Predicting Nigerian Stock Returns using Technical Analysis and Machine Learning

David Oyewola, Emmanuel Gbenga Dada, Ezekiel Olaoluwa Omole and K. A. Al-Mustapha

Abstract—Models of stock price prediction have customarily utilized technical indicators alone to produce trading signals. In this paper, we construct trading techniques by applying machine-learning methods to technical analysis indicators and stock market returns data. The resulting prediction models can be utilized as an artificial trader used to trade on any given stock trade. Here the issue of stock trading decision prediction is enunciated as a classification problem with two class values representing the buy and sell signals. The stacking technique utilized in this paper is to assist trader with applying the proposed algorithms in their trading using random forest which was staked with different algorithms which incorporates Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM) and Neural Network (NN). The experimental results indicated that Top Layer of Random Forest (TRF) produced the best performance among all the algorithms compared. This is an indication that it is a promising strategy for forecasting Nigerian stock returns.

Index Terms—Machine learning, stock price, trading signals, random forest, neural network.

I. INTRODUCTION

Securities exchange has pulled in much attention from academia community and business, it has the ability to procure substantial benefits if done wisely [7]. Predicting the patterns in stock market prices is an extremely difficult task due to the numerous uncertainties involved and numerous factors that influence the market value in a specific day, for example, financial conditions, speculators' notions towards a specific organization, political events and so forth. Because of this, stock markets are defenseless to brisk changes, causing arbitrary fluctuations in the stock price. Securities exchange arrangement are commonly unique, non-parametric, tumultuous and boisterous in nature and consequently, security exchange price movement is viewed as an irregular procedure with changes which are increasingly articulated for short time windows. Be that as it may, a few stocks generally tend to create direct patterns over long haul time windows. Due to the turbulent and profoundly unpredictable nature of stock behavior, investments in share market accompanies with high risk. So as to limit the hazard involved, propelled information of stock price movement in the future is required. Traders are bound to purchase a stock whose value is likely to increment in the future [20]. A paper by [10] propose a novel method to limit the danger of investment in stock market by foreseeing the returns of a stock utilizing a class of ground-breaking machine learning algorithms known as ensemble learning. Some of the technical indicators, for example Relative Strength Index (RSI), stochastic oscillator and so on are utilized as inputs to train our model. Their algorithm appeared to outperform existing algorithms found in the literature. Out of Bag (OOB) error estimates have been observed to energize. Reference [19] in their paper proposes a rough set model, a neural system model, and a hybrid neural network and rough set model to discover optimal buy and sell of a share on Dhaka stock exchange. Investigational discoveries show that their proposed hybrid model has higher exactness than the single rough set model and the neural network model. They believe this paper discoveries will help stock financial specialists to choose about optimal buy and/or sell time on Dhaka stock trade. There are numerous studies that compare diverse Machine Learning techniques to find the most reasonable for specific kinds of data. The analyst reported in [16] utilizing three algorithms which incorporates Naïve Bayes, feed forward Artificial Neural Networks with Backpropagation and Decision Trees learning utilizing C4.5 more than two datasets European organizations and Japanese organizations described by 59 financial features each. In view of their discoveries it was found that Decision Trees algorithm gives the best classification accuracy. The Decision Trees learning indicated the best outcomes as well as the most homogeneous outcomes throughout the whole experimental study. Several researcher had utilized support vector regression, for instance, [4] used Support Vector Regression (SVR) to predict stock prices for huge and small capitalizations and in three unique markets, utilizing prices with both daily and up-to-the-minute frequencies. Forecast errors are estimated, and the model is contrasted to the random walk model proposed by the Efficient Market Hypothesis (EMH). The outcomes recommend that the SVR has prescient power, particularly when utilizing a procedure of updating the model occasionally. There are additionally characteristic results of increased predictions precision during lower instability periods. The capacity of artificial neural network (ANN) in estimating the daily NASDAQ stock exchange rate was examined by [1]. A few feed forward ANNs that were trained by the back propagation algorithm have been surveyed. The procedure utilized in their investigation is to considered the short-term historical stock prices and also the day of week as inputs. Day by day stock trade rates of NASDAQ from January 28, 2015 to18 June, 2015 are utilized to build a robust model. Initial 70 days (January 28 to March 7) are chosen as training dataset and the last 29 days are utilized for testing the model forecast capacity. Networks for NASDAQ index forecast for two kind of

DOI: http://dx.doi.org/10.24018/ejece.2019.3.2.65
input dataset (four and nine earlier working days) were created and validated. The model yields show that there is no unmistakable distinction between the prediction capacity.

The paper by [15] proposes a machine learning model to predict securities exchange price. Their proposed algorithm includes Particle swarm optimization (PSO) and least square support vector machine (LS-SVM). The PSO algorithm is utilized to improve LS-SVM to predict the daily stock prices. The proposed model depends on the investigation of stocks historical data and technical indicators. PSO algorithm chooses best free parameters combination for LS-SVM to prevent over-fitting and local minima problems and enhance prediction accuracy. Their proposed model was applied and assessed utilizing thirteen benchmark financial datasets and contrasted with artificial neural network with Levenberg-Marquardt (LM) algorithm. The outcomes results demonstrated that the proposed model has better prediction accuracy and the capability of PSO algorithm in enhancing LS-SVM.

II. DATA AND METHODS

Stock data are collected from the Nigerian Stock Exchange (NSE) utilizing Guaranty Trust Bank traversing from January 2013 to August 2018 as shown in figure 1. In this study stock information for that period is taken to analyze the performance of the system. A return is then calculated and stored. Our data covers 5 financial years including 1390 trading days. For experimentation, the stock market datasets are divided in two sets, for example, training and testing dataset. January 2013 to December 2017 of the daily data were utilized for the training Dataset and the remaining of the daily data from January 2018 to August 2018 were utilized for the testing dataset as shown in figure 2.

Technical Indicators of stock market

Researchers have utilized diverse types of technical indicators to monitor the future development of stock prices and in setting up trading rules for buy-sell-hold decisions. In this study, ten well known technical indicators, for instance RSI, CMO, DPO, KST, MACD, ROC, EMA, SMA, STOC, TDI are chosen as input to the proposed model. The technical indicators are determined from historical prices as follows:

A. Relative Strength Index (RSI)

RSI is a mainstream momentum indicator which decides if the stock is overbought or on the other oversold. A stock is said to be overbought when the demand outlandishly pushes the price upwards. This condition is commonly translated as a sign that the stock is overvalued and the price is probably going to go down. A stock is said to be oversold when the price goes down strongly to a level below its actual value. This is a result caused because of panic selling. It is a price following an oscillator that ranges from 0 to 100, where RSI is above 70, it may demonstrate that the stock is overvalued and the price is possibly going to go down. A stock is said to be oversold if the stock is oversold [18]. The formula for calculating RSI is:

\[
RSI = 100 \left(\frac{\sum_{i=0}^{n-1} U_{p_{t-i}} - \sum_{i=0}^{n-1} D_{w_{t-i}}}{\sum_{i=0}^{n-1} U_{p_{t-i}} + \sum_{i=0}^{n-1} D_{w_{t-i}}}\right)
\]

Where \(U_{p_t}\) is the upward-price-change and \(D_{w_t}\) is the downward-price-change at time \(t\).

B. Chande Momentum Oscillator (CMO)

The Chande’s Momentum Oscillator was created by Tushar S. Chande [9]. The Chande Momentum Oscillator is an altered RSI. Where the RSI divides the upward movement by the net movement (up / (up + down)), the CMO separates the aggregate movement by the net movement ((up - down) / (up + down)). Values more than 50 shows a buy signal and values below 50 shows a sell signal. The equation for computing CMO is:

\[
CMO = 100 \times \frac{\sum_{i=0}^{n-1} U_{p_{t-i}} - \sum_{i=0}^{n-1} D_{w_{t-i}}}{\sum_{i=0}^{n-1} U_{p_{t-i}} + \sum_{i=0}^{n-1} D_{w_{t-i}}} \]

Where \(U_{p_t}\) is the upward-price-change and \(D_{w_t}\) is the downward-price-change at time \(t\).

C. Detrended Price Oscillator (DPO)

The Detrended Price Oscillator demonstrates the distinction between a past price and a simple moving average. As opposed to other price oscillators, DPO is not a momentum indicator. Rather, it is just intended to identify cycles with its peaks and troughs. Cycles can be evaluated by counting the periods between peaks or troughs. Detrended price oscillator is calculated by equation:

\[
DPO = C_t - SMA_{21} \left(\frac{C_t}{\text{close}_{21}} + 1\right)
\]

Where \(C_t\) is the closing price at time \(t\), \(SMA\) the simple moving average, \(n = 21\).
D. Know Sure Thing (KST)

Know Sure Thing (KST) Oscillator is a complex, smoothed price velocity indicator created by [11]. Rate of Change (ROC) indicator is the establishment of KST indicator. KST indicator is helpful to recognize real stock market cycle crossroads since it formula is weighed to have larger impact by the longer and progressively overwhelming time length to better reflecting the essential swings of securities exchange cycle. The idea behind the oscillator is that price patterns are determined by the association of wide range of time cycles and that important patterns reversals take place when various price patterns are at the same time altering course. The equation can be calculated as:

\[
RO{C}_1 = 1 - \frac{C_t}{C_t(x1)} \times 100
\]  

\[
RO{C}_2 = 1 - \frac{C_t}{C_t(x2)} \times 100
\]  

\[
RO{C}_3 = 1 - \frac{C_t}{C_t(x3)} \times 100
\]  

\[
RO{C}_4 = 1 - \frac{C_t}{C_t(x4)} \times 100
\]  

\[
KST = MA_t(RO{C}_1, A_1) \times W_1 + MA_t(RO{C}_2, A_2) \times W_2 + MA_t(RO{C}_3, A_3) \times W_3 + MA_t(RO{C}_4, A_4) \times W_4
\]  

Where \( C_t \) is the current closing price at time \( t \), \( C_t \) closing price at \( x1 \ldots x4 \), \( MA_t \) moving average at time \( t \), \( A_1 = 10 \), \( A_2 = 10, A_3 = 10, A_4 = 15 \), \( x1 = 10, x2 = 15, x3 = 20, x4 = 30 \), \( W_1 = 1, W_2 = 2, W_3 = 3 \) and \( W_4 = 4 \).

E. Moving Average Convergence Divergence (MACD)

MACD (Moving Average Convergence/Divergence) is a technical analysis indicator developed by Gerald Appel in the late 1970 [2]. It is used to spot changes in the strength, direction, momentum, and duration of a trend in a stock’s price. The MACD is a calculation of the contrast between two exponential moving averages (EMAs) of closing prices. This distinction is graphed after some time, alongside a two exponential moving averages (EMAs) of closing prices. Then by contrasting that difference to an average, an analyst can outline unobtrusive shifts in the stock's trend. The equation for calculating MACD is:

\[ MACD = EMA_{12}(C_t) - EMA_{26}(C_t) \]  

\[ \text{Signal Line} = EMA_{9}(MACD) \]  

Where \( C_t \) is the closing price at time \( t \), \( EMA_{n} = n = 9, 12 \) and 26 day Exponential Moving Average. When the \( MACD \) goes below the \( \text{Signal Line} \), it indicates a sell signal. When it goes above the \( \text{Signal Line} \), it indicates a buy signal.

F. Rate of Change (ROC)

Rate of Change shows the distinction between the present price and the price \( n \) days ago. It is computed as follows:

\[ ROC_t = \frac{C_t - C_{t-n}}{C_{t-n}} \times 100 \]  

Where \( ROC_t \) = Rate of Change at time \( t \), Closing price at time \( t \) It measures the most recent change in price with respect to the price in \( n \) days ago.

G. Exponential Moving Average (EMA)

Exponential Moving Average is utilized similarly as the Simple Moving Average, the only distinction being that EMA responds quicker to recent price values than the simple moving average. EMA is also used for computing numerous other indicators, such as Moving Average Convergence Divergence (MACD) and so forth. Exponential moving average of \( n \) days is determined as:

\[ EMA_t = C_t \times \alpha + EMA_{t-1} \times (1-\alpha) \]  

Where \( EMA_t \) is the exponential moving average at time \( t \), \( \alpha \) is the exponential smoothing Factor given as \( \alpha = 2 \div (n + 1) \).

H. Simple Moving Average (SMA)

An \( n \)-day Simple Moving Average is determined by calculating the average of \( n \) most recent prices. Simple Moving Average is determined as:

\[ SMA_t = \frac{\sum_{i=1}^{t} C_t}{t} \]  

Where \( SMA_t \) simple moving average at time \( t \), \( C_t \) closing price at time \( t \) and \( t \) is set at 25.

I. Stochastic Oscillator (Stoch)

The stochastic oscillator is a momentum indicator that utilizes support and resistance levels. Dr. George Lane advanced this indicator in the 1950s. The term stochastic alludes to the location of a current price in connection to its price range over a period of time. This strategy attempts to predict price turning points by comparing the closing price of a security to its price range [14]. The indicator is defined as follows:

\[ Stoch = \frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100 \]  

Where \( C_t \) closing price at time \( t \), \( LL_t \) is the lowest low and \( HH_t \) highest high in the last \( t \) days, respectively.

J. Trend Detection Index (TDI)

The Trend Detection Index (TDI) is utilized to identify when a trend has started and when it has come to an end. The TDI can be used as an independent indicator or combined with others; it will perform well in identifying the start of patterns [9]. An uptrend is flagged by a positive direction indicator value, while a downtrend is motioned by a negative value. If both TDI and direction indicator are
positive it is a buy signal and if the TDI is positive and the direction indicator is negative, it is a sell signal. Trend Detection Index of 20 days determined as:

$$TDI_{20} = |M_{20}| - \left( \sum |M_{40}| - \sum |M_{20}| \right)$$  \hspace{1cm} (15)

Where $TDI_{20}$ is the 20 days Trend Detection Index, $|M_{20}|$ absolute value of the sum of 20 – day momenta of the last 20 days, $\sum |M_{40}|$ is the sum of 20 – day absolute momenta of the last 40 days and $\sum |M_{20}|$ is the sum of 20 – day absolute momenta of the last 20 days.

### Classification methods in Predicting Nigerian Stock Market

In this section, the method utilized in predicting stock market are discussed.

**A. Logistic Regression (LR)**

Logistic regression (LR) was proposed by [3], the logic model considers a group of explanatory variables $X = \{X_1, ..., X_p\}$ and a response variable with two categories $Y = \{y_1, y_2\}$, the method of logistic regression comprises in the estimation of a linear combination among $X$ and the logit change of $Y$. In this manner, if we consider $y_2$ as the class of interest for investigation, the model can be represented as:

$$\log \frac{p(Y=y_2)}{1-p(Y=y_1)} = X\beta$$  \hspace{1cm} (17)

Where $\beta$ is the vector containing the model's coefficients.

**B. Neural Network (NN)**

Neural networks (NN). A neural network [17] is a framework dependent on input variables, otherwise known as explanatory variables, combined by linear and non-linear interactions through at least one hidden layers, bringing about the output variables, additionally called response variables. Neural networks were created in an attempt to recreate the human brain, since it depends on sending electronic signals between a large number of neurons. The NN structure have components which get an amount of stimuli (the input variables), create neurotransmitters in a few neurons (activation of neurons in hidden layers), and results in responses (output variables). Artificial Neural networks model can be represented by Fig. 3.

![Artificial Neural network](image)

The above figure demonstrates an artificial neuron which is a simple processing unit enlivened from the biological neuron [13]. The neuron has $n$ inputs $x_i$ and each input is associated with the neuron by weighted connection $w_i$. Here the neuron aggregates up the inputs multiplied by the weights utilizing the below equation

$$A = \sum x_i w_i + b$$  \hspace{1cm} (18)

Where $A$ is the net aggregate and $b$ is the threshold value. For getting the output this net sum is connected into a function called activation function $F(A)$.

$$output = F(A)$$  \hspace{1cm} (19)

Here the input values and weights are real numbers. In a few circumstances the threshold value $b$ is considered as an imaginary input $x_0 = +1$ and an association weight $w_0$ for the simplicity of calculation.

**C. Random Forest (RF)**

Random forest is an ensemble learning strategy for classification, that trains a model by making a multitude of decision trees and yielding the mode of the classes returned by the individual trees. The algorithm applies the method of bootstrapping aggregation, also known as bagging, to tree learners. Given a training set $X = x_1, ..., x_n$ and output $Y = y_1, ..., y_n$, bagging repeatedly chooses a sample at random with substitution from the training set and fits trees to the samples:

- For $b = 1, ..., B$:
  1. Sample, with substitution $n$ training models from $X_b, Y_b$
  2. A decision tree $f_b$ is train on the samples.

After training, predictions for concealed examples are made by averaging the predictions or taking the mode on account of decision trees.

**D. Support Vector Machine (SVM)**

This method is a statistical classification strategy and presented by [21]. Given a training set $\{x_i, y_i\}$, with $i = 1, ..., n$, where $x_i$ is the explanatory variable vector, and $y_i$ represents the binary category of interest, and $n$ means the quantity of dimensions of input vectors. SVM attempts to locate an optimal hyper-plane, making it a non-probabilistic binary linear classifier. The optimal hyper-plane could be written as follows:

$$\sum_{i=1}^{n} w_i x_i + b$$  \hspace{1cm} (20)

Where $w = w_1, ..., w_n$, is the normal of the hyper-plane, and $b$ is a scalar threshold. Considering the hyper-plane divisible concerning $y_i \in \{-1, 1\}$ and with geometric distance $\frac{2}{\|w\|}$, the strategy maximizes this distance, subject to the constraint $y_i(\sum_{i=1}^{n} w_i x_i + b) \geq 1$. Normally, this maximization may be done through the Lagrange multipliers and utilizing linear, polynomial, Gaussian or sigmoidal partitions.
E. Stacking of LR, RF, SVM and NN

Stacking is a method to ensemble various learning algorithms, where a meta-level algorithm is trained to make a final prediction utilizing the outputs of based-level algorithms as features. It is a non-generative ensemble algorithm. This is a combination algorithms having two layered architecture. All algorithms at the base layer are trained on the given training set, and after that a combiner algorithm at the top layer is trained utilizing predictions of the first level as inputs. This technique is utilized for increasing the accuracy from frail learners. The combiner algorithm is named as Meta learner. Rather than output from the base classifier, the probability predictions can likewise be passed as input to the meta layer.

F. Stock Returns

The real time idea of securities exchange requires real time investigation techniques. Prediction performed on prolonged stretch of time periods may not be the optimal tool for traders. For standardization purpose and to represent all stock data in an equivalent measurement, we calculated returns as follow:

\[ R = \frac{p_t}{p_{t-1}} \]  

(21)

where \( p_t \) denotes the stock closing price at time \( t \) and \( p_{t-1} \) the stock closing price at time \( t - 1 \). In this paper, the stock returns in equation (21) are utilized as the target value of all the algorithms while the input value is all the ten technical indicators used in this paper.

Algorithm 1

Input: \( M = \{m_1, m_2, m_3\} \) Machine learning classifiers as the base layer
\( X = \{x_{1}^d, ..., x_{n}^d\} \) Training set of size \( d \)
\( Z = \{z_{1}^d, ..., z_{p}^d\} \) Test set of size \( d \)
\( T \) Top learner decide from the top layer
Output: Prediction of every one of the \( M \) test vector every one of size \( d \) \( Y = \{y_{1}^d, ..., y_{M}^d\} \)
\( X' \) top training set for learning in top layer
\( Y' \) top test for learning in top layer
For each row \( r \) in \( Z \):
\( I \) Instance vector
For each classifier \( m \) in \( M \):
\( p = \text{Train the classifier with } y \) and predict on row vector \( r \)
Add prediction \( p \) to \( I \)
Add \( I \) to \( X' \)
For every row \( r \) in \( Y \):
\( I \) Instance vector
For every classifier \( m \) in \( M \):
\( p = \text{Train the classifier with } y \) and predict on row vector \( r \)
Add prediction to \( I \)
Add \( I \) to \( Y' \)

**prediction:** Train top learner \( M \) with top training set \( X' \) and predict on \( Y' \)

**Performance evaluation**

We evaluate prediction performance using five measures: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Mean Square Error (MSE) and Mean Absolute Scaled Error (MASE).

A. **Mean Absolute Error (MAE)**

Consider a set of target or actual returns \( r^n_t \) and their predicted values \( \hat{r}^n_t \). MAE is defined as follows:

\[
\frac{1}{n} \sum_{n=1}^{n} |r^n_t - \hat{r}^n_t|
\]

B. **Mean Absolute Percentage Error (MAPE)**

MAPE is defined as:

\[
\frac{1}{n} \sum_{n=1}^{n} \frac{|r^n_t - \hat{r}^n_t|}{r^n_t} \times 100
\]

C. **Root Mean Square Error (RMSE)**

RMSE is defined as:

\[
\sqrt{\frac{1}{n} \sum_{n=1}^{n} (r^n_t - \hat{r}^n_t)^2}
\]

D. **Mean Square Error (MSE)**

MSE is defined as:

\[
\frac{1}{n} \sum_{n=1}^{n} (r^n_t - \hat{r}^n_t)^2
\]

E. **Mean Absolute Scaled Error (MASE)**

MASE is defined as follows:

\[
\frac{1}{n} \sum_{n=1}^{n} \left( \frac{|r^n_t - \hat{r}^n_t|}{\frac{1}{n-m} \sum_{n=m+1}^{n} |r^n_t - r^{n-m}_t|} \right)
\]

Where \( m \) is the seasonal period of return \( r^n_t \)

### III. RESULTS

In this segment the execution of the stacking model is validated for stock exchanging problem by applying it on the return of the closing stock data sets. The model performance is also contrasted with some other classifiers like Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM) and Neural Network model. Five years of historical stock price values of Guaranty stock prices is utilized in this study. The data sets are divided into training and testing sets as described in Fig. 2. The training set comprises of 1235 patterns and remaining 155 patterns are utilized for testing. Ten technical indicators are utilized as the input after being used the return price as their input. Table 1 summarizes the statistical analysis of the selected technical indicators for both the stock returns. The primary aim of the study is to analyze the best top layer that can predict close to the actual value of the trading points from trend analysis, so additionally stock returns has utilized for discovering starting up- down movements of the stock prices. Rather than utilizing discrete value as output during training of the algorithms, continuous trading signals are produced from the trend and utilized during training process.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Technical Indicators</th>
<th>Minimum</th>
<th>Median</th>
<th>Mean</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guaranty Trust Bank</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rsi</td>
<td>-0.4354</td>
<td>50.9675</td>
<td>51.0844</td>
<td>98.66</td>
<td></td>
</tr>
<tr>
<td>cmo</td>
<td>-0.8190</td>
<td>0.016</td>
<td>0.0144</td>
<td>3.4890</td>
<td></td>
</tr>
<tr>
<td>kst</td>
<td>-239.611</td>
<td>0.3534</td>
<td>2.0288</td>
<td>94.1234</td>
<td></td>
</tr>
<tr>
<td>macd</td>
<td>-8.8416</td>
<td>0.1555</td>
<td>0.1856</td>
<td>7.2426</td>
<td></td>
</tr>
<tr>
<td>roc</td>
<td>-0.0125</td>
<td>0.0</td>
<td>0.0003</td>
<td>0.1397</td>
<td></td>
</tr>
<tr>
<td>ema</td>
<td>14.87</td>
<td>25.61</td>
<td>28.10</td>
<td>50.25</td>
<td></td>
</tr>
<tr>
<td>sma</td>
<td>14.88</td>
<td>25.55</td>
<td>28.10</td>
<td>50.25</td>
<td></td>
</tr>
<tr>
<td>stochRSI</td>
<td>0.0</td>
<td>0.4156</td>
<td>0.5031</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>tdi</td>
<td>-184.10</td>
<td>-5.1350</td>
<td>-0.2335</td>
<td>231.18</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 8. Output trading signal obtained from Logistic Regression (LR) model and Stock returns as the actual value.

Fig. 9. Output trading signal obtained from Random Forest (RF) model and Stock returns as the actual value.
Fig. 10. Output trading signal obtained from Support Vector Machine (SVM) model and Stock returns as the actual value.

Fig. 11. Output trading signal obtained from Neural Network (NN) model and Stock returns as the actual value.

Fig. 12. Output trading signal obtained from Top Layer of Logistic Regression (LR) model and Stock returns as the actual value.

Fig. 13. Output trading signal obtained from Top Layer of Random Forest (RF) model and Stock returns as the actual value.

Fig. 14. Output trading signal obtained from Top Layer of Support Vector Machine (SVM) model and Stock returns as the actual value.

Fig. 15. Output trading signal obtained from Top Layer of Neural Network (NN) model and Stock returns as the actual value.

In this research, we consider four algorithms to be specific Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM) and Neural Network (NN). The Stacking technique was proposed to determine among the four algorithms when every one of them are utilized as the top layer and the remaining as the second layer can effectively predict the stock returns as portrayed in Fig. 4 - 7. Fig. 8 – 11 is the machine learning algorithms utilized in this paper. The overall performance of every one of the stock return compared to other algorithms like Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM) and Neural Network for the dataset are shown in Tables 1 respectively.

Utilising the proposed method in Fig. 4 - 7, Fig.12-15 are the results of the each of the algorithms assigned as the top layer which incorporate Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM) and Neural Network. Figs. 8 – 15 represents the actual stock returns and predicted value of every one of the algorithms utilizing the ten technical indicators as the input. Through a series of experimental tests, the top layer of random forest outperform the remaining seven algorithms. There is no improvement in the top layer of Support Vector Machine, Logistic Regression and Neural Network. In this study, we also considered the Uptrend and Down trend Decision taken as Buy Signal and the later as sell signals utilizing the predicted value from all the eight algorithms. The results demonstrated that Top layer of the Random Forest can predict the Buy and Sell Signals when compared with the actual value of the stock returns. Tables 3 report MAE, RMSE, MSE and MASE in the test set. TRF outperform the remaining algorithms by achieving smaller than one, implying that the methods are more accurate than any other prediction. The results shows that fitting random forest as the top layer and the other algorithms as the base layer can greatly improve the prediction results of random forest.

<table>
<thead>
<tr>
<th>TABLE 2: COMPARISON OF STOCK TRADING MODELS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Metrics</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Actual</td>
</tr>
<tr>
<td>LR</td>
</tr>
<tr>
<td>RF</td>
</tr>
<tr>
<td>SVM</td>
</tr>
<tr>
<td>NN</td>
</tr>
<tr>
<td>TLR</td>
</tr>
<tr>
<td>TRF</td>
</tr>
<tr>
<td>TSVM</td>
</tr>
<tr>
<td>TNN</td>
</tr>
</tbody>
</table>

TABLE 3: PERFORMANCE EVALUATION
IV. CONCLUSION

This study has proposed a novel decision stacking framework for creating effective stock trading techniques, which may give appealing benefits for investors. The model has incorporated technical analysis with machine learning strategies for effective generation of stock trading decisions. We assessed the performance by contrasting the actual and the predicted value of the algorithms utilized in this paper. Our experimental studies demonstrate that at the point when the random forest is relegated as the top layer it can performs best in most cases compared to other algorithms designs for stock return prediction. In future works, we will focus on learning from sentiment news from facebook and whatsapp since they have high correlation with stock prices and market status.

REFERENCES


