Hybridization of Bat Algorithm with XGBOOST Model for Precise Prediction of Stock Market Directions

M. Jeyakarthic, S. Punitha

Abstract: ***In recent days, prediction of stock market returns is generally treated as a forecasting problem. The implicit volatile nature of stock market across the world makes the prediction process highly challenging. As a result, prediction and diffusion modeling undermine a wide range of issues present in the stock market prediction. The minimization in prediction error will greatly minimize the investment risks. This paper presents a new method to determine the direction of stock market variations indicating gain and loss. A new machine learning ML based model is applied to predict the direction of stock market prices. The presented model undergoes preprocessing, feature extraction and classification. Initially, preprocessing takes place using exponential smoothing. Then, required features are extracted from the preprocessed dataset. Afterwards, an effective Bat algorithm (BA) with the XGBoost model called BA-XGB is applied for forecasting the stock prices in market. The proposed model predicts whether the stock values gets increased or decreased based on the price existing n days in advance. The presented model is experimented using Apple (APPL) and Facebook (FB) stocks. The obtained simulation outcome stated that the BA-XGB model has offered superior outcome by achieving a maximum accuracy of 96.42.***

Keywords: Stock market, Prediction, Classification, Machine learning

I. INTRODUCTION

In past decades, it is strongly thought that variations in stock prices could not be predicted in earlier. Random Walk [1], and Efficient Market Hypothesis
(EMH) stated that a market will be effective based on the present information I(i), when it is not possible
predict the market flow because of arbitrary nature of stock prices, then if not a diverse (often excessive) types of risks are treated, financial profit could not get increased. The process of stock market price prediction is very difficult and that there is high return from enhancing value at risk forecast. On the other hand, the Wisdom of Crowd hypothesis stated that various individuals could offer a precise assessment when the details are drawn out in a proper way. It is, yet, not known to be helpful to predict the return of stock market; nevertheless, some individual, as well as institutional investors has the ability to strike the market for making gains. The incompetence of prediction gets emphasized because of different uncertainties involved and due to the existence of many attributes, which significantly affects the market value on a single day.

Revised Manuscript Received on February 26, 2020.
M.Jeyakarthic, Assistant Director, Tamil Virtual Academy, Chennnai E-mail: jeya_karthic@yahoo.com
S. Punitha, Assistant Professor, Department of Computer Science, DGGArts college for Woman , E-mail : spddeprof@gmail.com

Later, several explanatory parameters have been included in the literature. As a result, stock markets are vulnerable to rapid variations that frequently make into arbitrary fluctuations in stock prices.Several financial and statistical experts have started to think that the stock market prices can be foreseen in a partial way. Besides, a new of financial experts highlights the psychological and behavioral components of stock-price determination and volatility [2]. It is believed that the upcoming stock prices are slightly predictable based on the previous stock prices and some basic evaluation measures. Afterwards, some experts expressed that the forecasting patterns will allow the investor to get high profit risk-adjusted return rates. In recent days, stock market prediction models have been developed by the use of machine learning and text mining models. Some of the relevant works are discussed here. Predictive models are used for determining the upcoming patterns in stock market process which provides a way to enhance the prediction abilities and predetermine the effective market hypothesis and diffusion approaches [3]. ML, on the other hand, includes a collection of recent models. As a result, different models have been employed for predicting stock prices namely support vector machine (SVM), deep neural network [4], random forest (RF), naïve bayes (NB), etc to achieve the concept of predictability with high accurateness. At the same time, data mining models mainly utilizes the daily stock data. In general, autoregressive integrated moving average (ARIMA) model [5] has been employed to identify and predict time series modifications. The sensitivity of stock prices to exterior circumstances has been assumed. It includes explicit parameters into account covering daily updates of commodity cost like gold, crude oil, natural gas, corn, and cotton in two major foreign currencies (EUR, JPY). In [6], closing prices are applied for the analysis of the company 3M which holds the data in the interval of September 2008-August 2013. Many models were selected for training the predictive model and are utilized to predict the direction of the stocks on the succeeding day's equivalent to a provided data instance. It has the ability to predict the price for the upcoming n days. It defined that the US stock market is semi effective, indicating both technical and basic investigations could not be applied for achieved maximum profit. But, the long term predictive approach yield better accuracy which peaked when the time window was 44 days. In this case, SVM achieved a maximum accuracy of 79.3%. In [7], a set of three stocks are employed using different technical signs namely RSI, On Balance Volume, Williams %R, etc.

were used as features. Among the total of 84 features, an extremely randomized tree algorithm, as described by [8], is chosen for feature selection which is then given to the SVM with radial basis function (RBF) to train the model. It is familiar that stock market price rates are mainly non-static, non-parametric, chaotic and noisy in nature which results in risk investment. Besides, the movements of stock market prices are assumed to be an arbitrary process with modifications which are highly significant in short term. It is highly necessary to talk about that advanced facts of near future stock price movements be supposed to assist in reducing this risk. Dealers are likely to get a stock in the present duration whose value will be increased in the upcoming days. It is also noticed that the accurate forecast of the trends in stock market prices will definitely increase the capital gain and reduce the loss. Therefore, it is highly needed to develop a model which predicts the direction of the movement of stock prices in the market. This study focuses on the design of new ML based model to predict the direction of stock market prices. The proposed model predicts whether the stock values gets increased or decreased based on the price existing n days in advance. The presented model undergoes preprocessing, feature extraction and classification. Initially, preprocessing takes place using exponential smoothing. Then, required features are extracted from the preprocessed dataset. Afterwards, an effective Bat algorithm with the XGBoost (BA-XGB) model is applied for forecasting the stock prices in market. The organization of the study is given here. Section 2 reviews some of the existing works available in the literature. Section 3 discusses the presented BA-XGB model and the validation process takes place in Section 4. At last, paper is concluded in Section 5.

II. RELATED WORKS

In this section, a brief review of the existing work available in the literature has been provided. [9] presented a hybridization of cuckoo search (CS) with SVM model where CS algorithm is applied to tune the parameters of SVM. [10] presented a trading agent using a neural network ensemble which forecast when a particle stock will get increased or decreased. It is validated using North American and the Brazilian stock market data and it is resulted to the hit rate of only 56%. [11] employed a one-against all (OAA-NN) and one-against-one neural network (OAO-NN) for classifying buying, holding or selling data and undergo a comparison with the results offered by the classical NN. In [12], an optimized ANN using genetic algorithm (GA) is applied for predicting the stock market prices in an identical way. [13] applied different types of classification models for SPP in Saudi stock exchange in the due course of 2006-2013. Dow Jones dataset is applied along with 5-fold cross validation model. [14] proposed a ML technique to examine the integrity of upcoming cost for a prolonged time. This method helps to identify accurate increase in organizational values around 10% over a year. The key objective of employing ML technique is for training the predefined information which is more applicable in detecting the stock price as well as to explore the movement ratio for a specific duration. In this technique, various classification models have been compared for SPP. When related with alternate classifiers, RF attains slightly better outcome with respect to Precision, F-score as well as recall.

[15] presented a structural SVM (SSVM) for detecting the SSP. It classifies the tedious inputs, such as nodes of a graph. Therefore, the deployed technique enables SSVM to understand a forecasting method in case of difficult graph input including numerous edges of all nodes. Hence, the simulation outcome for the SPP where positive as well as negative class labels denotes the maximization and reduction in stock prices. Also, it is applied with three-fold cross validation for searching the original measure and SSVM feature C has been fixed to 1000. As a result, it attains the accuracy of 78%, which is proved that this method is trained without over-fitting. Therefore, the performance measure depicts that ML techniques are more suitable in predicting the stock prices.

[16] established an Artificial Neural Network (ANN) for SPP in Japanese stock exchange. It mainly concentrates in finding the future stock prices. In order to enhance the classification accuracy, ANN is combined with Genetic Algorithm (GA) which produces a GA-ANN model to obtain effective SPP. Here, GA is selected for improving the accuracy value of ANN and to remove the converging problem from back-propagation approach. Therefore, computational analysis has been declared as hybrid GA-ANN which helps to reach a hit value of 81.27% that is greater when compared to previous study. [17] executed a hybrid method that concatenates 2-D Principal Component Analysis (2D) (PCA) as well as RBF NN for SPP at Shanghai stock market. It selects about 36 stock market parameters as input values, where a sliding window has been applied to retrieve input data. Followed by, 2D-PCA is applicable to reduce the dimensions of data as well as filter of intrinsic attributes. Finally, RBFNN employs the information computed under the application of 2D-PCA to predict the next day's stock price. Thus, the final outcome denotes that the presented model performs far better when compared with MLP frameworks.

[18] utilizes k-Nearest Neighbor (KNN) technology as well as non-linear regression technique for SPP in a collection of 6 main firms which is tabulated in Jordanian stock exchange for the purpose of helping customers as well as suppliers to make proper decisions. According to the provided simulation outcome, kNN model should be robust and capable of attaining minimum error in predicting results which is adjacent to original stock prices. [19] introduced a SPP approach by applying PCA, canonical correlation analysis (CCA) as well as SVM. At initial stage, 2 attributes are extracted from past closing cost and 39 technical features derived from autonomous component analysis. Secondly, a CCA method is applied for combining the 2 kinds of features and obtains intrinsic parameters to improve the accuracy of prediction task. Finally, SVM is employed to forecast tomorrow's closing stock price. Thus, the process takes place under the help of Dow Jones index and produced outcome denotes that presented model performs well when compared with alternate techniques, respectively.

III. PROPOSED METHOD

The overall process of the presented BA-XGB model is shown in Fig. 1. As shown, the presented model involves preprocessing, feature extraction, and classification. These subprocesses are clearly explained in the following subsections.

3.1. Preprocessing

The extended smoothing provides higher weights for current observations and reduced in the same way of previous monitoring. The statistic series of y is estimated recursively by:

S0=y0 for

Z>0, Sz=a*y+(1-a)*Sz-1 (1)

where a denotes the smoothing factor and 0<a<1. Larger value of a helps to minimize the smoothing level. If a=1 smoothed statistic is similar to original computation. Also, smoothed statistic sz could be measured in a rapid manner as frequent observations are applicable. Therefore, the smoothing helps to avoid the arbitrary difference from historical information by enabling the method for identifying a longer term price in the behavior of stock price.

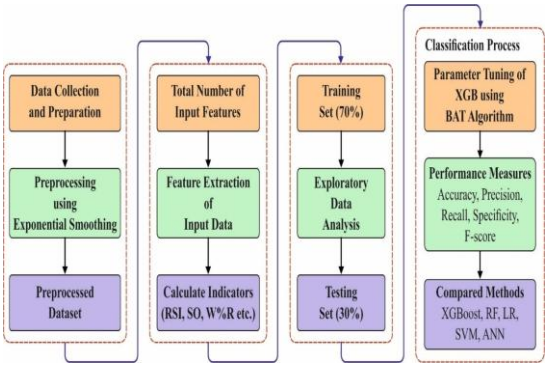

Fig. 1. Working Principle of BA-XGB model

The technical predictor has been estimated from exponentially smoothed duration that is deployed into feature matrix. Also, the target should be forecasted in u^{th} day which is measures in the following:

$$tar\ g_u = sign(close_{u+d} = close_u) \quad (2)$$

where d represent the count of days after completing the prediction task. If the measure of tar g_u is +1 it denotes a presence of positive shift in price after d days; whereas -1 implies an existence of negative shift after d days, that offers a clarity in the direction of prices for corresponding stock price. In addition, tar g_u metrics has been declared as labels for u^{th} row present in feature matrix.

3.2. Feature extraction process

From this solution, only closing price of a stock has been assumed as well as to gather such measures in last decades. Therefore, input data is defined in the form of (date,$price_{closing}$). Hence, the data is comprised with few indicators that undergo calculation are shown in Fig. 2 and also defined here.

RELATIVE strength INDEX (RSI)

It is a well-known momentum indicator that helps in computing the purchasing as well as selling rate of stock. Generally, a stock is defined as overbought at the time of demand increases the price for no reason. Hence, this state undergoes interpretation which represents the stock overvalue and minimized price respectively. Also, it is referred as oversold while there is a reduction of price under a positive value. The simulation outcome is presented because of the panic sold process. The RSI has the limits from 0 to 100 and if RSI is more than 70, it denotes that stock is overbought whereas RSI is lower than 30, it implies the stock oversold.

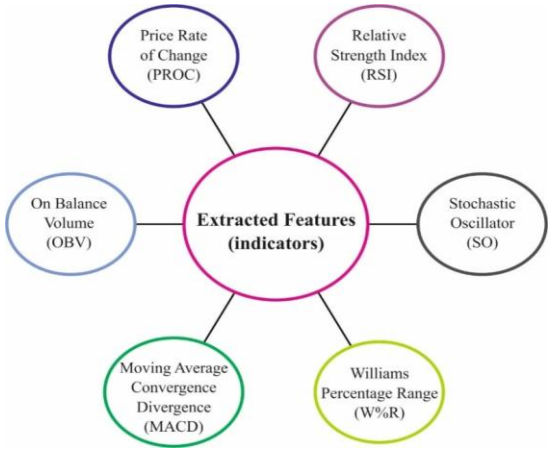

Fig. 2. Extracted features in stock market prediction as indicators

Stochastic oscillator (SO)

SO applies the momentum of price. According to the procedure, momentum is modified in prior to price alteration. It calculates the intensity of closing price that is relevant to low-high range across a time period.

Williams percentage range (W%R)

It is also known as Williams %R is an alternate momentum indicator, that is identical to the concept of SO. It represents the markets closing price level in correlation of maximum price for look-back duration, which might be 14 days. These value ranges from -100 to 0. If the value is more than -20, it is a sell point whereas value comes under-80, it is to buy signal.

Moving average convergence divergence (MACD)

It is defined as a momentum indicator that relates 2 moving averages of stock. Initial moving average is about 26-day Exponential Moving Average (EMA) and secondly moving average is of 12-day EMA. Hence, 26-day EMA is reduced from 12-day EM and 9-day EMA of MACD is declared as a signal line that facilitates as threshold for buying and selling signals.

Price rate of change (PROC)

It is said to be a technical indicator that shows the percentage modification in price among currently fixed price and the previous price which is observed for a specific time interval.

On balance volume (OBV)

It applies the modification in volume to evaluate the alteration present in stock prices. The technical indicator is employed to identify buy and sell trends of a stock, by assuming the aggregating volume: as it includes the volumes for days if the price has been increased, and minimize the volume when the price is decreased, and related with the previous day price respectively.

3.3. BA-XGB based stock market prediction

3.3.1. XGBoost

XGBoost classifier is a scalable and artificial intelligence (AI) model is applied for the improvement of tree. Actual technique of XGBoost is a gradient enhancement of decision trees (DT) that integrates massive number of "weak" learning methods with "strong" learning models by the application of iterative methodology. As depicted in Fig. 3, at the time of every iteration from XGBoost, is composed with predefined detectors which have been modified using the remaining values. Here, the algorithm might be computed in an independent manner with different varieties of loss functions that is applied in evaluating the method. In order to minimize the issue of over fitting, extra regularization is included for this technique. Hence, the mean value for every tree has been employed for regression. In case of m^{th} DT, the estimated function could be exposed as

$$\hat{y}_i = \sum_{i=1}^{m} f_m(x_i), f_m \in W, \quad (3)$$

where m implies the count of trees, f_m represent the ffexpression present in functional space W, and W denotes the space for each DT.

Objective function at t^{-th} round is signified by

$$\theta\theta(t) = \Phi(t) + \Omega(t),(t) = \sum_{i=1}^{n} \Phi\left(y_j, \hat{y}_i\right) + \sum_{k=1}^{t} \Omega(f_m), \quad (4)$$

where n indicates the n^{th} prediction and $\hat{y}$ could be provided as

$$\hat{y}_i^{(t)} = \sum_{m=1}^{t} f_m(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i). \quad (5)$$

The regularizing term $\Omega(f_k)$ is applied for DT which is discussed as provided:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2, \quad (6)$$

where C represents the difficulty of every leaf, T refers the count of leaves existing in tree, λ implies the parameter of scaling the penalty, as well as w shows vector measures on leaves. Followed by, the first-order and second-order Taylor expansions are consumed to loss function of XGBoost. For instance, while loss function is assumed to be Mean Square Error (MSE), then objective function is defined as follows:

$$\theta^{(t)} \approx \sum_{i=1}^{n} \left[g_i w_q(x_i) + \frac{1}{2}\left(h_j w_{q(x_i)}^2\right) \right] + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \quad (7)$$

where $q(\cdot)$ denotes the function which is capable of allocating the data points to respective leaves whereas g_i and h_i are assumed to be first and second order derivatives obtained from MSE loss function. From Eq. (7), loss function can be computed by addition of loss values for each every test data. As the sample data is relevant to particular leaf node, it is represented in the form of sum of loss measures for every leaf node that is described as

$$\theta^{(t)} \approx \sum_{j=1}^{T} \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2}\left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T, \quad (8)$$

where I_j is a sample data present in leaf node j. Thus, optimizing the objective function is converted to identify the lower value of quadratic function. On the other side, alteration in the performance is due to a definite node split occurs from DT which can be estimated according to the provided objective function. In case of enhanced model computation, then the modification is approved; else the split would be terminated. The architecture of XGBoost technique is depicted in Fig. 3. Apart from these techniques, while the objective function is optimized, a predictive classification is trained in contrary to over-fitting due to the existence of regularization terms. In this paper, 3 features were optimized, along with count of trees that is denoted by n iterations, and ratio of sub datasets for each data present in training method, and lower instance weight (hessian) requires a child (min_child_weight).

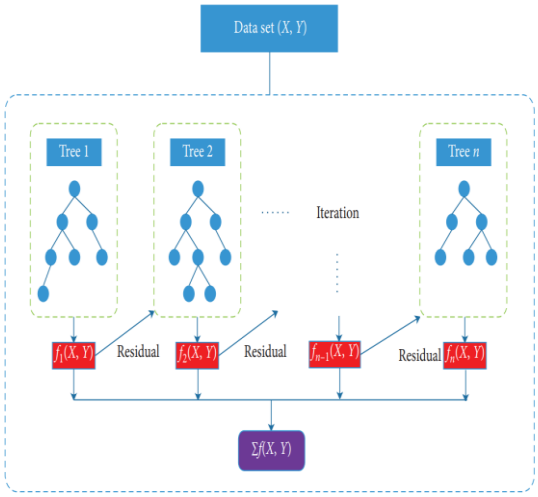

Fig. 3. The structure of the XGBoost algorithm

3.3.2. Bat algorithm (BA)

BA is assumed to be a meta-heuristic methodology designed that symbolizes the echolocation nature of microbats. These microbats produce sound pulses for identifying the prey, where every bat are comprised with corresponding echolocation behavior that is the pulse emission and sound, which is more applicable to develop diverse kinds of BA. It has the major operation to form a BA as consolidated in the following:

1) The populations of virtualized bats are produced for the purpose of simulation. In all virtual bat, there are random flies randomly with velocity v_i as well as frequency f_i at the location of x_i.
2) In the primary round, maximum process is computed for locations as well as velocities of bats in time t which are upgraded as

$$f_i = f_{min} + (f_{max} - f_{min})\beta,$$
$$v_i^t = v_i^{t-1} + (x_i^t - x_*)f_i, \quad (9)$$
$$x_i^t = x_i^{t-1} + v_i^t,$$

where $\in [0, 1)$ denotes the random vector derived from normal distribution, x_i^t and v_i^t are few upgraded locations and velocities of bats in time t, and x_* refers to be the present optimal position of a bat while comparing every solutions from all other bats inside the population.

3) After selecting the optimal position, bats would upgrade the positions on the basis of random walk:

$$x_{new} = x_{old} + rand\, A^t, \quad (10)$$

where rand $\in [-1, 1]$ represents the random value as well as A^t signifies mean loudness for every bat at time t correspondingly.

4) By repeating the process, the value of pulse emission r_i and loudness A_i of every bat present inside the population is upgraded by using the provided criteria:

$$A_i^{t+1} = \alpha A_i^t \quad (11)$$

$$r_i^{t+1} = r_i^0 [1 - \exp(-\gamma t)],$$

where α and C are constants. Thus, an emission rate as well as loudness of each bat is updated till an optimal solution is attained, and stops the iteration. The process involved in BA is shown in Fig. 4.

3.3.3. Hybridization of BA with XGB model

The hybrid technique is generated by combining BA with XGBoost (i.e., BA-XGB). The process of XGBoost classifier is greatly based on proper options of features. In BA-XGB, feature selection XGBoost has been computed as well as optimized using BA. It helps in optimizing 3 major attributes of XGBoost technique, like count of trees (n rounds), lower integrating of instance weight essential for child (Min_ Child _weight) and study value (eta) where the parameter measures are depicted in Table 1 [20].

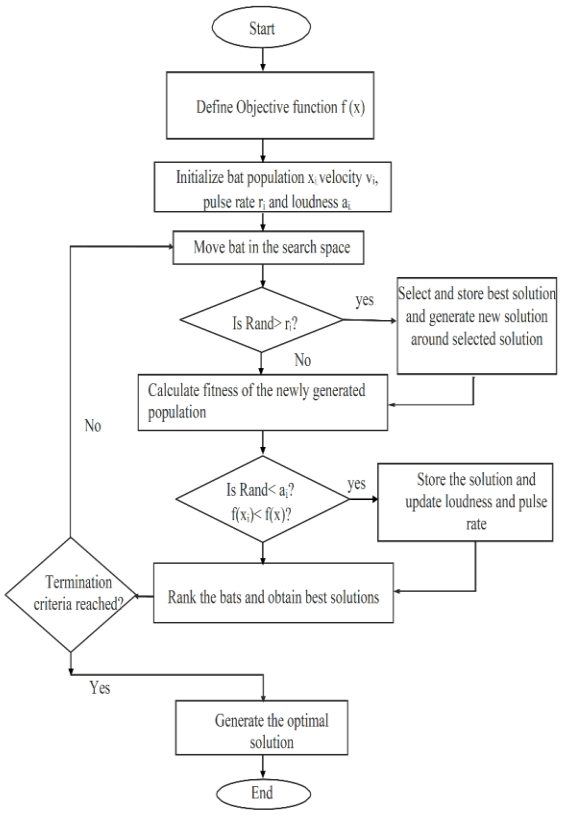

Fig. 4. Flowchart of BA

Table 1 Parameter values of the Bat-XGBoost algorithm

Parameter	Value
Population	50
Max generation	200
A	0.98
R	0.98
N rounds	[20, 2000]
Min_child_weight	[1, 15]
Eta	[0.01, 0.3]

IV. PERFORMANCE VALIDATION

4.1. Dataset used

For experimentation, the data from two organizations namely Facebook (FB) and Apple (APPL) as well as every available information is applied (the day they went to public yet 3rd February 2017). Such types of firms are arbitrarily sampled, with no rigorous assumptions from their background or any type of economic effect they provide on society. It has been emphasized that the diversity of these companies are selected to analyze the stock prices which is significant to assure the effectiveness of technique. The original measures are derived from data which is obtained from entry of data, closing price, volume, and so on. From the original format, size of corresponding data to stocks of various firms are differed among 10 kB to700 kB, by using the count of rows that is respective for closing prices and modifies among 1180 and 10,700. According to the closing prices, residual technical indicators can be computed. Some of the dataset does not obtain the categorical as well as ordinal parameters: every feature values are said to be consecutive. In general, the most observed features are known to be non-linear. Since it is a primary insight, the significance of features are projected and calculated by using the techniques in Table 2. This provides a clear point, as it is seldom in case, specifically for shorter trade, that has to be affected by smaller feature set. The combination of tiny features along with the observation, it is observed that, there is a requirement for reducing dimension which is more efficient. Also, feature space should be segmented as non-overlapping subspaces which tend to train the learning models to discriminate among classes. But, in specific class of issues, from where the number of features is not existed to serve the performance measure of ML models. As the basic requirement, each features indicators are sufficient. Therefore, the percentage existence of OBV enhances with drastic alteration along with increased size of trading window (see Table 2), by comparing the alternate feature indicators. Hence, it is said to be more reliable than all other stocks applied in this paper.

Table 2 Sample dataset

		Trading Window (in number of days) and Corresponding Importance of Features (in%)						
Company	**Indicator**	**3**	**5**	**10**	**15**	**30**	**60**	**90**
	RSI	17.98	17.34	17.49	16.88	16.19	13.81	12.66
	SO	14.23	13.51	12.82	12.43	11.61	10.3	9.41
	W%R	14.12	13.63	12.72	12.56	11.51	10.34	9.28
	MACO	17.8	17.89	18.2	18.16	17.7	16.48	16.34
	PROC	16.74	16.28	15.19	15.53	16	18.96	21.6
AAPL	OBV	19.12	21.36	23.57	24.43	26.99	30.11	30.72
	RSI	17.43	18.31	18.01	17.3	15.6	14.35	12.06
	SO	14.86	14.5	12.58	11.78	10.25	9.22	9.45
	W%R	14.8	13.13	12.34	11.64	11.42	9.28	10.77
	MACO	17.7	18.26	17.15	17.9	16.18	12.81	14.01
	PROC	16.17	15.77	15.55	16.11	17.19	22.44	22.13
FB	OBV	19.04	20.03	24.37	25.28	29.37	31.91	31.58

4.2. Results analysis

Table 2 and Fig. 5 offers the results attained by diverse models with respect to different measures on the two applied dataset namely FB and AAPL stocks. On measuring the predictive outcome on the AAPL dataset, a detailed analysis takes place under varying trading windows of 3-90. Under the training window of 3, the BA-XGB model offers an accuracy of 66.26%, recall of 73%, precision of 68%, specificity of 59% and F-score of 69%. Similarly, in case of 5 as training window, the BA-XGB model provides a slightly high accuracy of 74.55%, recall of 80%, precision of 75%, specificity of 58% and F-score of 76%. Likewise, under the presence of 10 as training window, the BA-XGB model provides a high accuracy of 79.80%, recall of 83%, precision of 83%, specificity of 77% and F-score of 82%.

Table 3 Results of classification using BA-XGB

Company Name	**Trading Window**	**Accuracy**	**Recall**	**Precision**	**Specificity**	**F-Score**
	3	66.26	73	68	59	69
	5	74.55	80	75	58	76
	10	79.8	83	83	77	82
	15	83.01	86	84	79	85
	30	86.34	89	88	82	88
	60	91.44	95	93	87	94
AAPL Stock	90	95.47	97	95	91	95
	3	68.59	74	69	63	73
	5	75.15	86	75	63	79
	10	82.39	92	83	70	86
	15	87.06	90	92	82	90
	30	90.05	97	94	83	94
	60	90.02	99	90	62	96
FB Stock	90	96.42	99	98	74	98

Equally, under the presence of 15 as training window, the BA-XGB model provides a high accuracy of 83.01%, recall of 86%, precision of 84%, specificity of 79% and F-score of 85%. Under the training window of 30, the BA-XGB model offers an accuracy of 86.34%, recall of 89%, precision of 88%, specificity of 82% and F-score of 88%.

In the same way, in case of 60 as training window, the BA-XGB model provides a slightly high accuracy of 91.44%, recall of 95%, precision of 93%, specificity of 87% and F-score of 94%.

On continuing with, under the presence of 60 as training window, the BA-XGB model provides a high accuracy of 91.44%, recall of 95%, precision of 93%, specificity of 87% and F-score of 94%. Equally, under the presence of 90 as training window, the BA-XGB model provides a high accuracy of 95.47%, recall of 97%, precision of 95%, specificity of 91% and F-score of 95%.

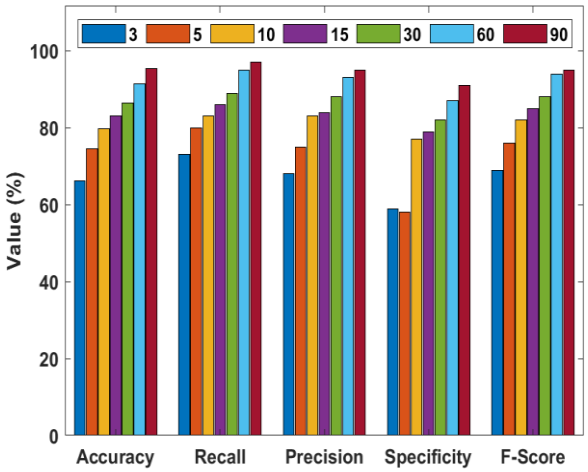

Fig. 5. Classification results of GD-XGB model on AAPL stock data

By determining the predictive outcome of the BA-XGB model on the FB dataset, an extensive examination is carried out under varying trading windows of 3-90 and the outcome is illustrated in Fig. 6. Under the training window of 3, the BA-XGB model provides an accuracy of 98.59%, recall of 74%, precision of 69%, specificity of 63% and F-score of 73%. On continuing with, in case of 5 as training window, the BA-XGB model gives high accuracy of 75.15%, recall of 86%, precision of 75%, specificity of 63% and F-score of 79%. In the same way, under the presence of 10 as training window, the BA-XGB model provides a high accuracy of 82.39%, recall of 92%, precision of 83%, specificity of 70% and F-score of 86%. Equally, under the presence of 15 as training window, the BA-XGB model provides a high accuracy of 87.06%, recall of 90%, precision of 92%, specificity of 82% and F-score of 90%. Under the training window of 30, the BA-XGB model offers an accuracy of 90.05%, recall of 97%, precision of 94%, specificity of 83% and F-score of 94%. At the same time, in case of 60 as training window, the BA-XGB model provides a slightly high accuracy of 90.02%, recall of 99%, precision of 90%, specificity of 62% and F-score of 96%. On continuing with, under the presence of 60 as training window, the BA-XGB model provides a high accuracy of 90.02%, recall of 99%, precision of 90%, specificity of 62% and F-score of 96%. Along with, under the presence of 90 as training window, the BA-XGB model provides a high accuracy of 96.42%, recall of 99%, precision of 98%, specificity of 74% and F-score of 98%. These values indicated that the BA-XGB model offers superior predictive outcome on the applied FB dataset. The above mentioned table values indicated that the results gets enhanced with an increase in window size.

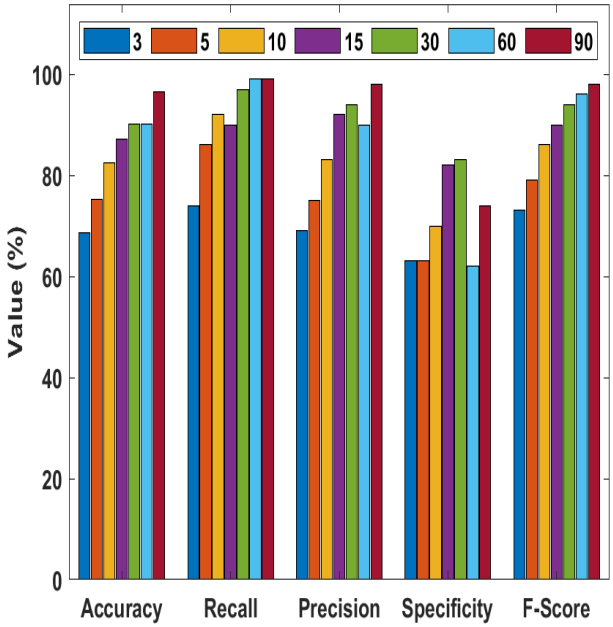

Fig. 6. Classification results of GD-XGB model on FB stock data

Table 4 Accuracy analysis of different models

Methods	Accuracy
BA-XGB	96.42
RF	92.00
XGboost	83.00
LR	55.00
SVM	58.00
NN	72.00

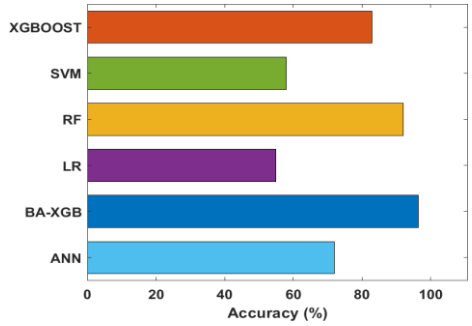

Fig. 7. Comparative analysis of diverse predictive models interms of accuracy

Afterwards, an accuracy analysis is made between the presented BA-XGB and existing models and is shown in Table 4 and Fig. 7. It is exhibited that the logistic regression (LR) model offers worse predictive performance by attaining a least accuracy of 55%. Next, the SVM model offers slightly higher predictive results by achieving an accuracy of 58%. Afterwards, the ANN model shows moderate predictive outcome by attaining an accuracy of 72%. Subsequently, the XGBoost model provides manageable predictive with an accuracy of 83%. At the same time, a competitive accuracy value of 92% is offered by the XGBoost model.

However, a highest accuracy value of 96.42% is provided by the presented BA-XGB model. This superior value ensured the optimal predictive performance of the presented models on the applied dataset.

V. CONCLUSION

This paper has presented an effective BA-XGB model for the accurate prediction of the direction of the stock prices. The presented model involves preprocessing, feature extraction and classification. The BA is incorporated to the XGB model to tune its parameters and thereby enhances the actual performance of it. The presented model undergoes extensive validation using the data from APPL and FB stocks. The obtained simulation outcome stated that the BA-XGB model has offered superior outcome by achieving a maximum accuracy of 96.42. It is also noted that the results gets enhanced with an increase in window size. In future, the presented model can be extended by the use of clustering techniques and deep learning based models.

REFERENCES

1. Malkiel, B. G. (2003). The efficient market hypothesis and its critics. The Journal of Economic Perspectives. 17(1), 59–82.
2. V. Veeramanikandan, M. Jeyakarthic, "Forecasting of Commodity Future Index using a Hybrid Regression Model based on Support Vector Machine and Grey Wolf Optimization Algorithm". International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-X, Issue-X, July 2019
3. V. Veeramanikandan, M. Jeyakarthic, "An Ensemble Model of Outlier Detection with Random Tree Data Classification for Financial Credit Scoring Prediction System". International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019
4. Murugan, S. and Jeyakarthic,"Optimal Deep Neural Network based Classification Model for Intrusion Detection in Mobile Adhoc Networks".Jour of Adv Research in Dynamical & Control Systems, Vol. 11, 10-Special Issue, 2019
5. Pai, P. F., & Lin, C. S. (2005). A hybrid arima and support vector machines model in stock price forecasting. Omega, 33(6), 497–505.
6. Dai, Y., & Zhang, Y. (2013). Machine learning in stock price trend forecasting. Stanford University http://cs229.stanford.edu/proj2013/DaiZhangMachineLearningInStock PriceTrendForecasting.pdf.
7. Di, X. (2014). Stock trend prediction with technical indicators using SVM. Stanford University.
8. Geurts, P., & Louppe, G. (2011). Learning to rank with extremely randomized tree. JMLR: Workshop and Conference Proceedings, 14, 49–61
9. Devi, K.N., Bhaskaran, V.M., & Kumar, G.P. (2015). Cuckoo optimized SVM for stock market prediction. In: IEEE Sponsored 2nd International Conference on Innovations in Information, Embedded and Communication systems (ICJJECS).
10. Giacomel ,F., Galante, R. & Pareira, A. (2015). An Algorithmic Trading Agent based on a Neural Network Ensemble: a Case of Study in North American and Brazilian Stock Markets. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.
11. Boonpeng, S., & Jeatrakul, P. (2016). Decision support system for investing in stock market by using OAA-neural network. In: 8th International Conference on Advanced Computational Intelligence Chiang Mai, Thailand.
12. Qiu, M., & Song, U. (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. PLoS One, 11(5).
13. M. Alrasheedi, A. Alghamdi, Comparison of Classification Methods for Predicting the Movement Direction of Saudi Stock Exchange Index, Journal of Applied Sciences, 14(16), 1883-1888, 2014.
14. Milosevic, Nikola. "Equity forecast: Predicting long term stock price movement using machine learning." arXiv preprint arXiv:1603.00751 (2016).
15. Leung, Carson Kai-Sang, Richard Kyle MacKinnon, and Yang Wang. "A machine learning approach for stock price prediction." Proceedings of the 18th International Database Engineering & Applications Symposium. ACM, 2014.
16. Qiu, Mingyue, and Yu Song. "Predicting the direction of stock market index movement using an optimized artificial neural network model." PloS one 11.5 (2016): e0155133.
17. Guo, Zhiqiang, et al. "A stock market forecasting model combining two-directional two-dimensional principal component analysis and radial basis function neural network." PloS one 10.4 (2015): e0122385.
18. Alkhatib, Khalid, Hassan Najadat, Ismail Hmeidi, and Mohammed K. Ali Shatnawi. "Stock price prediction using k-nearest neighbor (knn) algorithm." International Journal of Business, Humanities and Technology 3, no. 3 (2013): 32-44.
19. Guo, Zhiqiang, et al. "A feature fusion based forecasting model for financial time series." PloS one 9.6 (2014): e101113.