The Ensemble Model for Long-Term Stock Investment Is Based on Sentiment Analysis And Technical Analysis

Gauravkumarsingh Gaharwar[1] and Dr. Sharnil Pandya[2]
[1]Navrachana University, Vadodara, Gujarat, India
[2]Symbiosis International University, Pune, Maharashtra, India

ABSTRACT

Investors seek profitable possibilities while reducing risk in the complex stock market. Investors must find high-quality stocks with long-term growth potential. Traditional basic analysis can reveal a company's financial health, but it may not reflect short-term market sentiments. Stock selection using technical analysis and ensemble learning. Technical analysis, which analyses past price and volume data, is integrated with ensemble learning algorithms to improve stock identification. Using neural networks, gradient boosting, random forests, and decision trees, the proposed method makes investment decisions. The ensemble model's feature engineering, data preparation, and model optimization contribute to resilient performance. Moving averages, relative strength indices, and MACD are used to derive input features for the ensemble model from past stock prices and volume. Ensemble learning mitigates the drawbacks of several algorithms and reduces single-model prediction risk. The model uses various technical indicators and learning algorithms to find high-quality stocks with positive price movements and development prospects. The ensemble learning model can identify high-quality shares in various market scenarios. Compared to typical stock selection methods, high-quality commodity identification is more accurate. Technical analysis and ensemble learning are used to create a data-driven stock selection strategy for finance and investment. This approach can help investors and financial professionals make better investment decisions, boosting stock market portfolio performance and risk management.

KEYWORDS: Technical, Utilized, Ensemble, Stocks market

1. INTRODUCTION

The performance of a country's financial market plays a vital role in assessing its overall economic situation, allowing economists and financial professionals to evaluate the country's present economic well-being. Of all the financial markets, the stock market is particularly prominent and influential. The economic condition of a nation has a direct or indirect influence on several industries like finance, agriculture, metallurgy, and investment banking, among others. Financial markets encompass various types of market, including stock markets, derivatives markets, bond markets, and commodity markets. Stock prices fluctuate due to many reasons in the stochastic and demanding stock market [1, 2]. Every day, with the global stock market producing billions of structured and unstructured data points, its "volume", "velocity", "variety", and "veracity" increase, making evaluation more difficult [3]. Analysing stock market "Big Data" involves two ways, fundamental analysis and technical analysis, are generally accepted: fundamental analysis and technical analysis. National and international economic trends, public opinion, corporate assets and financial statements, political environments, and international business associations are the primary foci of the fundamental study [4]. The foundation of technical analysis is a statistical examination of stock price fluctuations over time. Moving average, dead cross, and golden cross are technical indicators that aid investors in making money in the stock market. Even with these methods, market analysis is complicated.

Technologies of machine learning and soft computing have surmounted obstacles in stock market trading, forecasting, and analysis. Artificial neural networks include Simple Bayes [6, 7], ANNs, Decision Trees, SVM, and neural network techniques [5] as examples. It has been determined that the accuracy and prediction error of logistic regression (LR) for stock market forecasting is inferior to that of [8, 9]. Ensemble learning (EL) outperforms individual classifiers and regressors [10, 11] via the integration of various learning techniques. It generates committees that improve predictions (stacking and blending), reduce variance (bagging), and eradicate boosting (bypassing).

Financial and intellectual economics, there is a growing interest in predicting stock prices. No novel approaches, schemes, or methods that have been developed thus far have proven to be highly effective. Modelling the stock market is difficult due to the numerous uncertain variables and market volatility. This data is also hard to anticipate because to its non-linear, non-stationary, and strong heteroscedasticity [11,12].

According to the efficient market hypothesis (EMH), past and future events are already accounted for in asset prices. The EMH claims that historical data cannot predict future pricing because it requires confidential information. Investors' psychological errors when faced with uncertainty lead to irrational and unpredictable behavior, according to EMH opponents [13]. No consensus exists on EMH, and discussion continues.

The efficient market hypothesis (EMH), which asserts that markets are driven more by fear and avarice than by logic, has been criticized for behavioral reasons. The adaptive markets hypothesis (AMH) was devised in response. AMH takes a biological approach to the stock market, employing an evolutionary theory positing that natural selection, adaptation, and competition cause price fluctuations in response to financial interactions.

According to AMH, sporadic periods of predictable patterns may emerge over time.

Technical analysis and fundamental analysis are the two most common methods for analyzing stock market data [14]. The concept of intrinsic value, which asserts that both quantitative and qualitative data are used to determine the current price, is the basis of fundamental analysis. This approach utilizes the EMH over time. In contrast, it suggests that inefficiencies may occur shortly. In contrast, technical analysis requires the use of specialized methodologies, makes predictions about future changes in stock prices by analyzing past data to discover patterns [15].

Due to the inherent unpredictability of market values, predicting trends in the stock market is generally regarded as a difficult endeavor. Supply and demand, government, economics, politics, and international trade drive Bulls and Bears. An intelligent prediction paradigm is needed due to the problem's seeming complexity. People predict stock market bulls and bears and trade using advanced algorithms and tactics. Intelligent prediction relies on intelligent problem-solving.

For this complex financial forecasting problem, it trains a subset of categorization models of labelled financial data that are classified as 1+ or -1 based on whether stock prices changed positively or negatively between initial and subsequent times. Two major financial forecasting advances are made in this article. The learning model's feature selection is a priori and automated, allowing it to "discover" which parameters are necessary for accurate prediction rather than accepting human-designated explanatory elements. This allows the ensemble to readjust its parameters based on GICS or location and time. Second, the ensemble's probabilistic ranking component approximates prediction confidence levels. Financial literature often formulates portfolios using a scoring model to distinguish favorable and undesirable equities. Instead of an absolute stock preference hierarchy, the ensemble returns a probabilistic stock desirability score. Probabilistic confidence criteria help in uncertain games like portfolio investment [16].

This study examines the use of Machine Learning approaches in stock trading to forecast the future movement of stock values and identify the quality stocks. The study specifically examines the implementation of machine learning algorithms such as Support Vector Machines, Linear Regression, ANN, Ensemble method. The study presents the characteristics and variables that may be used to identify quality stocks, which can aid in predicting future stock trends. Additionally, it explores the integration of Boosting with other learning methods to enhance the accuracy of these prediction systems.

The originality of this study is followed as:

This study presents a new framework that combines the knowledge from several technical analysis indicators and machine learning models to find stocks of superior quality, in contrast to conventional techniques that depend only on fundamental analysis or single-model approaches. The proposed model utilizes ensemble learning techniques. This approach provides a comprehensive and data-driven method for selecting stocks. This novel approach not only improves the precision and durability of stock forecasts but also offers significant understanding of market dynamics and trends, arming investors with practical knowledge for making well-informed investment choices. Furthermore, this study enhances the area of financial analytics by showcasing the efficacy of ensemble learning in amalgamating the advantages of various technical analysis indicators and models. As a result, it propels the advancement of stock selection techniques to a higher level. The main innovation of this research study is its interdisciplinary methodology that combines technical analysis, machine learning, and finance to create an advanced model for detecting high-quality equities in dynamic and intricate financial markets.

2. LITERATURE REVIEW

In the literature, technical analysis and ensemble learning are being used to create strong models for spotting excellent stocks. These models improve stock selection accuracy by combining technical indicators and classifiers, making them a potential strategy for lucrative investors.

Researchers have optimised technical analysis signal integration using ensemble techniques, developing excellent stock quality assessment tools.

2.1 Machine-Learning Techniques for stock Market Prediction and Ensemble Learning Strategies

Nti, Isaac et. al. (2020) [17] discussed stock-market forecast using machine learning to strive to improve accuracy. Various combination strategies have been used to predict stock markets with ensemble regressors and classifiers. When building ensemble classifiers and regressors, three risks arise. The basic regressor or classifier approach is the first issue. The second addresses combining strategies to build numerous the third is the ensemble number. There are few relevant studies on these issues. This study compared super learners (stacking), bagging, combining, and boosting in great detail. Using Neural Networks (NN), Support Vector Machines (SVM), and Decision Trees (DT), 25 ensembles of classifiers and regressors were created. GSE, JSE, and Bombay Stock Exchange stock data were compared to execution times, accuracy, and error measures. Stacking (0.0001–0.001) and blending (0.002–0.01) are more profitable for ensemble classifiers and more regressors than boosting (0.01-0.443) and bagging (0.01-0.11) in market analysis. As a result, ensemble algorithms should be applied to a new stock market trajectory research endeavour.

Asad, Muhammad (2015) [16] wanted a foolproof stock market prediction system. Machine learning methods like SVM extensively researched to anticipate market stock behaviour. It describes a machine learning-based portfolio trading approach. The strategy will be profitable if it can consistently detect stock indices, propose positive or negative returns, and allocate a portfolio using a trained model. Technical Indicators like MRA and data clustering are employed to train the system. Weighted SVM, RVM, random forest, and MLP classifiers form the trained model. A majority vote would determine the decision value. A boosting meta-algorithm enhances ensemble learning and a guided Relief algorithm selects features. ISE equities from Turkey are used to measure system performance. Ensemble committee produces fewer but compact rules and has a lower error rate than other methods.

Deshmukh et. al. (2023) [18] accurate stock market models that are driven by data can help investors make better, more timely decisions for more profitable transactions. These models can reduce risk by discouraging investments in riskier companies and increase the probability of selecting more profitable equities. Significant advances in soft computing techniques have aided the field of stock market forecasting over the past few decades, such as text mining (TM), deep learning (DL), machine learning (ML), and ensemble methods. In stock market forecasting models employing machine learning. Consideration has been given to the literature on various types of data sources, forecasting techniques, and effective assessment measures. It provides an overview of the most recent developments in stock market forecasting. It also summarises the analysis to emphasise opportunities for predicting stock market movement.

Shah, Aayush, et al. (2019) [19] proposed the significance of stock market price forecasting cannot be exaggerated. It aids investors and financial institutions in risk management, investment decisions, and financial system stability. Correct stock market estimates can help investors maximise returns and limit losses and help financial institutions manage risk. However, the stock market's complexity and many factors that affect stock values make predicting challenging. Deep learning is being used to evaluate massive volumes of data and gain stock market insights. Although deep learning can properly predict stock prices, much work remains.

Mohapatra, Sabyasachi, et al. (2022) [20] studied XGBoost, Gradient Boosting, AdaBoost, and Random Forest ensemble machine learning models to estimate Indian bank stock returns using technical indicators. Technical analysis' Price, Volume, and Turnover categories produce these indicators. The effectiveness of the models has been evaluated using MAE, MSE, MAPE, and RMSE. Out of four ensemble models, XGBoost performs best. Root-mean-square and absolute errors are 3–5%. Model feature importance graphs show how variables predict output. Using the suggested machine learning models, investors, speculators, and portfolio managers can make more accurate stock price forecasts. movements and returns, especially in banking stocks, so they are less dependent on macroeconomic factors. The methods let market players predict price-volume movements in all equities, independent of size, liquidity, turnover. Lastly, the approaches are robust as well as perform quite well at predicting trends, particularly for large deviations.

Nti, Isaac Kofi et.al (2020) [21] described every growing and flourishing economy relies heavily on the stock market, and every investment in the market aims to maximise profit while minimising risk. Many investigations use technical or basic analyses and soft computing formulas to forecast stock prices. Technical, fundamental, and integrated analyses were recognised in these publications. The quantity and quality of data sources served as the classification's foundation, data timeframe, task, accuracy, error, and modelling software based on machine learning. Technology accounted for 66% of reviewed documents, while fundamental and integrated analyses accounted for 23% and 11%. The majority of evaluated documents used one 8.2% and 2.46% utilised two and three data sources. The vast majority of stock market forecasting techniques employ artificial neural networks and support vector machines.

Sohangir, Sahar, et al. (2018) [22] studied big data analytics and deep learning are data science's main focus. Recently, deep learning models have demonstrated tremendous success in computer vision and voice recognition. Big data is essential for businesses that must collect enormous volumes of data, inclusive of social media One of the primary benefits of deep learning is its ability to analyse massive amounts of data. Deep learning is an essential tool for big data due to its utility. Deep learning may uncover astonishing insights from large data sets. The contemporary stock market exemplifies these social networks. Although they are a well-liked means of generating money, it is nevertheless challenging to know when to buy and sell shares. For nearly a decade, Salomon Brothers, Lehman Brothers, and Goldman Sachs dominated the financial advisory industry. Internet and social media sites for finance, such as Seeking Alpha and Stock Twits, have enabled investors from around the globe to communicate and exchange ideas. Experts on financial social networks are able to predict changes in the stock market, but what do the majority of these experts believe about various stocks? Deep Learning techniques may enhance Stock Twits' sentiment analysis. Various neural network models, including doc2vec, long short-term memory, and convolutional neural networks, were used to analyse the sentiments of the Stock Twits market.

The results indicate that the best model for predicting writers' sentiment in the Stock Twits dataset is a convolutional neural network, the application of a Deep Learning model to financial sentiment research can yield significant benefits.

Abe, Masaya et. al. (2018) [23] utilized neural networks and stock returns are forecast. In machine learning, Neural networks and stock returns are forecast. Deep learning is a machine learning technology that has recently received a great deal of attention due to its remarkable results in speech and image recognition. This study evaluates the efficacy of a deep learning technique for predicting stock returns on a cross-section of the Japanese stock market one month in advance. Deep neural networks are more effective than traditional machine learning. These results suggest that machine learning can use deep learning to predict cross-sectional stock return.

Krauss, Christopher et.al (2017) [24] described machine learning is gaining momentum: Deep learning permits numerous levels of abstraction and is replacing powerful tree-based methods that operate on the original feature space. Several industries, including finance, can utilise these methods. Deep neural networks (DNN), gradient-boosted trees (GBT), random forests (RAF), and an ENS mixture of these techniques are evaluated in the context of statistical arbitrage. To eliminate survivor bias, each model is trained using the returns of all S&P 500 equities without a lag. Daily trading signals were generated from 1992 to 2015 based on the likelihood that a stock would outperform the market. Converted long positions have the k highest probability, while short positions have the k lowest probability. Observational data indicates optimism. For k = 10, an ensemble consisting of a random forest, a gradient-boosted tree, and a deep neural network generates daily returns in excess of 0.45% before transaction costs.

Qiu, Xueheng et al. (2017) [25] studied that predicting stock prices is one of the most difficult aspects of time-series analysis due to the intrinsic nonlinearity and nonstationarity of the data. When forecasting the near-term cost of commodities, an ensemble method composed of v-Support Vector Regression (v-SVR) is used and the Empirical Mode Decomposition (EMD) algorithm is described. Initial identification of multiple intrinsic mode functions (IMFs) was accomplished by decomposing historical stock price data. Using a -SVR model, the expected IMF value for each IMF was then calculated. In the end, a stock price output was generated by aggregating the results of each IMF stock price estimate. The effectiveness of the proposed EMD-SVR method is evaluated using three stock market price datasets from the power industry. According to simulation results, this strategy is superior to six other forecasting techniques.

Table 2.1 Comparison of Reviews

Author and Year	Techniques	Key Findings
Nti, Isaac et al. (2020) [17]	Ensemble Learning, Decision Trees, Support Vector Machines, Neural Networks	Stacking and blending outperform bagging and boosting for stock market prediction.
Asad, Muhammad (2015) [16]	Weighted SVM, Relevance Vector Machine, Random Forest, MLP, Boosting	The ensemble committee produces compact rules with lower error rates.

Deshmukh et al. (2023) [18]	Soft computing techniques, ML, DL, Text Mining, Ensemble methods	Discusses advances in soft-computing techniques for stock market forecasting.
Shah, Aayush et al. (2019) [19]	Deep Learning, Stock Market Forecasting	Highlights the significance of stock market price forecasting with deep learning.
Mohapatra et al. (2022) [20]	Ensemble Machine Learning (XGBoost, Gradient Boosting, AdaBoost), Technical Indicators	XGBoost performs the best in predicting Indian bank stock returns.
Nti, Isaac Kofi et al. (2020) [21]	Machine Learning Algorithms, Technical, Fundamental, Combined Analysis	Classifies techniques based on data sources, algorithms, and accuracy metrics.
Sohangir et al. (2018) [22]	Deep Learning, Sentiment Analysis, Convolutional Neural Networks	Demonstrates the use of Deep Learning for financial sentiment analysis.
Abe, Masaya et al. (2018) [23]	Deep Learning, Neural Networks, Stock Return Prediction	Deep neural networks outperform shallow ones in stock return prediction.
Krauss et al. (2017) [24]	Deep Neural Networks, Gradient Boosted trees, Random Forest, Ensemble (ENS)	The ensemble of DNN, GBT, and RF generates significant returns in statistical arbitrage.
Qiu, Xueheng et al. (2017) [25]	Empirical Mode Decomposition, ν-SVR, Ensemble	Proposes an ensemble method for short-term stock price forecasting.

2.2 Ensemble Prediction Model for Stock Price Movement

Pasupulety, Ujjwal, et al. (2019) [26] discussed increasing numbers of financial companies use stochastic models to anticipate market behaviour due to AI. Stock return machine-learning algorithms are improved regularly by quantitative analysts. Regression models such as SVM and Random Forest predict the closing price with remarkable accuracy. utilising the aforementioned algorithms to develop a system for evaluating and projecting the stock value of a company. Preprocessed datasets from India's National Stock Exchange (NSE) include well-known technical indicators comprising fundamental market price data. Through feature selection, characteristics are prioritised based on their impact on the final price in order to reduce the size of the training dataset. Additionally, sentiment analysis is used to determine how the public feels about a business. A Word2Vec algorithm is trained to classify as positive or negative Twitter messages containing a company-specific hashtag. Next, we train our ensemble model using a new dataset that integrates a company's cumulative positive/negative tweets with technical indicator data. Experiments demonstrate that the kind and volume of training data influence the ensemble model's effectiveness. Technical indicator data and combined estimates of positive and negative tweets have no effect on the ensemble model.

Pathak, Ashish et. al. (2019) [27] studied the the stock market is an extremely erratic and unpredictable system with many levels and dimensions of influences affecting the trend's direction. EMH says the market is invincible. This makes predicting uptrends and downtrends difficult. It combines numerous methods to provide a more robust prediction model for investment-beneficial scenarios. In many circumstances, sentiment analysis and neural network algorithms are too limited and provide incorrect findings. Combining these methods gives More accurate and adaptable predictions from this model. Technical indicators help investors reduce risk and maximise returns.

Yang, Jian, et al. (2016) [28] discussed the essential for investors to accurately foresee the trend of stock price movement to generate enormous profits. However, it is difficult to generate accurate market forecasts due to the complexity of the financial environment. This type of prediction is a valuable application of machine learning. A successful investment strategy must include buy-and-sell signals, a highly accurate prediction model, and input quality options. Calculations of the maximal information coefficient (MIC) were used to determine which training elements were most crucial for predicting stock price changes. Constructed the SRA Voting ensemble prediction model for stock price trend movements (AB) utilising three exceptional classifiers: random forest (RF), support vector machine (SVM), and AdaBoost. Utilising technical indicators and indexes of the Chinese stock market, the proposed concept and method were validated. While SRAVoting outperformed SVM in terms of prediction accuracy, it did not always generate a higher annualised return than SVM-based buy-and-sell strategies. For various investment horizons, the SRAVoting-based methods may generate positive returns overall.

Gan, Kim Soon, et al. (2018) [29] examined the most important aspect of stock market forecasting, namely forecast precision. Several methods and algorithms have been proposed to improve the accuracy of forecasts. ANNs, or artificial neural networks, are a popular technique. The ANN algorithm, which draws inspiration from biology, simulates brain function using a network of artificial neurons. Even though artificial neural networks (ANN) have demonstrated promising forecasting outcomes, combining neural networks enhances prediction accuracy. Examine the predictions of single and multiple homogenous Announcement Networks (ANNs) for stock market closing prices. The empirical results indicate that ensembles of homogeneous ANNs outperform individual ANNs when predicting stock market prices.

Kohli, Pahul Preet Singh, et al. (2019) [30] described the complexity of market systems exceeds any individual's ability to forecast. To generate significant profits, however, Investors are responsible for accurately predicting stock market values. The ultimate objective of this initiative is to forecast activity on the Bombay Stock Exchange (BSE). The Bombay Stock Exchange (BSE) can be predicted using commodity prices (such as gold, silver, and petroleum oil prices), market history, and the foreign exchange rate (FEX). Next, the effectiveness of the models is compared to other standards. In addition, a hierarchical relationship was discovered between the employed qualities. It was discovered that the price of gold has the strongest correlation with market performance. The AdaBoost algorithm was the most effective of all methods.

Mabu, and Shingo (2015) [31] proposed an important pattern recognition subject with several ways to improve classification's generalisation capability. Ensemble learning is a technique for improving classification by constructing multiple classifiers and basing judgements on the sum of their classification outputs. It is essential to consider stock trading issues and market trends when deciding whether to purchase or sell equities. Trading techniques that adapt to varied trends help determine the best moment to purchase and sell given market conditions. To enhance the efficacy of the stock trading system, we introduce an evolving multi-layer perceptron with rule-based and ensemble learning. Using a rule-based evolutionary algorithm, Using this strategy, multiple rule pools for stock trading are created, the most successful rule pools are adaptively selected by MLP, and the selected rule pools then collaborate to make trading decisions. Simulations show that the suggested strategy outperforms no ensemble learning or buy-and-hold.

Wang, Qili, et al. (2019) [32] studied as stock markets in emerging nations grow, detecting stock price manipulation to protect regular investors is a major challenge. The experimental results of its machine learning stock price manipulation detection method were better than multivariate statistical methods. Statistically significant discrepancies between manipulated and unmanipulated securities have been disregarded by the manipulation detection algorithm. RNN-based ensemble learning (RNN-EL) is a novel technique that identifies instances of stock price manipulation using information about listed companies and trade-based features extracted from trading data. On the basis of China Securities Regulatory Commission (CSRC) reports of prosecuted manipulation instances, a specialised dataset consisting of labelled samples with trading data and identifying information was compiled for empirical studies. Using AUC, we determined that our method outperforms the best available techniques for detecting stock price manipulation by 29.8%. According to this study, government regulators can effectively identify dubious trading behaviours among vast quantities of trading activity in order to maintain fair trade.

Brzezinski et.al (2013) [33] proposed due to its widespread use in sensor networks, finance, and telecommunications, data stream mining is gaining popularity. Learning from data streams is hardest when dealing with idea drift, or unexpected changes in the stream's data distribution. Many concept drift categorization techniques exist, however most focus on one type of change. Accuracy Updated Ensemble (AUE2) is an additional new data stream classifier that takes drift into account. AUE2 combines the iterative structure of Hoeffding Trees with the accurate weighting mechanisms of block-based ensembles. Eleven novel stream techniques were empirically compared to the proposed algorithm under a variety of drift conditions. Among these methods were hybrid approaches, block-based and online ensembles, and singular classifiers. AUE2 obtained the highest average classification accuracy among all examined algorithms and used less memory than ensemble methods. The results of experiments indicate that AUE2 operates under a variety of drift and static conditions.

Agrawal, Manish et. al. (2019) [34] described stock market volatility and complexity. TA is often used on past price data by technical analysts, which is a time-consuming endeavour that may result in inaccurate forecasts. The combination of machine learning and fundamental and technical analysis provides satisfactory results. Learning with O-LSTM and adaptable Stock Technical

Indicators, an attempt is made in this work to forecast the price and price trend of equities. We also evaluated the model for making end-of-day buy-sell decisions. To optimise the deep learning task, we employed the Correlation-Tensor concept constructed with suitable STIs. The model receives the adaptive indicators tensor to increase prediction accuracy. Using predefined metrics, an existing deep learning classifier and two benchmark machine learning classifiers are compared. With a mean prediction accuracy of 59.25%, the proposed model significantly outperforms benchmark methods.

Agrawal, Manish et.al (2021) [35] studied stock Prediction Using a Deep Learning Model and Technical Indicators. 70, 287-304 Cmc -Tech Science Press-. 10.32604/cmc.2022.014637. Forecasting stock market trends is a hot topic and a difficult research task due to its volatility and dynamic character. In non-stationary stock data, attributes are uncorrelated. Stock market predictions may be inaccurate with traditional STIs. Using STIs, robust algorithms are developed to analyse the stock market and make prudent trading decisions. Based on STIs, EDLMs predict stock prices. The model demonstrates the Correlation-Tensor concept via Deep Learning (DL). The dataset of the three primary banks from the NSE-India cattle market is analysed using LSTM. STI-stock price correlations were extensively studied. EDLM-built model outperformed two benchmark machine learning models and deep learning model. On HDFC, Yes Bank, and SBI datasets, the proposed model provides trend-based forecasting and predicts with 63.59 %, 56.25 %, and 57.95 % accuracy, helping investors make lucrative investment decisions. The suggested EDLA with STIs often outperforms state-of-the-art algorithms.

Table 2.2: Comparison of Reviews

Author and Year	**Techniques Used**	**Key Findings**
Pasupulety et al. (2019) [26]	Support Vector Machine (SVM) and Random forest for stock price prediction. Ensemble model combining technical indicators and sentiment analysis.	Ensemble model performance varies with training data, and the effect of combining technical indicators with sentiment analysis is negligible.
Pathak et al. (2019) [27]	Combining sentiment analysis, neural network techniques, and technical indicators for stock market prediction.	Combining various techniques can provide more accurate and adaptable stock market predictions. Technical indicators help minimize risk.
Yang et al. (2016) [28]	Feature selection using maximal information coefficient (MIC). Ensemble prediction model (SRA Voting) based on SVM, Random Forest, and AdaBoost.	The SRA Voting model achieves higher prediction accuracy than SVM, but not necessarily higher annualized returns.

Gan et al. (2018) [29]	Comparison of single Artificial Neural Network (ANN) vs. multiple homogeneous ANNs for predicting stock prices.	A homogeneous ensemble ANN performs better at predicting stock prices than a single ANN.
Kohli et al. (2019) [30]	Use of input attributes like commodity prices, market history, and foreign exchange rates for stock market prediction. AdaBoost algorithm performs the best.	The correlation between gold price and market performance is the strongest. AdaBoost algorithm outperforms other methods.
Mabu (2015) [31]	Ensemble the rule-based evolutionary algorithm's stock trading learning process employing MLP.	Ensemble learning improves stock trading system performance compared to buy-and-hold.
Wang et al. (2019) [32]	RNN- based ensemble leaning framework for detecting stock price manipulation.	The proposed method outperforms other approaches for detecting manipulation an average of 29.8 percent.
Brzezinski et al. (2013) [33]	AUE2, an accuracy-updated data stream mining ensemble, handles drift.	AUE2 achieves the highest average classification accuracy among compared algorithms while using less memory.
Agrawal et al. (2019) [34]	LSTM deep learning and adaptive STIs for stock price prediction.	Proposed model has medium prediction accuracy of 59.25%, significantly higher than benchmark approaches.
Agrawal et al. (2021) [35]	Development of an Evolutionary Deep Learning Model (EDLM) using STIs for stock trend price identification.	EDLM outperforms benchmark machine learning models and a deep learning model in stock trend prediction.

2.3 Advancements in Stock Market Prediction Using Machine learning techniques

Rouf, Nusrat et. al. (2021) [36] discussed the a decade of Methodologies, Recent Developments, and Future Directions for Predicting the Stock Market with Machine Learning. 10.3390/electronics10212717. Electronics, 10, 2717. With the emergence of technical marvels such as global digitalization, which has transformed the traditional trading model, the era of technologically sophisticated stock market forecasting has arrived. Constant growth in market capitalization has made stock trading quite popular among investors. Many experts have developed algorithms to predict stock price fluctuations and help investors make smart investing decisions. Using non-traditional social media text, sophisticated trading models allow researchers to predict the market. Utilising sophisticated machine learning techniques such as ensemble methods and text data analytics has considerably improved the precision of predictions. Due to the volatile, irregular, and unpredictability of the data, it is difficult to forecast and analyse the stock market. This article discusses machine learning techniques and frameworks for stock market prediction. In addition, a comprehensive comparison study was conducted to determine the direction of importance.

Sun et al. (2012) [37] designed once unbalanced binary class data is converted to balanced multiclass data, exact coding and a fault prediction technique are used to develop a fault predictor. The method kept the class-imbalance information in SFP. In contrast to conventional sampling, ensemble learning methods, and cost-sensitive learning, the presented method did not alter the actual class distribution, nor did it suffer from unanticipated errors or data loss, while maximising the minority samples or minimising the majority samples. The given strategy addressed the issue of binary class imbalance in SFP by redefining it as a multi-classification issue that is independent of biassed data. To address the multi-classification problem, ensemble binary classifiers were constructed using a variety of coding techniques, among others, there are one-on-one, one-on-all, and random correction codes. In addition, the one-against-one coding scheme combined with the multi-classification learning strategy enhanced the performance of various classifier types, such as C4.5, Ripper, and RF, and produced superior results for binary classifiers using actual data. This method's primary impact is that it has not focused on domains with class imbalance issues, but rather on the SFP domain, which has facilitated the use of classification methods using ensembles.

Rathore et al. (2017a) [38] proposed an ensemble-based software defect prediction system. As a foundational learners' model for constructing the ensemble approach, the system utilised a variety of defect prediction methods to construct the ensemble method. The primary goal of this method is to determine whether ensemble methods are suitable for predicting a sufficient number of software module faults in software modules that are known to be malfunctioning and for which ensemble methods predict software module faults. The technology provided a heterogeneous ensemble method for predicting the occurrence of software bugs. Numerous defect prediction algorithms were included to enhance the learners' model for the heterogeneous ensemble. In order to incorporate the outcomes of the base learners' model into the ensemble method, the performance of ensemble methods with two distinct combination rules, such as a nonlinear and a linear rule, was evaluated. Several defect data sets, including the Eclipse bug data repository and the PROMISE data repository, contained these principles. This method's primary flaw is that system performance for the largest software projects is not analysed by various development and design methodologies. Moreover, the ensemble method cannot develop the ensemble system to analyse and Create additional combination principles that incorporate the outcomes of the learning procedure.

Rathore et al. (2017b) [39] developed ensemble techniques to predict the number of flawed software modules. Several attempts were made to classify software modules using ensemble methods to determine whether they were defective or not. However, the ensemble methods were unable to predict all of the software module's defects. Using ensemble approaches to predict defects has not been extensively investigated. In order to predict defects, the ensemble approach incorporated both homogeneous and heterogeneous ensemble methods. Using both linear and nonlinear combination principles, ensemble methods explored how to integrate the efforts of foundational students. In addition, these recommendations accounted for the efficacy of ensemble techniques in two distinct contexts: intra- and inter-release prediction. The ensemble method assumed that the prediction strategy had incorporated the results of diverse learning methods to improve the overall software prediction efforts. Therefore, the ensemble method utilised multiple learning models and focussed on achieving a higher level of performance. In the ensemble method, Some methods may have formed a similar theory; therefore, it was essential to select multiple strategies to ensure that the method did not perpetrate the same error twice. This type of ensemble method development has yielded positive results. This method cannot predict a substantial number of flaws in dissimilar-based learning approaches and combination rules.

Li et al. (2016) [40] suggested the modification of a three-way decision model to forecast software error problems. Three-way decision theory is more comprehensive than two-way decision theory. The SDP regarding a ternary classification has been determined by the crucial model of three-way decisions based on estimates of predetermined criteria. In two-way decision approaches, an object permitted the classification into an individual categorization model if the classification satisfied the condition, and it eliminated the individual categorization model if the classification did not satisfy the condition. The three-way decision models classified the objects based on the established criteria by analysing the ambiguity in a variety of situations. Utilising two-stage ranking and classification, a cost-sensitive, the feasibility of three-way decision-based software defect prediction was demonstrated. Therefore, three-way decision classification approaches have been implemented to overcome two-way decision classification challenges. Two-stage classifications, a three-way decision classification model is examined in the first phase, which indicates a maximal decision cost and maximal misclassification problem for ambiguous objects. The overall software modules were classified into several dissimilar regions, with the non-negative region including the FP software modules and the negative region including the nfp software modules. In the second phase, a method based on ensemble learning was utilised to classify the rescheduling software modules to achieve absolute two-way efforts. The primary benefit of three-way decisions over two-way decisions was the reduction of categorization errors and expense by identifying misclassified object possibilities for further examination. In three-way decision theory, converting three-way judgements to two-way decisions is prohibited by incrementally investigating additional samples or attributes. Nonetheless, this method is essential for predicting software defects.

Laradji et al. (2015) [41] developed a classification method for software defects, implemented by an APE (Average Probability Ensemble) learning module.. The primary objective of this strategy was to demonstrate how modern two-variable ensemble learning algorithms for defect classification perform more effectively when feature selection and ensemble learning are combined, including feature selection and without, were intended to demonstrate robustness against feature redundancy and data imbalance. The APE system contained various classifiersBNB, RF, LR, GB, MNB, W-SVMs, and SGD are examples. After the broad range of simulations is validated, the base classifiers have been chosen to incorporate the sequence of the APE system. The distinction between the base classifiers in the classification model allowed the classifiers in the ensemble learning model to obtain diverse statistical characteristics of the reputation data. By incorporating an efficient feature selection technique, the classification performance of the ensemble classifier was improved, and it even outperformed an upgraded version of the ensemble classifier. This procedure managed irrelevant and redundant software fault dataset features effectively. Two-variant ensemble techniques have demonstrated the robustness of irrelevant and redundant characteristics. The APE learning module is incapable of determining alternative feature selection strategies for the subsequent verification process, has not disclosed some of the features in publicly accessible software fault datasets, and is unable to determine whether the features are redundant or inconsequential.

Debata et al. (2018) [42] established a stock exchange prediction method for Indonesia based on fragment-based association mining and genetic algorithm prediction optimisation. Using the association rule-mining method to mine relationship rules during unstable stock prices, they analysed the fluctuating stock prices of companies. Various variables influencing the number of generated rules were also investigated and evaluated.

Ramezanian et al. (2019)[43] used a Hidden Markov Model (HMM) with daily share price values for stock price prediction. The trained HMM model identified and categorised identical samples using these data. Calculate the prices for the specified days and the subsequent trading day. Finally, projections of average price movements for the following day were made. They forecast daily changes in the S&P 500 price. Studies showed that neural networks' reasoning is superior to that of backpropagation and perceptron networks.

Li et al. (2020) [44] Using using a neural network with training to predict stock prices and a genetic algorithm to optimise neural network weights, a hybrid genetic strategy is proposed to improve the accuracy of the neural network. HCLTECH.NS data, figures for Wipro, Google, TCS, and Infosys. Yahoo Finance has compiled the information. This method classifies new data based on previously generated rules, as it is well-known that classification is a crucial task for effectively predicting an unknown instance. The associative classification made more accurate predictions than conventional classification methods, but its ability to manage data and its relationship with the data were inadequate. Due to this flaw, they were motivated to develop a suitable method and construct associative numerical data classifiers.

Han et al. (2020) [45] investigated technical trading principles' statistical properties. The focus of the study was on the numerous technical trading rules, including trend line breakouts, moving average crossovers, and charting patterns. Markov times standardise trading principles. Moving average crossovers were well-defined only. When prices breach the barrier or support level, traders buy or sell. Maximum and minimum values over the past 50, 100, 150, and 200 days represent resistance and support. Using returns for one-day and ten-day holding periods, each trading strategy's profitability is evaluated using bootstrapping and statistical testing. Technical trading beat buy-and-hold, but not when trading costs were considered. This strategy corrected data surveillance, making it better than bootstrap.

Table 2.3: Comparison of Reviews

Authors and Year	Techniques Used	Key Findings
Rouf, Nusrat et al. (2021) [36]	Machine Learning, Text Analytics, Ensemble Methods	Improved accuracy in stock market prediction using machine learning and ensemble methods.
Sun et al. (2012) [37]	Fault Prediction, Exact Coding Scheme, Ensemble Learning	Transformation of Unbalanced binary class data into balanced multiclass data for fault prediction.
Rathore et al. (2017a) [38]	Ensemble Methods, Heterogeneous Ensemble, Defect Prediction	Use of ensemble methods for predicting defects in software initiatives.
Rathore et al. (2017b) [39]	Ensemble methods, Heterogeneous and Homogeneous Ensembles	Ensemble methods for predicting defects in software modules.
Li et al. (2016) [40]	Three-Way Decisions Model, Ensemble Learning	Utilization of three-way decisions to predict software error issues.
Laradji et al. (2015) [41]	Average Probability Ensemble (APE), Feature Selection	Integration of ensemble learning and feature selection for defect classification.
Debata et al. (2018) [42]	Association Rule Mining, Genetic Algorithm	Prediction of stock market movements using association rule mining and genetic algorithm.
Ramezanian et al. (2019) [43]	HMM: Hidden Markov Model	Use of HMM with daily share price values for stock price prediction.
Li et al. (2020) [44]	Genetic Algorithm, Neural Network	Genetic algorithm optimization of neural network weights for stock price prediction.
Han et al. (2020) [45]	Technical Trading Principles, Markov Times, Bootstrap Methodology	Examination of statistical properties of technical trading rules and their profitability.

2.4 Advancements in Ensemble Learning Techniques for Various Applications

Tobek et al. (2021) [46] investigated if genetic algorithms can avoid local minimal difficulties by investing more computational resources in artificial neural network training, stock price predictions may be more accurate. ERIC B, a stock of the American Stock Exchange, was used to analyse S&P 500 closing price training data. It demonstrated that accuracy did not increase considerably although genetic algorithm training required additional resources. The training algorithm for integrating closing price with other attributes should have enough data to evaluate.

Wang, Qili et.al (2018) [47] used multiple information sources, including social media, it has attempted to forecast excess return trends on the financial market. Studies on public sentiment and stock price variations show that public opinion affects investment decisions. The approach combines crowd wisdom and technical analysis to predict financial markets using a revolutionary fusion strategy. Provide a deep random subspace ensembles (DRSE) machine learning proposal based on the prediction task's characteristics. Utilised are both ensemble learning and deep learning. The recommended strategy outperforms baseline stock market prediction models by 14.2% in AUC value., proving DRSE's financial market prediction potential.

Papadopoulos, Sokratis, et al. (2018) [48] studied the among the most reliable and extensively employed classes of machine learning techniques, tree-based ensemble learning has attracted a lot of attention. Due to this, its effectiveness and accuracy in this discipline have not received sufficient attention or evaluation. This study evaluates the performance of gradient-boosted regression trees, random forests, and extremely randomised trees, also known as extra trees, for predicting and forecasting the heating and cooling loads of buildings. The second-best algorithm's heating and cooling capacity predictions were 14% and 65% more accurate, respectively, after gradient boosting.

Dietterich, Thomas G. (2000) [49] described using algorithms for machine learning, ensemble approaches generate multiple classifiers and use their weighted predictions to classify new data points. Bayesian averaging existed before ensemble methods such as boosting, bagging, and error-correcting output coding. This strategy explains why ensembles outperform individual classifiers. New studies and ensemble methods are discussed to determine why Adaboost does not fast overfit.

Gu, Xiao-Feng, et al. (2014) [50] proposed with the transition from traditional machine learning to machine learning on massive datasets, the majority of stream classifiers must recognise and respond to concept drifts. Classifier dynamic ensemble and incremental learning are the two most commonly used adaptation strategies for concept drift. Ensemble classifiers are becoming increasingly popular due to their inherent adaptability to change. Nevertheless, many ensembles processing instances in blocks fail to respond precisely to small, incremental changes, and many ensembles processing streams fail to respond quickly enough to large, abrupt changes. Fortunately, Brzezinski and Stefanowski created the OAUE (Online Accuracy Updated Ensemble) algorithm. It has been demonstrated that the OAUE algorithm is a valuable ensemble for managing data streams with drift. Its ability to adapt to unexpected changes may be hindered by its fixed timeframe. To determine the window size of each candidate classifier, the ensemble includes a change detector and the Window-Adaptive Online Accuracy Updated Ensemble (WAOAUE) method. Using an experimental comparison with four cutting-edge online ensembles, such as OAUE, the proposed method identified the best practises for the stream mining of large data sets. **Wang, Xiao-Lin, et al. (2014) [51]** discussed innovative innovations Extreme Learning Machine (ELM) has demonstrated utility in a variety of settings. The M3-network, a parallelized Min–Max Modular network ELM ensemble, was developed to address the problem posed by so-called enormous data. Before incorporating classification challenges into the M3-network, the M3-ELM divides them into subproblems and trains an ELM for each subproblem. The proposed strategy is evaluated utilising eleven sets of actual and reference data. M-ELM reduces test error rates by 0.37–19.51 percent and accelerates the assimilation of training words by 1.60–4.6 times. M3-ELM scales well for massive tasks and improves the accuracy of unbalanced problems.

Table 2.4: Comparison of Reviews

Authors and Year	Technique	Key Findings
Tobek et al. (2021)	Genetic Algorithms, Neural Networks	Genetic algorithm training did not significantly improve stock price predictions despite increased computational resources. Adequate data is necessary for training algorithms with closing price data along with other attributes.
Wang, Qili et al. (2018)	Deep Random Subspace Ensembles (DRSE)	The fusion technique combining public sentiment and technical analysis utilising DRSE outperforms baseline stock market prediction models by 14.2% in AUC value.
Papadopoulos, Sokratis, et al. (2018)	Ensemble Learning (Random Forests, Extremely Randomize trees, Gradient boosted Regression trees)	Gradient boosting improved prediction accuracy for heating and cooling loads in buildings by an average of 14% and 65%, respectively.
Dietterich, Thomas G. (2000)	Ensemble Methods (Bagging, Boosting, etc.)	Ensemble methods, including Adaboost, can outperform individual classifiers due to their weighted voting mechanism. Experiments explore reasons why Adaboost does not overfit rapidly.
Gu, Xiao-Feng, et al. (2014)	Online Accuracy updated Ensemble (OAUE), Window-Adaptive OAUE	OAUE is efficient for drifting data streams but may not adapt quickly to sudden changes. WAOAUE, based on

		OAUE with a change detector, determines window size and offers improved performance for stream mining of large datasets.
Wang, Xiao-Lin, et al. (2014)	Parallelized Extreme Learning Machine Ensemble (M3-ELM)	M3-ELM breaks classification issues into subproblems, trains ELMs for each, and aggregates findings with its M3 network. It accelerates learning and reduces test errors, scalability on huge, imbalanced tasks.

3 PROBLEM STATEMENT

The issue in technical analysis for identifying quality stocks lies in creating a strong system to assess and choose high-quality stocks using historical market data and technical indicators. This entails establishing the specific standards and factors that distinguish a high-quality stock, such as robust performance, stability, and potential for growth, and determining the technical analysis methods and indicators that are most pertinent for evaluating these standards. Although there have been improvements in stock market intelligence, accurately predicting market trends continues to be a challenging task. The primary focus of this work is to develop and assess machine learning algorithms for predicting and analyzing trends in the stock market. The aim is to assess the effectiveness of various machine learning algorithms in predicting stock market fluctuations and providing reliable decision-making tools for investors in a dynamic market environment.

4 RESEARCH METHODOLOGY

Stock market prediction is forecasting the future value of a company's financial stocks. Currently, there is a growing inclination towards using machine learning in stock market prediction technologies. This approach involves making predictions by training the machine learning model on the past values of stock market indices and using the current values as input. Performed by the Technical Analysts, this method deals with the determination of the stock price based on the past patterns of the stock (using timeseries analysis.) When applying Machine Learning to Stock Data, we are more interested in doing a Technical Analysis to see if the algorithm can accurately learn the underlying patterns in the stock time series. Machine learning utilizes many models to provide accurate and reliable predictions. This study aims to forecast stock prices by using machine learning methods such as Artificial Neural Network and Support Vector Machines. This suggested method has the capability to train the machine using diverse historical data points in order to generate predictions about the future. The suggested technique has used data from the previous year's stock market to train the model. Two machine-learning libraries are mostly employed to tackle the challenge. The first library used was numpy, which facilitated the process of cleaning and manipulating the data, preparing it for further analysis. Another tool used was scikit, specifically for doing rigorous analysis and making accurate predictions. The data set used was obtained from the public internet database, consisting of stock market data from the previous year. 80% of the data was allocated for training the machine, while the remaining 20% was reserved for testing purposes.

The fundamental methodology of the supervised learning model involves acquiring knowledge of the patterns and correlations within the training set data, and then using this knowledge to accurately predict outcomes for the test data. The optimized data frame enabled us to preprocess the data for feature extraction. The data frame characteristics were date and the closing price for a given day. This approach utilizes all of these characteristics to train the machine on an Artificial Neural Network (ANN) model and forecast the dependent variable, which represents the price for a certain day. Also, it measured the accuracy by utilizing the predictions for the test set and the actual values. The suggested system encompasses several study domains, such as data pre-processing and random forest, among others. Figure 1 depict the architecture of proposed framework as shown below.

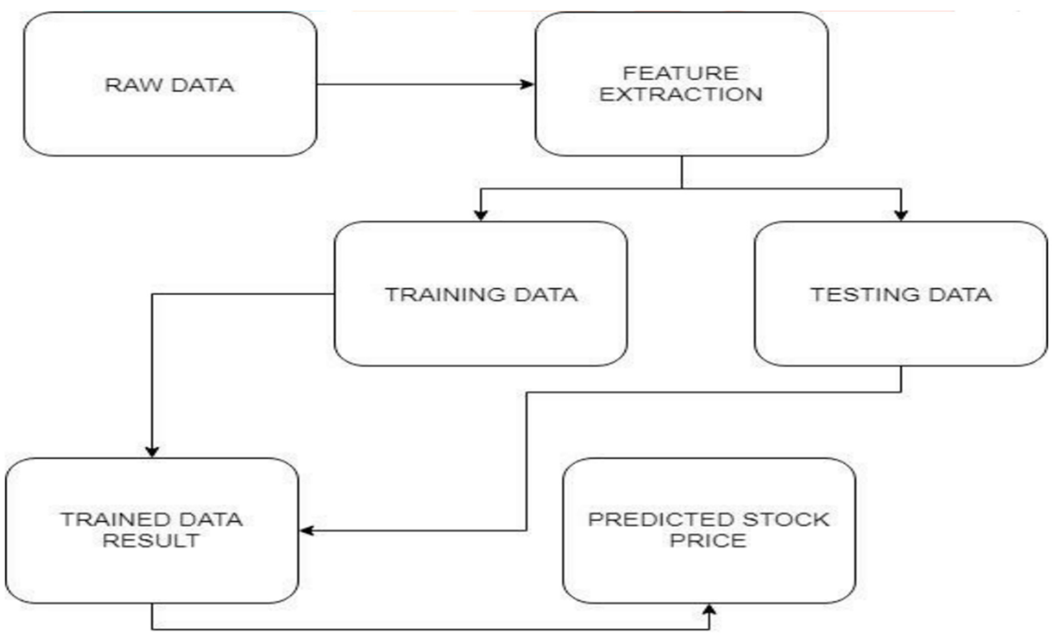

Figure 1. Proposed framework

The first stage involves transforming the unprocessed data into processed data. Feature extraction is used to filter out irrelevant features from the raw data gathered, since only a limited number of attributes are relevant for prediction purposes. The first stage involves feature extraction, whereby the essential characteristics are derived from the whole set of attributes included in the raw dataset. Feature extraction begins with an initial condition of measured data and constructs derived values or features. The purpose of these elements is to provide information and avoid unnecessary repetition, making it easier to learn and apply the knowledge in other situations. Feature extraction is a technique of reducing the number of variables in a dataset while maintaining the accuracy and completeness of the original data. It involves transforming the raw variables into more manageable features that effectively represent the original information. The feature extraction procedure is succeeded by a classification phase, during which the data acquired after feature extraction is divided into two separate and distinct segments. Classification is the task of determining the specific group or category to which a new observation belongs. The training dataset is used for model training, while the test dataset is employed to assess the correctness of the model. The splitting is performed in a manner that ensures the training data maintains a larger percentage compared to the test data. The random forest technique employs an ensemble of random decision trees to evaluate the data. Put simply, among the whole set of decision trees in the forest, a subset of the trees search for certain characteristics in the data. This process is often referred to as data splitting. Given the objective of this suggested system, which is to forecast the stock price by examining its past data.

4.1 Data collection

Data gathering is a fundamental component and the beginning phase of the project. It mostly involves the acquisition of the appropriate dataset. The dataset used for market prediction must undergo filtration based on many factors. Data collecting also serves to augment the dataset by including additional external data. The data mostly comprises the stock prices from the prior year. Firstly, we will evaluate the Kaggle dataset and choose the model that provides the highest accuracy. Then, we will use this model to properly assess the predictions based on the data.

4.2 Preprocessing

Data pre-processing is a crucial step in data mining that entails converting unprocessed data into a more organized and structured shape. Raw data often exhibits inconsistencies, incompleteness, and many inaccuracies. Data pre-processing encompasses many steps, including identifying missing values, identifying category values, dividing the data into training and test sets, and performing feature scaling to standardize the variables for fair comparison.

4.3 Training the machine

Training the machine involves inputting data into the algorithm to refine the test data. The training sets are used to calibrate and optimize the models. The test sets remain unaltered, since it is inappropriate to evaluate a model based on data that has not been previously seen. The model undergoes cross-validation during training to provide a reliable estimate of its performance using the training data.

4.4 Data scoring

Scoring the data is the term used to describe the process of applying a prediction model to a given dataset. The ensemble technique that is used for both classification and regression tasks. The system delivers intriguing outcomes based on the learning models. The last module explains how the model's outcome may be used to forecast the likelihood of a stock's increase or decrease, depending on certain criteria. Additionally, it demonstrates the susceptibilities of a certain stock or company. The user authentication system control is established to ensure that only authorized entities are able to access the results.

4.5 Machine learning techniques

- **Logistic regression**

Logistic regression is analogous to linear regression in that it constructs a curve by fitting a straight line, despite the fact that linear regression actually yields a curve. Logistic regression is a statistical technique used to model the relationship between a binary dependent variable and one or more independent variables. It uses logistic curves to depict the probability of the dependent variable falling within the range of 0 to 1. The shown curves are derived from one or more predictors or independent variables [52]. The logistic regression coefficients may be used to estimate the odds ratios for each independent variable in a model. This approach exhibits more adaptability compared to feature analysis. Additionally, it is the favored approach for resolving issues associated with binary categorization. Figure 2 depicts the schematic diagram of logistic regression.

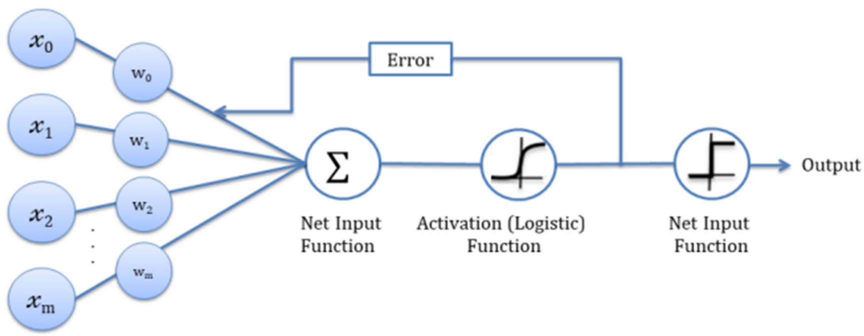

Figure 2. Classification of Logistic regression [53].

- **Support vector machine (SVM)**

SVM have become a potent and adaptable method in the field of machine learning, namely in the areas of classification and regression problems. SVM is primarily intended to determine an ideal hyperplane that effectively divides data points into discrete classes, while also maximizing the distance between them. This unique characteristic makes SVM particularly effective in scenarios with complex and nonlinear relationships between variables [54]. One of the strengths of SVM lies in its ability to handle high-dimensional data efficiently, making it well-suited for applications ranging from image recognition to sentiment analysis. The kernel trick, a key feature of SVM, allows it to implicitly map input data into higher-dimensional spaces, enabling the model to discern intricate patterns that might be challenging for other algorithms. Figure 3 illustrate the diagram of SVM as shown below.

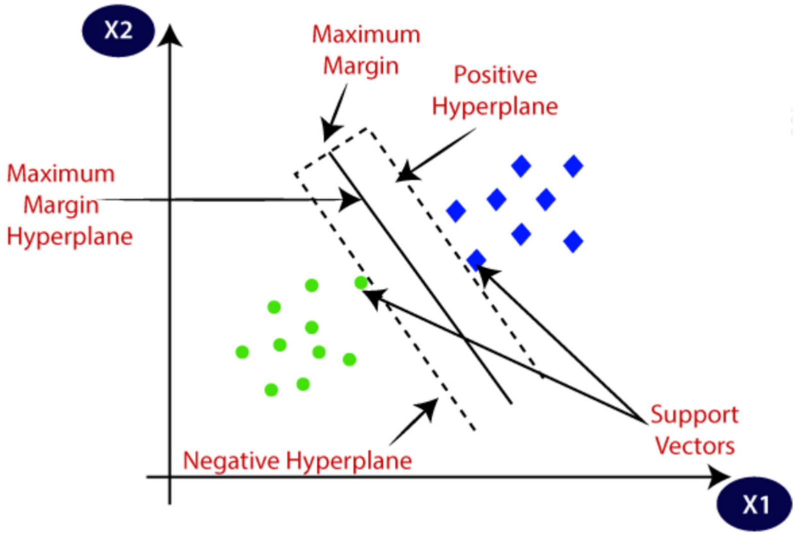

Figure 3. Diagram of SVM

- **Artificial Neural network (ANN)**

The Artificial Neural Network (ANN) method is a powerful machine learning technique inspired by the structure and functioning of the human brain. ANNs consist of interconnected nodes organized in layers, including an input layer, one or more hidden layers, and an output layer. Each node, or neuron, in the network receives inputs, applies a mathematical transformation to these inputs, and produces an output that is passed on to the next layer. Through a process known as training, ANNs learn to recognize patterns and relationships in data by adjusting the weights associated with connections between neurons. Figure 4 depicts the architecture of ANN as shown below.

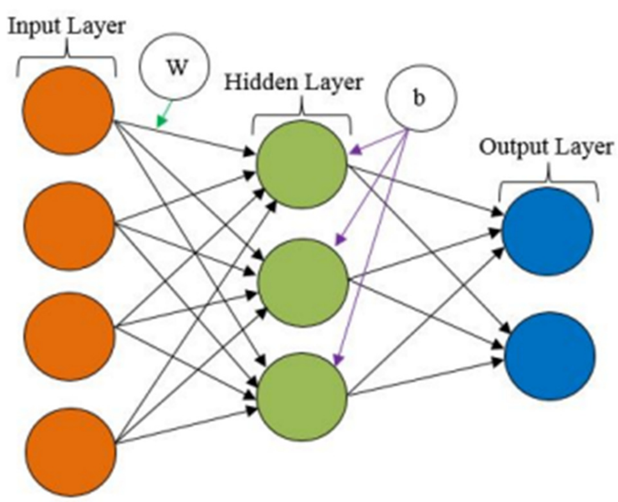

Figure 4. The basic ANN architecture [55].

- **Ensemble method**

Ensemble methods are a potent category of machine learning algorithms that amalgamate the predictions of numerous independent models to provide a more precise and resilient overall prediction. Ensemble approaches may surpass the constraints of individual models and achieve improved generalization performance by using the combined knowledge of varied models. Ensemble approaches include many techniques like as bagging, boosting, and stacking, each using a distinct strategy to amalgamate base models. Bagging, also known as Bootstrap Aggregating, is a technique that entails training many base models on distinct subsets of the training data and then averaging their predictions. This process is used to decrease variation and enhance stability. Boosting, in contrast, is a sequential training process where a succession of base models are trained. Each model focuses on the cases that were incorrectly categorized by the preceding models, resulting in a steady improvement of the overall prediction performance. Stacking is a technique that combines the predictions of many base models using a meta-learner. The meta-learner determines the weights for each base model's predictions based on their individual performance. Ensemble approaches have been extensively used in many fields such as classification, regression, and anomaly detection. They have repeatedly shown improved performance compared to individual models, thereby establishing themselves as an essential tool in the machine learning practitioner's arsenal.

4.6 Prediction and Evaluation

Model assessment is a crucial component of the model creation process. It is beneficial to identify the optimal model that accurately describes our data. The purpose is to assess the suggested categorization model and facilitate comparisons. The system will evaluate the prediction of quality stock performance and then compare the outcomes produced.

5 RESULT AND DISCUSSION

This section presented the findings of the study, which is defined in various stocks with their performance metrics such as accuracy, precision, recall and f-1score as shown below.

Table 3 depict the comparison of various algorithm based on TCS stock as well as seen in Figure 5. From table 3, linear regression algorithm obtained an accuracy 0.962, precision is 0.958, recall is 0.989 and f1 score is 0.973. K-NN method attained the accuracy 0.805, precision is 0.851, recall is 0.87 and f1 score is 0.86 on TCS stock. SVM technique attained the accuracy 0.97, precision is 0.958, recall is 1 and f1 score is 0.979 on TCS stock. The performance metrices of K-NN method is lower than the linear regression. The ANN method attained the accuracy 0.962, precision is 0.968, recall is 0.978, and f1-score is 0.973 and it is maximum as compared to LR, K-NN, and SVM methods. Therefore, it is clear that Ensemble method attained the maximum accuracy 0.97, precision 0.958, recall is 1, and f1-score is 0.979 and it gives superior performance as compared to all other methods as shown in Figure 5.

Table 3: Comparison of various algorithms based on TCS stock

S. no	Stock	Algorithm	Accuracy	Precision	Recall	F1_Score
0	TCS	LR	0.962	0.958	0.989	0.973
1	TCS	K-NN	0.805	0.851	0.87	0.86

2	TCS	SVM	0.97	0.958	1	0.979
3	TCS	ANN	0.962	0.968	0.978	0.973
4	TCS	Ensemble	0.97	0.958	1	0.979

As shown in Figure 5, the ensemble method attained high accuracy, precision, recall and f1-score as compared to all other methods.

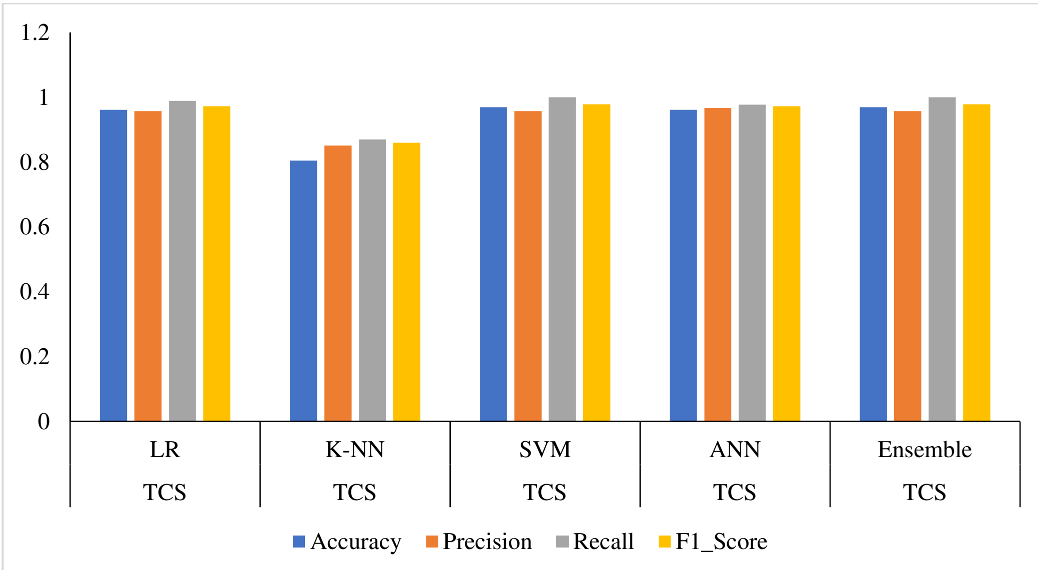

Figure 5. Comparison graph of Proposed method based on TCS stock

Table 4 displays a comparative analysis of several algorithms using the RELIANCE stock, as seen in Figure 6. The linear regression approach demonstrated exceptional performance with an accuracy of 0.994, precision of 0.992, recall of 1, and f1 score of 0.996, as shown in table 4. The K-NN method attained an accuracy of 0.92, a precision of 0.95, a recall of 0.929, and a f1 score of 0.944 for the RELIANCE stock. When the SVM technique was used to the RELIANCE stock, it obtained a flawless accuracy of 1. Additionally, both the precision and recall scores were also equal to 1, as well as the f1 score. The performance metrics of the K-NN technique are less favorable compared to those of linear regression. The Artificial Neural Network (ANN) attained a 0.994 accuracy, a 0.992 precision, a recall of 1, and a f1-score of 0.996, surpassing the LR, K-NN, and SVM methods. Therefore, it is clear that the Ensemble approach attained the best level of accuracy, precision, recall, and f1-score, all equal to 1, for the RELIANCE stock. It has superior performance compared to all other approaches, as seen in Figure 6.

Table 4: Comparison of various algorithms based on RELIANCE stock

S.no	Stock	Algorithm	Accuracy	Precision	Recall	F1_Score
0	RELIANCE	LR	0.994	0.992	1	0.996
1	RELIANCE	K-NN	0.92	0.959	0.929	0.944
2	RELIANCE	SVM	1	1	1	1
3	RELIANCE	ANN	0.994	0.992	1	0.996
4	RELIANCE	Ensemble	1	1	1	1

As shown in Figure 6, the proposed method attained high accuracy, precision, recall and f1-score as compared to all other methods.

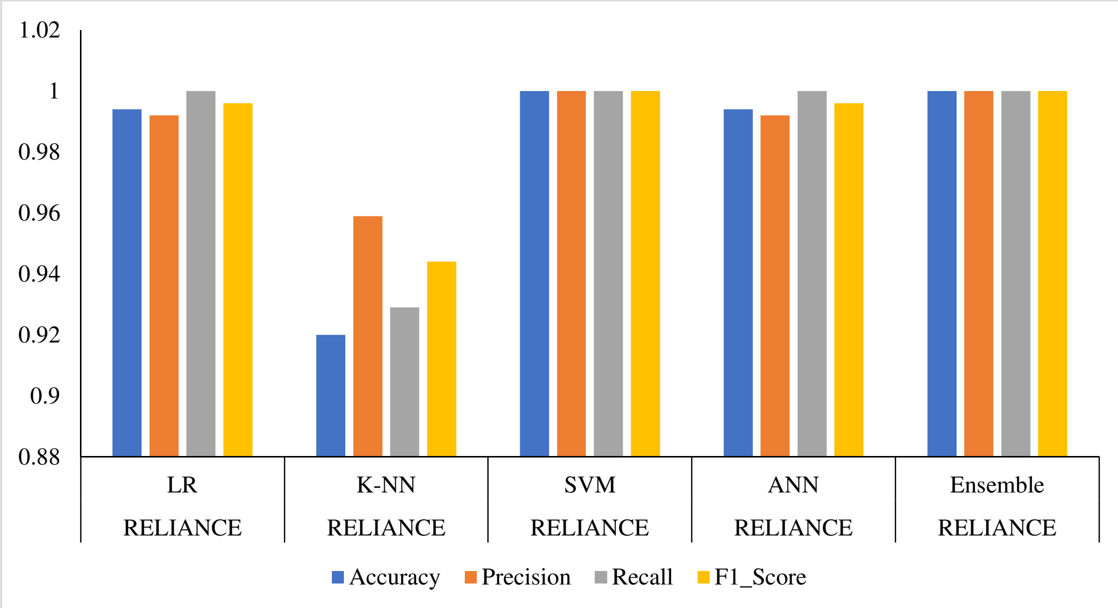

Figure 6. Comparison graph of proposed method based on RELIANCE stock

Table 5 provides a juxtaposition of several algorithms using ITC stock, as seen in Figure 7. The linear regression approach demonstrated an accuracy of 0.946, a precision of 0.931, a recall of 0.947, and a f1 score of 0.939, as shown in table 5. The K-NN method yielded an accuracy of 0.831, a precision of 0.787, a recall of 0.842, and a f1 score of 0.814 for the ITC stock. When the SVM technique was used to the ITC stock, it produced a flawless accuracy of 0.931, a precision of 0.887, a recall of 0.965, and a f1 score of 0.924. Therefore, it is clear that the Ensemble approach obtained the maximum accuracy of 0.931, precision of 0.914, recall of 0.93, and f1-score of 0.922 for the ITC stock. It has superior performance compared to all other approaches, as seen in Figure 7.

Table 5: Comparison of various algorithms based on ITC stock

S.no	Stock	Algorithm	Accuracy	Precision	Recall	F1_Score
0	ITC	LR	0.946	0.931	0.947	0.939
1	ITC	kNN	0.831	0.787	0.842	0.814
2	ITC	SVM	0.931	0.887	0.965	0.924
3	ITC	ANN	0.908	0.869	0.93	0.898
4	ITC	Ensemble	0.931	0.914	0.93	0.922

As shown in Figure 7, the proposed method attained high accuracy, precision, recall and f1-score as compared to all other methods.

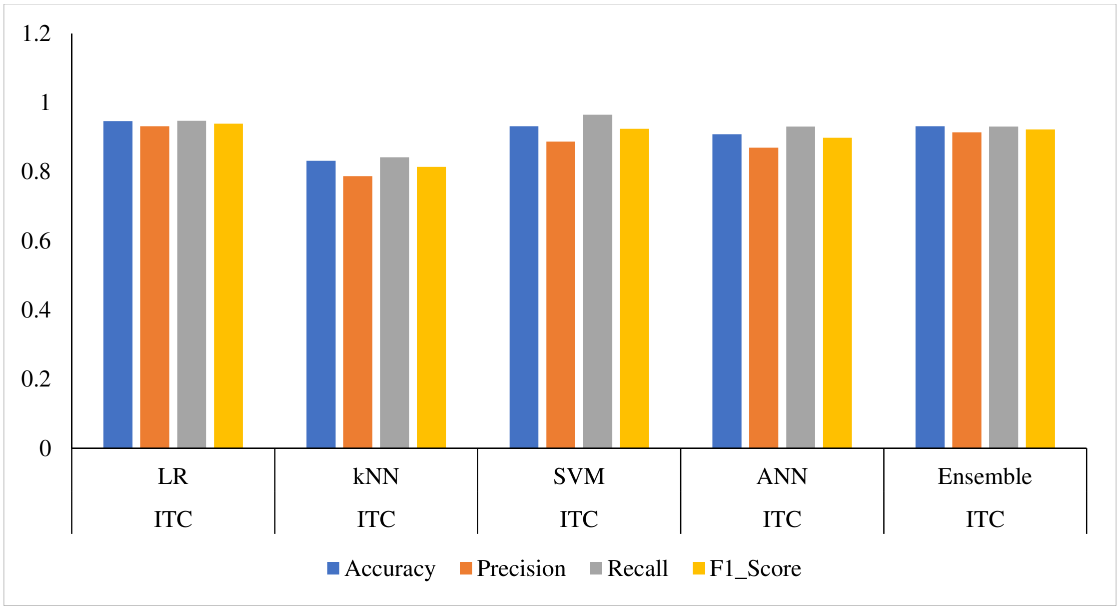

Figure 7. Comparison graph of proposed method based on ITC stock

Table 6 presents a comparative analysis of several algorithms using INFY stock data, as seen in Figure 8. The linear regression model demonstrated an accuracy of 0.861, precision of 0.838, recall of 0.954, and f1 score of 0.892, as shown in table 6. The K-NN method attained an accuracy of 0.847, a precision of 0.857, a recall of 0.897, and a f1 score of 0.876 for the INFY stock. The SVM technique demonstrated exceptional performance when used to the INFY stock, achieving a flawless accuracy of 0.875, precision of 0.879, recall of 0.92, and a f1 score of 0.899. Therefore, it is clear that the Artificial Neural Network (ANN) obtained the greatest level of accuracy, with a value of 0.924. Additionally, it acquired a precision of 0.904, a recall of 0.977, and a f1-score of 0.939 for the INFY stock. It has superior performance compared to all other approaches, as seen in Figure 8.

Table 6: Comparison of various algorithms based on INFY stock

S.no	Stock	Algorithm	Accuracy	Precision	Recall	F1_Score
0	INFY	LR	0.861	0.838	0.954	0.892
1	INFY	K-NN	0.847	0.857	0.897	0.876
2	INFY	SVM	0.875	0.879	0.92	0.899
3	INFY	ANN	0.924	0.904	0.977	0.939
4	INFY	Ensemble	0.903	0.884	0.966	0.923

As shown in Figure 8, the ANN method attained high accuracy, precision, recall and f1-score as compared to all other methods.

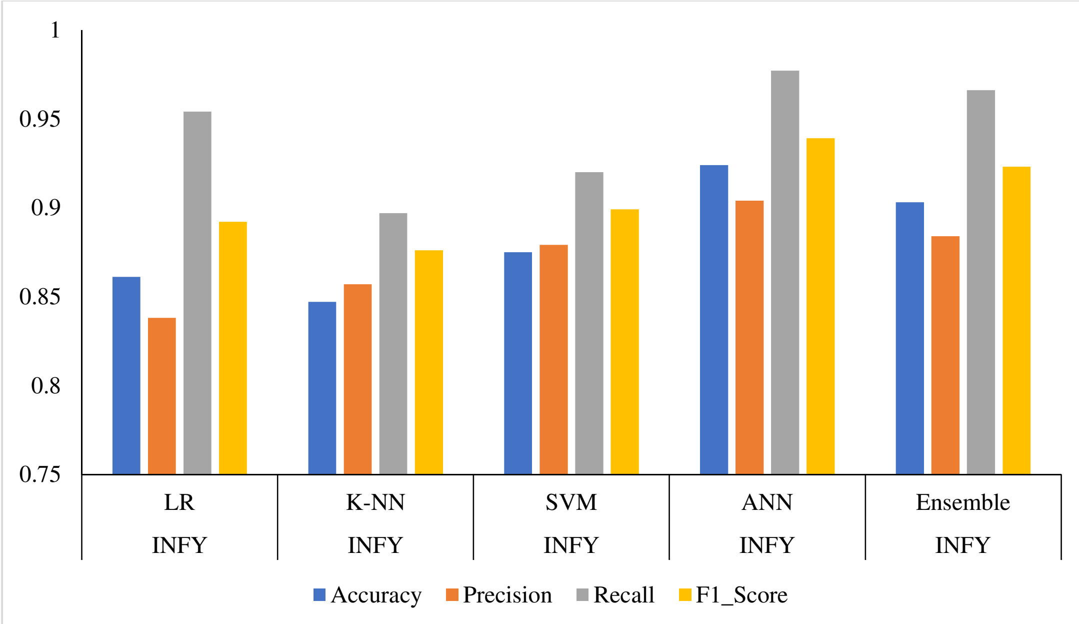

Figure 8. Comparison of Proposed method based on INFY stock

Table 7 presents a comparative analysis of several algorithms using HDFCBANK stock data, as seen in Figure 9. The linear regression model demonstrated an accuracy of 0.837, precision of 0.888, recall of 0.522, and a f1 score of 0.657, as shown in table 7. The K-NN method attained an accuracy of 0.835, a precision of 0.756, a recall of 0.662, and a f1 score of 0.706 for the HDFCBANK stock. The SVM technique yielded an accuracy of 0.885, precision of 0.882, recall of 9.713, and a f1 score of 0.789 when used on the HDFCBANK stock. Therefore, it is clear that the Ensemble model attained the maximum level of accuracy, with a value of 0.877. Additionally, it earned a precision of 0.926, a recall of 0.64, and a f1-score of 0.757 for the HDFCBANK stock. It has superior performance compared to all other approaches, as seen in Figure 9.

Table 7: Comparison of various algorithms based on HDFCBANK stock

S.no	Stock	Algorithm	Accuracy	Precision	Recall	F1_Score
0	HDFCBANK	LR	0.837	0.888	0.522	0.657
1	HDFCBANK	kNN	0.835	0.756	0.662	0.706
2	HDFCBANK	SVM	0.885	0.882	0.713	0.789
3	HDFCBANK	ANN	0.87	0.867	0.669	0.755
4	HDFCBANK	Ensemble	0.877	0.926	0.64	0.757

As shown in Figure 9, the proposed method attained high accuracy, precision, recall and f1-score as compared to all other methods.

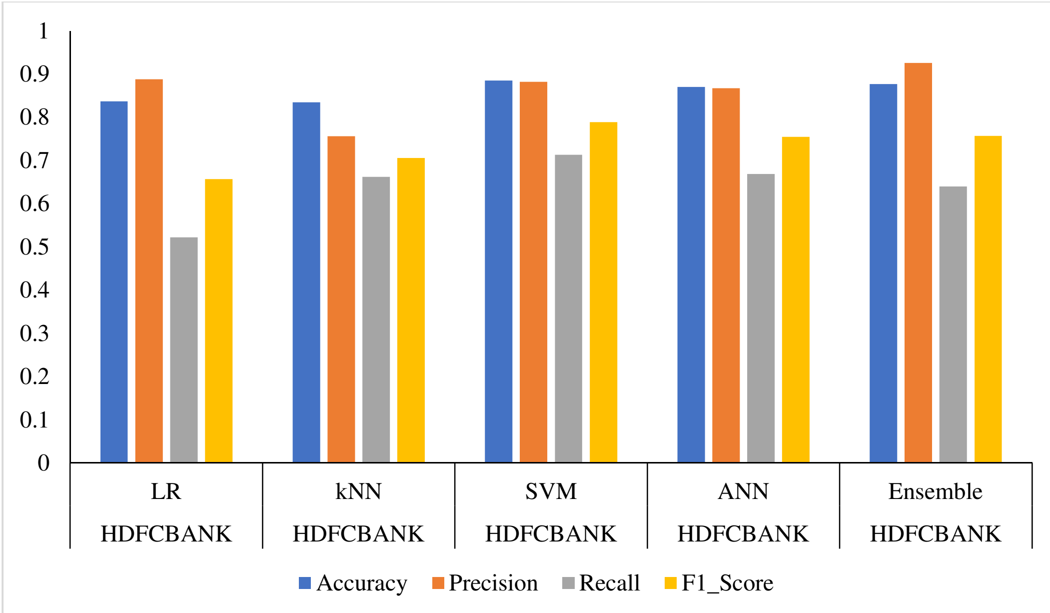

Figure 9. Comparison of proposed method based on HDFCBANK stock

6 CONCLUSION AND FUTURE SCOPE

In conclusion, the development and implementation of a technical analysis-based ensemble learning model for the identification of quality stocks offer promising avenues for investors and financial analysts alike. By leveraging the power of ensemble learning techniques and integrating technical analysis indicators, this model provides a robust framework for evaluating stock quality and making informed investment decisions. Through rigorous back testing and validation processes, the model demonstrates its effectiveness in identifying high-quality stocks with superior performance potential. Looking ahead, the future scope of this research extends towards several directions. Firstly, further refinement and optimization of the ensemble learning model can be pursued to enhance its predictive accuracy and robustness. This includes exploring additional technical indicators, refining feature selection methodologies, and fine-tuning ensemble algorithms to adapt to changing market conditions and dynamics.

REFERENCE

1. Nti IK, Adekoya AF, Weyori BA. A systematic review of fundamental and technical analysis of stock market predictions. Artif Intell Rev. 2019. https://doi.org/10.1007/s10462-019-09754-z.
2. Bousono-Calzon C, Bustarviejo-Munoz J, Aceituno-Aceituno P, Escudero-Garzas JJ. On the economic significance of stock market prediction and the no free lunch theorem. IEEE Access. 2019;7:75177–88. https://doi.org/10.1109/ ACCESS.2019.2921092.
3. Nti IK, Adekoya AF, Weyori BA. Random forest-based feature selection of macroeconomic variables for stock market prediction. Am J Appl Sci. 2019;16:200–12. https://doi.org/10.3844/ajassp.2019.200.212.

4. Wang Q, Xu W, Huang X, Yang K. Enhancing intraday stock price manipulation detection by leveraging recurrent neural networks with ensemble learning. Neurocomputing. 2019;347:46–58. https://doi.org/10.1016/j.neuco m.2019.03.006.
5. Billah M, Waheed S, Hanifa A. Stock market prediction using an improved training algorithm of neural network. In: 2016 2nd international conference on electrical, computer and telecommunication engineering. IEEE; 2016. pp. 1–4. http://doi.org/10.1109/ICECTE.2016.7879611.
6. Kraus M, Feuerriegel S. Decision support from financial disclosures with deep neural networks and transfer learning. Decis Supp Syst. 2017;104:38–48. https://doi.org/10.1016/j.dss.2017.10.001.
7. Pimprikar R, Ramachandran S, Senthilkumar K. Use of machine learning algorithms and Twitter sentiment analysis for stock market prediction. Int J Pure Appl Math. 2017;115:521–6.
8. Göçken M, Özçalici M, Boru A, Dosdołru AT. Integrating metaheuristics and artificial neural networks for improved stock price prediction. Expert Syst Appl. 2016;44:320–31. https://doi.org/10.1016/j.eswa.2015.09.029.
9. Dosdoğru AT, Boru A, Göçken M, Özçalici M, Göçken T. Assessment of hybrid artificial neural networks and metaheuristics for stock market forecasting Ç.Ü. Sos Bilim Enstitüsü Derg. 2018;24:63–78. 12
10. Akyuz AO, Uysal M, Bulbul BA, Uysal MO. Ensemble approach for time series analysis in demand forecasting: Ensemble learning. In: 2017 IEEE international conference on innovations in intelligent systems and applications. IEEE; 2017. pp. 7–12. https://doi.org/10.1109/inista.2017.8001123.
11. P.F. Pai, C.S. Lin, A hybrid ARIMA and support vector machines model in stock price forecasting, Omega 33 (6) (2005) 497–505.
12. B. Al-hnaity, M. Abbod, Intelligent Systems and Applications. Studies in Computational Intelligence, 650, Springer, 2016, pp. 19–41 (Chapter Predicting financial time series data using hybrid mode).
13. A.W. Lo, The New Palgrave: A Dictionary of Economics, second ed., Palgrave Macmillan Ltd., 2007 (Chapter efficient markets hypothesis).
14. E. Ahmadi, M. Jasemi, L. Monplaisir, M.A. Nabavi, A. Mahmoodi, P.A. Jam, New efficient hybrid candlestick technical analysis model for stock market timing based on the support vector machine and heuristic algorithms of imperialist competition and genetic, Expert Syst. Appl. 94 (2018) 21–31.
15. N. Petrusheva, I. Jordanoski, Comparative analysis between the fundamental and technical analysis of stocks, J. Process Manag. New Technol. Int. 4 (2) (2016) 26–31.
16. Asad, Muhammad. (2015). Optimized Stock market prediction using ensemble learning. 263-268. 10.1109/ICAICT.2015.7338559.
17. Nti, Isaac & Adekoya, Adebayo & Weyori, Benjamin. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. Journal Of Big Data. 7. 10.1186/s40537-020-00299-5.

18. Deshmukh, Sonal Sachin. "Progress in Machine Learning Techniques for Stock Market Movement Forecast." *Proceedings of the International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022)*. Vol. 105. Springer Nature, 2023.

19. Shah, Aayush, et al. "Identifying Trades Using Technical Analysis and ML/DL Models." *arXiv preprint arXiv:2304.09936* (2023).

20. Mohapatra, Sabyasachi, et al. "Can ensemble machine learning methods predict stock returns for Indian banks using technical indicators?." *Journal of Risk and Financial Management* 15.8 (2022): 350.

21. Nti, Isaac Kofi, Adebayo Felix Adekoya, and Benjamin Asubam Weyori. "A systematic review of fundamental and technical analysis of stock market predictions." *Artificial Intelligence Review* 53.4 (2020): 3007-3057.

22. Sohangir, Sahar, et al. "Big Data: Deep Learning for Financial Sentiment Analysis." *Journal of Big Data* 5.1 (2018): 1-25.

23. Abe, Masaya, and Hideki Nakayama. "Deep learning for forecasting stock returns in the cross-section." *Advances in Knowledge Discovery and Data Mining: 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018, Proceedings, Part I 22*. Springer International Publishing, 2018.

24. Krauss, Christopher, Xuan Anh Do, and Nicolas Huck. "Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500." *European Journal of Operational Research* 259.2 (2017): 689-702.

25. Qiu, Xueheng, et al. "Stock Price Forecasting with Empirical Mode Decomposition Based Ensemble-Support Vector Regression Model." *International Conference on Computational Intelligence, Communications, and Business Analytics*. Singapore: Springer Singapore, 2017.

26. Pasupulety, Ujjwal, et al. "Predicting stock prices using ensemble learning and sentiment analysis." *2019 IEEE second international conference on artificial intelligence and knowledge engineering (AIKE)*. IEEE, 2019.

27. Pathak, Ashish, and Nisha P. Shetty. "Indian stock market prediction using machine learning and sentiment analysis." *Computational Intelligence in Data Mining: Proceedings of the International Conference on CIDM 2017*. Springer Singapore, 2019.

28. Yang, Jian, et al. "Ensemble model for stock price movement trend prediction on different investing periods." *2016 12th International Conference on Computational Intelligence and Security (CIS)*. IEEE, 2016.

29. Gan, Kim Soon, et al. "Homogeneous ensemble feedforward neural network in CIMB stock price forecasting." *2018 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*. IEEE, 2018.

30. Kohli, Pahul Preet Singh, et al. "Stock prediction using machine learning algorithms." *Applications of Artificial Intelligence Techniques in Engineering: SIGMA 2018, Volume 1*. Springer Singapore, 2019.

31. Mabu, Shingo, Masanao Obayashi, and Takashi Kuremoto. "Ensemble learning of rule-based evolutionary algorithm using multi-layer perceptron for supporting decisions in stock trading problems." *Applied soft computing* 36 (2015): 357-367.

32. Wang, Qili, et al. "Enhancing intraday stock price manipulation detection by leveraging recurrent neural networks with ensemble learning." *Neurocomputing* 347 (2019): 46-58.

33. Brzezinski, Dariusz, and Jerzy Stefanowski. "Reacting to different types of concept drift: The accuracy updated ensemble algorithm." *IEEE transactions on neural networks and learning systems* 25.1 (2013): 81-94.

34. Agrawal, Manish & Khan, Asif & Shukla, Piyush. (2019). Stock Indices Price Prediction Based on Technical Indicators using Deep Learning Model. International Journal on Emerging Technologies. 10.

35. Agrawal, Manish & Shukla, Piyush & Nair, Rajit & Nayyar, Anand & Masud, Mehedi. (2021). Stock Prediction Based on Technical Indicators Using Deep Learning Model. Cmc -Tech Science Press-. 70. 287-304. 10.32604/cmc.2022.014637.

36. Rouf, Nusrat & Malik, Majid & Arif, Tasleem & Sharma, Sparsh & Singh, Saurabh & Aich, Satyabrata & Kim, Hee-Cheol. (2021). Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions. Electronics. 10. 2717. 10.3390/electronics10212717.

37. Sun, Z., Song, Q. and Zhu, X. (2012), ‘Using Coding-Based Ensemble Learning to Improve Software Defect Prediction’, IEEE Transactions on Systems, Man, and Cybernetics 42(6), 1806-1817.

38. Rathore, S.S. and Kumar, S. (2017a), ‘Towards an ensemble-based system for predicting the number of software faults’, Expert Systems with Applications 82, 357- 382

39. Rathore, S.S. and Kumar, S. (2017b), ‘Linear and non-linear heterogeneous ensemble methods to predict the number of faults in software systems’, Knowledge-Based System 119, 232-256.

40. Li, W., Huang, Z. and Li, Q. (2016), ‘Three-way decisions based software defect prediction, Knowledge-Based Systems, 91, 263-274

41. Laradji, I.H., Alshayeb, M. and Ghouti, L. (2015), ‘Software defect prediction using ensemble learning on selected features’, Information and Software Technology 58, 388-402

42. Debata, B, Dash, RS, Mahakud, J 2018, ‘Investor sentiment and emerging stock market liquidity’, Finance Research Letters, vol. 26, pp. 15 -31.

43. Ramezanain, R, Peymanfar, A & Ebrahimi, BS 2019, ‘An integrated framework of genetic network programming and multi-layer perceptron neural network for prediction of daily stock return: An application in Tehran stock exchange market’, Applied Soft Computing, vol. 82, pp.105551.

44. Li, J, Mirza, N, Rahat, B & Xiong, D 2020, ‘Machine learning and credit rating prediction in the age of fourth industrial revolution, Technological Forecasting and Social Change, vol. 161, pp. 120309

45. Han, J & Ge, Z 2020, ‘Effect of dimensionality reduction on stock selection with cluster analysis in different market situations’, Expert Systems with Applications, vol. 147, pp. 113226.

46. Tobek, O & Hronec, M. 2021, ‘Does it pay to follow anomalies research? Machine learning approach with international evidence, Journal of Financial Markets, vol. 56, pp. 100588.

47. Wang, Qili, Wei Xu, and Han Zheng. "Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles." *Neurocomputing* 299 (2018): 51-61.

48. Papadopoulos, Sokratis, et al. "Evaluation of tree-based ensemble learning algorithms for building energy performance estimation." *Journal of Building Performance Simulation* 11.3 (2018): 322-332.

49. Dietterich, Thomas G. "Ensemble methods in machine learning." *International workshop on multiple classifier systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000.

50. Gu, Xiao-Feng, et al. "An improving online accuracy updated ensemble method in learning from evolving data streams." *2014 11th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. IEEE, 2014.

51. Wang, Xiao-Lin, et al. "Parallelized extreme learning machine ensemble based in–max modular network." *Neurocomputing* 128 (2014): 31-41.

52. Hosmer Jr, David W., Stanley Lemeshow, and Rodney X. Sturdivant. Applied logistic regression. Vol. 398. John Wiley & Sons, 2013.

53. Hossain, Md Sabir, and Mohammad Shamsul Arefin. "Development of an Intelligent Job Recommender System for Freelancers using Client's Feedback Classification and Association Rule Mining Techniques." J. Softw. 14, no. 7 (2019): 312-339

54. Pisner, Derek A., and David M. Schnyer. "Support vector machine." In *Machine learning*, pp. 101-121. Academic Press, 2020.

55. Rahman, Md Akizur, Ravie Chandren Muniyandi, Kh Tohidul Islam, and Md Mokhlesur Rahman. "Ovarian Cancer Classification Accuracy Analysis Using 15-Neuron Artificial Neural Networks Model." In 2019 IEEE Student Conference on Research and Development (SCOReD), pp. 33-38. IEEE, 2019.