Original Research Paper

Random Forest Based Feature Selection of Macroeconomic Variables for Stock Market Prediction

[1,2]**Isaac Kofi Nti,** [2]**Adebayo Felix Adekoya and** [2]**Benjamin Asubam Weyori**

[1]*Department of Computer Science, Sunyani Technical University (STU), Sunyani Ghana*
[2]*Department of Computer Science and Informatics,*
University of Energy and Natural Resource, Sunyani-Fiapre, Ghana

Corresponding Author:
Isaac Kofi Nti
Department of Computer
Science, Sunyani Technical
University (STU), Sunyani
Ghana
Email: ntious1@gmail.com

Abstract: A firm's equity price on the stock-market is reported to be closely related to the Macroeconomic Variable (MVs) of the country in which the firm trades. For this reason, researchers, market traders, financial analysts and forecasters to examine the association between MVs and stock-price have carried out numerous studies, using time-series statistical analysis methods like Autoregressive Integrated Moving Average (ARIMA), Autoregressive Moving Average (ARMA) and Generalised Autoregressive Conditional Heteroscedasticity (GARCH). However, these techniques are reported to suffer from limited predictive power and restrictive assumptions. Besides, in pursuit of ways to remedy these paucities and limitations within these techniques, some researchers have examined uncountable machine learning techniques for measuring the stock-markets trends and making trading decisions using macroeconomic variables. On the other hand, a higher percentage of these studies paid attention to the stock index prediction and neglected the diversity of MVs that influence different sector indices. In addressing the issues above, this study seeks to examine the degree of significance between different sectors stock-price and MVs and predict a 30-day head stock-price using Random Forest (RF) with an improve leave-one-out cross-validation tactic and Long Short-Term Memory Recurrent Neural Network (LSTMRNN). An empirical analysis of the proposed model over the Ghana Stock Exchange (GSE) exhibits high prediction accuracy and better mean absolute error compared with other time-series techniques. It can, therefore, be inferred from the fallouts that the proposed stock-market prediction with MVs, provides an efficient approach to automatic identification and extraction of MVs that affect diverse sector stock and offer an accurate prediction of a stock's future price.

Keywords: Macroeconomic Variable, Inflation Rate, Recurrent Neural Network, Random Forest, Ghana Stock Exchange

Introduction

In every country, the financial markets play an essential role in the economy, in areas such as allocation, re-allocation, mobilisation, and pricing of capital resources to several stages of monetary intermediation (Asekome and Agbonkhese, 2015; Dosdoğru *et al.*, 2018). However, Macroeconomic Variables (MVs) such as the inflation rate and Gross Domestic Product (GDP) are reported to determine a country's growth (Ratti and Vespignani, 2015; 2016). Thus, the instability in a country's inflation rate or prices leads to the poor performance of the overall economy of a country (Siok *et al*., 2015).

As the market value of a company is believed to depend on a company's current economic status and imminent perspectives, the value of firms listed on the stock-market of any given country is believed to hinge on the overall economic status quo and imminent perspectives in the country (Peiró, 2015). With this, one can infer that the stock-prices are associated with economic changes happing or anticipated to happen within a country.

The remaining sections of this paper are structured as follows. Section 2 covers a brief description of machine learning, its application in a real-world situation, a review and comparison of stock-market prediction with macroeconomic variables in literature, and the anticipated

contribution of the current study. Section 3 discusses the methods and material employed in the study. Section 4 covers the empirical results and discussion. Lastly, Section 5 presents the conclusion and direction for future studies.

Literature Review

This section presents a brief understanding of machine learning, its application in the real world. A short discussion on macroeconomic variables and the stock market price movement and a summary of related works.

Machine Learning (ML)

ML is the practice where, a computer software or robot is said to study from knowledge (K) relating to some group of tasks (T) and evaluation metric (E_p), if E_p in task (T) improves with K. Thus making machines behave like humans (Nti *et al.*, 2019). ML have been applied in numerous sectors to improve productivity, profit, and efficiency. These sectors includes, healthcare (Choudhury, 2019; Choudhury and Eksioglu, 2019) for disease diagnosis, in education for predicting the academic performance of students (Devasia *et al.*, 2016; Nti and Quarcoo, 2019), in the energy industry for predicting electrical energy demand by consumers (Nti *et al.*, 2019) and in agriculture for predicting crop yields (Akshatha and Shreedhara, 2018; Crane-Droesch, 2018).

Economic Variables and Stock Market

Experimental evidence appears to support the view that global stock-markets have usually evolved in close association with local financial-prudence (Peiró, 2015). Therefore, the vicissitudes in macroeconomic variables can have a profound influence on the stock market. This impact of MVs on the economy and the stock market has resulted in many studies from researchers of all occupations, and this section discusses a few of them.

Using autoregressive distributed lag or ARDL model, the changes in global oil-price was found to have a direct effect on domestic inflation in low oil reliance nations and indirectly affecting the domestic inflation in high oil reliance nations (Siok *et al.*, 2015). The joint effect of Gross Domestic Product (GDP), Capacity Utilisation (CAU), an Exchange Rate (EXR), Money Supply (M2) and Inflation (INF) on All Share Index (ASI) in Nigeria was examined by (Asekome and Agbonkhese, 2015). Their results revealed that GDP and INF had a significant association with ASI. On the other hand, their study was silent on the various sectors in which these indicators performed significantly. The relationship between oil price and China's stock market was also carried out by (Wei and Guo, 2017). The study concluded that the price of oil affects the price of a stock, without explicitly indicating which sector of the Chinese economy is affected by the oil price.

The study by (Jareño and Negrut, 2016) in the United States using Pearson correlation shows a statistically significant relationship between GPD, the industrial production index, the unemployment rate, long-term interest rates, and stock return. A significant relationship between the MVs and stock returns in Kenya, Uganda, and Tanzania was also identified by (Laichena and Obwogi, 2015). In Nigeria (Asekome and Agbonkhese, 2015) discovered a significant relationship between GDP and money supply stock returns and insignificant relationship between exchange rate, capacity utilisation, and stock returns using Ordinary Least Squares technique.

The study by (Mutuku and Ng'eny, 2015) revealed a negative relationship between inflation rate and equity movement in Kenyan using Vector Autoregressive (VAR) and Vector Error Correction (VECM). A study, conducted in 2015 in Indian, Bangladesh, Pakistan, and Sri Lanka using VAR and Principal Components Analysis (PCA) revealed that economic factors could not explain stock returns in India. However, economic activity directly explains stock returns of Bangladesh and indirectly explains stock returns of Pakistan and Sri Lanka (Khan *et al.*, 2015).

Several studies globally have also examined the extent to which MVs affect the stock-market and concluded that there is an association between stock-price movement and MVs. These studies, however, did not say which sectors' stock is affected by which MVs.

In Ghana, some researchers' attempted to examine the association between MVs and stock-price on the Ghana Stock Exchange (GSE). Using the Autoregressive Distributed Lag Model (ARDL) and error correction parameterisation techniques, a significant relationship between MVs (exchange-rate, inflation-rate, and interest-rate) and stock returns in Ghana was discovered by (Kwofie and Ansah, 2018). A contrary short-run, decisive long-term, and unidirectional causality from inflation to stocks return in Ghana using VECM were revealed by (Issahaku *et al.*, 2013).

A unidirectional causality running from inflation to stock returns in Ghana was also revealed by (Frimpong, 2011). The weak-form efficacy of listed companies on the GSE was analysed by (Ayentimi *et al.*, 2013). The correlation between stock-price and inflation rate using Pearson correlations and linear regression was carried out by (Gyan, 2015). The effect of Liquidity growth and the interest rate on stock-price was examined in (Boachie *et al.*, 2016).

The correlation between exchange-rate and stock-price was evaluated using multiple causality tests and Bayesian Quantile Regression (QR) by (Gideon *et al.*, 2015) whose outcome confirms a positive relationship between exchange-rate and stock-price. The linkage between economic growth and stock-market performance

was carried out in (Manu, 2017), which confirms that the increase in the economy of a nation is evident in her stock-market prices. In fact, (Lahmiri, 2011) argues that MVs are better predictors of stock price movement than technical indicators.

Though several studies on MVs and the stock market both locally and globally concluded that MVs are associated either positively or negatively with stock-market movement, a high percentage of these studies did not explicitly revealed the actual sector stocks where these MVs were directly or indirectly associated.

On the other hand, a high percentage of these studies suffer from two teething troubles that are lack of direction of causative and multicollinearity problem among predictors. Most of these earlier studies were undertaken using a multivariate analysis approach, which lumped up macroeconomic variables such as exchange rate, interest rate, foreign direct investment, money supply, and inflation in one model. According to (Adusei, 2014) it leads to doubt on the acceptability of their outcomes and conclusions.

Secondly, most studies examining macroeconomic variables and stock-price movements were silent on how significant individual macroeconomic variables relate to stock-index from diverse sectors of the economy. Thus, generating an opening in the literature in that the MVs that possibly would affect stocks in the agriculture sector could diverge significantly from the MVs' swaying the banking or oil sector index.

Thirdly, a considerable number of these studies tried to measure the association between MVs and stock-price using time-series methodologies such as ARIMA, ARMA, and GARCH. These statistical techniques or methods are reported to have some defects, according to Weng (2017). They suffer from (a) limited predictive power and (b) restrictive assumptions (Weng, 2017).

Again Caiani *et al.* (2016) argue that despite the improvements introduced by some researchers to boost the realism of time-series models, their key drawbacks still emasculate their reliability (Caiani *et al.*, 2016).

In addition, these statistical techniques are reported to submit themselves to the theory of linear variation of the stock-prices during a definite period of time and they usually do not support easy-automation process, as they require alteration and adaptation at every single stage, calling for certain uniformities and static nature of the target-data (Bisoi and Dash, 2014).

Furthermore, (Vaisla and Bhatt, 2010) argue that the predictive capability of traditional arithmetical or statistical models has been frequently below par and as these models become complicated, the estimation error escalates.

In pursuit of ways to overcome these deficiencies and limitation of time-series methodologies, numerous market traders, researchers, and analysts have examined countless

intelligent-system and machine-learning techniques for measuring the stock-markets' predictions using macroeconomic variables. Some of these techniques used include Support Vector Machine (SVM), Artificial Neural Network (ANN), Probabilistic Neural Networks (PNN), Principal Component Analysis (PCA), Genetic Algorithms (GA), Autoencoders (AD) and decision trees (CART) (Tsai and Hsiao, 2010; Lahmiri, 2011; Ballings *et al.*, 2015).

Nevertheless, almost all these studies paid attention to the stock index prediction and neglected the diversity of MVs that influence different sector indices.

Hence, this paper seeks to address the aforementioned issues by employing a random forest-based feature selection techniques enhanced with leave-one-out cross-validation tactic and a Long Short-Term Memory Recurrent Neural Network (LSTMRNN) predictive model to: (i) investigate whether and by what means different macroeconomic variables' fluctuations affect the stock-market liquidity in various areas of the Ghanaian economy. (ii) Predict a 30-day (monthly) stock price in different economic sectors based on macroeconomic variables and (iii) Evaluate proposed prediction model against ARFIMA time-series model. The current study contributes to the body of knowledge as follows:

i. A mongrelised machine-learning model of RF and RNN for effective prediction of stock-market using macroeconomic variables

ii. A novel approach for eliminating the multicollinearity problem in stock-market prediction with MVs, through an improved leave-one-out cross-validation tactic to handpicked significant and key relevant macroeconomic predictors for a given stock index

iii. An increase to the scanty literature of the emerging stock market, to the measuring of association between macroeconomic variables and stock returns and

iv. A better explanation of which macroeconomic variables affects which sector stocks, to provide an investment risk indication to investors on the Ghana stock market

Materials and Methods

This section introduces the methods and techniques adopted for the development and modelling of the proposed feature selection approach for stock-market prediction based on macroeconomic variables.

Research Design

Figure 1 shows the study workflow diagram for predicting the stock market based on macroeconomic variables. The framework consists of 3 phases: Phase one, phase two, and phase three. In step 1, the empirical datasets

are downloaded from different online resources. In stage 2, a random forest with an enhanced leave-one-out cross-validation technique was introduced in selecting most significant-predictors. Phase 3 involves the application of the predictive model based on Long Short-Term Memory Recurrent Neural Network (LSTMRNN) over the most significant features from stage 2.

Dataset and Data-preparation

The study dataset was obtained from three sources. The historical stock prices were downloaded from the GSE official website from January 2002, through December 2018, which includes year high, year low, previous closing price, opening price, closing price, price-change, closing bid price and closing-offer. Stock-data in four economy sectors (Petroleum, Banking, Technology, Pharmaceuticals, and Telecommunication industries) were selected. The macroeconomic variables dataset was obtained from the Bank of Ghana (BoG) official site and https://knoema.com/atlas/Ghana/Poverty-rate from January 2002 to December 2018. Table 1 (Appendix 1) shows the forty-two (42) initially selected macroeconomic indicators. These 42 features were selected based on their association with the stock market price movement, as reported in the literature, discussed under section 2 of this study.

Equation (3.1), (3.2) and (3.3) represent a matrix notation of our initial dataset of macroeconomic variables. Let A represent matrix, where each column represents each initially selected feature, which in the current study is forty-two (42) and (n) be the total records (number of rows):

$$A = [n \times 42] \, matrix \tag{3.1}$$

$$A \in \mathbb{R}^{n \times 42} \tag{3.2}$$

$$A = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots\dots x_{42}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots\dots x_{42}^{(2)} \\ x_1^{(3)} & x_2^{(3)} & x_3^{(3)} & \dots\dots x_{42}^{(3)} \\ x_1^{(4)} & x_2^{(4)} & x_3^{(4)} & \dots\dots x_{42}^{(4)} \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{(n)} & x_2^{(n)} & x_3^{(n)} & \dots\dots x_{42}^{(n)} \end{bmatrix} \tag{3.3}$$

Data Preprocessing

The downloaded datasets were preprocessed for better and accurate prediction of the machine-learning algorithm. The data pre-processing includes (i) Data integration, (ii) data cleaning, and (iii) data transformation.

Data Cleaning

The complexity and stochastic nature of stock data make it always prone to noise, which might disturb the ML algorithm from studying the structure and trends in data. The Wavelet Transform (WT) as given in Equation (3.4) was applied to the obtained data to free it from noise and resolving data inconsistency. We transformed the data (X_ω), using WT, remove coefficients (a, b) with values more than full a Standard Deviation (STD). We then inverse the transformed coefficients to get new data, free from noise:

$$X_\omega(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\varphi\left(\frac{t-b}{a}\right)dt \tag{3.4}$$
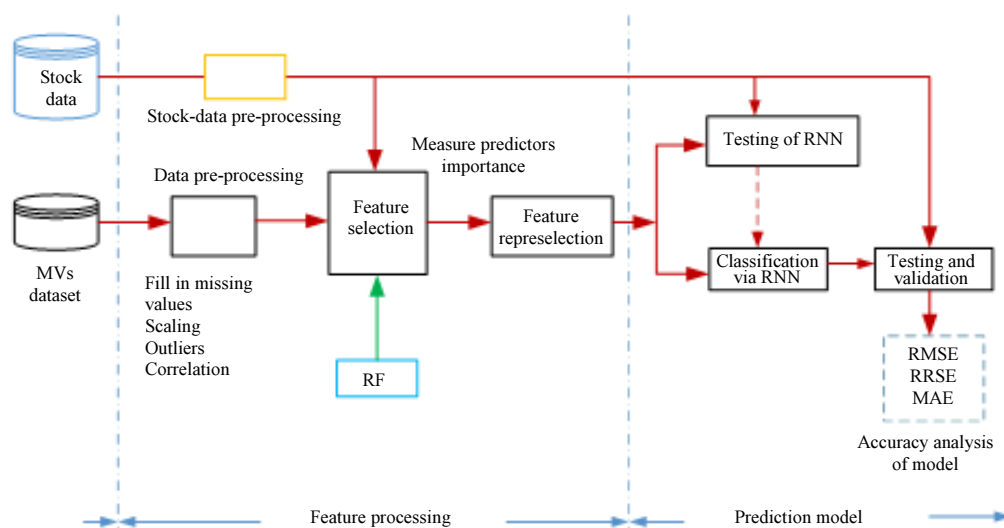

Fig. 1: Workflow diagram of the proposed prediction model

Data Transformation

The clean dataset was normalised within the range of [0, 1], using Equation (3.5) to enhance the performance of the ML algorithm.

Where:
$x^{'}$ = The new value obtained after normalisation
x = The value to be normalised, x_{min} and x_{max} are the minimum and maximum value of the dataset

$$x' = \frac{x - x_{(min)}}{x_{(max)} - x_{(min)}} \qquad (3.5)$$

Dimensionality Reduction

The Dataset (DS) had forty-two (42) initially selected macroeconomic variable (K). DS was passed through feature selection techniques to reduce dimensionality. The aim was to find a subset (U) to maximise the performance criterion C. $U = argmaxC(U)$, $s.t|U| = m$, $m<<d$, where d and m are the feature dimension of the original and selected, respectively. Random Forest (RF) algorithm with an improved leave-one-out cross-validation tactic was used to elite significant macroeconomic predictors.

Random Forest (RF)

RF combines the performance of several decision tree algorithms to predict the value of a variable (Wakefield, 2013; Rodriguez-Galiano *et al.*, 2015). The RF builds several regression trees (N) and averages their results from a received input vector Dataset (*DS*) of unique features. The RF was applied to the Dataset (*DS*) to rank the features. The root Means Squared Error (RMSE) obtained by each tree is recorded. The process is repetitive after arbitrarily permuting the values apiece of the independent variables. The resulted difference in errors is average between the non-permuted and permuted trees and then standardised by the Standard Deviation (STD) of their differences. The significant feature is calculated using Equation (3.6).

$$BG_k = \frac{\sum_{\theta=1}^{N}(e_{\theta k} - e_{\theta,\pi k})}{N.\hat{\sigma}} \qquad (3.6)$$

Where, N represents the total number of a tree, the RMSE of the trees (θ) represented by $e_{\theta k}$ afore permuting (k), RMSE of trees (θ) after permutation of (k) is $e_{\theta,\pi k}$. While δ being the STD of the variances concerning $e_{\theta k}$ and $e_{\theta,\pi k}$. An independent variable that resulted in a lager *GB* is considered more important than the one that resulted in a lower *GB*. The following algorithm was used for the feature selection:

Feature Selection Algorithm

Train the RF with the Dataset (DS) *including all k features*
Calculate the RMSE average of the model on cross-validation data
Use Equation (3.6) to rank the features
for *each subset of features* $k_i = \{(k-1),(k-2),(k-3),(k-4),....1\}$

    do
        Train a new RF using the feature k_i with the highest GB
        Calculate the RMSE average of the model on cross-validation data
        Use Equation (3.6) to rank the features
end for
Determine which k_i generated the lowest RMSE

Recurrent Neural Network (RNN)

RNN is one of the classes of Artificial Neural Network (ANN). The RNN can be seen as a Feedforward Neural Network (FFNN) enhanced with feedback loops or backpropagation through time. In RNN an input data cycles via a loop and every decision by the RNN takes into accounts the current input and the knowledge it acquired from the inputs it encountered earlier (Appiah *et al.*, 2019). The decision arrived by RNN at time-step (*t-1*) affects the decision; it will arrive a moment far-ahead at time step *t*. That is, the RNN has two input sources, the recent and the present, they amalgamate to estimate how they react to new data.

Basic RNN Algorithm

1. Data (X) is input into the RNN model
2. X is processed using the current context
3. After processing data is output, alongside a new context for the model
4. Repeat 1 to 3 until all data points are processed

The simple RNN falls typically into the vanishing gradient and the exploding grading problem. To overcome this, the Long-Short-Term Memory (LSTM) RNN was adopted for the study. Figure 2 shows the network of the RNN enhanced with LSTM. The LSTM is an abstraction of computer memory that works as intended with the RNN; it aids to maintain, update, and regulate the state of the network model outside its normal execution flow. When an RNN is a train-employing backpropagation through time and time, it overcomes the vanishing gradient and the exploding grading problem. Three gates, namely keep, write, and read gates are associated with the RNN. The keep gate retains the data in the information cell, it receives the same input-data and stays as the network, and it then computes how much of the present data should be remembered.

204

The write gate is accountable for inputting new data into the network, receiving the same input-data and stay at the network; it also receives the RNN output-data from the most current time set, and it uses the input-data to decide how much of the output should be written into the cell. The read gate sends data from the LSTM back into the RNN while the information cell is responsible for holding the data.

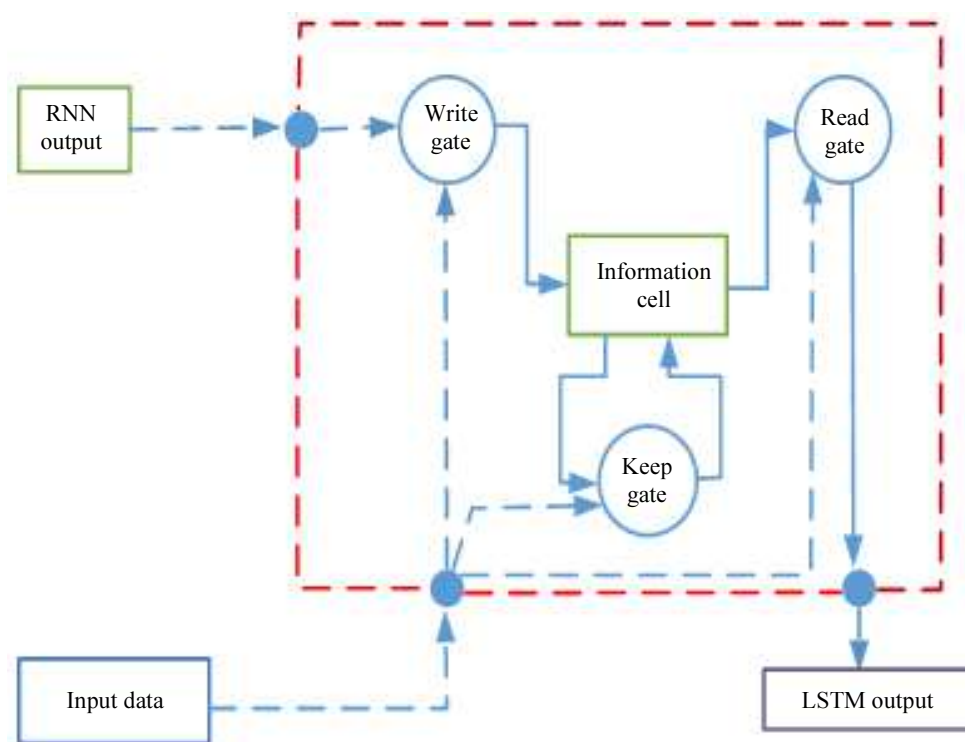### RNN is One of the classes of Artificial Neural Net

The choice of optimiser for an LSTM model has a significant effect on how fast the algorithm converges to the minimum value and the ideal of uncertainty to prevent getting trapped in a local minimum and not getting to the global minimum. The Adam (Adaptive Moment Estimation) optimiser was adopted for model optimisation. The adopted optimiser put-together the advantages of two other optimists, namely RMSprop and ADAgrad. The ADAgrad optimiser fundamentally adopts different Studying Rate (SR) for every single parameter and each time-step. The ADAgrad apportions more significant studying rates to irregular parameter while a lesser studying rate is apportioned to recurrent parameters. In an attempt to prevent overfitting of the proposed model during training, the Tikhonov regularisation proposed by (Golub *et al.*, 1999) was adopted and implemented for the study.

Evaluation Metrics

The predictive power of the proposed model was determined using well-known performance metrics in literature which includes (i) Mean Accuracy Error (MAE), as given in Equation (3.7), measures the average size of the errors in a set of predictions, without making an allowance for their direction. It is the average over the test sample of the absolute variances between ($\hat{y}_i$) and (y_i) observation where all individual differences have equal weight. (ii) Root Mean Square Error (RMSE). The RMSE is used for measuring model performance, calculating the difference between predicted ($\hat{y}_i$) and the actual (y_i) values as given in Equation 3.8. (iii) Relative Absolute Error (RAE) as given in Equation (3.9). (iv) The correlation coefficient (R) as given in Equation (3.10), is a performance index that measures the degree of associations between predicted values and actual values, its value ranges from 0 to 1 and the bigger the *R* is the better the model performance. (v) Root Relative Squared Error (RRSE).

Where (y_i) are the actual values and ($\hat{y}_i$) are the predicted values:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right) \tag{3.7}$$

Fig. 2: Proposed LSTM for enhancing RNN for stock prediction

205

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)}$$

$$)\hspace{6cm}(3.8$$

$$RAE = \frac{\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\frac{1}{n}\sum_{i=1}^{n}|y_i - \overline{y}|}\hspace{2cm}(3.9)$$

$$R = \frac{\sum_{i=1}^{n}\left(t_i - \overline{t}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(t_i - \overline{t}\right)^2 \cdot \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}}\hspace{1cm}(3.10)$$

where $\overline{t} = \frac{1}{n}\sum_{i=1}^{n}t_i$ and $\overline{y} = \frac{1}{n}\sum_{i=1}^{n}t_i$ are the average values of t_i and y_i, respectively.

Experimental Results and Discussion

The empirical results of the proposed predictive model with RF and LSTMRNN and comparison with ARIMA model are discussed in this section. All experiments were carried out with R studio.

Essential Factors for Ghana Stock Major Sectors

Forty-two (42) features were selected initially, their significance concerning stock-prices in the petroleum, banking, technology, pharmaceuticals and telecommunication industries were measured and presented. The study revealed that the exchange rate and inflation rate were associated with the stock price movement. On the contrary, individual components of macroeconomic variables associated with different sectors of the economy differ. The following sections discuss in details.

Significant Features on Telecommunication Sector Stock

Figure 3 shows the ranking of the features as regards petroleum stock-prices. It was revealed that Net Foreign Assets, (i.e., DMBs Net Foreign Asset Monthly) under Monetary Survey recorded the highest importance factor. This result confirms (Asekome and Agbonkhese, 2015) outcome that a positively high association exists between stock-price and money supply. Nonetheless, this study has explicitly revealed which component of money supply relates well with the petroleum sector stock.

The U.S Dollar was found to be most significant as compared to currencies like the Danish Kroner, Canadian Dollar, Swiss France, and Swedish Kroner in terms of the exchange rate and stock-price, confirming study (Kwofie and Ansah, 2018) that exchange rate associated with stock on the GSE. The results revealed that the U.S dollar is the most significant exchange rate factor as also argue by (Kwofie and Ansah, 2018). The

significance ranking of exchange rate variable (U.S dollar) in the petroleum sector, reveals how the U. S dollar impacts petroleum commodity price locally and globally and this can be attributed to the benchmark pricing mechanism of the U.S dollar for most commodities.

The British Pound Sterling (PD) was found to be the thirteenth significant predictors of stocks in the petroleum sector. The study revealed that the Minimum Daily Wage (MDW) was a useful predictor of petroleum sector stocks as compared with the British Pound Sterling. The high positive association of MDW and petroleum stock movement reveals that as the MDW goes up the tendency of people demanding more petroleum products goes up, which in a way affects the prices of petroleum products.

Furthermore, the study revealed that the most significant inflation rate components (Food, Beverages (FB), Alcohol and Tobacco (AT)) were the ninth and eighteenth significant feature, respectively, in the petroleum sector. The results disagree with the outcomes of (Kwofie and Ansah, 2018) that inflation rate is of higher significance compared with the exchange rate.
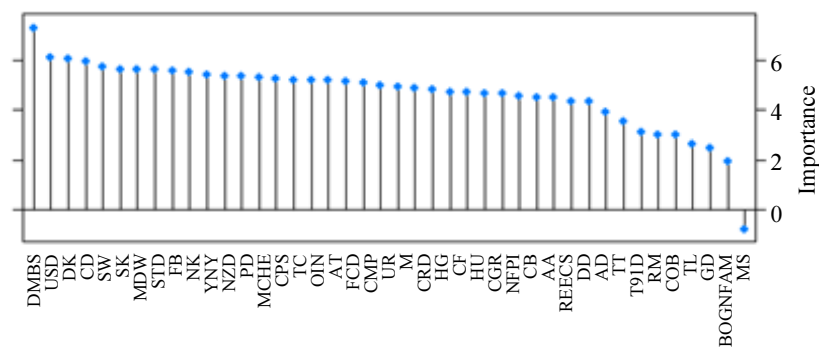
Significant Features on Banking Sector Stock

Figure 4 shows the importance features ranking over banking sector stock-prices. The Year-on-Year (YNY) is the higher essential macroeconomic factor in the bank sector sock, followed by the Danish Kroner (DK), Minimum Daily Wages (DMW), the Swedish Kroner (SK) and U.S dollar, as shown in Fig. 4. The study revealed that stock from the banking sectors is highly associated with change rate. This outcome disagrees with (Gyan, 2015) and agrees with (Gideon *et al.*, 2015) that the link between stock in the banking sectors and exchange rate follows the international trade. The results revealed that despite the higher rate of the U.S dollar to the Ghana Cedis (GH¢) as compared with the Danish Kroner and Swedish Kroner, the movement of the stock price in the banking sector is higher associated with Danish Kroner and Swedish Kroner than the U.S dollar.

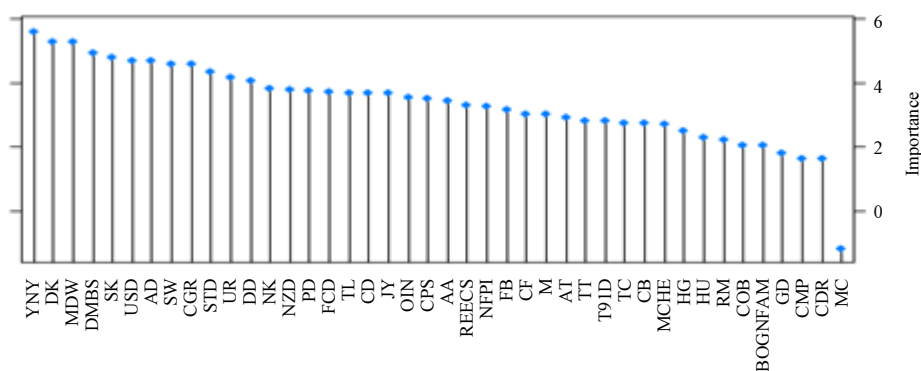### *Significant Features on Technology Sector Stock*

Figure 5 shows the importance of features ranking over the stock price movement in the telecommunication industries. The study reveals that Housing and Utilities (HU) and Savings and Time Deposits (STD) were the highest predictor of the stock price in the telecom industry, followed by transport and communications TC. Again, Foreign Currency Deposits (FCD) was the fifth feature of significance.

The study further revealed that an exchange rate component Norwegian Kroner (NK) was more of significance in the telecommunication sector stocks than the U.S Dollar (USD). The inflation rate on Clothing and Footwear (CF) was of more significance than the USD.
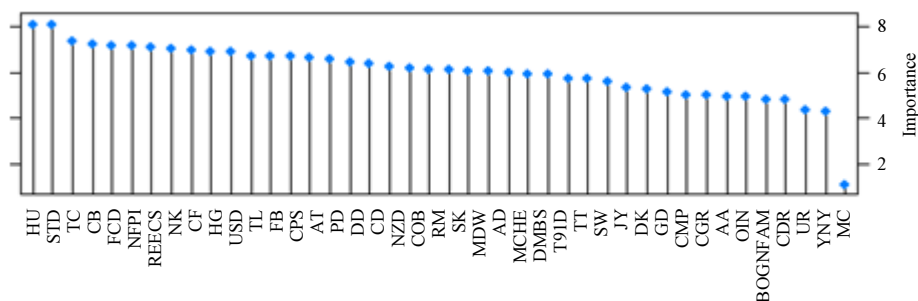
206

Fig. 3: Important features ranking over the petroleum industry

Fig. 4: Important features ranking over the banking industry

Fig. 5: Important features ranking over the telecommunication industry

Significant Features on Pharmaceutical Sector Stock

Figure 6 shows the ranking of the essential features of stock price movement from the technology industries. The results revealed that Total Government Liabilities (TL) were the most significant feature, followed by Savings and Time Deposits (STD), Daily Minimum Wages (DMW), Currency Outside the Bank (COB) and Treasury bill rate for 91 days. The results imply that as the inflow of foreign currency changes, the stock-prices in the technology sectors also changes. It was observed that the top five feature that affects stock price in the technology sector are managed

locally. The outcome reveals that the daily minimum wages are a good indicator of variation in stock price movement in the technology industries. Thus, the technological industries do well as more people patronise their services, and people can patronise these services if they earn a higher income.

Essential Factors for Ghana Stock Major Sectors

The essential features of the pharmaceutical industry are as shown in Fig. 7. Exchange rate components (U.S Dollar, Pound Sterling, Swedish Kroner, Swedish Kroner, Danish Kroner, Swiss Franc,

207

and New Zealand Dollar) were the top seven (7) most significant features respectively. The results come the reports that most pharmaceutical products are imported into the country and when the exchange rate changes stock within the sector changes. The results reveal a negative correlation between Claims on Government (CGR), Year on Year (YNY), Currency/Deposit Ratio (CDR) and stock price from the pharmaceutical sectors. Again, the results revealed that there is no correction between pharmaceutical sectors stock and government Treasury bill rate (T91D and Japanese Yen (JY), as shown in Fig. 7.

Essential Factors for Ghana Major Indices

Figure 8 shows the top 20 most significant macroeconomic features that are associated with the stock market movement in banking, petroleum, pharmaceutical, telecom, and technology sectors of Ghana's economy.

The results reveal that out of the top twenty (20) selected feature from 42 initial feature in each sector understudied in this study, only two (2) features British Pound Sterling (PD) and the Canadian Dollar (CD) representing 10% are associated with stock price movement in all studied sectors of Ghana's economy. While MVs

(MDW, SK, SW, STD, and NZD) affects four sectors stock price movement out of the five studies sectors stocks, represent 25% of the top 20 significant features from all sectors stock movement. Again, MVs features (DK, USD, NK, FCD, and MCHE) were found to be associated with at most three of studied sectors in this study.

The study reveals that the exchange rate is the most significant features the predicts the movement of the stock on the GSE, confirming (Mireku *et al.*, 2013; Asekome and Agbonkhese, 2015; Gideon, Maurice, and Magnus, 2015; Siok *et al.*, 2015) results. On the other hand, these factors were not the most significant in all economic sectors in the current study.

Even though the US Dollar and the Danish Kroner recorded high significance in petroleum, pharmaceuticals and banking industries; the British Pound Sterling is of more effects on the movement of the stock price in all sectors as compared with the USD and DK.

It was also revealed that the Minimum Daily Wages (MDW) were of significant in four sectors as compared with the U.S dollar this implies that for every thousand stock on the GSE, the MDW is a good predictor of eight hundred of these listed stock movement than the USD. Furthermore, it was revealed that the MDW was a common feature in the telecom and technology industry.
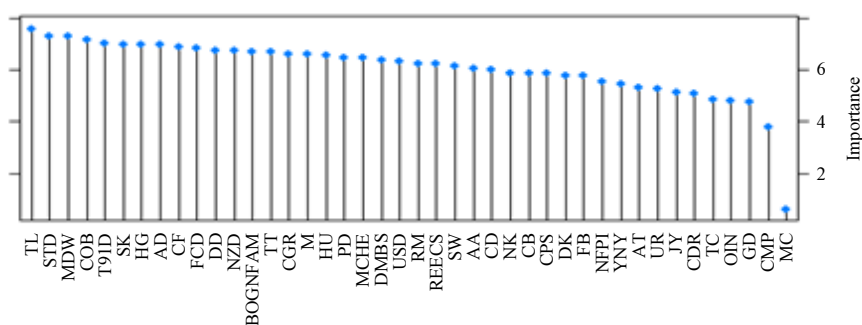

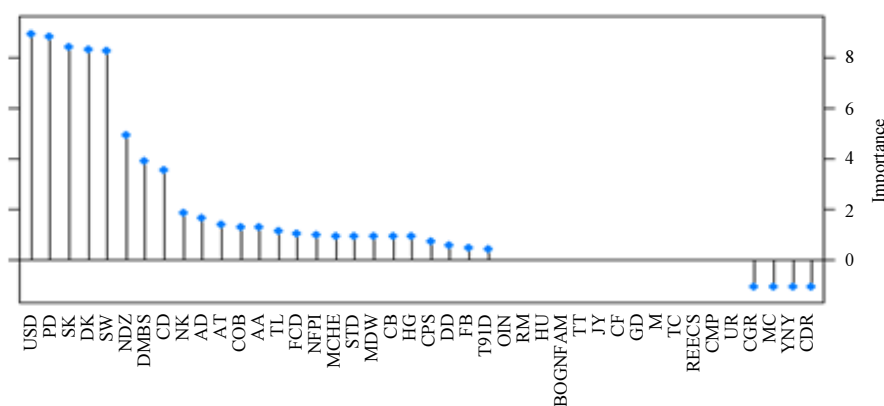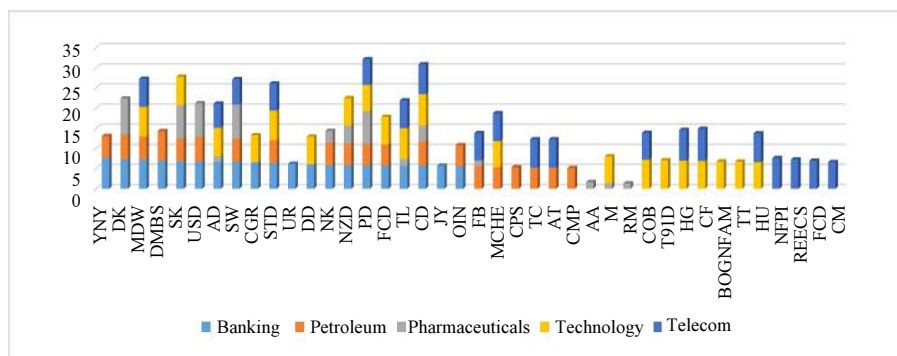**Fig. 6:** Important features ranking over the technology industry

Fig. 7: Important features ranking over the pharmaceutical industry

208

Fig. 8: Important macroeconomic feature for stock price movement on the GSE

Table 2: Results of LSTMRNN and ARIMA models

Error Metrics	Banking		Petroleum		Pharmaceuticals		Telecommunication		Technology	
	A[a]	L[b]	A[a]	L[b]	A[a]	L[b]	A[a]	L[b]	A[a]	L[b]
R	0.9588	0.9803	0.7956	0.9382	0.4591	0.9992	0.8456	0.9786	0.9213	0.9968
MAE	0.2557	0.1784	0.1373	0.0559	0.001	0.0001	0.0286	0.0115	0.1802	0.0527
RMSE	0.398	0.2547	0.2126	0.1191	0.002	0.0001	0.0362	0.0157	0.2273	0.0667
RAE (%)	24.970	18.0600	63.5800	25.8900	122.400	6.4700	47.7000	19.1400	38.5100	11.2700
RRSE (%)	28.110	19.6900	61.3100	34.3300	96.610	4.6200	51.9000	22.5000	46.7000	13.7100

A[a]: ARIMA model; L[b]: LSTMRNN proposed model

Predicting the GSE

The twenty (20) top features of each sector were selected from the feature importance ranking, given a new subset data (DS_{new}) for each selector economic sector for our prediction experiment. Every DS_{new} was then partitioned into three datasets, training Dataset (DS_{Train}) which constitute 75% of DS_{new}, 20% was for testing (DS_{Test}), and the remaining 5% was for validation. The DS_{Train} of each sector was fed separately into the LSTM RNN to train the network, and the pattern learned from the model was used on DS_{Test} to predict a 30-day stock price and compared with ARIMA time-series model on the same dataset based on the selected error metrics. The outcomes are as shown in Table 2.

The results, as shown in Table 2, point out the superiority of the proposed (LSTMRNN) model over the ARIMA model in predicting the stock-market price using macroeconomic variables. About R, MAE, RMSE, RAE, and RRSE, the LSTMRNN outperformed the ARIMA model comparing its error performance metric values to that of the ARIMA. They are respectively smaller in all sector stock compared with that of the ARIMA model. The outcomes imply that the proposed model achieved a smaller variance in individual errors.

Subsequently, the proposed (LSTMRNN) model predicted approximately 89.57% accuracy side-by-side 62.34% for the ARIMA. These results confirmed the critiques of (Vaisla and Bhatt, 2010; Bisoi and Dash, 2014; Caiani *et al.*, 2016; Weng, 2017) on time series model for stock-market predictions explained in section 2 of this study.

A decreased in simulation run time and increased in success rate was observed, when the RF algorithm was applied to selected features with the highest importance rate, which enhanced the efficiency of the (*LSTMRNN*) model. Errors in prediction decreased by 7.1%. Applying all 42, initially selected MVs caused a decrease in success. All the same, quantitatively, not much prediction accuracy was realised with a very small sub-dataset. Finally, the results revealed that macroeconomic factors are good predictors of stock-market volatility.

Conclusion and Recommendations for Future Works

Several previous studies aimed at finding the relationship between stock market volatility and MVs. Contrariwise, most of these studies have relied on a time-series model, which are reported to have some deficiencies in stock prediction. Works that applied computational intelligence and machine learning techniques for predicting the stock market using macroeconomic variables also did focus on predicting the stock-price movement, ignoring the relationship among the macroeconomic variables and sector stocks.

The current study was aimed at identifying the degree of association between MVs and stock price movement in the various economic areas such as banking, pharmaceuticals, telecommunication, technology, and petroleum on the GSE. The results revealed that exchange rate components Pound Sterling (PD) and the Canadian Dollar (CD) were the most MVs features that

affect all four studied sectors stock price movement in the GSE. The Net-Foreign-Assets under monitory survey was highly related to stock prices within the petroleum sector as against exchange rate (US dollar and Canadian dollar) which was found to be higher associated with stock-prices in the banking and pharmaceutical sectors. The daily minimum wage was found to be higher associated with stock from the technology. Unemployment Rate (UR) was found to be a significant factor associated with only banking sector stocks.

These results confirm that fluctuations in macroeconomic variables influence the stock volatility on the GSE. On the other hand, the weights of the influence of each macroeconomic variable on the stock price movement from different sectors of the Ghanaian economy differ from one macroeconomic variable to the other.

The proposed hybrid predictive model of enhanced RF and LSTMRNN was benchmarked with the ARIMA model, in an experiment with the real-world stock market and macroeconomic variables dataset from the GSE and BoG. The low values of error metrics (R, MAE, RMSE, RAE, and RRSE) recorded by the proposed model as against the ARIMA, indicated better accuracy in prediction by the proposed model.

The reduction in computational time and a slight increase in accuracy after the application of the enhanced leave-one-out feature selection algorithm clearly shows how vital feature selection is in machine learning.

Notwithstanding the degree of association revealed in the current study between MVs and stock-market volatility, the effect of socioeconomic and behavioural economics theory of finance as argued in (Chen *et al.*, 2014) cannot be overlooked. Hence, future studies should include other factors such as social media sentiment and web financial news to examine a great feature-pool to the treasure trove. To determine if the prediction accuracy upsurges by picking more features as input to a predictive model.

Acknowledgement

Funding Information

Author's Contributions

All authors equally contributed to this work.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript, and there are no ethical issues involved.

References

Adusei, M., 2014. The inflation-stock market returns nexus: Evidence from the ghana stock exchange. J. Econ. Int. Finance, 6: 38-46. DOI: 10.5958/2321-5763.2016.00010.X

Akshatha, K.R. and K.S. Shreedhara, 2018. Implementation of machine learning algorithms for crop recommendation using precision agriculture. Int. J. Res. Eng., Sci. Manage., 6: 58-60.

Appiah, A.Y., X. Zhang and B.B.K. Ayawli, 2019. Long short-term memory networks based automatic feature extraction for photovoltaic array fault diagnosis. IEEE Access. 7: 1-1. DOI: 10.1109/ACCESS.2019.2902949

Asekome, M.O. and A.O. Agbonkhese, 2015. Macroeconomic variables, stock market bubble, meltdown and recovery: Evidence from Nigeria. J. Fin. Bank Manage., 3: 25-34. DOI: 10.15640/jfbm.v3n2a3

Ayentimi, D.T., E.A. Mensah and F. Naa-Idar, 2013. Stock market efficiency of Ghana stock exchange: An objective analysis. International J. Manage. Econ. Soc. Sci., 2: 54-75.

Ballings, M., D.V. den Poel N. Hespeels and R. Gryp, 2015. Evaluating multiple classifiers for stock price direction prediction. Expert Syst. Applic., 42: 7046-7056. DOI: 10.1016/j.eswa.2015.05.013

Bisoi, R. and P.K. Dash, 2014. A hybrid evolutionary dynamic neural network for stock market trend analysis and prediction using unscented Kalman filter. Applied Soft Comput. J., 19: 41-56. DOI: 10.1016/j.asoc.2014.01.039

Boachie, M.K., S.O. Mensah, A.O. Frimpong and M. Ruzima, 2016. Interest rate, liquidity and stock market performance in Ghana. Int. J. Account. Econ. Stud., 4: 46-46. DOI: 10.14419/ijaes.v4i1.5990

Caiani, A., A. Godin E. Caverzasi M. Gallegatia and S. Kinsella *et al.*, 2016. Agent based-stock flow consistent macroeconomics: Towards a benchmark model. J. Econ. Dynam. Control, 69: 375-408. DOI: 10.1016/j.jedc.2016.06.001

Chen, C., W. Dongxing, H. Chunyan and Y. Xiaojie, 2014. Exploiting social media for stock market prediction with factorization machine. Proceedings of the IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, Aug. 11-14, IEEE Xplore Press, Warsaw, Poland, pp: 49-56. DOI: 10.1109/WI-IAT.2014.91

Choudhury, A., 2019 Classification of functioning, disability and health: ICF-CY self-care (SCADI dataset) using predictive analytics. Proceedings of the IISE Annual Conference, (IAC' 2019), pp: 1-8. DOI: 10.2139/ssrn.3307719

Choudhury, A. and B. Eksioglu, 2019. Using predictive analytics for cancer identification. Proceedings of the 2019 IISE Annual Conference, May 6, Orlando, pp: 6-6.

Crane-Droesch, A., 2018. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environ. Res. Lett., 3: 6-15. DOI: 10.1088/1748-9326/aae159

Devasia, T., T.P. Vinushree and V. Hegde, 2016. Prediction of students performance using educational data mining. Proceedings of the International Conference on Data Mining and Advanced Computing, Mar. 16-18, IEEE Xplore Press, Ernakulam, India, pp: 1-16. DOI: 10.1109/SAPIENCE.2016.7684167

Dosdoğru, A.T., A. Boru, M. Göçken, M. Özçalıcı and T. Göçken, 2018. Assessment of hybrid artificial neural networks and metaheuristics for stock market forecasting. Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi, 24: 63-78.

Frimpong, S., 2011. Speed of adjustment of stock prices to macroeconomic information. Int. Bus. Manage., 2: 151-156. www.cscanada.net%0A

Gideon, B., O. Maurice and F.J. Magnus, 2015. Stock returns and exchange rate nexus in Ghana: A Bayesian quantile regression approach. South African J. Economic, 84: 149-179. DOI: 10.1111/saje.12096

Golub, G.H., P.E.R. Christian and D.P.O. Leary, 1999. Tikhonov regularization and total least squares. SIAM J. Matrix Anal. Applic., 21: 185-194.

Gyan, M.K., 2015. Factors influencing the patronage of stocks, Knu. Kwame Nkrumah University of Science Technology.

Issahaku, H., Y. Ustarz and P.B. Domanban, 2013. Macroeconomic variables and stock market returns in Ghana: Any causal link? Asian Economic Financial Rev., 3: 1044-1062.

Jareño, F. and L. Negrut, 2016. US stock market and macroeconomic factors. J. Applied Bus. Res., 32: 325-340.

Khan, M.N., N. Tantisantiwong, S.G.M. Fifield and D.M. Power, 2015. The relationship between South Asian stock returns and macroeconomic variables. Applied Economic Routledge, 47: 1298-1313. DOI: 10.1080/00036846.2014.995360

Kwofie, C. and R.K. Ansah, 2018. A study of the effect of inflation and exchange rate on stock market returns in Ghana. Int. J. Math. Math. Sci., 2018: 6-6. DOI: 10.1155/2018/7016792

Lahmiri, S., 2011. A comparison of PNN and SVM for stock market trend prediction using economic and technical information. Int. J. Comput. Applic., 29: 975-8887.

Laichena, K.E. and T.N. Obwogi, 2015. Effects of macroeconomic variables on stock returns in the East African community stock exchange market. Int. J. Educ. Res., 3: 305-320.

Manu, F., 2017. Performance of the Ghana stock exchange and economic growth. Ashesi University College.

Mireku, K., K. Sarkodie and K. Poku, 2013. Effect of macroeconomic factors on stock prices in Ghana: A vector error correction model approach. Int. J. Academic Res. Account. Fin. Manage. Sci., 3: 32-43.

Mutuku, C. and K.L. Ng'eny, 2015. Macroeconomic variables and the kenyan equity market: A time series analysis. Bus. Econ. Res., 5: 1-10. DOI: 10.5296/ber.v5i1.6733

Nti, I.K. and J.A. Quarcoo, 2019. Self-motivation and academic performance in computer programming language using a hybridised machine learning technique. Int. J. Art. Intel. Expert Syst., 8: 12-30.

Nti, I.K., S. Asafo-Adjei and M. Agyemang, 2019. Predicting monthly electricity demand using soft-computing technique. Int. Res. J. Eng. Technol., 06: 1967-1973.

Peiró, A., 2015. Stock prices and macroeconomic factors: Some European evidence. Int. Rev. Econ. Fin., 41: 287-294. DOI: 10.1016/j.iref.2015.08.004

Ratti, R.A. and J.L. Vespignani, 2015. OPEC and non-OPEC oil production and the global economy. Energy Econ., 50: 364-378. DOI: 10.1016/j.eneco.2014.12.001

Ratti, R.A. and J.L. Vespignani, 2016. Oil prices and global factor macroeconomic variables. Energy Econ., King Abdullah Petroleum Stud. Res. Cen. 59: 198-212. DOI: 10.1016/j.eneco.2016.06.002

Rodriguez-Galiano, V., M. Sanchez-Castillo, M. Chica-Olmoc and M. Chica-Rivas, 2015. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol. Rev., 71: 804-818. DOI: 10.1016/j.oregeorev.2015.01.001

Siok, K.S., Q.T. Xue and N.W. Yen, 2015. A Comparative Study on the Effects of Oil Price Changes on Inflation. Procedia Economic Fin., 26: 630-636. DOI: 10.1016/S2212-5671(15)00800-X

Tsai, C.F. and Y.C. Hsiao, 2010. Combining multiple feature selection methods for stock prediction: Union, intersection and multi-intersection approaches. Decis. Support Syst., 50: 258-269. DOI: 10.1016/j.dss.2010.08.028

Vaisla, S.K. and K.A. Bhatt, 2010. An analysis of the performance of artificial neural network technique for stock market forecasting. Int. J. Comput. Sci. Eng., 02: 2104-2109.

211

Wakefield, K., 2013. A guide to machine learning algorithms and their applications, SAS UK.

Wei, Y. and X. Guo, 2017. Oil price shocks and China's stock market. Energy, 140: 185-197.

DOI: 10.1016/j.energy.2017.07.137

Weng, B., 2017. Application of machine learning techniques for stock market prediction. Auburn University.

Appendix 1

Table 1: The initial selected economic indicators

MONETARY SURVEY

S/N	MACROECONOMIC VARIABLE	ABBREVIATION	SOURCE
Net Foreign Assets			
1.	BOG Net Foreign Asset Monthly	BOGNFAM	BoG[a]
2.	DMBs Net Foreign Asset Monthly	DMBS	
3.	Total	TT	
Net Domestic Assets			
4.	Claims on Government	CGR	BoG[a]
5.	Govt. Deposits	GD	
6.	Claims on Private Sector	CPS	
7.	Other Items Net	OIN	
Money Supply M2+			
8.	Currency outside bank	COB	BoG[a]
9.	Demand deposits	DD	
10.	Savings & Time Deposits	STD	
11.	Foreign Currency Deposits	FCD	
12.	TBR-91 day	T91D	
13.	Total Liabilities	TL	
Memorandum Items			
14.	Reserve Money	RM	BoG[a]
15.	Currency /Deposit Ratio	CDR	
16.	Currency / M2 PLUS	CMP	
INFLATION			
17.	Combined	CB	BoG[a]
18.	Food and Beverages	FB	
19.	Non-Food Price Index	NFPI	
20.	Alcohol and Tobacco	AT	
21.	Clothing and Footwear	CF	
22.	Housing and Utilities	HU	
23.	Household Goods	HG	
24.	Medical Care and Health Expenses	MCHE	
25.	Transport and Communications	TC	
26.	Recr. Enter. Educ & Cultural Services	REECS	
27.	Miscellaneous	M	
28.	Monthly Changes (%)	MC	
29.	Year-on-Year	YNY	
30.	Annual Average (%)	AA	
31.	Minimum Daily Wages (¢)	MDW	
32.	Unemployment rate	UR	Kn[c]
EXCHANGE RATE			
33.	U.S Dollar	USD	BoG[a]
34.	Pound Sterling	PD	
35.	Swiss Franc	SW	
36.	Australian Dollar	AD	
37.	Canadian Dollar	CD	
38.	Danish Kroner	DK	
39.	Japanese Yen	JY	
40.	New Zealand Dollar	NZD	
41.	Norwegian Kroner	NK	
42.	Swedish Kroner	SK	

BoG[a] = Bank of Ghana (https://www.bog.gov.gh) GSE = https://gse.com.gh/daily-shares-and-etfs-trades
Kn[c] = https://knoema.com/atlas/Ghana/Poverty-rate