Progress in Machine Learning Techniques for Stock Market Movement Forecast

S. S. S. Shameem[1](✉) and Sonal Sachin Deshmukh[2]

[1] Manipal Institute of Technology, MAHE, Bangalore, Karnataka, India
shameem.u4@gmail.com
[2] Jawaharlal Nehru Engineering College, MGM University, Aurangabad, Maharashtra, India

Abstract. Data-driven accurate stock market models can lead to timely, better decision making by the investors for a more profitable transaction. Such models can increase the chances of selecting more profitable stocks and reduce risk by avoiding risky investment. Last few decades of advances in soft-computing techniques in machine learning (ML), deep learning (DL), text mining (TM), and ensemble methods have positively reflected in the forecasting of stock market as well. In our work, we have reviewed some recent machine learning models for stock market forecasting. We have considered the works that cover various types for data sources, forecasting techniques, and efficient evaluation metrics. With our paper, we aim to provide a brief idea on the latest progress in stock market forecasting. We also summarize our analysis to highlight future research scopes in stock market movement forecasting.

Keywords: Machine Learning · Stock Market Prediction · Literature Review · Financial Modelling

1 Introduction

In this twenty first century, a country's economy and financial growth are deeply interrelated with the financial market and stock-prices of top-listed companies. Its hence very crucial to study, analyze and understand the financial market. Over the last few decades, there has been exponential growth in the interest of investors in the stock market. The global digitization has empowered millions of common-man to be potential investors and active player in stock market, with a deep desire of making good profit from the stock market which witnesses assets worth billions of dollars traded on a daily basis on stock exchanges. A more accurate forecast of the stock market behavior would definitely enable these common, as well as institutional, investors to consistently make low-risk, high-return portfolios [1]. A number of factors and uncertainties (like general economic conditions, global events, socio-political factors, natural disasters) makes it highly difficult to make prediction in financial markets [2, 3]. This has been a major motivation for the use of artificial intelligent techniques to build more accurate models for stock market forecasting.

© The Author(s) 2023
S. Tamane et al. (Eds.): ICAMIDA 2022, ACSR 105, pp. 69–77, 2023.
https://doi.org/10.2991/978-94-6463-136-4_9

Stock market forecasting is considered as a very relevant but challenging tasks in financial and computational research field. Fama's efficient-market hypothesis (EMH) suggests that, stock price behavior is very random, making it impossible to predict changes in stock direction and/or magnitude. They proposed 3 categories of forms of prediction efficiency. In the weak form its proposed that, past price movements can't be used to predict future price. Semi-strong form tells that, neither past price movements nor any public information is relevant for market prediction. And, the strong-form declines the use of either public or private information to forecast the market. Contrary to Fama's hypothesis, a large set of research community has been continuously proposing various models to forecast stock prices, and to some extend the results are in their favor. For their analysis & market forecasting, they consider many factors that can be majorly categorized as fundamental analysis, technical analysis, & macro-economic. Fundamental analysis covers underlying factors that affect company's overall performance and financial records. Technical analysis mainly comprised of historical prices and volumes.

In this paper, we present a review of the recent works done in field of stock market forecasting based on fundamental and technical analysis. The remaining sections of this paper are organized as follows. Section 2 provides insight to stock market, data sources, market forecasting, technical and fundamental analysis, machine-learning algorithms, and performance evaluation of predictive model. Section 3 presents a summary of recent review works. Section 4 presents the review methodology that we have adopted for our work. Finally, Sect. 5 concludes this work while also outlining future research opportunities.

2 Overview of Stock Market Data Sources and Machine Learning Techniques Used

Stocks and other financial instruments of public listed companies are traded in stock market. Investor bids to forecast the price movement direction for particular stock, and aims to make profit with the prediction accuracy. Such predictions are based on historical data analysis, company's internal financial information, and other environmental/policy related factors [4].

2.1 Types of Data Sources

The popularly considered approaches for stock market forecasting are categorized as fundamental analysis and technical analysis (see Fig. 1).

- **Technical Analysis** is the most common approach that uses stock prices (& other derived indicators) as input. Its popularly believed that all other information (news, macroeconomic variables, etc.) are inherently represented in the stock price itself. So, price pattern analysis is enough to predict stock movement. Technical indicators are popularly used stock signals to buy, sell or hold a stock.

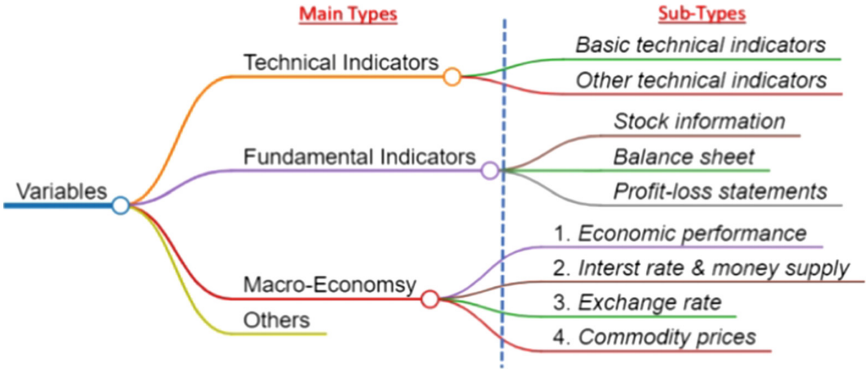

Fig. 1. Variable categories for stock price prediction [5]

- **Fundamental analysis** is harder to build models to capture stock fluctuation pattern, It commonly includes company information and other related macroeconomic (like interest rates, currency exchange rates, GDP, commodity prices, customer price index, etc.). These are unstructured in nature and shows non-continuous behavior. Text mining, sentiment analysis, and other natural language processing techniques have been applied to handle their complexity.

2.2 Technical Analysis

In this review, we are limiting our work to technical analysis only, and not exposing to any form of natural language processing-based predictions. According to most prediction studies, technical indicators are applied as input variables, since they have provide buy or sell signals. These technical indicators (TI) are categorized as: basic TI and other TI.

Basic technical indicators: Basic TI are a set of variables for important information of stocks related to price and volume traded [5]. Such variables include the open price and close price of stock in last trading time-period (minute/hour/day/week), high and low prices during that time period, and volume of stock traded during this period, and other variations in these variables. All these basic TI are broadly categorized into seven fields: Close, Volume, Range, Open Price, High Price, Low Price and Bid Price, various prices with lag-1 (close, high, low, open prices with lag-1), and the volume traded in a single time period are some commonly used basic variables. These lag-1 represent most recent information for next time-period, which makes these five variables more commonly used than other higher lag variables.

Datetime	Open	High	Low	Close	Volume
2021-07-02 09:30:00-04:00	272.820007	274.570007	272.500000	274.459991	2064942
2021-07-02 09:40:00-04:00	274.359985	275.179993	274.179993	275.000000	1198341
2021-07-02 09:50:00-04:00	275.000000	275.056091	274.709991	274.750000	646755
2021-07-02 10:00:00-04:00	274.739990	275.039886	274.434998	275.015015	672251
2021-07-02 10:10:00-04:00	275.010010	275.049988	274.559998	274.720001	405374

Fig. 2. Stock Price dataset (sample) [12]

SYMBOL	OPEN	HIGH	LOW	PREV. CLOSE	LTP	CHNG	%CHNG	VOLUME (shares)	VALUE (₹ Lakhs)	52W H	52W L
NIFTY 50	18,376.40	18,399.45	18,327.20	18,349.70	18,335.05	-14.65	-0.08	12,90,36,891	9,06,478.97	18,362.30	15,183.40
HINDALCO	435.00	452.85	434.05	429.85	450.05	20.20	4.70	1,39,60,607	62,603.55	636.00	308.95
TATAMOTORS	424.00	437.70	423.50	423.50	435.50	12.00	2.83	96,59,661	41,837.92	536.70	366.20
APOLLOHOSP	4,430.00	4,568.00	4,425.20	4,421.40	4,541.10	119.70	2.71	5,23,246	23,707.07	5,935.40	3,361.55
GRASIM	1,711.00	1,748.85	1,709.05	1,708.45	1,747.50	39.05	2.29	2,83,357	4,908.62	1,929.80	1,276.60
TATASTEEL	108.50	110.00	108.25	107.55	109.20	1.65	1.53	2,93,66,775	32,059.71	138.67	82.70

Fig. 3. NSE Nifty-50 Stock dataset (sample) [12]

Other technical indicators: Other technical indicators are acquired by processing the basic technical indicators. It comprises a large set of 1208 variables which are grouped into four major categories: Momentum Indicator (measures speed with which stock price is changing over time), Trend Indicator (focuses on direction and strength of change in stock), Volatility Indicator (measures the magnitude of change in variable during a certain trading time period), and Volume Indicator (focuses on volume associated with these changes).

Technical Chart: The raw data with respective timestamp for a particular stock is shown below (see Fig. 2). The complete information for a stock/exchange contains various fields is shown below (see Fig. 3). A Candlestick is a popular form of graphical representation of important basic technical indicators for a single traded period for better visual understanding (see Fig. 4). The technical chart for a stock movement is shown below (see Fig. 5). Data can be in form of daily or intraday stock price movement information. Popular timeframes for stock price movement observations include 5–15 min, 1 h and 1 day interval. It may further be transformed into any other time-frame suitable for analysis.

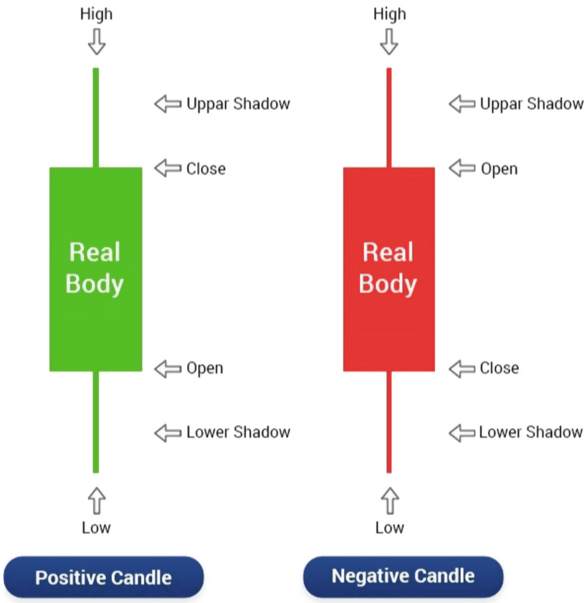

Fig. 4. Stock Price in visual representation (Candlestick) [13]

Trend indicators: Moving Averages and Momentum are most commonly used trend indicators (formulae are shown in Table 1). Simple Moving Average (SMA) summarizes performance of last trading period. Weighted moving average (WMA) weights recent prices more than other earlier prices [4]. Momentum indicator measures price differences over a time period. Momentum (MOM) is calculated as difference between two SMAs. Moving Average Convergence Divergence (MACD) is calculated as difference between two EMAs. Relative Strength Index (RSI), Commodity Channel Index (CCI), Williams R, and Stochastic Oscillators are most commonly used oscillators. Stochastic Oscillators (K and D) acts as a momentum indicator to highlight oversold or overbought stocks. Similar to K and D, RSI indicator takes values between 0 and 100 to alert overbought and oversold conditions.

2.3 Machine Learning

Machine learning techniques are applied to extract underlying pattern from available historical data. In stock market prediction supervised learning is widely used. A general framework of ML model for stock market forecast is shown below (see Fig. 6). The process starts with selecting historic data for the stock price in time-series format and/or other relevant information (e.g., stock information and financial news, etc.) for a particular time period. The collected data is pre-processed to clean it from noise, inconsistency, and missing values. Feature extraction, feature selection, feature transformation dimensionality reduction and data normalization are applied for better performance and accuracy.

Fig. 5. Indian Stock Market NIFTY-50 cd Chart [13]

Table 1. Technical Indicator equations [2]

Indicator	Equation
SMA_t	$\frac{C_t+C_{t-1}+\cdots+C_{t-9}}{n}$
WMA_t	$\frac{nC_t+(n-1)C_{t-1}+\cdots+C_{t-9}}{n+(n-1)+\cdots+1}$
K_t	$\frac{C_t-LL_{t-n}}{HH_{t-n}-LL_{t-n}} \times 100$
D_t	$\frac{\sum_{i=0}^{n} K_t}{n}$
MOM_t	$C_t + C_{t-10}$
RSI_t	$100 - \frac{100}{1+(\sum_{i=0}^{n} \frac{UP_{t-i}}{n})/(\sum_{i=0}^{n} \frac{DOWN_{t-i}}{n})}$
$MACD_t$	$EMA12\_t - EMA26\_t$
LW_t	$\frac{H_n-C_t}{H_n-L_n} \times 100$
AD_t	$\frac{H_t-C_{t-1}}{H_t-L_t}$
CCI_t	$\frac{M_t-SM_t}{0.015D_t}$

C_t: closing stock price for time t, H_n: highest stock price, L_n: lowest stock price.

Model evaluation: Model evaluation is important to check the model accuracy. Some commonly used performance metrics include: mean absolute error (MAE), mean square error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) etc. [4, 6] are shown in Table 2.

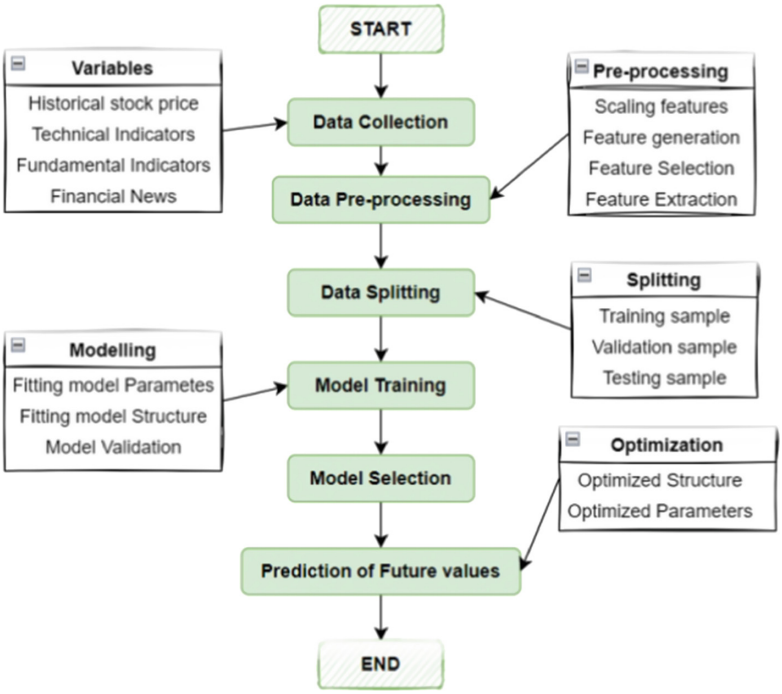

Fig. 6. Workflow of stock market prediction model

3 Related Work

Initially, for this review paper, we focused on survey articles on stock market prediction. Some of the works review relevant ML techniques, and others focused on specific technique that includes Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Text Mining, etc. In their work [7], researchers reviewed articles by observing that neural networks and fuzzy networks deliver better results for stock market forecasting. They review the predictive variables, modeling techniques, type of transfer function, benchmarks, and performance measures, etc. The comparative research review paper [3] focus on artificial intelligence techniques like artificial neural network (ANN), expert system (ES) and hybrid intelligence systems which report high percentage of accuracy and seem to be more superior to those of traditional & statistical methods in dealing with financial market related predictions. Review work of [8] and [9] focus on text-mining based stock market prediction to handle the unstructured type of textual news-based information. The latest advances in social network analysis makes it possible to include news-driven sentiments and emotions as a crucial predictive input in model building. Their analysis covers different type of text sources employed, natural language processing (NLP) algorithms, and reported results to categorize and standardize the evaluation metrics. In [2, 4, 10] and [11] review of technical, fundamental, and combined analysis techniques by grouping them based on dataset nature, data sources usage, data timeframe, machine learning algorithms used, performance evaluation metrics are focused.

Table 2. Performance metrics equations [4]

Metric	Equation
R	$R = \frac{\sum_{i=1}^{n}(t_i-\bar{t})(y_i-\bar{y})}{\sqrt{\left(\sum_{i=1}^{n}(t_i-\bar{t})^2\right).(\sum_{i=1}^{n}(y_i-\bar{y})^2)}}$
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(t_i - y_i)}$
MAPE	$\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{t_i-y_i}{t_i}\right\rvert$
MAE	$\frac{1}{n}\sum_{i=1}^{n}(t_i - y_i)$
Accuracy	$\frac{TR+TF}{TR+FR+FF+TF} \times 100$
Precision	$P^R = \frac{TR}{TR+FR} \times 100$
Recall	$R^R = \frac{TR}{TR+FF} \times 100$
F-score	$\frac{2\times P^R \times R^R}{P^R+R^R}$

They find ANN, SVM and hybrid ensemble algorithms to have comparative advantage for stock forecasting.

4 Review Methodology

In this review work, we specifically focused only on other primary review works published in last 15 years. Primary scientific paper databases like Web of Science and Scopus were considered for literature selection. Articles with the relevant filter keywords as stock market, financial market forecast, prediction, etc. were considered. A total of 10 published review work were covered for this research on stock prediction.

5 Conclusions

From the literature study so far, many gaps were observed as opportunities for future studies. We realize the importance and vacuum in the field of algo-trading, specifically for Indian stock market. Focus on the performance of ensemble techniques for stock prediction offers a lot of research scope. In a live robust environment, predictive modelling can play pivot role to build the bots that take quick and accurate decision for high frequency trading. Through this research work, we are hopeful of contributing in this field of computational data science for stock forecasting.

References

1. “NIFM Report on Algo trading”.

2. O. Bustos and A. Pomares-Quimbaya, "Stock market movement forecast: A Systematic review," *Expert Systems with Applications*, vol. 156. Elsevier Ltd, Oct. 15, 2020. doi: https://doi.org/10.1016/j.eswa.2020.113464.
3. A. Bahrammirzaee, "A comparative survey of artificial intelligence applications in finance: Artificial neural networks, expert system and hybrid intelligent systems," *Neural Comput Appl*, vol. 19, no. 8, pp. 1165–1195, 2010, doi: https://doi.org/10.1007/s00521-010-0362-z.
4. I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," *Artif Intell Rev*, vol. 53, no. 4, pp. 3007–3057, Apr. 2020, doi: https://doi.org/10.1007/s10462-019-09754-z.
5. M. M. Kumbure, C. Lohrmann, P. Luukka, and J. Porras, "Machine learning techniques and data for stock market forecasting: A literature review," *Expert Systems with Applications*, vol. 197. Elsevier Ltd, Jul. 01, 2022. doi: https://doi.org/10.1016/j.eswa.2022.116659.
6. M. Umer, M. Awais, and M. Muzammul, "Stock Market Prediction Using Machine Learning(ML)Algorithms," *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 8, no. 4, pp. 97–116, Sep. 2019, doi: https://doi.org/10.14201/ADCAIJ20198497116.
7. G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques - Part II: Soft computing methods," *Expert Syst Appl*, vol. 36, no. 3 PART 2, pp. 5932–5941, 2009, doi: https://doi.org/10.1016/j.eswa.2008.07.006.
8. A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, "Text mining for market prediction: A systematic review," *Expert Systems with Applications*, vol. 41, no. 16. Elsevier Ltd, pp. 7653–7670, Nov. 15, 2014. doi: https://doi.org/10.1016/j.eswa.2014.06.009.
9. F. Z. Xing, E. Cambria, and R. E. Welsch, "Natural language based financial forecasting: a survey," *Artif Intell Rev*, vol. 50, no. 1, pp. 49–73, Jun. 2018, doi: https://doi.org/10.1007/s10462-017-9588-9.
10. T. J. Strader, J. J. Rozycki, T. H. Root, and Y.-H. Huang, "Machine Learning Stock Market Prediction Studies: Review and Research Directions," 2020.
11. D. Kumar, P. K. Sarangi, and R. Verma, "A systematic review of stock market prediction using machine learning and statistical techniques," in *Materials Today: Proceedings*, 2020, vol. 49, pp. 3187–3191. doi: https://doi.org/10.1016/j.matpr.2020.11.399.
12. NSE India, https://www.nseindia.com/, last accessed 2022/10/23.
13. Investing.com, https://in.investing.com/, last accessed 2022/10/25.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.