Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 10 Issue: 05 | Oct-2021

Effectiveness of Artificial Intelligence in Stock Market Prediction Based on Machine Learning

[1] Kanchan Singh

[1] Kkwagh Engineering , Nashik

---***---

Abstract - This paper tries to address the problem of stock request vaticination using artificial intelligence (AI) strategies. The stock request vaticination can be modeled grounded on two top analyses called specialized and abecedarian. In the specialized analysis approach, the retrogression machine literacy (ML) algorithms are employed to prognosticate the stock price trend at the end of a business day grounded on the literal price data. In discrepancy, in the abecedarian analysis, the bracket ML algorithms are applied to classify the public sentiment grounded on news and social media. In the specialized analysis, the literal price data is exploited from Yahoo Finance, and in abecedarian analysis, public tweets on Twitter associated with the stock request are delved to assess the impact of sentiments on the stock request's cast. The results show a median performance, inferring that with the current technology of AI, it's too soon to claim AI can beat the stock requests.

***Key Words*:** Machine literacy, time series vaticination, specialized analysis, sentiment embedding, fiscal request

1. INTRODUCTION

This Stock requests are always an seductive investment way to grow capital. With the development of communication technology, the stock requests are getting further popular among individual investors in recent decades. While time by time, the number of shareholders and companies is growing in the stock requests, numerous try to find a result to prognosticate a stock request's unborn trend. This is a grueling problem with a multitude of complex factors that are impacting the price changes. Then, vaticination algorithms similar as Kalman sludge (1) and optimization styles similar as Nash equilibrium (2) can be helpful; but, for this specific problem, AI can play a significant part. For this, ML styles are developed in numerous exploration papers to estimate the vaticination power of AI in the stock requests. The ML algorithms that are enforced for this purpose substantially try to figure out patterns of data, measure the investment threat, or prognosticate the investment future.

This field's sweats have led to two central theoretical suppositions Effective Request Thesis (EHM) and Adaptive Request

Thesis (AMH). The EMH (3) claims that the spot request price is eventually a response to lately published news aggregation. Since news vaticination is an impracticable miracle, request prices are always following an changeable trend. This thesis implies that there's no possible result to' beat the request.'On the other hand, the AMH (4) is trying to find a correlation between the evidential EMH and the extravagant behavioral finance principles. Behavioral finance tries to describe the request trend by psychology- grounded propositions. Regarding the AMH, investors can work the request effectiveness weakness to gain profit from share trading.

Counting on the AMH statement, there should be possible results to prognosticate the future of request geste. Considering this fact, along with the Dow proposition (5), leads to the creation of two introductory stock request analysis principles abecedarian and specialized. Abecedarian analysis tries to probe a stock's natural value by assessing affiliated factors similar as the balance distance,micro-economic pointers, and consumer geste. Whenever the stock value reckoned by this strategy is advanced/ lower than the request price, investors are attracted to buy/ vend it. On the other hand, the specialized analysis only examines the stock's price history and makes the trading opinions grounded on the fine pointers exploited from the stock price. These pointers include relative strength indicator (RSI), moving average confluence/ divergence (MACD), and plutocrat inflow indicator (MFI) (6).

Decades agone, the proposed request analysis was performed by fiscal judges; but through the development of calculating power and artificial intelligence, this process could also be done by data scientists. Currently, the power of ML strategies in addressing the stock request vaticination problem is strengthening fleetly in both abecedarian and specialized analyses. In an early study using ML for stock request vaticination, Piotroski etal.

(7) introduced an ML model called F- Score to estimate companies' factual share values. Their system was grounded on nine factors exploited by a company's fiscal reports divided into three main orders profitability, liquidity, and operating effectiveness. They enforced the F- Score algorithm on the literal companies' fiscal reports of theU.S. stock request for twenty times from 1976 to 1996 and presented remarkable issues. Some times latterly, Mohanram etal. (8) proposed a developed ML

algorithm named G- Score to decide on trading stocks. Their approach was grounded on abecedarian analysis applying fiscal reports to estimate three criteria profitability, naive extrapolation, and counting traditionalism. They also showed

their algorithm adequacy by back- testing theU.S. stock request trend between 1978 and 2001.

Principally, in the abecedarian analysis, unlike the specialized analysis, the data is unshaped and hard to be reused for training an ML model. Nonetheless, numerous studies by using this type of analysis proved it can lead to a rational vaticination of the request price. Conversely, to dissect the request grounded on a specialized system, only literal price data is needed. This data is a structured type of data that's directly available to the public. This redounded in a far advanced volume of exploration papers studying the vaticination of the stock request grounded on the specialized analysis approaches. As one of the early studies in this field, at the morning of the 90s, Kimoto etal. (9) worked on a feed-forward neural network (NN) algorithm (10) to prognosticate the stock request exploiting the literal fiscal pointers similar as interest rate, and foreign exchange rate. Their model was a decision- making tool in generating the signal of buying/ selling shares. Although their model could be successful in the steal-and- hold strategy, it couldn't prognosticate the signal of dealing sufficiently. Thus, other ML algorithms were examined to assess the vaticination power of ML using specialized data similar as artificial neural network (ANN), arbitrary timber (RF), support vector machine (SVM), and naive Bayesian. In (11), Patel etal. enforced four distinct ML algorithms on this problem, including artificial neural network (ANN), arbitrary timber (RF), support vector machine (SVM), and naive Bayesian. Ten- time- period testbed results clarified that the arbitrary timber algorithm can be more effective among other algorithms, especially while the input data is discretized. Also, in a veritably recent study, Zhong etal. (12) studied a comprehensive big data logical procedure applying ML to prognosticate a diurnal return in the stock request. They employed both deep neural network (DNN) and ANN to fit and prognosticate sixty fiscal input features of the model. They concluded that the ANN performs better than the DNN, and applying principle element analysis (PCA) in thepre-processing step can ameliorate vaticination delicacy.

This paper attempts to probe the effectiveness of AI,

. particularly ML, in addressing stock request vaticination. In this exploration, both specialized and abecedarian stock request analyses are applied to measure ML algorithms' delicacy in prognosticating request trends. Also, a multitude of ML algorithms similar as logistic retrogression, k-nearest neighbor, arbitrary timber, decision tree, and ANN are employed to find the most accurate algorithms to be named as a result to this problem. The data used to induce ML models are acquired in real- time, and the purpose of this exploration is to estimate the delicacy of a selling/ buying/ holding signal of a specific share for investors.

The remainder of the paper is organized as follows Section II is devoted to the problem description and provocations; Section III contains methodology; Section IV measures the performance of ML algorithms on a stock request testbed; Section V points out the conclusion and unborn workshop.

2. PROBLEM DESCRIPTION AND MOTIVATION

The Since the 90s, early studies tried to prognosticate the stock requests using AI strategies. Numerous exploration studies are published to estimate the performance of AI approaches in the stock request vaticination. Researchers' enthusiasm in studying the stock request vaticination problem is due to the tremendous diurnal volume of traded plutocrat in the stock requests.

Generally, assaying the stock requests is grounded on two primary strategies specialized analysis and abecedarian analysis. In the first one, stockholders try to estimate the stock requests regarding the literal price data and probing the generated pointers exploited from this data, similar as the RSI and the MACD. An ML model can do the same. It can be trained to find a logical pattern between the fiscal pointers and the stock's ending price. This can lead to a vaticination model that estimates the stock price at the end of a business day. On the other hand, in the abecedarian analysis, stockholders essay to calculate an factual stock value grounded on its proprietor company's fiscal reports, similar as the request cap or thedividends.However, the stockholders admit a selling signal, while if the estimated price is lower than the stock price, If the estimated price value is advanced than the stock price. It's apparent that any changes in a company's fiscal report can incontinently affect the public sentiment on the news and social media. An ML model can probe news and social media through the Internet to prognosticate a positive/ negative impact of the stock prices' abecedarian pointers. Also, give the action signal for the stockholders grounded on the public sentiment. But, the question is how important these approaches can be effective in the vaticination of the request. In other words," Can AI beat the stock request?". This study is trying to employ the ML algorithms and estimate their performance in prognosticating the stock request to answer this question. The retrogression models are employed to prognosticate the stock ending prices, and the bracket models are used to prognosticate the action signal for stockholders. In the following section, the methodology for addressing this evaluation is explained.

3. METHODOLOGY

In this study, the stock request's vaticination, using ML tools, includes four main way dataset structure, data engineering, model training, and vaticination. This section is devoted to explaining each of these way in detail.

3.1 Dataset

The first step of erecting an ML model is having access to a dataset. This dataset includes some features that train the ML model. The training procedure can be done with or without a set of labeled data called targetvalues.However, the training procedure is called supervised literacy; while, unsupervised literacy doesn't need any target values and tries to find the retired patterns in the training dataset, If the training is grounded on a set of labeled data.

In the problem of prognosticating the stock request, utmost datasets are labeled. For case, the dataset includes some fiscal pointers similar as RSI and MACD as features and the stock's ending price as the target value in the specialized analysis approach. It's apparent that the data associated with the specialized analysis is nonstop figures, which is shown in time-series format data. On the other hand, in the abecedarian analysis strategy, the features are some statements similar as fiscal reports or investors' sentiments, and the target value is the signal of decision- making in buying/ dealing the stock. In this type of analysis, the data includes generally alphabetic inputs similar as reports and sentiments.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 10 Issue: 05 | Oct-2021

Hopefully, utmost of the essential data needed for this problem is available online similar as literal stock prices or public sentiments in the news. The data employed in this study was acquired from two sources specialized analysis data available on Y ahoo Finance1, and sentiment analysis data available on Twitter2. The Yahoo Finance data includes the open, close, medial, high, low price, and volume values without missing samples; while, the Twitter dataset contains tweets from the public comprising news agencies and individualities that can have missing samples.

Data Engineering

The data attained from the proposed datasets requires to bepre-processed before being exploited in model training. Either specialized analysis or abecedarian analysis has several pointers applied in the model training step, and the most significant bones are explained in the following.

3.2.1 Specialized Analysis. The literal stock prices are used to calculate applicable fiscal pointers similar as simple moving normal (SMA), exponential moving normal (EMA), RSI, MACD, and on- balance- volume (OBV) to make the input features of an ML training model. These pointers are explained in the following. SMA. This index is the normal of the most recent ending prices of a stock in a particular period. The fine computation of the SMA is shown as below

$$SMA(t, N) = \sum_{k=1}^{N} \frac{CP(t-k)}{N} \quad (1)$$

where CP is the ending price, N indicates the number of days. that the CP is estimated, and k shows the days associated with a particular CP.
EMA. This index tracks a stock price the same as the SMA,. but it pays further attention to the recent ending prices by weighting
them. Equation 2 indicates the weighting process of this index

$$EMA(t, \Delta) = (CP(t) - EMA(t-1)) * \Gamma + EMA(t-1)$$

$$\Gamma = \frac{2}{\Delta+1} \qquad \Delta = \text{Time period EMA}$$

MACD. This indicator tries to compare the short-term and the long-term trends of a stock price. Equation 3 describes this indicator as follow:

$$MACD = EMA(t, k) - EMA(t, d) \quad (3)$$

where where k and d are the ages of short- term and long-term trends. Typically, these values are considered as k = 12 and d = 16 days. OBV. This index uses a stock volume inflow to show the price trend and indicates whether this volume is flowing in or out. The following equation explains the OBV conception

$$OBV = OBV_{pr} + \begin{cases} \text{volume}, & \text{if } CP > CP_{pr} \\ 0, & \text{if } CP = CP_{pr} \\ -\text{volume}, & \text{if } CP < CP_{pr} \end{cases} \quad (4)$$

where OBVpr is the previous OBV, volume is the latest trading volume amount, and CPpr is the previous closing price.
RSI. This indicator is measuring the oversold or the overbought characteristic of a stock. Indeed, it shows the trend of buying/selling a stock. The RSI is described as:

$$RSI = \frac{100}{1 + RS(t)}, \quad RS(t) = \frac{AvgGain(t)}{AvgLoss(t)} \quad (5)$$

where RS(t) shows the rate of profitability of stock, AvgGain(t) is the average gained profit of stock at time t, and AvgLoss(t) indicates the average loss on that price.

3.2.2 Fundamental Analysis. Due to the unshaped nature of abecedarian pointers, rooting data for abecedarian analysis isn't easy. Still, the development of AI makes it possible to exploit data from the Internet for this purpose, leading to a more accurate stock request vaticination. This data can be information related to the fiscal report of a company or the sentiment of investors. Literally, companies' fiscal reports incontinently impact public sentiment and present themselves on social media, particularly Twitter. Therefore, one way of assessing the impact of abecedarian data on request trends is by looking at public tweets. This strategy is called sentiment analysis of the stock request. In the sentiment analysis, the input data for training a model is principally unshaped, imported as textbook format to the model. The target of abecedarian datasets is a double value indicating the textbook's positive/ negative impact on a specific stock. Either, grounded on the types of data, thepre-processing step differs. In the specialized analysis, due to the data's numeric nature, it's essential to homogenize the data before employing them for model training. The data normalization step is significant when the ML model wants to find a logical pattern in the inputdata.However, the vaticination process would not directly perform, If the data aren't on the same scale. Therefore, numerous functions are applied to homogenize the data, similar as Min Max Scaler, Standard Scaler, and Robust Scaler. In this paper, Min Max Scaler is used to gauge the data and is described as below

$$a^{m}_{scaled} = \frac{(a^{m} - a_{min})}{(a_{max} - a_{min})}$$

where am is the ith point (index) from mth trial (time sample), amin and amax are the minimum and the outside values of the point among the trials, independently. Also m gauged

indicates the gauged value for the ith point of mth trial. On the other hand, in the fundamental analysis, the data is not numeric. The thing is to probe the impact of a judgment – that can be a tweet on Twitter – on public sentiment. Whenever usingnon-numerical data in training an ML model, the input data should be restated into numeric data. Therefore, one way to do so is data labeling.
Point selection means chancing the most precious features that lead to a more accurate ML model in a smaller calculation

time. This fashion can be classified as a sludge, wrapper, bedded, and mongrel styles (13). In the sludge system, correlation criterion plays a significant part. Correlation is a measure of the direct relationship between two or further parameters. In this system, features showing the most correlation with the target are named to make the model. Likewise, to avoid spare calculation, the named features shouldn't be largely identified to each other. To do so, the Pearson correlation fashion is one of the most useful styles, which is described as below

$$Corr(i) = \sqrt{\frac{cov(a_i\ b)}{var(a\ ) * var(b)}} \quad (7)$$

where ai is the ith point, b is the target marker, cov () and var (). represent the covariance and the friction functions, independently. The reused data could be employed to train the ML model, as shown inFig. 1.

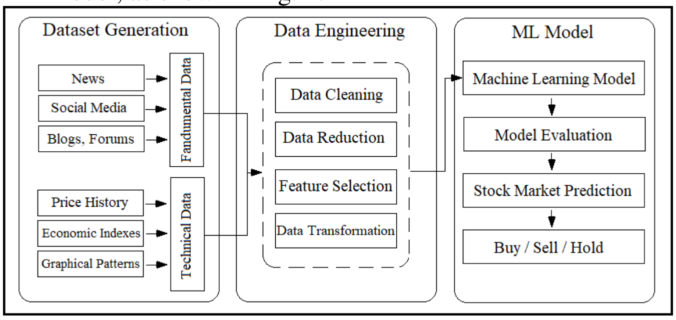


Fig. 1: The framework of model training to predict the stock market

.

3.3. Machine Learning Model Training

Numerous ML algorithms have been employed to prognosticate stock requests in exploration studies. Principally, there are two main orders of models to address this problem bracket models that try to help the investors in the decision- making process of buying, selling, or holding stock, and retrogression models that essay to prognosticate stock price movements similar as the ending price of a stock. In exploration studies, over 90 of the algorithms abused in prognosticating the stock request are bracket models (14). Still, many studies tried to prognosticate the exact stock prices using the retrogression models (15, 16, 17).

Among ML algorithms, the decision tree (DT), support vector machine (SVM), and artificial neural networks (ANN) are the most popular bones employed to prognosticate stock requests (18). In this study, besides using the ANN, DT, and SVM models, logistic retrogression (LR), Gaussian naive Bayes (GNB), Bernoulli Naive Bayes (BNB), arbitrary timber (RF), k-nearest neighbor (KNN), and XGboost (XGB) are employed for bracket strategy; also, direct retrogression and long short-term memory (LSTM) are used in retrogression problems. In the following, these algorithms are compactly explained.

ANN. Firstly came from the generalities in biology and comported of colorful processing rudiments called neurons. Theseinter-connected neurons' task is majorly casting up the values of input parameters corresponding to their specified weights and also adding a bias. The number of neurons in the input should equal the number of neurons in the affair. In the end, the affair values are calculated after the transfer function is applied.

DT. Decision tree owns a structure analogous to a tree, where each branch represents the test outgrowth, and each splint indicates a class marker. The structure also includes internal bumps, which represent the test on a particular trait. The outgrowth is a final decision that provides the stylish fitting of calculated attributes of the stylish class. SVM. In the SVM model, exemplifications are counterplotted as separated points in the space as vast as possible concerning each other. Hence, the prognosticated exemplifications are also counterplotted to the same space and also distributed.

LR. Logistic Retrogression algorithm is one of the most suitable algorithms in retrogression analysis, especially when the dependent variable is double, where a logistic function is abused for modeling.

GNB, and BNB. Gaussian Naive Bayes and Bernoulli Naive Bayes are considered supervised literacy algorithms which are simple but veritably functional. Gaussian Naive Bayes includes previous and posterior chances of the dataset classes, while Bernoulli Naive Bayes only applies to data with double-valued variables.

RF. Random timber algorithm includes a series of decision trees whose ideal is to induce an uncorrelated group of trees whose vaticination is more accurate than any single tree in the group.

KNN. The KNN is a well- known algorithm for bracket problems, in which test data is used to determine what an unclassified point should be classified as. Manhattan distance and Euclidean distance are the styles that are used in this algorithm to measure the distance of the unclassified point to its analogous points. XGB. A popular and open- source interpretation of the grade boosted trees algorithm, XGBoost is a supervised literacy algorithm for the accurate vaticination of an aimed variable grounded on its simpler and weaker models estimation.

Linear Retrogression. A subset of supervised literacy, Linear Retrogression, is principally a first- order vaticination,e.g., a line or a aeroplane that stylish fits the dataset's data points. Any new point as the vaticination will be located on that line or aeroplane.

LSTM. Unlike standard feed-forward neural networks, the Long Term Short Memory algorithm owns feedback connections and is employed in deep literacy. This algorithm is extensively used to classify problems and make prognostications grounded on data in the time sphere. All the proposed algorithms are used to perform a stock request vaticination, and their performance is compared to estimate the adequacy of ML in this problem. The following subsection explains the criteria that are applied in the comparison procedure.

3.2 Model Evaluation Metrics

All vaticination models bear some evaluation criteria to investigates their delicacy in the vaticination procedure. In ML algorithms, a multitude of criteria are available to measure the models' performance, including confusion matrix, and receiver driver characteristic (ROC) wind for bracket models; and R- squared, explanation variation, mean absolute chance error (MAPE), root mean squared error (RMSE), and mean absolute error (MAE) for retrogression (19). The rest of this subsection is devoted to explaining the conception of these criteria.

3.4.1 Confusion matrix. This measure evaluates the delicacy of an ML model using apre-known set of targeted data. Also, some other criteria, including perceptivity, particularity, perfection, and F1- score, are redounded regarding this matrix. The perceptivity or recall is the liability of prognosticating true positive, while the particularity shows the true negative rate. Also, the perfection indicates the delicacy of the true positive prognosticated classes. The F1-Score computes the balance between perceptivity and perfection. Eventually, the delicacy of the model would be the evaluation of the true prognosticated classes. Figure 2 shows the confusion matrix conception.

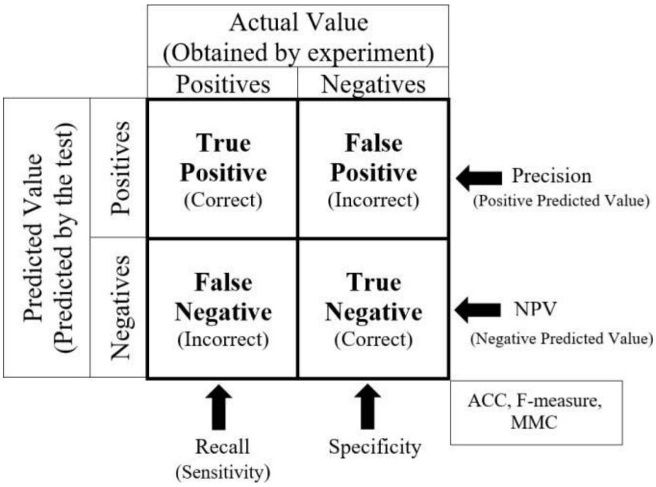


Fig. 2: Confusion matrix explanation.

3.4.2 ROC, AUC. The receiver driver characteristic (ROC) wind includes two values true-positive and false-positive rates. The ROC investigates the classifiers' performance among the whole range of class distributions and error costs. ROC angles are compared by the area under the wind (AUC) metric. The further values of AUC mention more accurate prognosticated labors (20).

3.4.3 R-squared (R2). The R2 is a statistical measure indicating the friction portion for a dependent variable that's explained by an independent variable or variables in a retrogression model. It's also known as the measure of determination or the measure

of multiple determination for multiple retrogression. Using retrogression analysis, advanced R2 is always better to explain changes in your outgrowth variable. If the R- squared value is lower than0.3, this value is generally considered a fragile effect size; if the R- squared value is between0.3 and0.5, this value is generally considered a low effect size; if the R-squared value is bigger than0.7, this value is generally considered strong effect size. The following equation presents the formula for calculating the R2 metric.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (8)$$

where yi, and yˆi are the ith factual and awaited value, independently, and y ¯ shows the mean of factual values.

3.4.4 Explanation Variation. The explained friction is used to measure the distinction between a model and factual data. In other words, it's the part of the model's total friction that's explained by factors that are actually present and aren't due to error friction. The explained variation is the sum of the squared of the differences between each prognosticated value and the mean of factual values. Equation (9) shows the conception of explanation variation as below

$$EV = \sum (\hat{y}_i - \bar{y})^2 \quad (9)$$

Where EV is the explanation variation, yˆi is the prognosticated value, and y ¯ indicates the mean of factual values.

3.4.5 MAPE. The MAPE is how far the model's prognostications are out from their corresponding labors on average. The MAPE is asymmetric and reports advanced crimes if the vaticination is further than the factual value and lower crimes when the vaticination is lower than the factual value. Equation (10) explains the fine expression of this metric.

$$\frac{1}{n} \sum_{i=1} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (10)$$

where n is the number of trials, yˆi is the prognosticated value, and yi is the factual value for the ith trial.

3.4.6 RMSE. The reckoned standard divagation for vaticination crimes in an ML model is called RMSE. The vaticination error or residual shows how far are the data from the retrogression line. Indeed, RMSE is a measure of how spread out these residuals are (21). In other words, it shows how concentrated the data is around the line of stylish fit, as shown in Equation (11). The lower value of this metric represents a better vaticination of the model.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \quad (11)$$

where n is the number of trials, yˆi is the prognosticated value, and yi is the factual value for the ith trial.

3.4.7 MAE. The MAE is the sum of absolute differences between the target and the prognosticated variables. Therefore, it evaluates the average magnitude of crimes in a set of prognostications without considering their directions.

The lower values of this metric mean a better vaticination model. The following equation presents the fine MAE formula.

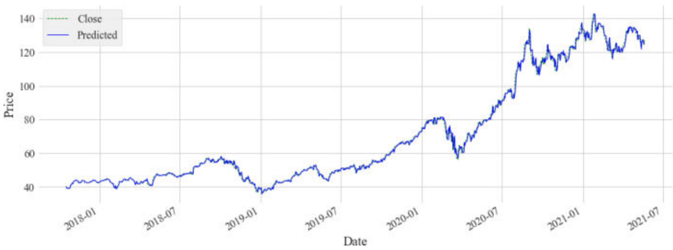
$$MAE = \frac{1}{n}\sum_{i=1}(y_i - \hat{y}_i) \qquad (12)$$

where n is the number of trials, yˆi is the prognosticated value, and yi is the factual value for the ith trial.

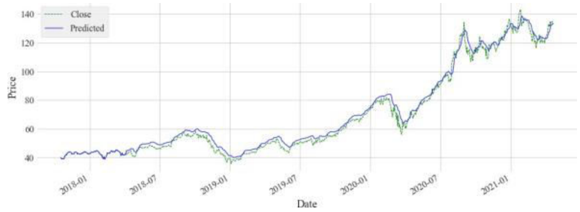
Regarding the proposed frame, the performance of ML algorithms on the vaticination of stock requests can be estimated. The following section tools ML algorithms on the real- life problem of theU.S. stock request vaticination.

4. RESULTS AND DISCUSSION

This section tries to illustrate the performance of the proposed methodology on the vaticination of stock requests. For this, Python software is used to train the ML models and prognosticate unlooked-for



(a) Prediction of the linear regression model.



(b) Prediction of the LSTM model.

Fig. 3: AAPL price prediction with the technical analysis approach.

Table1. : Models performance comparison, in technical analysis approach.

Metric	Linear Regression	LSTM
R^2	1.0	0.99
Explained Variation	1.0	0.99
MAPE	1.56	2.99
RMSE	1.82	3.42
MAE	1.18	2.3

data. First, the market prediction based on the technical analysis is evaluated, and then the fundamental analysis is investigated in this problem.

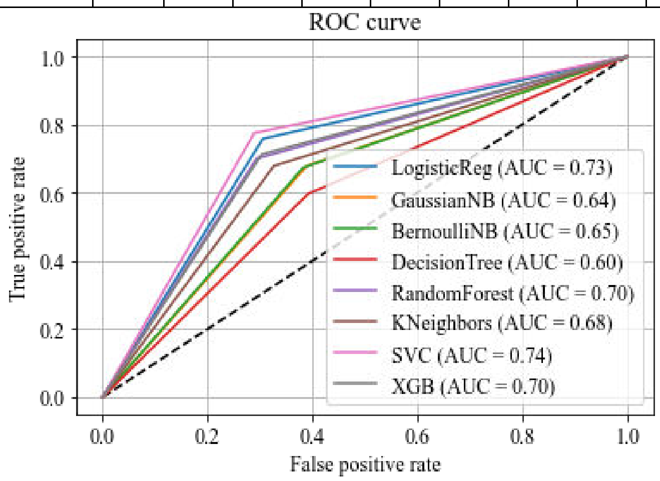
4.1 Technical Analysis Performance

In this paper, the dataset for erecting a predictor model grounded on the specialized analysis is exploited from the" Y ahoo Finance" website. Indeed, it contains the literal data for a well- known stock called AAPL, which indicates Apple company information through a period of further than ten times between 2010 to 2021. The dataset includes 60 features similar as open, high, low prices, the moving normal, MACD, and RSI. The target is the close price, representing the final price of AAPL at the end of a business day. Also, the most correlated features to the target are named, and also the spare features that show a high correlation together are intermingled. Eventually, the data is gauged by the MinMaxScaler function explained in Section 3.

The dataset is divided into three corridor of the training data, confirmation data, and testing data to make the ML model. A large portion of the data is devoted to the training process, and the rest belongs to confirmation and testing. In the training process, the algorithm uses the training data to learn how to prognosticate the target value accessible to the algorithm. Also, the model evaluates the performance of the vaticination regarding the confirmation data. Eventually, it can prognosticate the unlooked-for target of the testing dataset to compare with the true target values. In the end, by using the prognosticated and factual values of the ending price, the evaluation criteria can be measured. Table

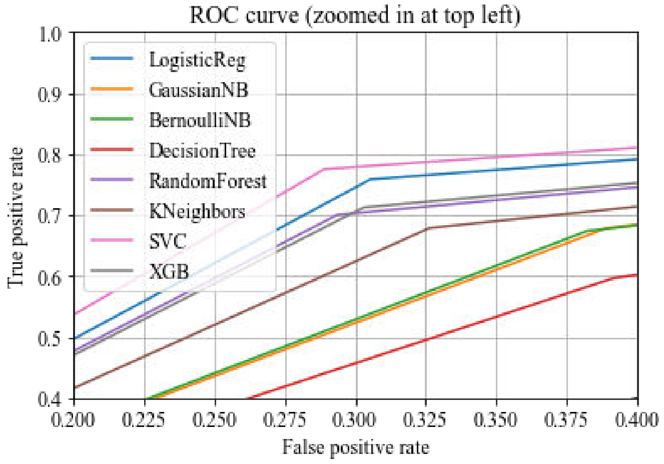
Table 2. : Models performance comparison in fundamental analysis.

Metrics	LR	GNB	BNB	DT	RF	KNN	SVM	XGB	ANN
Precision	0.729	0.636	0.644	0.620	0.727	0.684	0.757	0.710	0.684
Recall	0.727	0.634	0.644	0.620	0.727	0.684	0.755	0.709	0.684
F1-score	0.726	0.632	0.644	0.620	0.727	0.684	0.755	0.709	0.684
Accuracy	0.727	0.634	0.644	0.620	0.727	0.684	**0.755**	0.709	0.684
AUC	0.73	0.63	0.64	0.62	0.73	0.68	0.76	0.71	0.68



(a) ROC curves.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 10 Issue: 05 | Oct-2021



(b) ROC curves from a closer view.
Fig. 4: ROC curves for classification algorithms.

1 shows the comparison of the evaluation criteria. Also,Fig. 3 shows the vaticination of stock price grounded on the LR and the LSTM algorithms. Regarding Table 1, the LR model is far better in prognosticating the AAPL ending price compared to the LSTM model. Also, to illustrate the ending price's prognosticated and factual values,Fig. 3 shows these values since 2018. The solid blue line shows the prognosticated value, and the dashed green line is the factual bone.

4.2 Fundamental Analysis Performance

In this paper, a set of public tweets associated with Apple company is employed to induce the needed dataset available at (22). In this case, the features are textbooks in Twitter, and the target is a double value of impactedsentiment.However, the sentiment value would be 1, while a negative impact would give a-1 value to the sentiment, If the content of the tweet has a positive impact on the stock request. Also, the impacts of tweets on the specific stock are estimated. Eventually, the performance of ML in the vaticination of buying/ selling/ holding signal is delved.
The dataset includes nearly 6000 tweets, and thepre-processing data includes labeling the target values and employing the principle element analysis (PCA) (23) to reduce features dimension that show a high correlation. Also, the proposed algorithms in Section 3 are used to classify the outgrowth of the model by negative or positive sentiment. Grounded on the evaluation criteria explained in the former section, the performance of ML algorithms is compared and showed in Table 2. This table indicates that in this paper, the vaticination of the public sentiment using ML algorithms doesn't show promising results. The most accurate algorithm is the SVM, with an delicacy of 76. Also, the performance of these algorithms is illustrated inFig. 4 that
compares the ROC curves and also shows the AUC for each algorithm. In this figure, the SVM algorithm has the best AUC score.

5. CONCLUSIONS

This study tries to address the problem of stock request vaticination using ML algorithms. To do so, two main orders of stock request analysis (specialized and abecedarian) are considered. The performance of ML algorithms on the cast of the stock request is delved grounded on both of these orders. For this, labeled datasets are used to train the supervised literacy algorithms, and evaluation criteria are employed to examine the delicacy of ML algorithms in the vaticination process. The results show that the direct retrogression model predicts the ending price remarkably with a shallow error value in the specialized analysis. Also, in the abecedarian analysis, the SVM model can prognosticate public sentiment with an delicacy of 76. These results indicate that although AI can prognosticate the stock price trends or public sentiment about the stock requests, its delicacy isn't good enough. Likewise, while the direct retrogression can prognosticate the ending price with a sensible range of error, it can not precisely prognosticate the same value for the coming business day. Therefore, this model isn't sufficient for long- term investments. On the other hand, the delicacy of bracket algorithms in prognosticating buying, selling, or holding a stock isn't satisfying enough and can affect in loss of capital.

Nonetheless, numerous exploration studies on this content are using a mongrel model that employs both the specialized analysis and the abecedarian analysis in one ML model to compensate for the individual algorithms'downsides. This could increase the delicacy in the vaticination process that implies an instigative content for unborn studies. Grounded on this study, it seems that AI isn't close to the vaticination of the stock request with dependable delicacy. Perhaps in the future, with AI development and especially computation power, a more precise model of stock request vaticination can be available. Still, so far, there's no estimable model that can beat the stock request.

REFERENCES

[1] Sohrab Mokhtari and Kang K Yen. "A Novel Bilateral Fuzzy Adaptive Unscented Kalman Filter and its Implementation to Nonlinear Systems with Additive Noise". In: 2020 IEEE Industry Applications Society Annual Meeting. IEEE. 2020, pp. 1–6.
[2] Sohrab Mokhtari and Kang K Yen. "Impact of large-scale wind power penetration on incentive of individual investors, a supply function equilibrium approach". In: Electric Power Systems Research 194 (2021), p. 107014.
[3] Eugene F Fama. "Efficient capital markets: II". In: The journal of finance 46.5 (1991), pp. 1575–1617.
[4] Andrew W Lo. "The adaptive markets hypothesis". In: The Journal of Portfolio Management 30.5 (2004), pp. 15–29.
[5] From Charles D Kirkpatrick II and R Julie. "Dow Theory". In: CMT Level I 2019: An Introduction to Technical Analysis (2019), p. 15.

[6] Robert D Edwards, WHC Bassetti, and John Magee. Technical analysis of stock trends. CRC press, 2007.
[7] Joseph D Piotroski. “Value investing: The use of historical financial statement information to separate winners from losers”. In: Journal of Accounting Research (2000), pp. 1–41.
[8] Partha S Mohanram. “Separating winners from losers among lowbook-to-market stocks using financial statement analysis”. In: Review of accounting studies 10.2-3 (2005), pp. 133–170.
[9] Takashi Kimoto et al. “Stock market prediction system with modular neural networks”. In: 1990 IJCNN international joint conference on neural networks. IEEE. 1990, pp. 1–6.
[10] Alireza Abbaspour et al. “A Survey on Active Fault-Tolerant Control Systems”. In: Electronics 9.9 (2020). ISSN: 2079-9292.
[11] Jigar Patel et al. “Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques”. In: Expert systems with applications 42.1 (2015), pp. 259–268.
[12] Xiao Zhong and David Enke. “Predicting the daily return direction of the stock market using hybrid machine learning algorithms”. In: Financial Innovation 5.1 (2019), p. 4.
[13] Girish Chandrashekar and Ferat Sahin. “A survey on feature selection methods”. In: Computers & Electrical Engineering 40.1 (2014), pp. 16–28.

[14] D Van Thanh, HN Minh, and DD Hieu. “Building unconditional forecast model of Stock Market Indexes using combined leading indicators and principal components: application to Vietnamese Stock Market”. In: Indian Journal of Science and Technology 11 (2018).
[15] Ahmad Kazem et al. “Support vector regression with chaos-based firefly algorithm for stock market price forecasting”. In: Applied soft computing 13.2 (2013), pp. 947–958.
[16] Haiqin Yang, Laiwan Chan, and Irwin King. “Support vector machine regression for volatile stock market prediction”. In: International Conference on Intelligent Data Engineering and Automated Learning. Springer. 2002, pp. 391–396.
[17] Riswan Efendi, Nureize Arbaiy, and Mustafa Mat Deris. “A new procedure in stock market forecasting based on fuzzy random auto-regression time series model”. In: Information Sciences 441 (2018), pp. 113–132.
[18] Isaac Kofi Nti, Adebayo Felix Adekoya, and Benjamin Asubam Weyori. “A systematic review of fundamental and technical analysis of stock market predictions”. In: Artificial Intelligence Review (2019), pp. 1–51.
[19] Sohrab Mokhtari et al. “A Machine Learning Approach for Anomaly Detection in Industrial Control Systems Based on Measurement Data”. In: Electronics 10.4 (2021). ISSN: 2079-9292.
[20] Caren Marzban. “The ROC curve and the area under it as performance measures”. In: Weather and Forecasting 19.6 (2004), pp. 1106–1114.
[21] Mohammad Abedin, Sohrab Mokhtari, and Armin B Mehrabi. “Bridge damage detection using machine learning algorithms”. In: Health Monitoring of Structural and Biological Systems XV. Vol. 11593. International Society for Optics and Photonics. 2021, 115932P.
[22] Yash Chaudhary. Stock-Market Sentiment Dataset. 2020.
DOI: 10.34740/KAGGLE/DSV/1217821.
[23] Tom Howley et al. “The effect of principal component analysis on machine learning accuracy with high dimensional spectral data”. In: International Conference on Innovative Techniques and Applications of Artificial Intelligence. Springer. 2005, pp. 209–222.