Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

Kendall's Coefficient of Concordance Ranking of the Effectiveness of Single Machine Learning Models in Predicting Stock Price Movement

Ampomah Ernest Kwame[1]
School of Information and Software Engineering
University of Electronic Science and Technology of China
Chengdu, China

Eric Ahudey[3]
School of Business Administration
Christian Service University College
Kumasi, Ghana

Qin Zhiguang[2]
School of Information and Software Engineering
University of Electronic Science and Technology of China
Chengdu, China

Addo Prince Clement[4]
School of Management and Economics
University of Electronic Science and Technology of China
Chengdu, China

***Abstract*—The prediction of stock price behavior is extremely important in the world of finance. Generally, stock investment decisions are made by investors by predicting the future movement of prices of stocks. An accurate prediction of the movement of stock price is needed by investors to make decisions regarding buying and selling of stocks, and minimize the risk associated with such investment. In this study, we adopt Kendall's coefficient of concordance technique to rank the effectiveness of five single machine models (which are Decision Tree (DT), Gaussian Naïve Bayes (GNB), Logistic Regression (LR), Support Vector Machine (SVM) and k-Nearest Neighbor (kNN)) in forecasting the direction of movement of stock price. The experiment is conducted utilizing ten different stock data sets randomly collected from three stock markets. Each data set is split into training and test sets. The models are trained on the training data set, and the test data set is used to evaluate the models. The models are evaluated using the six classical evaluation metrics. The experimental results recorded illustrate that LR model had the highest rank when using Kendall's coefficient of concordance (W) with accuracy, precision, f1-score, specificity, and AUC metrics. However, SVM model achieved the best rank when using the Kendall's coefficient of concordance (W) with recall metric.**

Keywords—Stock price; machine learning; technical indicators; feature extraction, Kendall's coefficient of concordance

I. INTRODUCTION

Predicting the behavior of stock price is of extreme importance in the world of finance. In general, stock investment decisions are made by investors by predicting the future movement of prices of stocks. An accurate prediction of the movement of stock price enables investors to make decisions regarding buying and selling of stocks, and minimize the risk associated with such investment [1-3]. Predicting how stock price will behave is a challenging task as the stock market is influence by factors such as the general macroeconomic conditions of a country, political events, investors sentiments, etc. and these factors make the stock market very dynamic, non-linear, and noisy in nature [4]. Different prediction techniques have been used to predict the stock market and reduce its associated uncertainty. These prediction techniques can generally be grouped into three major categories: (i) fundamental analysis, (ii) technical analysis, and (iii) machine learning (ML) approaches. The fundamental and technical approaches are not able to deal effectively with the dynamic, non-linear, and chaotic nature of the stock market time series data. On the contrary, machine learning models have the ability to handle the noisy, dynamic, and non-linear stock market time series data effectively and efficiently [5-7].

The discovery of machine learning models having the ability to capture the stock market dynamics very well, with the goal to reduce uncertainties and risk, has been the topic of research and has attracted the interest of academic researchers from various fields of study and market professionals. Fischer and Krauss, [8] compared the performance random forest, deep neural net, logistic regression and LSTM network in forecasting the direction of movement of constituent stocks of the S&P 500 from 1992 to 2015. The experimental results suggested that LSTM network outperformed the other models. Nguyen & Yoon [9] compared the effectiveness of ensembled tree-based machine learning models in forecasting the direction of movement of stock prices. The models considered included random forest, XGBoost, Bagging classifier, AdaBoost, Extra Trees and Voting Classifier. The authors used eight different stock data sets in the study. Classical evaluation metrics including accuracy, precision, recall, F1-score, specificity, and area under receiver operating characteristics curve were used to evaluate the models. The performance of the models across the all the data sets were ranked using Kendall W test of concordance. The results presented indicated that extra trees model outperformed all the other models. Vijh el at [10] used Artificial Neural Network and Random Forest techniques to predict the next day closing price of stocks of five companies selected from different sectors of operation. The authors generated new features from Open, High, Low and Close prices of stock and used them as inputs to the model. RMSE and MAPE were used to evaluated the models. The outcome of the study showed that ANN produced better prediction of stock prices than random forest.

In this research work, we conduct a comparative assessment of the efficacy of five different single machine learning classifier models in forecasting the movement of stock price using Kendall's coefficient of concordance to rank the performance of the models.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

II. MATERIALS AND METHODS

A. Machine learning algorithms

The study uses Kendall's coefficient of concordance technique to rank the effectiveness of five single machine learning models (which are Decision Tree (DT), Gaussian Naïve Bayes (GNB), Logistic Regression (LR), Support Vector Machine (SVM) and k-Nearest Neighbor (kNN)) in predicting the direction of future movement of stock price. A discussion of these ML algorithms and the Kendall's coefficient of concordance is presented here.

1. Decision Tree (DT)

Decision tree is a predictive model which has flowchart-like structure. It has non-leaf node(s) which denote test on an attribute, branches which represent possible results of the test, and leaf nodes denoting target classes. The paths from root to leaf nodes present the classification rules. A new instance is labeled according to it attributes values. DT assigns class label to new instances by moving them down the tree from the root to some leaf node, according to the outcomes of the tests along the path [11]. DT has the capability to extract decision-making knowledge from the supplied data set. It is computationally efficient and has a high adaptability to deal with diverse datasets. DT models are intuitive and easy to explain.

2. Gaussian Naïve Bayes (GNB)

The GNB is a variant of Naïve Bayes that is based on Gaussian normal distribution and supports continuous data. It is a probabilistic model which is based on application of Bayes' theorem with naive independence assumptions. GNB treats all input features as being independent from each other. The classifier assigns data points to the closest class, however, instead of using Euclidean distance from the class-means to determine the nearness, the GNB considers the distance from the mean, in addition to, how it compares to the variance of the class. For each dimension, the z-score (which is computed as the distance from the mean divided by the standard deviation) is determine. GNB operates with the assumption that the classes have Gaussian normal distributions. This enables each z-score to be directly transformed into a p-value (the probability of getting a specific data point (x), if x was taken from the distribution of a certain class). However, what is being sought after is the probability of a class, given the data, and not the probability of the data given a specific class [12]. The Bayes' Theorem (eq.2) permits us to get each one from the other.

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)} \quad (1)$$

A, B = events
P(A|B) = probability of A given B occurs
P(B|A) = probability of B given A occurs
P(A) = probabilities of A occurring
P(B) = probabilities of A occurring

The likelihood of the features is assumed to be Gaussian. Hence, the conditional probability is given as:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma^2 y}}\exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma^2 y}\right) \quad (2)$$

The maximum likelihood estimation is used to estimate the parameters σ_y and μ_y. x_i = feature vector, y = class variable

3. Support Vector Machine (SVM)

SVM is machine learning technique derived from statistic learning theory, and it is based on is the structural risk minimization principle. It is kernel learning discriminative classifier that constructs a hyperplane(s) to classify new instances [13]. The objective of SVM is to find a hyperplane that has the maximum margin (distance from the hyperplane to the support vectors). Maximizing the margin gives some reinforcement so that new instances can be classified with confidence. A kernel trick technique is used by SVM to do a complex data transformation and on the basis of the transformation, it determines an optimal hyperplane separating the classes. The kernel trick technique permits us to efficiently add extra dimensions to the margin through the introduction of kernel matrices such as radial basis function kernel (RBF), polynomial kernel and sigmoid kernel. The generalization performance of SVM is very remarkable [14]. SVM models are able to deal effectively with very complex relationships among data points, however, it is computationally exhaustive and takes long time to train it.

4. Logistic Regression (LR)

Logistic Regression is a supervised learning predictive algorithm which is based on the concept of probability. It assigns a new instance to a target class by analyzing and determining the relationship between a binary target class and the predictor variables. LR uses logistic sigmoid function to convert its output to a probability which is mapped to the target classes. The logistic regression's prediction function produces the probability of an observation belonging to a positive class [15]. The sigmoid function eq. (3) is used to map predictions to probabilities.

$$y = \sigma(z) = \frac{1}{1-e^{-z}} \quad (3)$$

To map the probability value to a target class, we set a threshold value above which we assign an observation to one class or the other as: $p \geq 0.5$, class =1, and $p < 0.5$, class = 0.

5. k-Nearest Neighbor (kNN)

kNN is a non-parametric and lazy learning algorithm classifier which uses feature similarity to predict the class of a new instance. To predict the class of a new instance, the kNN classifier computes the distances between the test point and points in the training data set. It then classifies the new instance by majority vote of its k closest training points (where k is an integer). Each of the k closest training points votes for its class, and the class that gets the most votes is taken as the prediction [16]. kNN learns from training data only at the time of making predictions. kNN is very simple to comprehend and easy to implement. It is able to respond quickly to changes in the input data during real-time use.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

However, kNN is sensitive to outliers, and cannot deal with missing values. Also, it struggles with high dimensional data.

B. Hyperparameter Optimization

Hyperparameters are essential for machine learning algorithms as they have direct control over the behaviors of learning algorithms and have a significant impact on the performance of machine learning models [17]. For this study, the values of hyperparameters the machine learning algorithms are set using Bayesian hyperparameter optimization (BHPO) approach. BHPO is an iterative algorithm which operates on the basis of the Bayesian theorem. BHPO has two major constituents: a probabilistic surrogate function and an acquisition function. It uses the probability surrogate function of the objective function to pick the most promising hyperparameters to examine in the actual objective function. BHPO keeps track of previous evaluation results and use them to build the surrogate. After every evaluation of the objective function, the surrogate function is updated by the algorithm incorporating the new results. The next set of values to try in the objective function are chosen by the acquisition function (algorithm optimizing the surrogate function which is usually Expected Improvement criteria) trading off exploration and exploitation. Determining the values that will provide the highest expected improvement in the surrogate function is extremely cheaper than evaluating the objective function itself [18]. Hence, BHPO provides an efficient and cheap way to select good hyperparameter for ML models.

C. Evaluation Metric

The evaluation of the performance of the ML models are done by using the following six classical evaluation metrics: (a) accuracy, (b) precision, (c) recall, (d) F1-score (e) specificity (f) area under receiver operating characteristics curve (AUC-ROC). Below are their definitions:

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{4}$$

$$precision = \frac{tp}{tp + fp} \tag{5}$$

$$recall = \frac{tp}{tp + fn} \tag{6}$$

$$f1\_score = \frac{2 \times precision \times recall}{precision + recall} \tag{7}$$

$$specificity = \frac{tn}{tn + fn} \tag{8}$$

where tp, tn, fp, and fn denote the 'true positive', 'true negative', 'false positive', and 'false negative' respectively. AUC measures the ability of the classifier to discriminate between the different classes. The best AUC value is one (1), and the worst AUC value is 0.5.

ROC is a probability curve that illustrates the trade-off between recall and specificity in a graphical way. A model with a perfect separability (100% recall, 100% specificity) will have ROC curve that passes through the upper left corner.

D. Kendall's coefficient of concordance (W)

The Kendall's coefficient of concordance (W) is a measure that applies ranks to establish an agreement among raters. It is a measure of the agreement among different raters who are evaluating a given set of n objects [19]. Depending on the area where it is being applied, the raters can be variables, characters, and so on. The raters are the different data sets in this article. The Kendall's W statistic can be computed in two ways (the upper and lower forms of equation (9) and (10)), and they both produce the same outcome. S' or S is calculated first from the row marginal sums of ranks R_i received by the objects.

$$S = \sum_{i=1}^{n} \left(R_i - \bar{R}\right)^2 \tag{9}$$

$$S' = \sum_{i=1}^{n} R_i^2 = SSR \tag{10}$$

S is the sum of squares statistic over the row sum of ranks R_i. $\bar{R}$ is the mean of the R_i. The Kendall's W can be computed from either of the following formulas:

$$W = \frac{12S}{m^2(n^3 - n) - mT} \tag{11}$$

$$W = \frac{12S' - 3p^2 n(n+1)^2}{m^2(n^3 - n) - mT} \tag{12}$$

Where n is the number of objects m the number of raters, T is a correction factor for tied ranks.

$$T = \sum_{k=1}^{g} \left(t_k^3 - t_k\right) \tag{13}$$

In which t_k is the number of tied ranks in each (k) of g groups of ties. Kendall's W is the variance estimation of the row sums of ranks R_i divided by the highest possible value the variance can assume. This happens when all the variances are in complete agreement. Hence $0 \le W \le 1$, with 1 representing total concordance. Friedman's chi-square statistic is derived as

$$\chi^2 = m(n-1)W \tag{14}$$

E. Feature Scaling

Standardization scaling (z-score) is applied to the data sets to ensure that the values of all the features are within the same range. The purpose for scaling of input data is to ensure that the bigger value input features do not overwhelm smaller value input features, and also to help in the reduction of prediction errors [20]. The z-score converts the values of each feature to have the characteristics of a Gaussian distribution with the values of each feature centered around zero, and having a unit-variance.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

$$z(x) = \left(x[:,i] - \mu_i \right) / \sigma_i \quad (15)$$

Where μ_i = mean of the ith feature, σ_i = standard deviation of the ith feature

F. Dimensionality Reduction

In order to reduce the amount of redundant data, the Principal Component Analysis (PCA) is used to extract the most relevant features for the machine learning models. The total number of features in the final data set is forty-five (45) which is made up of the forty computed technical indicators and the initial five OHLCV (open, high, low, close & volume) variables of the original data. When the dimensionality of data is high, machine learning models tend to suffer from the curse of dimensionality causing their performance to reduce. Hence, dimensionality reduction process is relevant in this study. The stability and performance of machine learning models in stock price prediction is improve by PCA [21, 22]. PCA extracts and keep only the most important features from the original dataset. It generates new uncorrelated features that successively maximize variance. It applies an orthogonal transformation to transform values of possibly correlated features into values of features that are linearly uncorrelated. The new features are referred to as principal components (PC). The first PC is chosen in such a way that, it decreases the distance between the data and its projection onto the PC. By decreasing the distance, the variance of the projected points is increased. The succeeding PCs are selected in a similar manner, but with an extra obligation that they must be uncorrelated with the preceding PCs. More often than not, most variance within the dataset are taken care of by the initial few PCs, therefore, the remaining PCs can be overlooked with only a minor information loss. There seems to be many highly correlated features in our final data set, hence, an application of PCA helps us lessen the effect of strong correlations among features, while reducing the dimensionality of the feature space. We adopt PCs that retain most of the variance of the original data, hence, we set a threshold of 95%.

G. Data and Features

The stock datasets used in the study are randomly collected from three stock exchanges (NYSE, NASDAQ, and NSE) through yahoo API. The study is carried out using ten different stock data sets. These stock data sets are Apple Inc. (‘AAPL’), Abbott Laboratories (‘ABT’), Bank of America Corp (‘BAC’), CarMax Inc. (‘KMX’), S&P_500 Index, Microsoft Corporation (‘MSFT’), Exxon Mobil Corporation (‘XOM’), Tata Steel Limited (‘TATASTEEL’), Hindustan Petroleum Corporation Limited (HPCL), and HCL Technologies Ltd (‘HCLTECH’). Table I presents a description of the data sets used. From the original data sets, forty (40) technical indicators are computed and used as input features. Table XIV-XVII in the appendix section provide the details of these technical indicators. Each of the dataset is divided into two sets, the training and test sets for the purpose of this study. The training set consist of the initial 70% of the data set, and the test set constitute the final 30% of the data set. Each model is train with the training set and the test set is used to evaluate them.

TABLE I. DESCRIPTION OF THE STOCK DATA SETS

Data Set	Stock Market	Time Frame	Number of Sample
BAC	NYSE	2005-01-01 to 2019-12-30	3773
ABT	NYSE	2005-01-01 to 2019-12-30	3773
TATASTEEL	NSE	2005-01-01 to 2019-12-30	3278
HCLTECH	NSE	2005-01-01 to 2019-12-30	3476
KMX	NYSE	2005-01-01 to 2019-12-30	3773
MSFT	NASDAQ	2005-01-01 to 2019-12-30	3773
S&P_500	INDEXSP	2005-01-01 to 2019-12-30	3773
XOM	NYSE	2005-01-01 to 2019-12-30	3773
HPCL	NSE	2005-01-01 to 2019-12-30	3476
AAPL	NASDAQ	2005-01-01 to 2019-12-30	3773

III. RESULTS AND ANALYSIS

Table II presents the accuracy values recorded by the single ML models on the various stock data sets. Overall, the mean accuracy scores of LR, and SVM are equal and the best mean accuracy score compared with the other ML models. Fig. 1 below presents a boxplot of the accuracy values of the ML models.

Table III shows the precision values obtained by the single ML models on the various stock data sets. In general, the mean precision value of LR is the highest mean precision score compared with the other ML models. Fig. 2 below presents a boxplot of the precision values of the ML models. Table IV provides the recall outcomes of the single ML models on the ten different stock data sets. The mean recall value of SVM is the best mean recall score among the ML models. Boxplot displaying the recall values recorded by the ML models are displayed by Fig 3. Table V gives the F1 scores evaluation metric results of the ML models on the ten different stock data sets. Overall, the mean F1 score of LR model is the highest mean F1 score among the ML models. Fig 4 shows boxplot of f1 scores achieved by the ML models in the experiment. Table VI presents the specificity values achieved by the ML models on the ten different stock data sets On the whole, the mean specificity score of LR model is the highest mean specificity score among the ML models. Fig 5 presents boxplot of specificity scores achieved by the ML models in the experiment. Table VII presents the AUC values recorded by the ML models on the ten different stock data sets. In general, the mean AUC score of SVM model is the highest mean specificity score among the ML models.

Fig 6 presents boxplot of AUC scores achieved by the ML models in the experiment. Fig. 7-16 present the ROC curves of the ML models on the XOM, ABT, BAC, AAPL, HCLTECH, HPCL, MSFT, KMX, S&P_500 and TATASTEEL stock data sets respectively. These curves demonstrate how well the ML models perform on the different stock data sets.

The rankings of the performances of the ML models by means of Kendall’s coefficient of concordance (W) using the various evaluation metrics is displayed in Table VIII. A cut-off value of 0.05 for the significance level (p-value) is used. The Kendall’s coefficient is taken to be significant and possessing the ability of giving an overall ranking when

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

$p \leq 0.05$. The critical value of chi-square (χ^2) for four (4) degrees of freedom is 9.488 at $p = 0.05$. The degrees of freedom equal the total number of ML algorithms (which is five) minus one.

The Tables VIII show that Kendall's coefficient is significant using the accuracy, precision, recall, F1, specificity, and AUC metrics ($p<0.05$, $\chi^2 > 9.488$). The performance of LR model is the best ranked among the ML models using the Kendall's coefficient of concordance (W) with accuracy, precision, F1, specificity and AUC metrics. Similarly, the SVM model is the highest ranked among the ML models using the Kendall's coefficient of concordance (W) with recall metric. In general, the Kendall's coefficient of concordance (W) ranking of the ML models using accuracy, F1 score, and AUC metrics is **LR > SVM > GNB > kNN > DT**. Similarly, the Kendall's coefficient of concordance (W) ranking of the ML models using precision and specificity metrics is **LR > GNB > SVM > kNN > DT**. Additionally, the Kendall's coefficient of concordance (W) ranking of the ML models using recall metric is **SVM > LR > GNB > kNN > DT.**

TABLE II. ACCURACY MEASURE OF THE ML LEARNING ALGORITHMS ON THE TEST DATASETS

Data set	DT	GNB	LR	SVM	KNN
XOM	0.7194	0.8593	**0.8667**	0.8648	0.8065
ABT	0.6157	0.8370	**0.8741**	0.8713	0.7241
BAC	0.7611	**0.8482**	0.8453	0.8435	0.7880
AAPL	0.6278	0.7120	0.8657	**0.8667**	0.7333
HCLTECH	0.7144	0.8203	**0.8789**	0.8749	0.7568
HPCL	0.5944	0.6943	0.8739	**0.8799**	0.6831
MSFT	0.6213	0.5990	0.8093	**0.8194**	0.6324
KMX	0.6537	0.8482	**0.8851**	0.8806	0.7463
S&P_500	0.7056	0.7472	**0.8556**	0.8519	0.7593
TATASTEEL	0.7318	0.8863	0.8949	**0.8959**	0.8144
Mean	0.6745	0.7852	**0.8650**	**0.8650**	0.7444

TABLE III. PRECISION MEASURE OF THE ML LEARNING ALGORITHMS ON THE TEST DATA SETS

Data set	DT	GNB	LR	SVM	KNN
XOM	0.7485	**0.9078**	0.8934	0.8648	0.8350
ABT	0.6571	**0.8838**	0.8812	0.8767	0.7707
BAC	0.7765	**0.8422**	0.8330	0.8256	0.7725
AAPL	0.6500	**0.9006**	0.8830	0.8739	0.7227
HCLTECH	0.7702	0.8258	**0.8912**	0.8748	0.7526
HPCL	0.6256	0.6289	**0.8540**	0.8479	0.7354
MSFT	0.6824	**0.9500**	0.9219	0.9255	0.7912
KMX	0.6789	0.8805	**0.9006**	0.8996	0.7725
S&P_500	0.7561	**0.9517**	0.9223	0.9354	0.8117
TATASTEEL	0.7293	0.8944	**0.9065**	0.9050	0.8147
Mean	0.7075	0.8666	**0.8887**	0.8829	0.7779

TABLE IV. RECALL SCORE OF THE LEARNING ALGORITHMS ON THE TEST DATA SETS

Data set	DT	GNB	LR	SVM	KNN
XOM	0.6764	0.8055	**0.8382**	0.8255	0.7727
ABT	0.5924	0.8014	**0.8843**	**0.8843**	0.6909
BAC	0.7455	0.8636	0.8709	**0.8782**	0.8273
AAPL	0.6257	0.5182	0.8631	**0.8770**	0.8128
HCLTECH	0.6741	0.8473	0.8845	**0.8976**	0.8212
HPCL	0.5197	**0.9843**	0.9095	0.9331	0.5965
MSFT	0.5849	0.2874	**0.7143**	0.7310	0.4521
KMX	0.6279	0.8175	**0.8748**	0.8658	0.7227
S&P_500	0.6819	0.5668	**0.8037**	0.7834	0.7293
TATASTEEL	0.7479	0.8792	0.8834	**0.8877**	0.8199
Mean	0.6476	0.7371	0.8527	**0.8564**	0.7245

TABLE V. F1 SCORES OF THE LEARNING ALGORITHMS ON THE TEST DATA SETS

Data set	DT	GNB	LR	SVM	KNN
XOM	0.7106	0.8536	**0.8649**	0.8255	0.8026
ABT	0.6231	0.8406	**0.8828**	0.8805	0.7286
BAC	0.7607	**0.8528**	0.8516	0.8511	0.7990
AAPL	0.6424	0.6579	0.8729	**0.8754**	0.7651
HCLTECH	0.7190	0.8364	**0.8879**	0.8860	0.7854
HPCL	0.5677	0.7675	0.8808	**0.8885**	0.6587
MSFT	0.6299	0.4413	0.8049	**0.8169**	0.5754
KMX	0.6524	0.8478	**0.8875**	0.8824	0.7468
S&P_500	0.7171	0.7105	**0.8590**	0.8527	0.7683
TATASTEEL	0.7385	0.8868	0.8949	**0.8963**	0.8173
Mean	0.6761	0.7695	**0.8687**	0.8655	0.7447

TABLE VI. SPECIFICITY MEASURE OF THE LEARNING ALGORITHMS ON THE TEST DATA SETS

Data set	DT	GNB	LR	SVM	KNN
XOM	0.7642	**0.9151**	0.8649	0.9057	0.8415
ABT	0.6427	**0.8782**	0.8623	0.8563	0.7625
BAC	0.7774	**0.8321**	0.8189	0.8076	0.7472
AAPL	0.6302	**0.9344**	0.8688	0.8549	0.6422
HCLTECH	0.7621	0.7886	**0.8723**	0.8480	0.6806
HPCL	0.6729	0.3892	**0.8364**	0.8240	0.7743
MSFT	0.6660	**0.9814**	0.9258	0.9278	0.8536
KMX	0.6814	0.8810	**0.8964**	**0.8964**	0.7716
S&P_500	0.7342	**0.9652**	0.9182	0.9346	0.7955
TATASTEEL	0.7152	0.8935	**0.9065**	0.9044	0.8087
Mean	0.7046	0.8459	**0.8771**	0.8760	0.7678

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

TABLE VII. AUC OF THE TREE-BASED ENSEMBLE ML MODELS ON THE TEST DATA SETS

Data set	DT	GNB	LR	SVM	KNN
XOM	0.7203	0.9407	**0.9470**	0.9464	0.8728
ABT	0.6176	0.9269	**0.9523**	0.9508	0.7981
BAC	0.7614	0.9335	**0.9385**	0.9381	0.8610
AAPL	0.6279	0.8217	0.9474	**0.9486**	0.7797
HCLTECH	0.7181	0.8668	**0.9505**	0.9495	0.8205
HPCL	0.5963	0.7442	0.9311	**0.9410**	0.7637
MSFT	0.6254	0.6588	0.9259	**0.9268**	0.7191
KMX	0.6546	0.9309	**0.9527**	0.9522	0.8130
S&P_500	0.7080	0.8900	**0.9445**	0.9431	0.8316
TATASTEEL	0.7316	0.9571	**0.9679**	0.9665	0.8887
Mean	0.6761	0.8671	0.9458	**0.9463**	0.8148

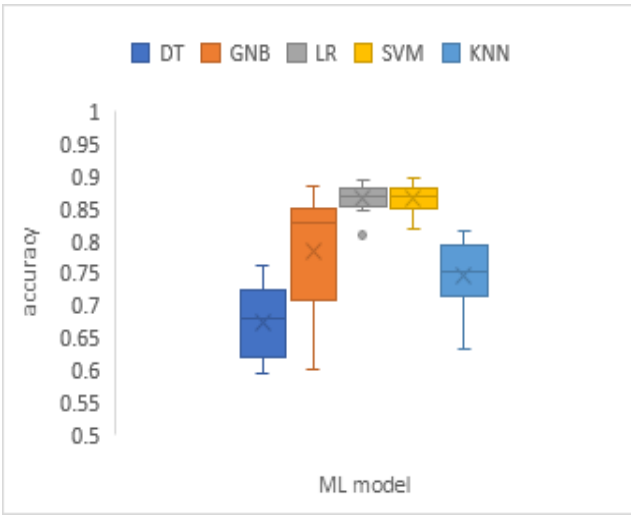

Fig. 1. Boxplot of accuracy values recorded by single ML models on the test data sets

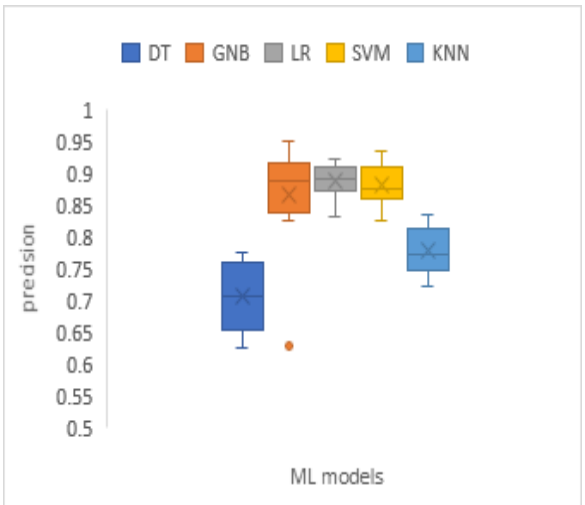

Fig. 2. Boxplot of precision values obtained by single ML models on the test data sets

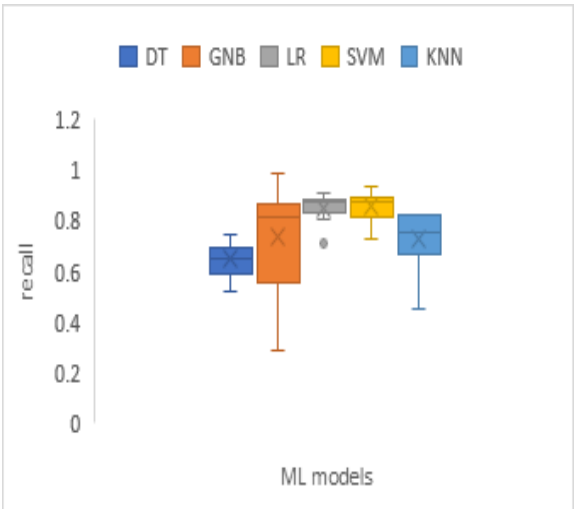

Fig. 3. Boxplot of recall values achieved by single ML models on the test data sets

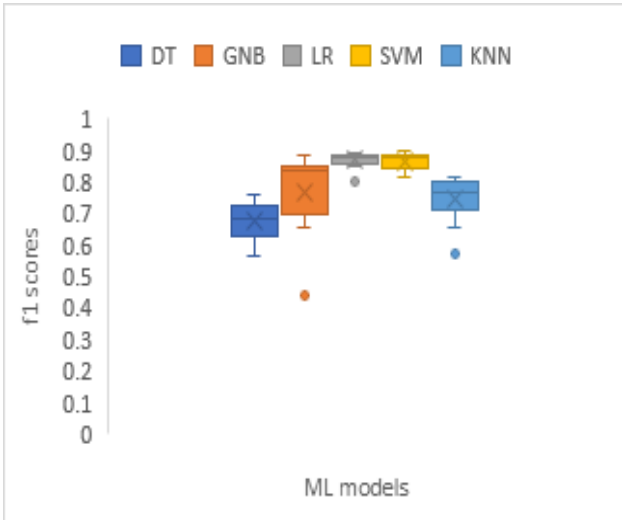

Fig. 4. Boxplot of F1 scores achieved by single ML models on the test data sets

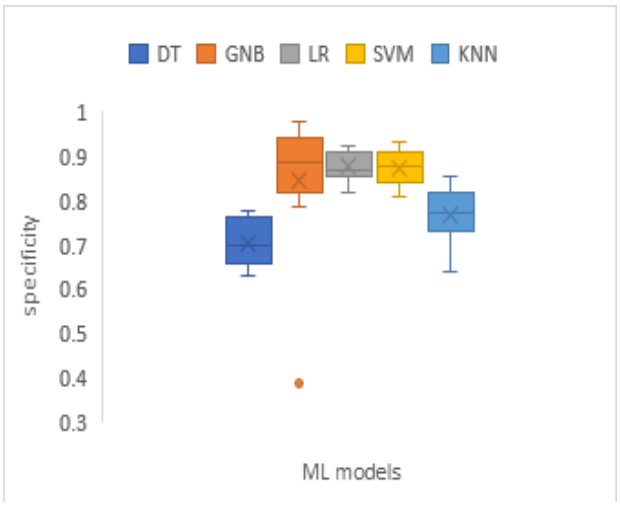

Fig. 5. Boxplot of specificity values recorded by single ML models on the test data sets

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

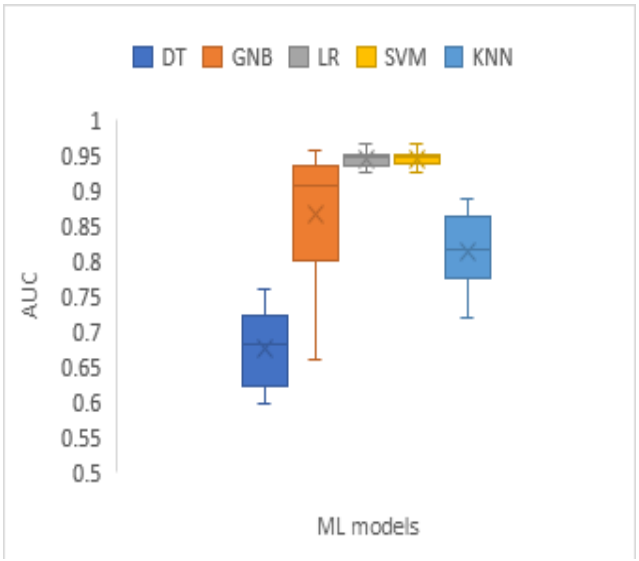

Fig. 6. Boxplot of AUC values achieved by single ML models on the test data sets

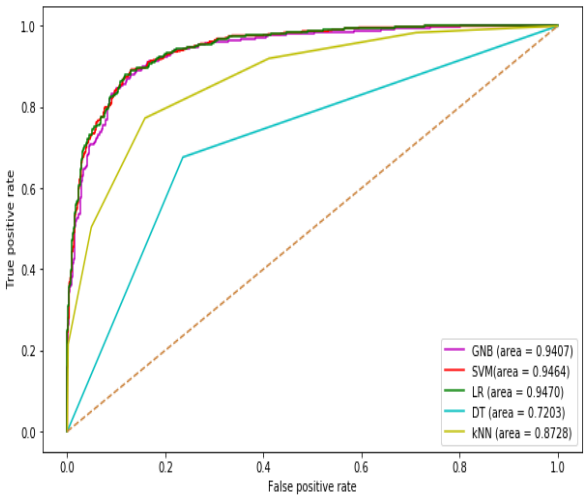

Fig. 7. ROC Curves of the single machine learning models on XOM stock dataset

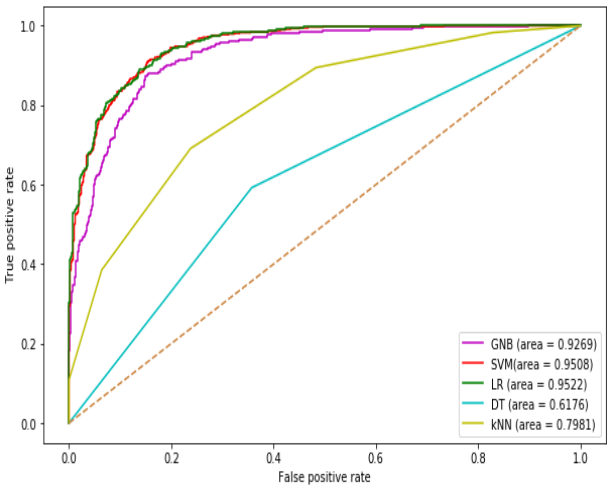

Fig. 8. ROC Curves of the single machine learning models on ABT stock dataset

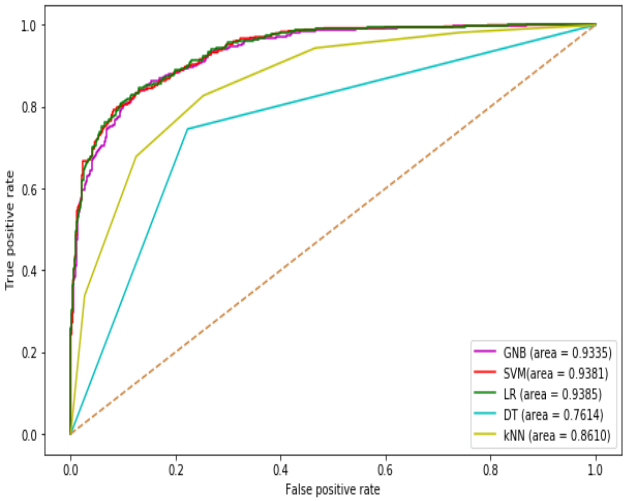

Fig. 9. ROC Curves of the single machine learning models on BAC stock dataset

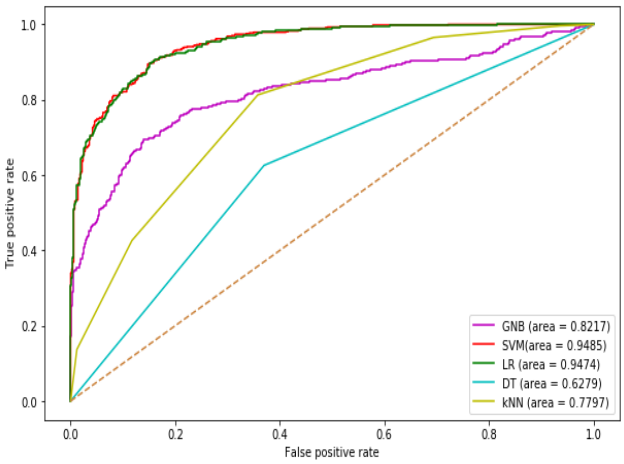

Fig. 10. ROC Curves of the single machine learning models on AAPL stock dataset

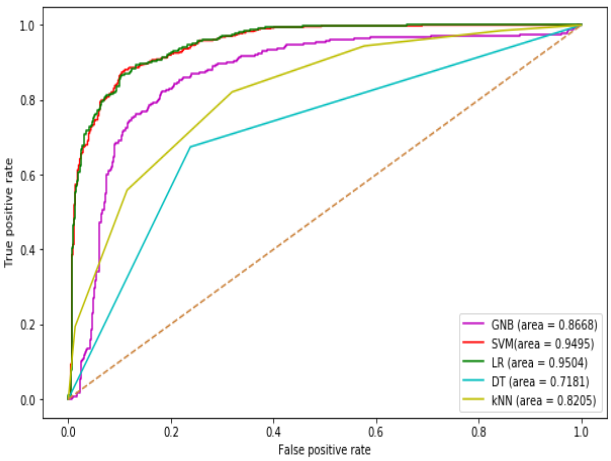

Fig. 11. ROC Curves of the single machine learning models on HCLTECH stock dataset

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

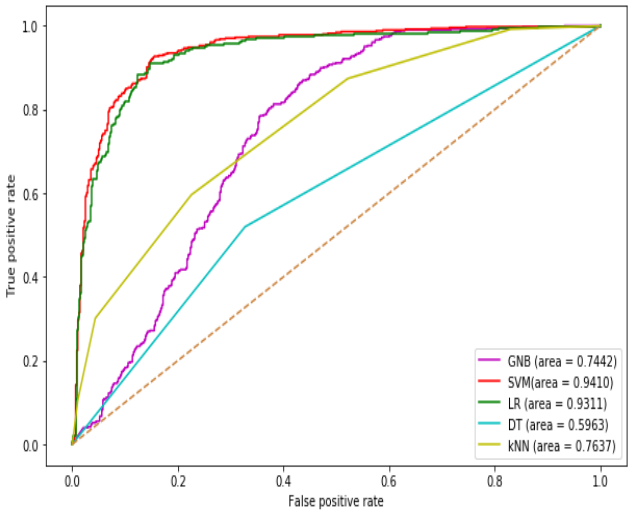

Fig. 12. ROC Curves of the single machine learning models on HPCL stock dataset

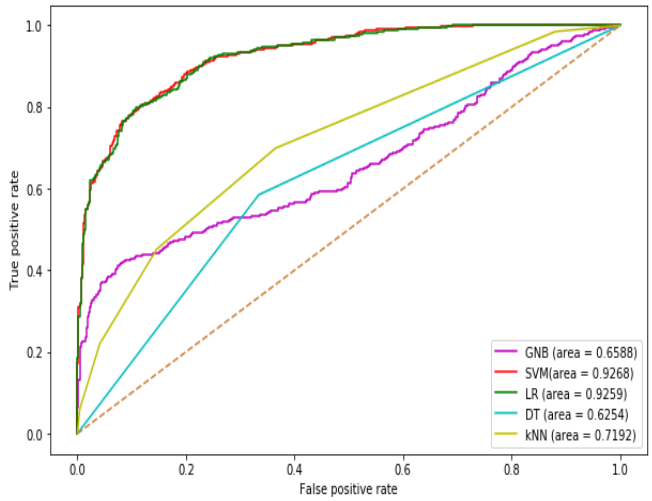

Fig. 13. ROC Curves of the single machine learning models on MSFT stock dataset

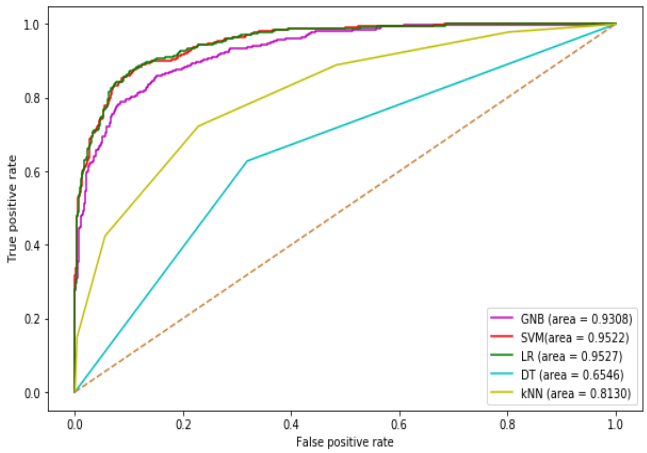

Fig. 14. ROC Curves of the single machine learning models on KMX stock dataset

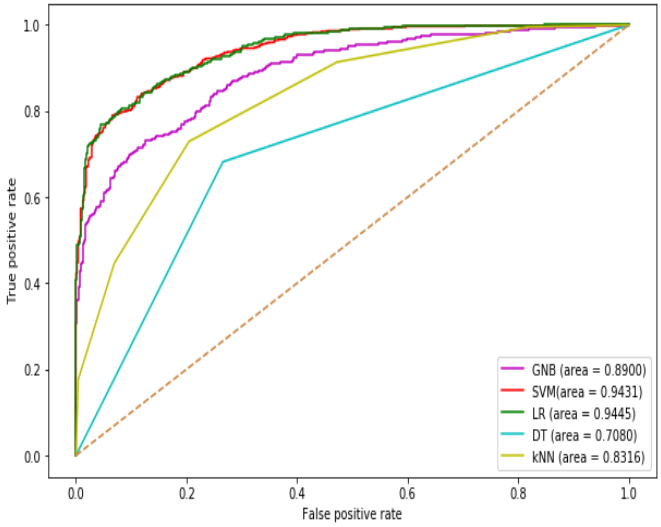

Fig. 15. ROC Curves of the single machine learning models on KMX stock dataset

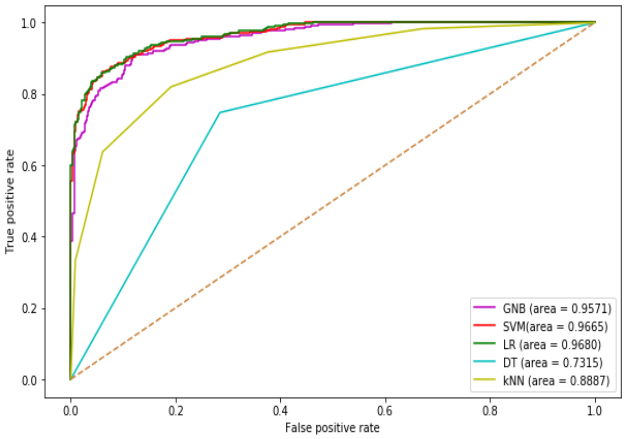

Fig. 16. ROC Curves of the single machine learning models on TATASTEEL stock dataset

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

TABLE VIII. RANKINGS OF THE MACHINE LEARNING MODELS BASED ON KENDALL W TEST RESULTS USING EVALUATION METRICS

Measure	**W**	$\chi 2$	**p**	**Ranks**					
Accuracy	0.808	32.32	0.000	**Technique**	DT	GNB	LR	SVM	kNN
				Mean Rank	1.1	2.8	**4.5**	4.3	2.3
Precision	0.746	29.84	0.000	**Technique**	DT	GNB	**LR**	SVM	kNN
				Mean Rank	1.2	4.1	**4.2**	3.6	1.9
Recall	0.7285	29.14	0.000	**Technique**	DT	GNB	**LR**	SVM	kNN
				Mean Rank	1.4	2.6	**4.4**	4.5	2.2
F1	0.7100	28.40	0.000	**Technique**	DT	GNB	**LR**	SVM	kNN
				Mean Rank	1.4	2.8	**4.5**	4.2	2.1
Specificity	0.677	27.06	0.000	**Technique**	DT	GNB	**LR**	SVM	kNN
				Mean Rank	1.3	4.0	**4.1**	3.8	1.9
AUC	0.926	37.04	0.000	**Technique**	DT	GNB	**LR**	SVM	kNN
				Mean Rank	1.0	2.8	**4.7**	4.3	2.2

IV. CONCLUSION

In this study, we have presented performance of five single machine learning models in predicting the direction of movement of stock prices by applying the models to ten different stock data sets from three stock markets. The machines learning models are evaluated using six classical evaluation metrics, and these models are ranked using Kendall's coefficient of concordance (W). In the experiment, each data set is split into training and test sets. The models are fit on the training data set, and the evaluation is done using the test data set. The experimental results recorded illustrate that LR model is the highest rank when using Kendall's coefficient of concordance (W) with accuracy, precision, f1-score, specificity, and AUC metrics. Also, SVM model achieved the best rank when using the Kendall's coefficient of concordance (W) with recall. This study is limited to only single machine learning models without considering ensemble machine learning models, Hence, in our subsequent work, we will incorporate ensemble machine learning models.

Funding: This work was supported by the NSFC-Guangdong Joint Fund (Grant No. U1401257), National Natural Science Foundation of China (Grant Nos. 61300090, 61133016, and 61272527), science and technology plan projects in Sichuan Province (Grant No. 2014JY0172) and the opening project of Guangdong Provincial Key Laboratory of Electronic Information Products Reliability Technology (Grant No. 2013A061401003).

APPENDIX

TABLE IX. VOLUME INDICATORS USED IN THE STUDY

Volume Indicator
Chaikin A/D Line (ADL)
Chaikin A/D Oscillator (ADOSC)
On Balance Volume (OBV)

TABLE X. PRICE TRANSFORM FUNCTION USED IN THE STUDY

Price Transform Indicator
Median Price (MEDPRICE)
Typical Price (TYPPRICE)
Weighted Close Price (WCLPRICE)

TABLE XI. OVERLAP STUDIES INDICATORS USED IN THE STUDY

Overlap Studies Indicators
Bollinger Bands (BBANDS)
Weighted Moving Average (WMA)
Exponential Moving Average (EMA)
Double Exponential Moving Average (DEMA)
Kaufman Adaptive Moving Average (KAMA)
MESA Adaptive Moving Average (MAMA)
Midpoint Price over period (MIDPRICE)
Parabolic SAR (SAR)
Simple Moving Average (SMA)
Triple Exponential Moving Average (T3)
Triple Exponential Moving Average (TEMA)
Triangular Moving Average (TRIMA)

TABLE XII. MOMENTUM INDICATORS USED IN THE STUDY.

Momentum Indicators
Average Directional Movement Index (ADX)
Average Directional Movement Index Rating (ADXR)
Absolute Price Oscillator (APO)
Aroon
Aroon Oscillator (AROONOSC)
Balance of Power (BOP)
Commodity Channel Index (CCI)
Chande Momentum Oscillator (CMO)
Directional Movement Index (DMI)
Moving Average Convergence /Divergence (MACD)
Money Flow Index (MFI)
Minus Directional Indicator (MINUS_DI)
Momentum (MOM)
Plus Directional Indicator (PLUS_DI)
Log Return
Percentage Price Oscillator (PPO)
Rate of change (ROC)
Relative Strength Index (RSI)
Stochastic (STOCH)
Stochastic Relative Strength Index (STOCHRSI)
Ultimate Oscillator (ULTOSC)
Williams' %R (WILLR)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 09, September-2020

REFERENCES

[1] R. Al-Hmouz, W. Pedrycz, A. Balamash, A. Description and prediction of time series: A general framework of granular computing, Expert Systems with Applications, vol. 42, pp. 4830–4839, 2015.
[2] A. Booth, E. Gerding, F. McGroarty, Automated trading with performance weighted random forests and seasonality. Expert Systems with Applications, vol. 41, pp. 3651–3661, 2014.
[3] M. Kumar, M. Thenmozhi, Forecasting Stock index movement: A comparison of support vector machines and random forest, SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, January 24, 2006.
[4] Y. S. Abu-Mostafa, A. F. Atiya, Introduction to financial forecasting, Applied. Intelligence, vol. 6, 205–213, 1996.
[5] G. S. Atsalakis, K. P. Valavanis, Surveying stock market forecasting techniques part ii: soft computing methods, Expert Systems with Applications, vol. 36, no.3, pp. 5932–5941, 2009.
[6] P. Meesad, R. I. Rasel, Predicting stock market price using support vector regression, In Informatics, electronics & vision (iciev), 2013 international conference on (pp. 1–6). IEEE, 2013.
[7] Y. Zhang, L. Wu, Stock market prediction of s&p 500 via combination of improved bco approach and bp neural network, Expert Systems with Applications, vol. 36, no.5, pp. 8849–8854, 2009.
[8] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, European Journal of Operational Research, vol. 270, pp. 654–669, 2018.
[9] E. K. Ampomah, Z. Qin, G. Nyame, Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement, Information, vol. 11, 2020.
[10] M. Vijh, D. Chandola, V. A. Tikkiwal. A. Kumar, Stock Closing Price Prediction using Machine Learning Techniques, Procedia Comput. Sci. vol. 167, pp. 599–606, 2020.
[11] L. Rokach, O. Maimon, Data mining with decision trees: theory and applications, World Scientific, Singapore. 2008.
[12] R. D. S. Raizada, Y. S. Lee, Smoothness without Smoothing: Why Gaussian Naive Bayes Is Not Naive for Multi-Subject Searchlight Studies, PLoS ONE, vol. 8, no. 7, 2013.
[13] C. Cortes, V. N. Vapnik, Support vector networks. Machine Learning, vol. 20, pp. 1–25, 1995.
[14] J. Cao, M. Wang, Y. Li, Q. Zhang, Improved support vector machine classification algorithm based on adaptive feature weight updating in the Hadoop cluster environment. PLoS ONE, vol. 14, no. 4, 2019.
[15] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, Logistic regression. New York: Springer-Verlag, 2002.
[16] L. Hu, M. Huang, S. Ke, C. Tsai, The distance function effect on k-nearest neighbor classification for medical datasets. SpringerPlus 5, 2016.
[17] J. Wu, X. Chen, H. Zhang, L. Xiong, H. Lei, S. Deng, Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. Journal of Electronic, Science and Technology, vol. 17, no. 1, pp.26-40, 2019.
[18] J. S. Bergstra, R. Bardenet, Y. Bengio, B. Kęgl, Algorithms for hyperparameter optimization, 2011. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F.
[19] M. G. Kendall, S. B. Babington, The Problem of m Rankings, The Annals of Mathematical Statistics, vol. 10, pp. 275-287, 1939.
[20] K. Kim, Financial time series forecasting using support vector machines, Neurocomputing, vol. 55, pp. 307–319, 2003.
[21] X. Lin, Z. Yang, Y. Song, Short-term stock price prediction based on echo state networks, Expert Systems with Applications, vol. 36, no. 3, 7313–7317, 2009.
[22] C. F. Tsai, Y. C. Hsiao, Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches, Decision Support Systems, vol. 50, no.1, pp. 258–269, 2010.