International Journal of Financial, Accounting, and Management (IJFAM)
ISSN: 2656-3355, Vol 3, No 3, 2021, 275-287 https://doi.org/10.35912/ijfam.v3i3.604

Study of machine learning algorithms for potential stock trading strategy frameworks

Aakash Agarwal
Doon University, India
aakashagrawal778@gmail.com

Article History
Received on 31 May 2021
1st Revision on 14 June 2021
2nd Revision on 19 July 2021
3rd Revision on 30 July 2021
4th Revision on 6 September 2021
Accepted on 19 September 2021

Abstract
Purpose: This paper discusses major stock market trends and provides information on stock market forecasting. Stock market forecasting is essentially an attempt to forecast the future value of the stock market. Doing this manually can be a strenuous task, and thus we need some software and algorithms to make our task easier. This paper also lists a few of those algorithms, formulas, and calculations associated with them. These algorithms and models primarily revolve around the concept of Machine Learning (ML) and Deep Learning.
Research Methodology: This study is based on descriptive, quantitative, and cross-sectional research design. We used a multivariate algorithm model and indicators to examine stocks for investing or trading and their efficiency. It concludes with the recommendations for enhancing trading strategies using machine learning algorithms.
Results: This study suggests that after comparing and combining the various algorithms using experimental analysis, the random forest algorithm is the most suitable algorithm for forecasting a stock's market prices based on various data points from historical data.
Limitations: The applicability of the study was only hampered by unforeseeable tragic events such as economic crisis, market collapse, etc
Contribution: Successful stock prediction will be a substantial benefit for stock market institutions and provide real-world answers to the challenges that stock investors face. As a result, gaining significant knowledge on the subject is quite beneficial for us.
Keywords: Algorithms, Algo-trading, Deep learning, Machine learning, Price prediction, Stock market, Trading, Trends
How to Cite: Agarwal, A. (2021). Study of machine learning algorithms for potential stock trading strategy frameworks. *International Journal of Financial, Accounting, and Management*, 3(3), 275-287.

1. Introduction

The ownership or equity in a corporation is termed as a stock. Now, here comes another term called shares, in the form of which; stocks are issued. Thus, combining the two, we can say that a 'share of 'ownership' precisely defines a stock.

A stock market, sometimes known as a share market, is a gathering of stock buyers and sellers. These stocks indicate an individual's or a group's claim to ownership of a specific enterprise. The primary aim of any investor pre-investing in any stock is to check whether that investment is profitable to him. The era or generation of economic growth and the endowment of digital technology has led to the agglomeration of financial data. This data's rapid and sudden growth has made it impossible for humans to access it manually. Thus, this inexhaustible expansion of inconsistent data has made it necessary to encourage an automated approach to financial data. (Prerana et al. 2020)
There are primarily five functions of a stock market, namely-

1) Control- A stock-market listing impacts the relationship between the management control over the allocation of resources in the company and the ownership of the shares. It usually results in the separation of ownership and control. Simultaneously, it can also cause the reintegration of ownership and control through the accretion of shares with voting rights, both directly or through proxies. Now, those people who gain control over the corporate resource allocation may use it to their benefit.
2) Cash- The stock market is a medium through which the shareholders gain from the distribution of corporate cash to shareholders through repurchases and dividends. Value extraction may be enabled by the cash function of the stock market in the absence of appropriate regulation, which in turn drains the company funds, which are essential for investment in value creation.
3) Creation- The anticipation of a stock market listing can activate venture capital to bolster new firm establishment and advancement by enabling private equity to egress from an investment in the dynamic proficiencies of a company.
4) Combination- The 'companies' shares are made into a currency that can be used as payment for another companies shares in M&A (Mergers and Acquisitions) by stock market listing. The new combination will build productive capabilities that support value creation with the help of these M&A deals.
5) Compensation- The 'companies' shares are turned into a currency in a stock market listing. Thus, these shares can be issued to employees as a form of compensation in the form of stock awards and stock options.

In the literature review, we have discussed the stock market, its features and roles, and the affiliations in more detail. The role of investors in the market and how they make investment decisions. Investing methodology and tactics for analyzing profitable trades are the focus. The use case for machine learning, problems it can solve, and standard machine learning activities that can be performed have all been thoroughly addressed. Deep learning is further described in terms of its architectures, hierarchies, and layers. Based on bringing the degree of fundamental understanding of the topics mentioned above up to par, then continued with algorithms and strategies that can be created and played out according to the requirements and stepping out of the traditional way of trading and exploring in the new realm of the algorithm trading touching new limits of the stock market.

This study aimed to improve understanding of machine learning algorithms and financial models and see the potential through the combinations of the different algorithms presented in the paper. Specifically, the study contributes to the understanding of the conception of algorithm models and indicators in the stock market and how it combines them both in creating favorable opportunities for investors.

The remainder of this paper is organized as follows: section 3 outlines related research and methodology, section 4 presents, section 4 reports results and develops discussions, section 5 highlights main conclusions, and section 6 presents the limitations.

2. Literature review and hypothesis development

Stock market

Surveys were conducted by the NYSE in 1954 and again in 1959 to learn more about why people invest in the stock market. Even after 40 long years, the reasons outlined at the time remain unchanged now: long-term capital growth, dividends, and a hedge against the accelerated erosion of purchasing power.
Simultaneously, stock investment comes with a risk factor as well. This risk has to be examined carefully, but at the same time, it should not prove to be a disadvantage when compared with a risk-free Treasury Bond. There is always some risk involved in all kinds of investments. The buyer of this Treasury Bond, for example, receives both explicit and implicit guarantees. The Treasury confirms to give a stipulated rate of interest, say 8%, and that it also agrees to refund the original amount at maturity. As for the implicit one, the Treasury will not pay the investor 10% and will not reimburse more than the amount that was originally paid. Thus, in simple terms, safety comes with a price. This puts the investors seeking a greater rate of return in a bind because they must seize the opportunity to invest in common stocks. The stock prices could fall and sometimes for long periods, which is a risk that the new

investors should consider. Even if you own the highest-quality stocks, it does not guarantee a superior return or any return. The Johns-Manville Corporation and International Harvester, previously Dow Jones Industrial Average fundamentals, have filed for bankruptcy in recent years. Union Carbide, Chrysler, and Texaco, and many of the country's most significant banking institutions, such as Citicorp and Chase Manhattan, were among those hit hard by the financial crisis. Some of these businesses survived and even prospered, but investors who sold their stock when the future looked bleak lost a lot of money. Some of these companies survived, and some even prospered but the investors who sold their shares when the future seemed the most uncertain lost heavily.

As a result, investors should keep a constant eye on their investments. Evaluation of companies solely on their past results should not be done, no matter how convenient those may be in showing superior products as well as management. Constant vigilance over investments does not imply active trading, but it keeps us informed of the 'companies' business nature as well as with the continuous changes in the economic conditions of the world, it tells us the 'companies' place in the global economy. Investors probably cannot protect themselves from the kind of disaster that wiped out Union Carbide in Bhopal, India, but intelligent investors that are informed can discern that will enable them to avoid declining industries and invest in those industries that have a greater potential to grow. (Nayak et al. 2016)

The general lack of trust also affects stock market participation. While deciding whether to buy stocks or not, investors fear the fact of being cheated upon. This risk is a function, i.e., it depends on the characteristics of the investor and the stocks. Individuals who do not have faith in the stock market are less likely to buy stocks, and even if they do, they will invest less. As a result, a lack of trust is a significant component in explaining the low stock participation.

Stock market forecasting is essentially an attempt to forecast the stock market's future value. The main goal of stock market prediction using machine learning is to develop effective and efficient models capable of delivering a higher rate of forecast accuracy. Stock price forecasting is essential since both business people and ordinary people utilize it. People will either make money or lose their entire life savings in stock market activity. Therefore it can be highly lucrative for some while also creating irreversible losses for others. As a result, it is a chaotic system. Simultaneously, building an effective model for stock prediction is difficult since price fluctuation is influenced by various elements such as news, social media data, fundamentals, corporate production, government bonds, historical pricing, and national economics. As a result, a prediction model that includes one element may not be accurate. As a result, combining several inputs such as news, social media data, and historical pricing data might improve the model's accuracy. (Sadia, Sharma, & Sanyal, 2019)

Fundamental analysis, market mimicry, technical analysis, machine learning, and time-series aspect structure are some of the methodologies and strategies to create the prediction system. The forecast has gone up into the technical sphere as a result of the ongoing advancement in the sphere of technology. The most well-known and promising strategy is the usage of Artificial Neural Networks, also known as Recurrent Neural Networks, which are essentially machine learning implementations. (Dash & Dash, 2016)

Machine Learning

Machine Learning is a branch of Artificial Intelligence that primarily focuses on building applications that automatically learn from the data without being explicitly programmed or simply without any human intervention. Machine learning is described using three parameters that are P, E, and T, where T is the task learned, E is the experience through which T is learned, and performance P varies with E. (Prerana, Mahishi, Taj, & Shetty, 2020)

Problems that can be tackled using ML

1) *Document or Text Classification*- It mainly includes problems like assigning a topic to a document or text file or automatically determining whether a certain web page's content is too explicit or inappropriate. This includes spam detection as well.

2) *Natural Language Processing(NLP)*- The tasks in this field like named-entity recognition, part-of-speech tagging, context-free parsing are classified as learning problems.
3) *Speech Processing Applications*- speech synthesis, speech recognition, speaker identification, speaker verification, speaker verification, and sub-problems such as acoustic and language modeling are included in this.
4) *Computer vision Applications*- This mainly comprises object identification, face detection and object recognition, optical character recognition(OCR), pose estimation, or content-based image retrieval.
5) *Computational Biology Applications*- This comprises analysis of protein and gene networks, protein function prediction, and identification of key sites.
6) Numerous other problems, including network intrusion, medical diagnosis, search engines, learning to play games like go, chess and backgammon, fraud detection through credit cards, unaccompanied control of vehicles such as robots and cars, are dealt with using machine learning algorithms and techniques.

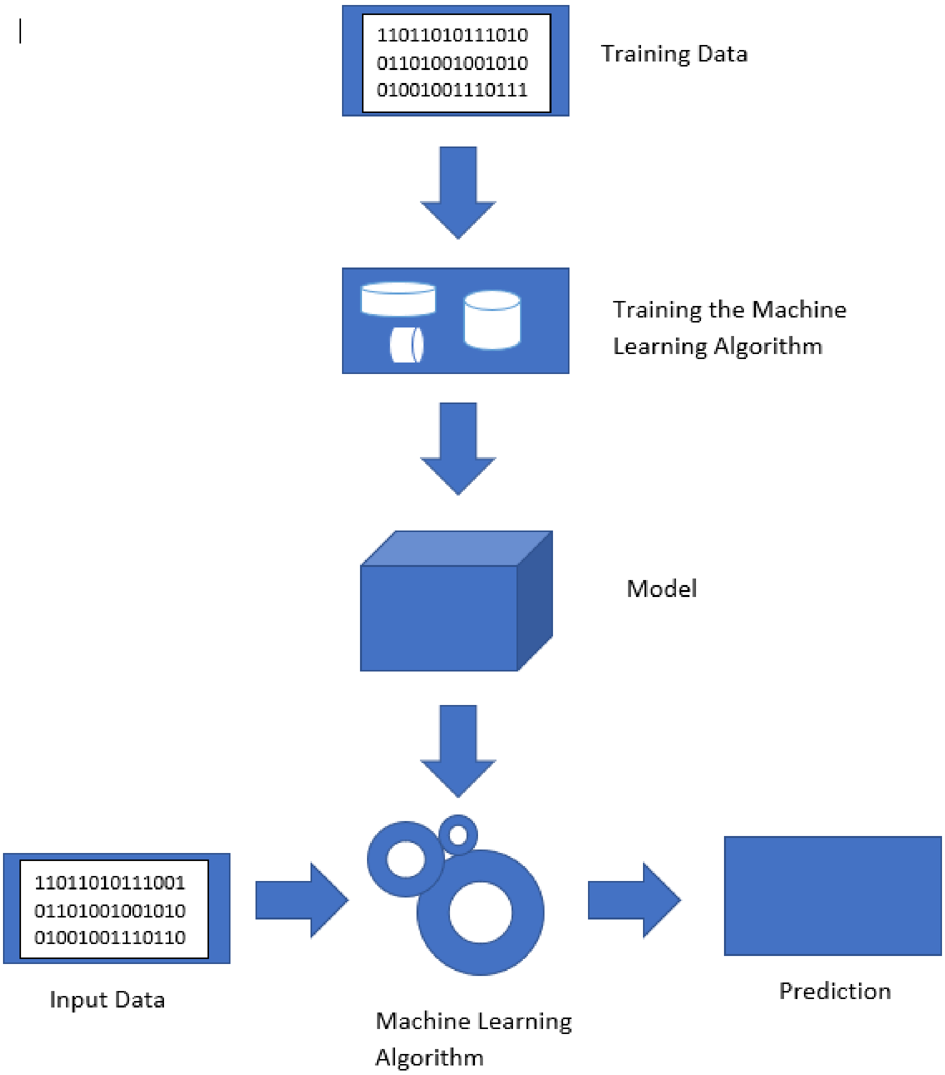

Figure 1. Data structure in machine learning | (Prerana et. al. 2020)

The following are some standard machine learning tasks that have been studied extensively-
1) *Classification*- This problem implies assigning a particular category to each item. For example, in document classification, we can assign a particular category to each document like business, sports, politics, or weather. In image classification, we can assign a particular category such as

a train, car, or plane. The number of such categories in these tasks is often less than a few hundred, but it can be very large in some complex tasks and sometimes even unbounded, like in speech recognition.

2) *Regression*- The real value of each item is predicted using this problem. Some examples of regression are the prediction of stock values or variations of economic variables. In the classification problem, there is no concept of closeness between the various categories, but in regression, the penalty for an incorrect prediction depends on the magnitude of the difference between the predicted and true values.
3) *Ranking*- If we want to order the items according to some criterion, we use this problem. For example, in web search, the search engine usually returns web pages relevant to a search query, and this is the canonical ranking example.
4) *Clustering*- This includes partitioning a set of items into homogenous subsets. This is usually used to analyze extensive data sets. For example, clustering algorithms attempt to identify natural communities within large groups of people in the context of social network analysis.
5) *Dimensionality reduction or manifold learning* involves converting an introductory representation of items into a lower-dimensional representation while conserving some properties of the initial representation. For example, it is used in pre-processing digital images in computer vision tasks.

Deep Learning

Deep Learning is an Artificial Intelligence (A.I.) subset of Machine Learning that employs multiple layers to extract higher-level characteristics from raw data. It is essentially a three- or more-layered neural network. A single-layer neural network can still make approximate predictions, but more layers can help to optimize and refine for accuracy, resulting in a more favorable outcome. It is a type of machine learning that revolves around making computers learn from prior experience and understand the world in terms of conceptual hierarchy. There is no need for a human-computer operator to cite all the expertise needed by the computer since the computer gathers knowledge from experience. The conceptual hierarchy enables the computer to build simpler concepts to understand a more complex one. Thus, if we draw a graph of all these hierarchies, it would be several layers deep. (Kelleher, 2019)

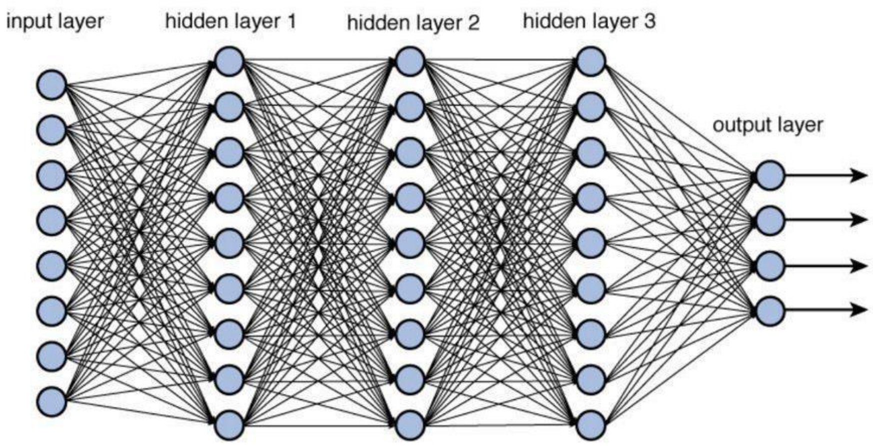

Figure 2. Architectures of deep neural network with multiple layers | (Ravindra Parmar, 2018)

Deep learning is establishing innovation and bringing in change across all areas of our modern world. Most of the Artificial Intelligence breakthroughs that we hear about in the news are a consequence of deep learning or are based on the deep learning methodology. So, whether you are interested in improving the efficiency of your organization, you want to work with complex data, you are anxious about privacy and ethics in a Big Data world, or you want to learn a bit or get a gist of the potential of Artificial Intelligence and how it is going to impact your future, you must have some prior knowledge and understanding of the concept of deep learning. The amazing fact about deep learning is not the complex math it is built upon but rather the diverse set of impressive and exciting tasks it can perform with the help of simple calculations. A Deep Learning model comprises a lot of calculations, multiplications, and additions with a few non-linear mappings. Yet, despite all these, it was still able to beat the World Go Champion, drive a car, and define state of art in machine translation and computer vision.

Google, Microsoft, and Baidu all use Deep Learning for machine translation and also for image search. Smartphones that are used in today's world have deep learning systems running on them. The cameras which use face detection technology and the apps which use speech recognition all work on deep learning. Processing medical images (X-Rays, MRI, and T.C.T. Scans) and diagnosing health conditions are some other areas that utilize deep learning. This world is progressing day by day, the old ones are replacing self-driving cars. All these also work on deep learning. (Domowitz & Steil, 1999)

The best-known example of this is DeepMind's AlphaGo. AlphaGo was the first computer program to beat a professional Go player. Lee Sedol, the top Korean professional Go player, lost to this program in March 2016, in a match watched by over two hundred million people. The following year, China's Ke Jie, the world's No. one ranking player was also beaten by AlphaGo. This program's success in 2016 was quite astonishing as most of the people were convinced that it would still take many years of research before a computer would be able to work at par with the top human Go players since it had been a well-known fact for quite a long time that programming a computer to play Go is much more difficult than programming it to play chess mainly owing to the simpler rules and larger board in Go when compared to chess. Another interesting fact is that there are more possible Go configurations in this game than the number of atoms in the universe. As a result, making a program like AlphaGo is a huge achievement and a step further towards the progress of mankind.

Algo-trading

Developments in the field of algorithm trading have improved recently. Numerous advanced fields of research have gained popularity and most importantly, the consolidation of financial studies with algorithms has made it feasible to organize research that would have been impossible a few years ago. Since the electronic financial sector has advanced significantly, the bulk of transactions is now completed electronically. Using market forecasting methods and trading algorithms, an additional interest as a new area of research has been obtained by the electronic financial market. (Colianni et al., 2015)

Only one specific strategy is implemented using traditional trading systems, whereas a computer makes a specific investment instead of a human in the Algo-trading method. Many studies show that traditional approaches are inferior when compared to algorithmic approaches.

A trading strategy is a critical financial method. It can be defined as the gathering of data to achieve a favorable return on investment and profit. Trading strategies when standalone, are not always outright profitable. It is a well-known and undeniable fact that financial markets are unpredictable and change continuously and essentially, and at times quite dramatically. Trading strategies that worked well for quite some time may sometimes die abruptly as a consequence of the aforementioned fact. Trading strategy results are affected by myriad factors thus, we can essentially have no universal model that can make perfect predictions for all problems or even be a single best trading method for all situations.

However, advancements in technology have given rise to contemporary types of trading, such as trading strategies dealing with machine learning and data mining. Algo-trading strategies can make the desired judgments based on the strategy without requiring any human intervention. The transaction can then be transmitted for execution to the exchange from the computer. Aside from that, complex analysis can also be done in real-time using Algo-trading. Hundreds of issues can be easily traded simultaneously with the help of advanced laws with layers of predefined rules using an algorithm. Algo-trading also seeks to identify quite transitory trends or signals by analyzing large volumes of diverse data types. A lot of these signals being very faint cannot be easily traded individually. A suitable way to tackle this problem is to cluster numerous such signals with non-trivial weights that would enhance and amplify the signal so that it becomes tradable on its own. (Chihab et al. 2019)

The foreign exchange market is currently the most liquid and the most substantial in the world. However, for the survival of both the buy and sell sides, algo-trading has become essential. The nature of the data is noisy, non-stationary, and chaotic has coerced the major traders to switch to the use of auto-algo-trading to stay competitive.

There has been an increase in the dependence on prediction using artificial intelligence past the last decade in distinct markets including Foreign Exchange Market (Forex) by using algorithms like Random Forest (R.F.), Genetic Algorithms (G.A.), Linear Regression, Support Vector Machine (SVM) and Artificial Neural Networks (ANN) approach to analysis and forecasting of financial time series. (C, K. et al. 2021)

3. Research methodology

This is a descriptive study in which we have adopted an explanatory approach to the machine learning algorithms and indicators for stock trading. The existing studies have been examined. Research papers and studies available on secondary sources were used for the study. The scope of research was trading and price prediction. The study concludes with recommendations for enhancing trading strategies using machine learning algorithms.

Algorithms

Linear regression

In a particular equation consisting of dependent and independent variables,

$$Z = m_1X_1 + m_2X_2 + m_1X_3 + \cdots m_nX_n$$

depending on the type of variables $X_{1,} X_2, X_3, \dots X_n$ classification of linear models can be done in the following way-

1) Analysis of Variance Model- The linear model will be called the Analysis of Variance Model when the matrix X consists of only ' '0's and 1's. The variables $X_1, X_2, X_3, \dots X_n$ in this case, are qualitative.
2) Regression Model- The model will be a regression model if all the variables $X_1, X_2, X_3, \dots X_n$ are quantitative. Also, the model is called regression with intercept when there exists one variable which is a constant and is also equal to 1. Besides that, all the other variables are quantitative.
3) Analysis of Covariance Model- The model is called analysis of covariance model if some of the variables are quantitative whereas the others are qualitative.

Linear regression is the simplest and the most basic Machine Learning algorithm that can be enforced. This algorithm returns an equation ship used to determine the relationship between the independent and dependent variables. Before attempting to fit a model to the given data, the modeler must first ensure that the two variables have a relationship. But this does not imply that one variable causes the other as causation and co-relation are different. However, instead of that, it conveys a strong relationship between the two variables. For linear regression, we generally use a scatter plot as it helps to find out the degree of association between the two variables. Fitting a linear regression model to the data will be an exercise in futility if there is no relationship between the two variables. (Pimprikar, Ramachadran, & Senthilkumar, 2017)

Linear Regression can be represented by the equation:
$Z = m_1X_1 + m_2X_2 + m_1X_3 + \cdots m_nX_n$, where, $X_1, X_2, X_3, \dots X_n$ represent the independent variables and $m_1, m_2, m_3, \dots m_n$ represent the weights and Z represents the dependent variable.
Although we may be able to specify such a hypothesis for various reasons, we would still not be able to find out the exact nature of the considered linear relationship, or in simple words the parameters $m_1, m_2, m_3, \dots m_n$ will be unknown. We can draw information about them from a given set of observations of the variables Z and $X_1, X_2, X_3, \dots X_n$. (Groß, 2012)

In matrix form, the above equation can be represented as-

$$\boldsymbol{Z} = \boldsymbol{X}m + \in$$

$$\text{where} \in \sim (0, \sigma^2 I_n)$$

$$\text{and } \boldsymbol{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ \vdots \\ Z_n \end{pmatrix},\ \boldsymbol{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}, m = \begin{pmatrix} m_1 \\ \vdots \\ \vdots \\ m_p \end{pmatrix},\ \in = \begin{pmatrix} \in_1 \\ \vdots \\ \vdots \\ \in_n \end{pmatrix}$$

Support vector machine algorithms

Vapnik and co-workers formulated the Support Vector Machine(SVM) in the early 1990s. It has been used to classify and recognize patterns in many different areas of science, business, and industry. Forecasting the movement and direction of financial time series is one of the prime areas which utilize this machine effectively. Various researchers proposed theories, performed experiments, and compared this with distinct models to forecast the financial time series. For example, some of them concluded that SVM provided a very good and significant alternative to stock market prediction, whereas others concluded that Support Vector Regression(SVR) has good predicting power, especially when the strategy we are using is to update the model periodically. (Harries & Horn, 1995)

It is a machine learning model that solves two-group classification with algorithms. The fundamental purpose of the Support Vector Machine Algorithm is to find an N-dimensional space that categorizes input points. The letter N stands for the number of attributes. The objective of this algorithm is to find a hyperplane among the multiple possible hyperplanes between two classes of data points, that have the maximum margin. The maximal margin is the distance between data points from both classes. This is useful since it gives some reinforcement, making subsequent data points easier to classify. (Nti, Adekoya, & Weyori, 2020)

Let DS be the training dataset, DS={(x_i,y_i,…,(x_n,y_n))}∈X.R where i = (1,2,3,…,n).DS={(x_i,y_i,…,(x_n,y_n))}∈X.R where i = (1,2,3,…,n).

The formula utilized in the SVM optimization technique is shown in equations (a) and (b).

$$\min_{d,bw,} \frac{1}{2} W^T W + C \sum_{i=1}^{n} \omega i \qquad (a)$$

Subject to $$y_i(W^T \theta(x_i + b) \geq 1 - \omega i), \ \ \omega i > 0 \qquad (b)$$

The function θ of vectors x_i(D.S.) is mapped in a higher spatial dimension. In this dimension, the SVM finds a linear separating hyperplane with the best margin. The kernel function can be formulated as $K(x_i, x_j) \equiv \theta(x_i)^T \theta(x_j)$. This study used the Radial Basis Function (RBF) kernel, which is represented in Eq. (c).

$$RBF{:}K(x_i, x_j) = \exp\left(-y \left|\left|x_i - x_j\right|\right|^2\right),\ \ y > 0 \qquad (c)$$

where $(x_i - x_j)$ is the Euclidian distance between two data points.

Random forest algorithm

Breiman proposed the Random Forest Algorithm in 2001, and it is known to generate accurate predictive models. An advantage of using this algorithm is that when the data consists of many variables and we are uncertain about which variables to include in the model, we can easily use this algorithm as it automatically identifies the important predictors. Random forest Algorithm is one among the many useful applications of machine learning since it helps us figure out the financial markets' functioning. By forecasting changes in price, this algorithm has been used in numerous works to beat the market.

Since predicting price fluctuations impact the investment, it is a very difficult and complex area. We can use a Random Forest classifier to build a predictive system with parameters such as precision, recall, accuracy, and specificity to evaluate the robustness of the system. Experiments were done by researchers using Linear Discriminant Analysis (LDA) Artificial Neural Network (ANN), Logit, SVM, and Random Forest Algorithm. From the results, they concluded that the Random Forest Algorithm surpasses the other ones.

It is been dubbed one of the most user-friendly machine learning algorithms, with high accuracy in stock market forecasting. It is most commonly used in jobs involving categorization. A Random Forest Classifier with the same hyperparameters as a decision tree is used. The decision tool also uses a tree-like model to make decisions based on potential outcomes, such as resource cost, event result, and utility. The Random Forest Algorithm employs a method in which it picks different observations and characteristics at random to build several decision trees. It then sums the results of these decision trees. (Sadia, Sharma, & Sanyal, 2019).

Auto ARIMA

It is the abbreviated form of Auto-Regressive Integrated Moving Average. It is a very popular model used primarily for time series forecasting. ARIMA model takes the past values and data into account and predicts the future values using that data. It is a type of regression analysis that takes the strength of one dependent variable which is related to other changing variables. (Prerana, Mahishi, Taj, & Shetty, 2020)

Forecasting time series is necessary for the financial industry as well as other disciplines, both economic and non-economic. The ability to draw these models requires a thorough understanding of current events. However, sometimes it is quite hard for the one producing these time-series models to keep track of the trends, so an automatic forecast algorithm is quite an essential tool. Any autonomous prediction method must be capable of selecting a suitable time series model, estimate the ' 'model's parameters, and calculate the prediction. The most often used automatic forecast methods are exponential smoothing or ARIMA models, which makes ARIMA particularly suitable for time-series forecasting. We can use programming languages like R and ARIMA to study time series forecasting more effectively. (Dhamo & Puka, 2010)

Extreme Learning Machine (ELM)

ELM is a recently introduced learning algorithm for single-hidden layer feed-forward neural networks (SLFNs) that, instead of iterative tuning, chooses at random the weights of connections between input variables and neurons in the hidden layer, as well as the bias of neurons in the hidden layer, and calculates the output weights analytically. ELM tends to produce superior generalization performance in addition to having incredibly quick learning and testing speeds. The SLFNs' hidden layer does not need to be tweaked in ELM, and it can handle a broad range of piecewise continuous functions. If we are given a set of N training dataset D = (x_i, y_i), i = 1 to N where each x_i is a d dimensional input pattern and y_i is the desired output, activation function for hidden layer nodes, H is the number of hidden layer nodes and a linear activation function in the output neuron, then the output function of ELM for SLFN can be represented as:

$$y_i = \sum_{j=0}^{H} \omega_j h_j(x_i)$$

W is the weight vector connecting the hidden layer neurons to the output layer neuron, and hj(x) is the activation function of the hidden layer. (Wen et. al. 2010)

Indicators

Moving Average (M.A.)

Each prediction employs the most recent set of data, and for each succeeding step, the oldest observation is eliminated and the anticipated values are considered.

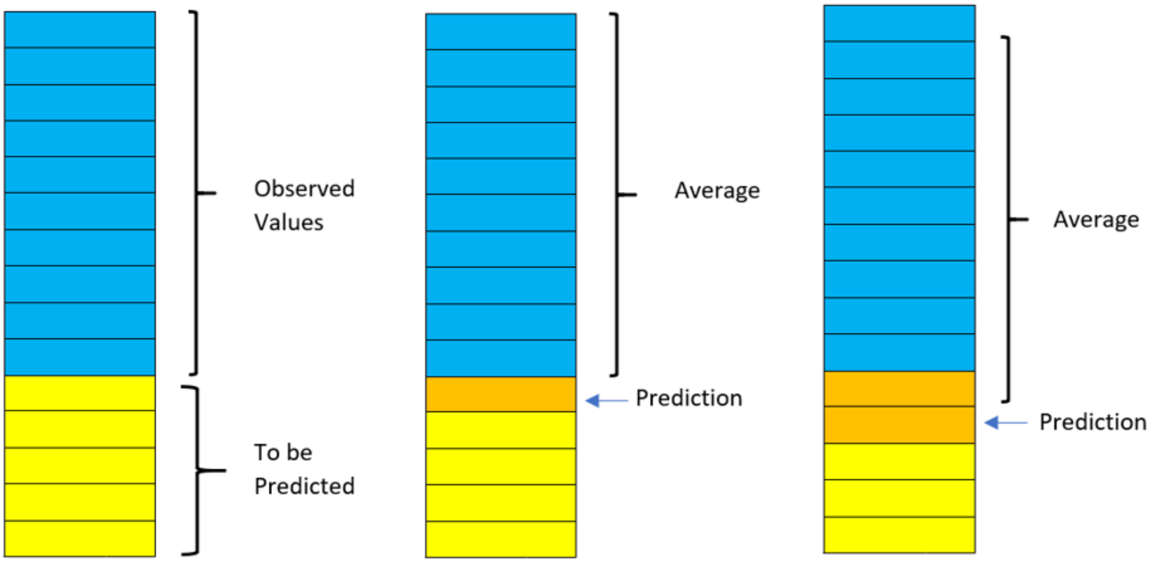

Figure 3. Moving average | (C, K. et. al. 2021)

$$MA_t = \frac{1}{t}\sum_{i=1}^{t} cp(i)$$

The closing price is denoted by cp(i). (Prerana, Mahishi, Taj & Shetty, 2020)

Moving Average Convergence and Divergence (MACD)

The MACD is a price indicator that shows the connection between two exponential moving averages (EMAs). (Nti et. al. 2016)

$$MACD = EMA_{15} - EMA_{22}$$

$$EMA(i) = \big(CP(i) - EMA(i-1)\big) * Multiplier + EMA(i-1)$$

$$where, \qquad Multiplier = \frac{2}{no. of\ days\ to\ be\ considered + 1}$$

Stochastic K.D.

It calculates the average of the measuring price movement velocity, where D percent is the three-day moving average of K percent, and K percent is the relative position of the current closing price in a given period. (Nti et. al. 2016)

$$K\%(i) = (cp(i) - L_t)/(H_t - L_t) \times 100$$

$$D\%(i) = \frac{\big(K\%(i-2) + K\%(i-1) + K\%(i)\big)}{3}$$

Where L_t is the last t days' lowest price, cp(i) is the closing price, and H_t is the last t days' highest price.

Relative Strength Index (RSI)

This is a momentum indicator, and it is computed like this:

$$RSI = 100 - \frac{100}{1 + RS}$$

Where

$$RS = \frac{Average\ of\ t\ day's\ up\ closes}{Average\ of\ t\ day's\ down\ closes}$$

4. Results and discussion

Over 70 empirical studies of entrance and exit trends from eleven different countries agree that entry is more prevalent in more profitable and rapidly growing businesses. In contrast, exit is slower in sectors where the absolute capital expenditures of building a minimum efficient size plant are prohibitive. As entry hurdles, excess capacity, scale economies, and limit pricing all find some empirical evidence. It is difficult to separate the effects of advertising and R&D intensity since the evidence is so strong. Exit is the inverse of entry, and it occurs faster when profits are smaller and slower when long-term particular (sunk) capital expenses are more relevant. Exit and entrance are inextricably linked, owing to displacement and vacuum effects (in which entrants are enticed by the prospects of selling to uncommitted customers abandoned by a recent exit). (Siegfried & Evans, 1994)

Both oligopoly and monopoly models have entry requirements. Potential entry limits the price alternatives available to sector leaders, whereas the actual entrance of new businesses and new items erodes incumbent businesses' market strength at different rates. If actually entry responds rapidly to profit opportunities, then the persistent exercise of monopoly power would be rare. However, analyses of both these models i.e., oligopoly and monopoly typically approach entry indirectly, by inferring that persistent profit differentials reflect high barriers to entry. There are a few direct tests of the effects of such barriers on actual entry (MacDonald, 1986). There are only a few empirical studies of entry and exit that exist, despite their importance. Severe data limitations restrict both the number of studies and the scope of extant analysis.

By comparing the accuracy of the various algorithms using experimental analysis, we can say that the random forest algorithm, which is explained in the earlier part of this paper, is the most suitable algorithm for predicting the market price of a stock based on various data points from historical data. Entry and Exit points can also be predicted with more accuracy than other algorithm models and, with its help, can place a large number of orders as well in both types of orders.

5. Conclusion

Among various algorithms, the random forest algorithm will prove to be a valuable tool for brokers and investors looking to trade in the stock market because it was chosen after being tested on a sample dataset and is also trained on an extensive collection of historical data. As a result, the Random Forest Algorithm reveals that the machine learning model can accurately forecast stock value when compared to previously developed machine learning models.

At the same, we also learned about the essence of Machine Learning and how it can be helpful to us. The various models that are used to tackle problems like *Document or Text Classification, Natural Language Processing (NLP), Speech Processing Applications, Computer vision Applications, Computational Biology Applications,* and so on. We also got a brief understanding of the different machine learning tasks that can be performed, including classification, regression, ranking, dimensionality reduction, manifold learning, and clustering. As for stock markets, the five important functions of stocks were discussed and the risks involved while making investments were also pondered upon. A bit further we also got a gist of Algo-trading and deep learning. Methodologies like Linear Regression, Random Forest Algorithm, Support Vector Machine, and so on were explained in detail and the mathematical calculations involved in some of them were listed. Indicators like Moving Average, MACD, Stochastic. D.K.D. and Relative Strength Index were stressed as well.

Additional testing on more recent futures data will be done in the future and an examination of the approaches' application to other domains. We will also put the system through its paces in a trading system, simulating real-world futures buying and selling to see whether a profit can be made. Synthetic data sets are particularly valuable for testing alternative detection strategies since the amount of concept drift may be adjusted. The computational cost-benefits of other strategies for detecting idea drift, such as closest neighbor and model reinforcement, will be compared. Additional studies in the field of error analysis may be done to increase the accuracy of the learning algorithms even further.

Limitations and study forward

There are a few limitations of Machine Learning Algorithms for example in ethics. Environmental, social, and governance (ESG) industries are currently the most widely debated instance – how do we pick or how the stocks will behave in a tragic event such as a crash or crisis? Will we have to choose the ethical framework we want our self-driving car to follow in the future when we buy it? While all of these are intriguing questions, they are not the focus of this paper. On the other hand, machine learning clearly cannot tell us what normative ideals we should embrace or how we should act in the world in a particular scenario.

One cannot "derive an ought from an is," as David Hume famously stated. Machine learning is not deterministic but stochastic. There are no physical limits on a neural network. Hence it does not grasp Newton's second law or that density cannot be negative. This, however, may not be a long-term constraint. Several academics are investigating the addition of physical limitations to neural networks and other algorithms so that they may be utilized for applications like this. Also, when you feed a model bad data, it will provide bad outcomes. This can take two forms: data scarcity and data quality scarcity. One of the most serious issues with machine learning is its interpretability. If an A.I. consulting business tries to sell to a company that solely employs traditional statistical approaches, they may be turned down since the model is not interpretable. How likely are they to believe you and your knowledge if you cannot persuade them that you understand how the algorithm arrived at the choice it did?

References

Chihab, Y., Bousbaa, Z., Chihab, M., Bencharef, O., & Ziti, S. (2019). Algo-Trading Strategy for Intraweek Foreign Exchange Speculation Based on Random Forest and Probit Regression. Applied Computational Intelligence and Soft Computing, 2019. https://www.hindawi.com/journals/acisc/2019/8342461/

C, K., A, S., P, A., R, P., D, M. & D, V. (2021). *Predicting Stock Prices Using Machine Learning Techniques. International Conference on Inventive Computation Technologies (ICICT).* https://ieeexplore.ieee.org/abstract/document/9358537

Colianni, S., Rosales, S., & Signorotti, M. (2015). Algorithmic trading of cryptocurrency based on Twitter sentiment analysis. *CS229 Project*. http://cs229.stanford.edu/proj2015/029_report.pdf

Dash, R. & Dash, P.K. (2016). *A hybrid stock trading framework integrating technical analysis with machine learning techniques*. The Journal of Finance and Data Science. https://doi.org/10.1016/j.jfds.2016.03.002

Dhamo, E., & Puka, L. (2010). *Using the R-package to forecast time series: ARIMA models and Application*. In *INTERNATIONAL CONFERENCE Economic & Social Challenges and Problems*. https://www.researchgate.net/profile/Eralda-Dhamo-Gjika/publication/274249061_Using_the_R-package_to_forecast_time_series_ARIMA_models_and_Application/links/55193e570cf2d241f355e6b7/Using-the-R-package-to-forecast-time-series-ARIMA-models-and-Application.pdf

Domowitz, I., & Steil, B. (1999). *Automation, trading costs, and the structure of the securities trading industry. Brookings-Wharton papers on financial services*. https://www.researchgate.net/profile/Ian-Domowitz/publication/241369793_Automation_Trading_Costs_and_the_Structure_of_the_Securities_Trading_Industry/links/545259310cf2bccc4908dccf/Automation-Trading-Costs-and-the-Structure-of-the-Securities-Trading-Industry.pdf

Groß, J. (2012). *Linear regression*. Springer Science & Business Media. https://books.google.co.in/books?hl=en&lr=&id=enwQBwAAQBAJ&oi=fnd&pg=PA4&dq=line

ar+regression&ots=KH9QBZZMZk&sig=uJQk9-N4kxANWxHfkzNcWFzIzFU#v=onepage&q=linear%20regression&f=false

Harries, M., & Horn, K. (1995). Detecting concept drift in financial time series prediction using symbolic machine learning. World Scientific Publishing. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.5260&rep=rep1&type=pdf

Kelleher, J. D. (2019). *Deep learning*. MIT press. https://books.google.co.in/books?hl=en&lr=&id=b06qDwAAQBAJ&oi=fnd&pg=PP9&dq=deep+learning&ots=_oBVUMlT_Q&sig=_Bhmf14kAhctFTpE8Bkdl5ZHIZM#v=onepage&q=deep%20learning&f=false

MacDonald, J. (1986). *Entry and Exit on the Competitive Frine. Southern Economic Journal*. doi:10.2307/1059263 https://doi.org/10.2307/1059263

Nayak, P., Pai, M.M. and Pai, Radhika. (2016) Prediction Models or Indian Stock Market. Procedia Computer Science. https://doi.org/10.1016/j.procs.2016.06.096.

Nti, I. K., Adekoya, A. F. & Weyori, B. A. (2019). Random Forest Based Feature Selection of Macroeconomic Variables for Stock Market Prediction. *American Journal of Applied Sciences*. https://doi.org/10.3844/ajassp.2019.200.212

Nti, I.K., Adekoya, A.F. & Weyori, B.A. (2020). *A comprehensive evaluation of ensemble learning for stock-market prediction*. *J Big Data*. https://doi.org/10.1186/s40537-020-00299-5

Parmar, R. (2018). *Training Deep Neural Networks*. Medium. https://towardsdatascience.com/training-deep-neural-networks-9fdb1964b964.

Pimprikar, R., Ramachadran, S., & Senthilkumar, K. (2017). *Use of machine learning algorithms and Twitter sentiment analysis for stock market prediction. Int J Pure Applied Maths*. https://acadpubl.eu/jsi/2017-115-6-7/articles/6/69.pdf

Prerana, C., Mahishi, P., Taj, N., & Shetty, A. (2020). *STOCK MARKET PREDICTION USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES*. https://www.irjet.net/archives/V7/i4/IRJET-V7I4720.pdf

Sadia, K., Sharma, A., & Sanyal, S. (2019). *Stock Market Prediction Using Machine Learning Algorithms*. https://www.ijeat.org/wp-content/uploads/papers/v8i4/D6321048419.pdf

Siegfried, J.J., Evans, L.B. *(1994). Empirical studies of entry and exit: A survey of the evidence*. *Review of Industrial Organisation*. https://doi.org/10.1007/BF01035654

Wen, Q., Yang, Z., Song, Y., & Jia, P. (2010). *Automatic stock decision support system based on box theory and SVM algorithm*, Expert Systems with Applications. https://doi.org/10.1016/j.eswa.2009.05.093.