Stock Market Prediction with High Accuracy using Machine Learning Techniques

Manjusha Nagpure
Computer Science, SCMIRT Bavdhan Pune.

Prajakta Kamble
Computer Science, SCMIRT Bavdhan Pune.

Guided by
Dr. Archana Wafgaonkar
Assistant Professor, SIBMT, Bavdhan Pune.

Dr. Deepak Singh
Vice-Principal, SCMIRT Bavdhan Pune.

Abstract

Technical analysis of stock markets has remained a focal area of interest in recent years due to huge profit making opportunities. In recent years with the advent of various ML techniques, predicting models are more advanced and dynamic to handle the patterns within large sets of data. This research work is centered on a highly robust stock market prediction model using different types of machine learning algorithms such as decision tree, support vector machines, and neural networks. It is based on historical stock price records, different technical indicators and macro economic variables to predict the future stock price in an accurate manner.

In this study, we discuss data preprocessing like; Normalization, feature selection, and handling missing values to guarantee the good input data. Accuracy of the predictions is the major criterion for comparing different ML algorithms, and a method for combining these algorithms is also proposed. The experiments presented here demonstrate that adding feature engineering on top of the neural network provides a much higher accuracy compared to other approaches.

However, difficulties like overfitting of models, the requirement for prescient data, vagaries of stock markets due to unprecedented circumstances globally are also considered by this research. This analysis indicates that whilst machine learning does provide potential solutions in relation to stock predictions there is still a need for on-going improvement and updating for the techniques to become more effective if used in the long-term. The model described hereby gives high accuracy and can be very beneficial for investors who seek to make rational choices in the stock exchange market.

Keywords
Stock price prediction, Machine learning algorithms, Decision tree, Support Vector Machine, Neural Networks, pre-processing of data, ensemble techniques, financial forecasting

Introduction

However, there are difficulties for machine learning in the stock market prediction to date. These are challenges such as over-fitting where the models do well within previous data, but they do poorly within subsequent data, as well as issues to deal with noisy and missing data. To offer solutions to the above challenges, this research intends to implement a high accuracy stock market prediction model, which integrates several machine learning approaches, with main emphasis on data pre-processing, feature extraction, and model enhancement.

To this end, this paper aims at demonstrating the applicability of applying a machine learning approach to the training of a model that predicts the stock prices with much higher precision. In this paper, as a benchmark for differentiating between algorithms' effectiveness and for utilizing ensemble learning techniques to evaluate the extent of their synchronization, the focus is given to the stock market prediction problem. The potential knowledge derived from this research will not only advance the scholarship on financial forecasting but also provide useful advice for investors and institutions that demand trustworthy and efficient means and strategies to compete effectively in such a rapidly growing market environment.

The market for equities is vast and unsteady, with many synergistic factors that are capable of shaping the operations and the stock exchange. Stock price forecasting has always been a focus in the finance industry and amongst investors as even minimal increases in forecasting efficiencies would lead to substantial financial returns. But fundamental and technical analysis for share prices proved to be inefficient in extracting and identifying correlation between numerous factors due to their sequential, non-linear relationships. In this regard, the use of the ML techniques has become effective for post processing of large datasets and identification of other patterns with an improved degree of distinction.

Machine learning has brought great changes in many sectors and the same applies to stock market prediction. Machine learning, therefore, makes use of decision trees, support vector machines (SVM), and neural networks to therefore analyze the previous performance of the stock, available technical data, and other possibly influencing variables to arrive at high probability results. These models do not get outdated when new data is added, they also allow the new inputs like sentiment analysis of the social media or news.

The nature of the stock market is uncertain by definition owing to external factors like political instability or natural disaster and therefore applying a machine learning model that can learn and update itself has an edge on a rule based system. Also, high-frequency operations such as high-frequency trading and algostocks require more frequent, exact forecasts making scalability of Machine/Artificial Learning even more critical.

Objectives

1. The importance of stock market prediction for investors, analysts and for the economy, to illustrate the effects on decision making and economic stability.
2. To review the methods of technical and fundamental analysis with purpose to conclude about their appropriateness for forecasting the stock price directions.
3. Algorithms are much more useful for research in stock prediction than traditional statistics because they help to work with big data.

4. To acquaint the reader with various tools of machine learning that have been used in the prediction of stock markets and to describe the capabilities and uses of the tools in the field of finance.
5. The relevance of feature selection and data preprocessing is explained and examples are provided to demonstrate the improvement that can be made in the working of machine learning models.
6. To identify certain composite approaches that incorporate both conventional and machine learning methods and evaluate their utility regarding accuracy of prediction.
7. To describe the strategy of measuring performance in the context of building and testing machine learning models for stock prediction so that there is no confusion in the backtesting of their efficiency.
8. To review the difficulties encountered when it comes to using a model for stock prediction and general challenges that affect models such as over-fitting and external influences.
9. In regards to the work done so far in the field of machine learning, including explainable AI as well as sentiment analysis, and how these can be applied to the field of stock market prediction.
10. In this context, to offer the examples of how it actually works on real-world data to predict stocks, and reveal some aspects of high-frequency trading, based on machine learning.
11. For more about current research directions in the direction of raising the precision of forecasts, the issues of real-time prediction, and the ethical issues related to the usage of algorithmic trading.

Literature Review

Introduction to Stock Market Prediction

Stock market prediction is the process by which future stock prices in the stock market are sought to be forecasted or predicted. A correct guess of future share price could translate into big gains. The efficient market hypothesis provides that stock prices provide the information that is currently available and any change in the stimulation whose causes are not related to the newly released information is thus random. Some people don't agree with this and those people who think so have countless ways and tools which allegedly can provide them with the future price data.[1]

Stock market prediction is important for investors and financial analysts, because it helps to make decisions that may influence the big picture and the amount of risk that investors are willing to shoulder. Market trends can be used to determine the right time to invest or divest in the shares hence reasonable portfolio allocation. Finance professionals, however, make use of historical information and trends together with economic forecasts as they inform investment and business choices. This minimizes risk exposure preventable through knowledge and enables business entities, as well as individuals, to manage market risks better.[1][2]

From a microbusiness perspective, the precise prognosis for stock exchange results defines other macroeconomic planning. These forecasts are then deployed within businesses to decide where resources need to be invested, where expansions need to occur or to manage risks. Markets also help governments to assess the status of the economy and hence switch from one policy of fiscal or monetary discipline to another. Erroneous forecasts enable poor decisions to be made, and as a result, we may have economic fluctuation or failure to exploit growth chances.[1][2]

Challenges in Stock Market Prediction

To begin with we must appreciate the idea that predicting the stock market is not an easy task as this market is volatile, full of random events and its behavior tends to be influenced by external factors. Fluctuations of prices in the market, short-term hikes and drops, make accuracy of future trends' predictions impossible. This is because prices can change crazily in response to random events, therefore a prediction can be misleading. Investopedia provides

elaboration on complexities that instability brings about due to unpredictable high or low movement in prices due to factors such as news, sentiment or data.[2]

Another challenge that greatly affects the overall business is the price fluctuation of the markets for the particular product. Pricing of stocks is affected by so many factors, of which most of them are erratic. For example, minor and unplanned occurrences will cause the direction of markets to change. In addition, such factors from outside the business environment, general issues like political risks or instabilities, social issues like movements etc or change in policy issues are crucial. Wikipedia also points out that in some cases major geopolitical events or an economic liberalization could occur which would observe the markets up and which makes former accurate predictions invalid .[1][2]

Finally, cultural factors such as the attitude of consumers or impact by the media are sometimes hard to estimate and such can change the markets erratic.[1]

Traditional Methods of Stock Market Prediction

There are some systematic ways to make predictions which can be divided into three general classes that may (and frequently do) intersect. They are fundamental analysis, technical analysis or charting. [1]

1. Fundamental Analysis

The fundamental analysis focuses on a company that is behind the particular stock in question. They express opinions about prognosis and about the quality of that company's accounts. There are many performance ratios developed that help the fundamental analyst with the verification of a stock, including the P/E ratio. Warren Buffett, the company's scion and chairman, is arguably the epitome of a profuse fundamental analyst. He uses the overall market capitalization to GDP ratio to show the relative value of the market in general and that's why this ratio is commonly referred to as the Buffett indicator.[4][5][6]

What fundamental analysis in the stock market is in effect seeking to accomplish is to determine the price which a certain stock is truly worth and comparing that with the price it is offered in stock exchange and thus establish if the stock on the market is value-starved. Learning the true value can be done by different types of methods with a similar idea as the main goal. The idea is that the worth of any business equals the sum of all the profit that a company will earn in the future. These future profits also need to be brought down to the present current currency value. This principle compliments the business idea that business is all about money and has no other moral compass.[1]

However, fundamental analysis is considered a longer-term approach rather than the tendency of technical analysis.[1]
Fundamental analysis goes hand in hand with the concept that for the progression of human society capital is required and if a company is running good then it should be provided capital and the stock images will go up. Fundamental analysis is popular because it would be silly to do otherwise, it is logical and quantitative and has many public domain data points like financial statement analysis.[1]

Another meaning of fundamental analysis takes a broader view which involves a top down approach, where one starts with an assessment of the world economy followed by country level analysis and then the sector level and finally the company level analysis.[1]

2. Technical analysis

Technical analysis is the analysis technique of price and volume trends in the market used in making forecasts of the trend through analysis of past tendencies. The efficiency of this approach is being questioned by the efficient-market hypothesis which assumes that price movements in stock exchanges cannot be forecasted,[3] and the results of studies on the effectiveness of the technical approach are ambiguous.[7][8][9]

Technical analysts or chartists do not necessarily pay any attention to any of a company's fundamentals. They try to find probabilities of future movement of stock price mainly categorized by past price movement (time series analysis). Many structures are used for instance the head and shoulders or cup and saucer. In addition to the patterns, other techniques include EMA, oscillators or S/R levels or momentum/ volume indicators. Japanese rice merchants are credited for developing candlestick patterns which nowadays are used by technical analysts. But technical analysis is much more suitable for short-term planning than the long-term one. And therefore it is much more rife in commodities and foreign exchange where most traders rely on price fluctuations in the short term. The analysis made here is based on a number of assumptions, these being that everything there is to know about a company is reflected in the price of the stock, the price of the stock moves in trends, infact price action is usually in line with what has happened somewhere in the world before because of market psychology.[1]

Limitations of Traditional Methods

A lot of research has been done in order to predict the stock prices and expert advice such as fundamental and technical analysis have already been used traditionally with some big overhangs that has to do with the accuracy as well as the compatibility with modern data setups. An obvious limitation of these approaches is that they tend to base decisions on history and precedent, which may provide little means of addressing unique market occurrences or abrupt changes in perception among the investment markets. Because of this, an investor may be led to misjudge the market situation, more so during fluctuating volatility. For instance, Investopedia points out that traditional models may not work when one is trying to address an economic upheaval, meaning that today's markets require more dynamic analytical models.[10]

Further, conventional approaches prove to be inadequate in coping with the large volumes of data in today's environment. Methods like technical analysis are concerned with price and volume, but do not include fresh data feeds from numerous sources, including social media and global macroeconomic data. This limitation makes it difficult for them to make or expound detailed forecasts. According to Forbes magazine, the application of Big Data analytics is limited because of statistical methods applied in the market, which, in turn, undermines Market prediction enhancement.[1]

Therefore, as the financial markets complicate, simple approaches may be outperformed by sophisticated methods that include the use of Artificial intelligence and machine learning to consider datasets and revolutionize markets.

Introduction to Machine Learning (ML) in Stock Market Prediction

Machine learning as a subset of artificial intelligence is a form of programming that allows algorithms to naturally recognize patterns within a set of data and provide solutions on the similar data points without necessarily being programmed for each of them. Machine learning can be best described as a process that enhances data analysis by applying statistical models. So traditional machine learning incorporates data with statistical algorithms to generate outputs useful to decision makers. This technology is used in various domains that include image/speech retrieval,

text comprehension, recommendation services, identification of fraud, portfolio allocation and others, process automation.[11]

Relevance of ML:

Due to its ability to work with big data, identify nonlinear relationships, and control the trading rules in the HFT zone, machine learning (ML) has assumed prominence in stock market prediction. There exists a massive amount of data available in the financial markets such as price trends, other quantitative data such as trading volumes, economic indicators, and qualitative data from social media. Quantitative techniques used in conventional statistical methods can prove to be inadequate in their analysis of the data particularly where non linear relationships exist in the financial markets. Since ML algorithms are data-driven, they can analyze past data and apply the findings to present day results making their use a useful tool in stock price and market trend predictions. [12][13]

A major benefit of ML in stock market prediction is the ability to deal with big data. Better strategies such as deep learning have the capability of identifying features from raw data and are thus very effective. Also, the ML models can be trained with a variety of data structures such as structured and unstructured data increasing the capacity of the models to capture the market. This capability is vital for making consistent and high-quality analytical assessments regarding patterns and trends that do not become evident with standard analysis procedures. [12][13]

High frequency trading, in particular requires decision making in subseconds and, therefore, with the help of ML algorithms it is possible to define the best ways of trade execution taking into account real-time market characteristics. They can read news articles, social media feeds and economic reports to make instant trading decisions all of which they do. It means all these get accomplished at great speed and efficiency, and all the trading firms stand to gain higher profitability and lower risks. In the future, as changes in the field of finance and the amount of data in this area change, the use of ML in predicting the stock market will only increase. [14]

Stock Market Prediction with Machine Learning VS Traditional Way

Stock Market Prediction with ML	**Traditional Stock market prediction**
Incorporation of algorithms to make analysis of the big data more automated. [15]	Requires the manually in-analysis of the financial information. [15]
Finite, grows with new information, and gets enhanced as new data is received. [15]	Usually used infrequently; therefore, the changes must be made manually. [15]
They both can take care of complex and nonlinear relationships fairly well. [15]	Encompasses difficulty and multitasking. [15]
Can generate real-time predictions speedily.[15]	Delayed due to the conduct of manual analysis.[15]
Options to achieve higher accuracy because of pattern recognition. [15]	May have lower capacity of accuracy due to the involvement of the subjective judgments.[15]
Uses advanced models that enable the estimation of the advertisement success factor which might be hard to understand. [15]	Less complex than when using the other transportation models and also easier to interpret.[15]
Cannot be performed directly by users	Often demand less technical know-how than other

themselves as it demands computational resources along with data proficiency.[15]	types of work.[15]

Popular Machine Learning Techniques Used

1. Supervised Learning:

Supervised learning is the kind of machine learning that involves operating on well "labelled" training data and making predictions on the basis of that data. This means that certain of the input data is already classified with the right output or we're referred to as labeled data. The outcome of supervised learning is that in the training process, the data given to the work of machines act as the supervisor to help the machines make the right prediction. It uses the principles as a student uses in the learning process under the direction of the teacher.[16]

In the real world, the concept of supervised learning in Risk Assessment, Image classification, Fraud Detection, spam filtering, etc.[16]

Regression

Regression algorithms are employed in cases when the input variable is somehow connected with the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc. Popular Regression algorithms which come under supervised learning are Linear Regression, Regression Trees, Non-Linear Regression, Bayesian Linear Regression, Polynomial Regression

Classification

In cases where the output variable represents a categorical construct, then classification algorithms will be applied and the classes most common are 2, including for example Yes-No, Male-Female, True-False , Spam Filtering etc. [16]. Random Forest, Decision Trees, Logistic Regression, Support vector Machines algorithm comes under classification.

Advantages of Supervised learning:

- Using a supervised learning approach, it becomes much easier to determine the output since it is a function of the previous experiences.
- Supervised learning can help to have the perfect notion of the classes of objects to make predictions.[16]
- Supervised learning model enables you to solve several practical tasks like fraud detection, spam filtering, etc[16]

Disadvantages of supervised learning:

- For complex tasks, the use of supervised learning models is not possible.[16]
- Supervised learning misses the capacity to predict the right output in case the test data set is other than the training data set.
- Training required a lot of computation times.[16]

- The requirement of sufficient knowledge about the classes of objects is a prerequisite of supervised learning.[16]

2. Unsupervised Learning:

Clustering

Clustering therefore involves partitioning the objects in such a way that the objects with the highest degree of similarity are put into a group and do not have any, or least, similarity with objects in another group. Cluster analysis makes several groups out of the data objects based on similarities, and the presence or absence of similarities between the data objects.[16]

Association

An association rule serves as an unsupervised learning model with prioritized application in ventures used for defining relations between the variables stored in a massive database. It defines what items co-occur in the dataset which is important for facet modeling. Marketing strategy is made more effective by association rule. For example, clients who purchase X product (say a bread) are also likely to purchase a Y product (Butter or Jam). An excellent example of Association rule is Market Basket Analysis.[16]

Unsupervised learning algorithms includes K-means clustering, KNN (k-nearest neighbors), Hierarchal clustering, Anomaly detection, Neural Networks, Principle Component Analysis, Independent Component Analysis, Apriori algorithm, Singular value decomposition

Advantages of Unsupervised Learning

- Unsupervised learning is employed in the more complicated tasks than in supervised learning since in the latter problem there are no input data labeled.[16]
- Unsupervised learning is better because it is easy to obtain data that are not labeled as compared to those that are labeled.[16]

Disadvantages of Unsupervised Learning

- Unsupervised learning is inherently more challenging than supervised learning since the latter has its corresponding output.
- The final of unsupervised learning algorithms can be less precise because input data is not marked and algorithms do not know the right output at the same time.[16]

- **Reinforcement Learning**:

Reinforcement Learning (RL) is a subfield of machine learning that operates based on selecting actions in a particular environment in order to achieve the greatest amount of overall reward possible. While in supervised learning there is a training dataset with answers to which the algorithm has to approximate, in RL there is no training dataset. In RL, an agent adapts to gain a certain goal in an arbitrary, possibly intricate surroundings where he takes some action and gets specific feedback either in the form of a reward or a

penalty. RL solves a specific type of problem where decision making is sequential, and the goal is long-term, such as game-playing, robotics, etc.[16]

RL algorithms like Q-learning or Deep Q-Networks (DQN) can be applied to develop trading strategies that maximize returns by learning optimal actions over time.[16]

- **Deep Learning**:

Discuss the use of advanced models like Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN) for time-series forecasting and stock price prediction.

Recurrent Neural Network (RNN)

Recurrent neural networks are another type of feed-forward network, which are described in detail in the forthcoming section. Here all the neurons existing in the hidden layers get an input with a certain amount of delay in time. By definition, the use of the Recurrent neural network is highly dependent on the previous information of existing iterations. For instance, in any given value of the position of words in any sentence, they require a pointing on the former words. Not only does it integrate and algorithmic the inputs but also provides the length and weights across time. It does not allow the size of the model to grow with the size of inputs fed to the model. However, the function of this recurrent neural network has a limitation of slow computation rate and no provision of future input to the current state. It has an issue of back-referencing previous information.

Recurrent Neural Network used for Machine Translation, Robot Control, Time Series Prediction, Speech Recognition, Speech Synthesis, Time Series Anomaly Detection, Rhythm Learning, Music Composition

Long Short-Term Memory (LSTM)

Released by Hochreiter & Schmidhuber, Long Short-Term Memory is an updated version of recurrent neural networks. Traditional RNN works with one hidden state which is propagated across the time steps and makes it hard for this network to capture long term dependencies. LSTMs model address this problem by introducing a memory cell, which is the vessel that can hold information for as long as possible.

Therefore, LSTM frameworks have the ability to address difficult issues involving long-term dependency studying of sequential data for different assignments like language translation, speech recognition and time series predictions.

Convolutional Neural Networks (CNN)

Convolutional Neural Networks are a special kind of neural network mainly used for image classification, clustering of images and object recognition. DNNs enable unsupervised construction of hierarchical image representations. To achieve the best accuracy, deep convolutional neural networks are preferred more than any other neural network.

Convolutional Neural Networks used for Identify Faces, Street Signs, Tumors, Image Recognition, Video Analysis, NLP, Anomaly Detection, Drug Discovery, Checkers Game, Time Series Forecasting.

Challenges and Limitations of Machine Learning Models

Data Quality and Quantity:
Any machine learning model requires a massive amount of accurate data to learn and develop from. Lack of quality data or lack of data at all results in making wrong forecasts.
Overfitting and Underfitting:
Getting too intimate with the noise is termed overfitting and making the model too simplistic to learn the data patterns is termed underfitting.
Model Interpretability:
Advanced solutions, such as deep learning, are primarily opaque, which means that it is challenging to understand the basis for the decision made in a particular situation P Tanzania, essential in the provision of health care, for instance.
Computational Power:
Training Large models are highly demanding, computationally and, thus, expensive or at least beyond the resources of many users.
Bias and Fairness:
Such models built on biased datasets may produce bias based results and hence require constant checking plus application of bias reduction methods.
Scalability:
When dealing with large data sets, machine learning models become harder to scale up, and some of the methods that may need to be deployed are; distributed learning.
Data Privacy and Security:
There is usually a privacy issue when one needs to use personal or sensitive data, hence the need for privacy mechanisms such as Differential privacy.

Future Directions

However, future developments in applying ML for stock market prediction can be grouped in several directions. Current work is being conducted to enhance the usefulness of the models by researching better algorithms, using data other than accounts and experimenting with combinations of models. One of the main issues for accomplishing the set task is to obtain accurate real-time forecasts, as it is challenging to make decisions quickly based on large amounts of data to respond to market changes within minutes. Moreover, there is an ethical concern regarding the manipulation of the market through algorithms and using AI/ML effect, thus a need for reconsideration of the current laws governing algorithms for the fairness of the automated system. Juggling between driving innovation and maintaining the deserved ethical concerns on the development of ML for financial institutions, will then be central towards responsible development of this technology.

References

1. Stock Market Prediction https://en.wikipedia.org/wiki/Stock_market_prediction
2. Forecasting: What It Is, How It's Used in Business and Investing https://www.investopedia.com/terms/f/forecasting.asp
3. Andrew W. Lo; Jasmina Hasanhodzic (2010). *The Evolution of Technical Analysis: Financial Prediction from Babylonian Tablets to Bloomberg Terminals*. Bloomberg Press. p. 150. ISBN 978-1576603499. Retrieved 8 August 2011.
4. "Buffett Indicator: Where Are We with Market Valuations?".
5. Mislinski, Jill (3 March 2020). "Market Cap to GDP: An Updated Look at the Buffett Valuation Indicator".
6. "Warren Buffett On The Stock Market What's in the future for investors--another roaring bull market or more upset stomach? Amazingly, the answer may come down to three simple factors. Here, the world's most celebrated investor talks about what really makes the market tick--and whether that ticking should make you nervous. - December 10, 2001". *archive.fortune.com*. Fortune Magazine. 2001. Archived from the original on 8 March 2020.
7. Irwin, Scott H.; Park, Cheol-Ho (2007). "What Do We Know About the Profitability of Technical Analysis?". *Journal of Economic Surveys*. **21** (4): 786–826. doi:10.1111/j.1467-6419.2007.00519.x. S2CID 154488391.
8. Osler, Karen (July 2000). "Support for Resistance: Technical Analysis and Intraday Exchange Rates," FRBNY Economic Policy Review (abstract and paper here).
9. Lo, Andrew W.; Mamaysky, Harry; Wang, Jiang (2000). "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation". *Journal of Finance*. **55** (4): 1705–1765. CiteSeerX 10.1.1.134.1546. doi:10.1111/0022-1082.00265.
10. Investopedia - Forecasting in Investing https://www.investopedia.com/terms/f/forecasting.asp
11. What is machine learning? https://www.geeksforgeeks.org/ml-machine-learning/#what-is-machine-learning
12. NIPS 2014 Workshop on Machine Learning for Trading: https://nips.cc/virtual/2014/events/workshop
13. A Survey on Machine Learning Techniques for Stock Market Prediction: https://arxiv.org/abs/2212.12717
14. Machine Learning in Quantitative Finance: A Review: https://www.sciencedirect.com/journal/machine-learning-with-applications/special-issue/10B54VDFBX1
15. Machine Learning Forecasting vs Traditional Methods: Which is better? https://datakulture.com/blog/why-is-machine-learning-forecasting-better-in-traditional-methods/#:~:text=ML%20forecasting%20models%20have%20always,are%20exposed%20to%20different%20datasets.&text=The%20prediction%20looks%20best%20with,%2C%20social%20media%20trends%2C%20etc.
16. Machine Learning: What It is, Tutorial, Definition, Types https://www.javatpoint.com/machine-learning
17. 7 Major Challenges Faced By Machine Learning https://www.geeksforgeeks.org/7-major-challenges-faced-by-machine-learning-professionals/