners in recent times. Abundant research is been carried out in predicting stock prices with the aim to offer investment tool for decision making process. Machine learning is prevalent in the development of these predictive models.

Gocken et al (2011) developed a hybrid model combining ANN and Metaheuristics. Genetic algorithm (GA) and Harmony search (HS) are used to facilitate the best selection of technical indicators, where the ANN is for the prediction.

Hegazy et al (2013) proposed the combination of particle swarm optimization and LS-SVM to predict stock prices. Technical indicators used in the model are stochastic oscillator, money flow index, exponential moving average, moving average convergence/divergence and relative strength index.

Creighton and Zulermine (2017**)** presented a hybrid stock price prediction using ARIMA, Backpropagation and exponential smoothing model (ESM). The S&P 400 and 500 data were both tested with the model

Hossain et al (2018**)** proposed a hybrid stock prediction model with LSTM and GRU. Both the GRU and LSTM were implemented using the python programming language with the library keras.

Ojo et al (2019) proposed a model to predict stock market behavior. LSTM is used as the method as the predictive method. Further, Gupta et al (2019) designed a stock price prediction model with the regression method. The model is tested with the Indian stock exchange. The paper failed to include any evaluation metrics to the performance of the model.

Also, Yang et al (2019) proposed a hybrid system using particle swam optimization (PSO) and Brain-Storm-Optimization (BSO). The proposed model is experimented with dataset from NASDAQ and DOW JONES. The performance of the model is ascertained using MAPE. RMSE, R2, ARV, VAF (%).

Mangampali et al (2020**)** proposed a hybrid stock prediction model combining ARIMA and Gated Recurrent neural network (GRU). ARIMA is first modeled with the dataset to obtain the linear component and the residual is modeled with the GRU. The result is then summed at the final prediction model

3. PRELIMINARIES

3.1 Long short term memory (LSTM)

The LSTM is an improved version of the recurrent neural network designed to remember information thereby avoiding long term dependency (M et al, 2019). It structure differs from other form of neural network.

3.2 ARIMA

ARIMA is a linear modeling technique (Babu and Reddy, 2014) which accepts only a single variable as input. ARIMA work with stationary data and therefore non-stationary data has to be continuously differenced until stationarity is obtained. Further it is to be fitted with the best ***p,d,q*** as in ARIMA**(p,d,q)** where ***p*** is the auto regressive terms, ***d*** is the differencing order, and ***q*** is the lagged errors.

The ARIMA forecast equations is presented as in **Eq. 1**

$$Y_1 = \emptyset_0 + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + .. + \emptyset_p Y_{t-p} + \epsilon_1 - \emptyset_1 \in_{t-1} - \emptyset_2 \in_{t-2} \quad \text{Eq.1}$$

where

Y_1 is the real value and ϵ_1 being the error term at ***t.*** $\emptyset_0$ *and* $\emptyset_j$ representing the coefficients of p and q.

3.3 Discrete Fourier transform

Fourier transform is a mathematical theorem which is been applied in engineering, time series application and the medical field. The Fourier theorem states that, for any periodic function f(x) that has frequency w_0 , it can be transcribed as sum of sinusoids which can be mathematically represented as in eq. 2

$$f(x)=\sum_{k=0}^{\infty}(A_k cos(kw_0x) + B_k sin(kw_0x)) \quad \text{Eq. 2}$$

where the above can be termed as Fourier series with sums of cosine and sines which are also multiples of the frequency and A_k and B_k being the Fourier coefficients. Fourier uses lower pass filter, upper filters and circular pass filter for data decomposition

4. PROPOSED MODEL

4.1 Data source

The data for this study was obtained from the Ghana stock exchange (https://gse.com). It is the stock price of a bank which has one thousand three hundred and ninety eight (1398) instances with the headers date, opening price, previous opening price, closing prices and previous closing prices for the period 1st February-21st September, 2020. However, for the development of the prediction model, only the closing price is chosen

4.2 Algorithm of the proposed model

1. Express the stock price data series as of sine and cosine functions as
   $$f(x)= \sum_{k=0}^{\infty}(A_k cos(kw_0x) + B_k sin(kw_0x)) \quad (3)$$
2. Decompose the series into linear (L) and non-linear (NL) components using discrete Fourier low-pass filter
   2.1 use a cutoff point of **0.3**
   $$\hat{h}(w) = \begin{cases} 1 & if|w| \leq w_c \\ 0 & if|w| \geq w_c \end{cases}$$
   Where $\boldsymbol{w_c}$ is the cut-off frequency of the low-pass, ***w*** being the amplitude spectrum.
3. Perform an inverse transform of the frequency
4. Enter the total in-sample [train (60%) and validation (40%)]
5. Get the linear prediction using ARIMA modeling (p=1, q=2, d=1)
6. Get the non-linear prediction using LSTM modeling
7. Obtain the final prediction by summing ARIMA and LSTM models

4.3 Development environment

Development of the model was carried out with the Python programming language using libraries Pandas, Numpy, with Jupiter notebook as the IDE. This was implemented on Core i7 machine with RAM of 8GB

Table 1: Sample Simulation result

DATE	Actual	Predicted		
	Closing price	ARIMA	LSTM	Proposed
5/24/2018	6.4	5.809199	5.978621	6.448111
5/28/2018	6.4	5.799461	5.956918	6.432296
5/29/2018	6.38	5.785728	5.932334	6.410884
5/30/2018	6.38	5.76922	5.906898	6.386118
5/31/2018	6.38	5.753216	5.885513	6.362689
6/1/2018	6.3	5.739385	5.870068	6.343186
6/4/2018	6.3	5.719682	5.847395	6.31772
6/5/2018	6.3	5.699532	5.826007	6.292536
8/5/2020	3.8	3.594179	3.690433	3.807311
8/6/2020	3.8	3.591362	3.687356	3.803491
8/7/2020	3.8	3.588579	3.685556	3.80039
8/10/2020	3.8	3.58605	3.684875	3.798139
8/11/2020	3.8	3.583894	3.685081	3.79673
8/12/2020	3.8	3.582153	3.685918	3.796065
8/13/2020	3.8	3.580819	3.687142	3.796002
8/14/2020	3.8	3.579854	3.688546	3.796382
8/17/2020	3.8	3.579202	3.68997	3.797057
8/18/2020	3.81	3.578806	3.691298	3.797895
8/19/2020	3.81	3.579614	3.693928	3.799964
8/20/2020	3.81	3.581007	3.696875	3.802586
8/21/2020	3.81	3.582582	3.69957	3.805282

To evaluate the performance of developed models, the root mean square error (RMSE) is used. The RSME is given by the formula

$$R.M.S.E. = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t + F_t)^2}$$

Equation 5

Table 2 shows the root means square error (RMSE) values of the different models. The lower the value, the better the performance. It can be concluded that the proposed hybrid ARIMA-LSTM model performs better than the single models and that the LSTM model also performs better than the ARIMA.

Table 2: Performance results of models

TECHNIQUE	RMSE
ARIMA	0.5202
LSTM	0.1625
HYBRID(ARIMA +LSTM)	0.0017

Fig 1, Fig 2 and Fig 3 are the graphical representation of ARIMA, LSTM and Hybrid models respectively. It shows the plot of the train, validation and prediction of the three (3) models.

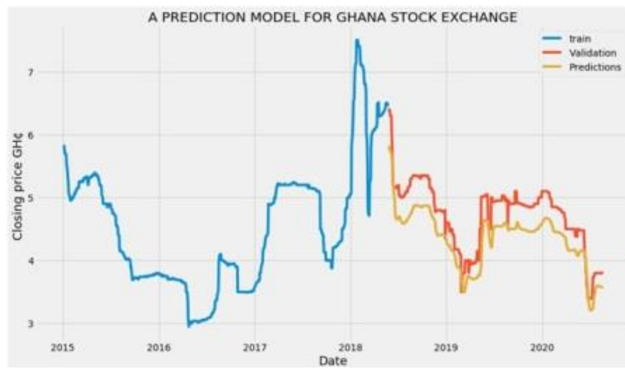

Fig . 1. ARIMA Model

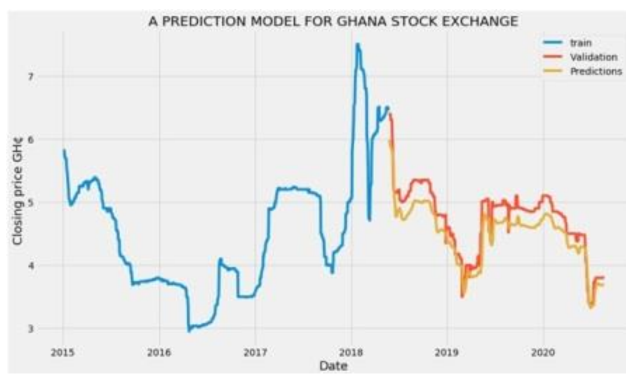

Fig. 2. LSTM Model

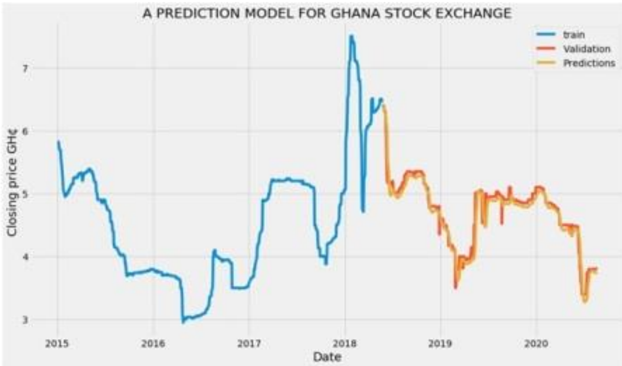

Fig. 3. Hybrid ARIMA-LSTM Model

4. CONCLUSIONS

In this paper, we have presented a new prediction model by combining the strength of ARIMA and LSTM based on DFT low-pass filter decomposition. Given that every time series data is a combination of linear and non-linear components, a DFT could be used to decompose any time series data and further used for prediction.

The prediction results of the hybrid model is obtained by summing the results of the individual methods which offers a better results than the individual models as can be seen from table 2 from the RMSE values

REFERENCES

[1] Göçken, M., Özçalıcı, M., Boru, A., & Dosdoğru, A. T. (2016). Integrating metaheuristics and artificial neural networks for improved stock price prediction. Expert Systems with Applications, 44, 320-331.

[2] Hegazy, O., Soliman, O. S., & Salam, M. A. (2014). A machine learning model for stock market prediction. arXiv preprint arXiv:1402.7351.

[3] Ojo, S. O., Owolawi, P. A., Mphahlele, M., & Adisa, J. A. (2019, November). Stock Market Behaviour Prediction using Stacked LSTM Networks. In 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC) (pp. 1-5). IEEE.

[4] Yang, B., Zhang, W., & Wang, H. (2019). Stock Market Forecasting Using Restricted Gene Expression Programming. Computational Intelligence and Neuroscience, 2019.

[5] Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018). NSE stock market prediction using deep-learning models. Procedia computer science, 132, 1351-1362.

[6] Babu, C. N., & Reddy, B. E. (2014). A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. Applied Soft Computing, 23, 27-38.

[7] Rustam, Z., & Kintandani, P. (2019). Application of Support Vector Regression in Indonesian Stock Price Prediction with Feature Selection Using Particle Swarm Optimisation. Modelling and Simulation in Engineering, 2019.

[8] Hamed, I. M., Hussein, A. S., & Tolba, M. F. (2012). An intelligent model for stock market prediction. International Journal of Computational Intelligence Systems, 5(4), 639-652.

[9] Hossain, M. A., Karim, R., THulasiram, R., Bruce, N. D., & Wang, Y. (2018, November). Hybrid Deep Learning Model for Stock Price Prediction. In 2018 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1837-1844). IEEE.

[10] Creighton, J., & Zulkernine, F. H. (2017, December). Towards building a hybrid model for predicting stock indexes. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 4128-4133). IEEE.

[11] Mangalampalli, R., Pandey, V., Khetre, P., & Malviya, V. Stock Prediction using Hybrid ARIMA and GRU Models.