MJSAT

Malaysian Journal of Science and Advanced Technology

journal homepage: https://mjsat.com.my/

A Comparative Review on Stock Market Prediction Using Artificial Intelligence

Pulok Sarker[1], Adnan Sayed[1], Abu Bakar Siddique[1], Avijit Saha Apu[1], Syeda Anika Tasnim*[1], and Rifath Mahmud[1]

[1] *Computer Science Department, American International University-Bangladesh, Dhaka, Bangladesh.*

KEYWORDS

Artificial Neural Network
Long-Short-Term Memory (LSTM)
Predictive Modelling Data

ARTICLE HISTORY

Received 1 May 2024
Received in revised form 6 August 2024
Accepted 13 August 2024
Available online 2 September 2024

ABSTRACT

The global financial landscape has undergone unprecedented transformations in recent decades, characterized by increased complexity, volatility, and interconnectivity. In this dynamic environment, the ability to anticipate stock market trends has become a paramount concern for investors, financial analysts, and policymakers alike. This research aims to distil insights and contribute to advanced predictive models for the dynamic global financial landscape. The exploration encompasses diverse approaches, including artificial neural networks, convolutional neural networks, LSTM, and traditional machine learning algorithms. Emphasis is placed on data pre-processing, numerical analyses, and the efficacy of LSTM models. The significance of this research lies in its synthesis of existing knowledge, offering a holistic view of methodologies and outcomes in Share Market Prediction. The model signifies a foundation for further innovation in predictive modeling, addressing real-time data challenges and dynamic market conditions. This work advances the understanding and forecasting of stock market trends.

1. INTRODUCTION

The global financial landscape has undergone unprecedented transformations in recent decades, characterized by increased complexity, volatility, and interconnectivity. In this dynamic environment, the ability to anticipate stock market trends has become a paramount concern for investors, financial analysts, and policymakers alike. The advent of artificial intelligence (AI) has ushered in a new era of predictive modeling, offering advanced tools and techniques to navigate the intricate patterns inherent in financial markets.

This thesis explores the relationship between artificial intelligence (AI), classical machine learning algorithms, and stock market prediction, with a particular emphasis on LSTM, ANN, and other similar technologies. The need for more complex and precise prediction models that can understand the many elements impacting market fluctuations is the driving force behind this study. The use of LSTM, a specialized form of recurrent neural networks designed to capture long-term dependencies, represents a novel approach to discerning subtle patterns embedded in historical stock market data.

Concurrently, ANN, renowned for its adaptability and ability to learn from complex datasets, offers a dynamic perspective on predictive modeling in the financial domain. Additionally, traditional machine learning algorithms serve as a benchmark, providing a comparative lens for evaluating the efficacy of AI-driven methodologies. As financial markets continue to evolve, traditional models often struggle to adapt to the intricacies of real-time data and dynamic market conditions. The overarching goal of this research is to assess the comparative performance of LSTM, ANN, and machine learning algorithms in predicting stock market trends.

By understanding the strengths and limitations of each approach, this study seeks to contribute valuable insights to the ongoing discourse on the role of AI in financial markets. The significance of this research extends beyond the realm of academia, reaching practitioners, policymakers, and industry professionals. The outcomes of this study are poised to inform investment strategies, risk management practices, and policy decisions, thereby shaping the landscape of financial decision-making in the context of an increasingly AI-driven future. In the subsequent chapters, we delve into the methodologies, data

*Corresponding author:
E-mail address: Syeda Anika Tasnim <anika.tasnim@aiub.edu>.
https://doi.org/10.56532/mjsat.v4i4.316
2785-8901/

sets, and analyses undertaken to unravel the potential of AI in enhancing stock market prediction.

2. Background Study

In the current situation, predicting the stock market is a very complex task. However, people are constantly working on this sector. According to a report, a stock vector concept using neural networks and deep learning techniques is being developed to improve forecasting precision in the Shanghai A-shares market. This will address the limited research in developing countries and contribute to the Internet of Multimedia of Things [1].

Support vector machines (SVM) and artificial neural networks (ANN) were used to compare the fluctuations of stock prices in the Turkish market [2]. We can improve the accuracy by combining deep network analysis, correlated stock evaluation, and sentiment analysis of news articles [3]. Creating a machine learning system that leverages user input for the best possible profit and risk management would help stock market investors [4].

The Turkish stock market is a great example of how ML systems can predict stock values. A real-time stock value prediction system, recurrent neural network, and Holt-Winters algorithm might be possible by looking at historical patterns and manipulating data more accurately. Not only does this approach include user feedback, but it also challenges the Efficient Market Hypothesis, which states that outperforming the market is impossible since stock prices represent all relevant information. Investors may get a clear edge in the stock market with this system's blend of real-time machine learning algorithms, historical data analysis, and customized user input.

3. Comparative Review Based on Framework

Figure 1 shows the framework for the stock market prediction conducted in this study.

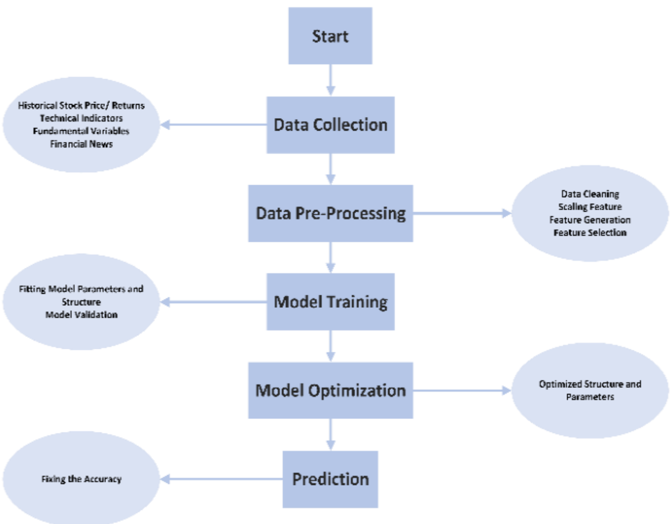

Fig. 1. Framework of Stock Market Prediction using Artificial Intelligence

3.1 *Deep Neural Network Framework*

The research paper [1], authored by Xiongwen et el. proposes a novel neural network approach to enhance the accuracy of stock market predictions. The study utilized real-time and offline data obtained from the livestock market for analysis. The researchers used visualizations and analytics to demonstrate how the Internet of Multimedia Things (IoMT) can be used for stock analysis. Traditional neural network algorithms can make incorrect predictions about the stock market due to the random selection problem in the initial weight. Therefore, the study aimed to investigate the impact of market characteristics on stock prices to improve prediction accuracy.

Deep learning, a technique being introduced by author Rafael Konstantinou, promises to transform the way that data processing issues in the stock market are handled. This approach entails forecasting the stock market's behavior, allowing traders to queue up their transactions earlier and be faster than others, thereby maximizing their profits. This can be done without having to be physically located close to the data sources [5].

To forecast stock values, author Abhinav Tipirisetty has used a deep learning system that combines long-short-term, wavelet transformations (WT), and stacked auto-encoders. Using a mix of numerical and textual research, the author of this study examines a new method for forecasting stock prices and compares it to previous tactics. The study's goal is to develop a hybrid model that can better forecast stock movements by combining the current trend in the market with public sentiment as measured by internet news sources and blogs. [6].

A deep learning framework was presented by Jingyi Shen et al. with the intention of forecasting stock market price trends. The framework is made to offer complete feature engineering modification, which includes deep learning-based system customization, multiple feature engineering methodologies, and pre-processing of the stock market dataset. The authors' exhaustive testing of popular machine learning models revealed that their suggested approach performed better than the others because of the extensive feature engineering that was applied. [7].

Jinan Zou et al. presented a deep learning framework in paper [8] that deals with stock prediction-related tasks. The framework includes various types of models such as deep learning-based models, convolutional neural network-based models, recurrent neural network-based models, graph-based neural network-based models, and reinforcement learning models, transformer-based models. The researchers utilized these models to obtain their expected research output.

3.2 *Machine Learning Framework*

In paper [9], Víctor Rubio Jornet, integrated machine learning with finance and used machine learning algorithms to achieve two main objectives. The first objective was to create a predictor that could determine whether a stock's value would increase or decrease in the medium term (3 to 5 months) using technical analysis. The second objective was to identify the algorithm that would offer the best possible prediction of the stock's movement.

The research paper [2], authored by Yakup Kara et al. introduce a machine learning framework that predicts the movement of stock price indexes. The study's overarching objective is to develop two efficient models that make use of SVM and ANN classification techniques. Every day, the National 100 Index of the Istanbul Stock Exchange (ISE) is tracked to see how well the models can predict its future movement. The proposed models use 10 technical indicators as inputs and employ two comprehensive parameter settings to improve the models' forecasting abilities. Experiments were carried out to enhance the precision of both models.

In paper [3], the author Hari Kiran Sai Surayagari introduces a machine learning framework that utilizes sentiment analysis and deep learning techniques to predict stock prices. The author attempts to establish the impact of news articles and correlated stocks on a particular stock. A machine learning framework that combines deep learning methods, machine learning, and sentiment analysis has been suggested by Isaac Kofi Nti et al. to forecast stock values. The goal of this study is to provide a thorough and critical analysis of 122 pertinent research studies on machine learning-based stock market prediction that have been published in academic publications over the course of 11 years (2007–2018). Three kinds of methodologies were recognized in these reports: combined analyses, technical analyses, and fundamental analyses [10].

In order to accomplish the primary goal of the article, Iyyappan M. et al. have used a framework for machine learning [4]. Predicting the closing prices of stocks in different sectors is the main goal of the study, which aims to help people with their stock market investments. In paper [11], the author of the research, Prince Vipulbhai Patel, analyzed and explained how machine learning is used to forecast stocks using an open-source framework. To train machine learning models to provide reliable stock market forecasts, the article suggests a method that makes use of publicly accessible stock data. The research forecasts stock values using two ML methods: Support Vector Machine (SVM) and Long Short-Term Memory (LSTM).

To forecast stock market movements using microblogging platforms, author Nousi Christina built a machine learning framework that employs sentiment analysis and machine learning methods. Using sentiment and historical data, the framework was used to examine the stock price fluctuations of Microsoft. The writers combed the Finance Yahoo! website from that time period to get historical data. A sentiment Reasoner, Text Blob, and VADER (Valence Aware Dictionary) are two Python packages used for sentiment analysis. Moreover, a multitude of machine learning models were used by the authors, including logistic regression, KNN, SVM, Naïve Bayes, decision trees, random forests, and MLP [12].

Alice Zheng and Jack Jin conducted a study on a machine learning framework that can predict stock prices using only time-series data. This framework specifically focuses on short-term price predictions for general stocks. Their approach involves analyzing time-series data on stock prices to make predictions on the stock market [13].

Utilizing comprehensive valuation processes and statistical methodologies, author Erhan Beyaz introduces machine learning frameworks for effective stock price forecasting utilizing machine learning techniques, all while considering the current state of the market [14].

In paper [15], author Debalina Barua has developed a technique for forecasting stock prices on the Dhaka Stock Exchange using models generated by deep learning and machine learning. Several models using various machine learning approaches are proposed by the framework. These methods include ARIMA, bidirectional LSTM, and multi-head attention-based LSTM. These models are trained using historical data that may be found on the official website of the Dhaka Stock Exchange.

By combining optimization methods with deep learning, Dixit and Soni created a complex framework for predicting stock market outcomes. Using technical indicators to examine stock market data and Natural Language Processing (NLP) to analyze news data, this system efficiently integrates the use of news feeds with historical stock market data. By combining a three-phase classifier with Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memories (Bi-LSTM), and Auto encoders, their approach aims to improve the accuracy of stock market predictions. The hybrid optimization model they use is called Exchange Market Emperor Pigeon Optimization (EM-EPO) [16].

In their proposal for a thorough review framework, Sonkavde et al. cover all the various ML and DL models used in the financial industry for classifying and predicting stock prices. This framework highlights the practical uses of several algorithms in finance, including supervised and unsupervised ML, ensemble, time series analysis, and DL methods [17].

3.3 LSTM Neural Network Framework

Combining CNNs with long short-term memory (LSTM), researchers Theyazn H. H. et al. created a neural network architecture that can forecast stock market prices. The main objective of this research was to develop an intelligent framework that can assess financial time series inputs and predict the movement of stock prices. A growing number of market trend forecasting models are using AI. [18].

In this study [19], author Yixin Guo introduces a machine learning method using the LSTM neural network framework. The article covers the theoretical knowledge of time series models and LSTM neural networks. It also includes the modelling analysis and prediction of stock prices for real stocks in the stock market. Finally, the article compares the prediction results of several models using the root means square error.

A subtype of Recurrent Neural Networks (RNNs) which is Long Short-Term Memory (LSTM) neural networks are introduced by author Khalid Bin et al. They established their prediction model and transformed the data using the Scikit-learn min-max scaler. Furthermore, they contend that linear approaches are less successful than the ARIMA (Autoregressive Integrated Moving Average) method [20].

3.4 SVM Machine Learning Framework

In the paper [21] Mehar Vijha et al. showcase a system that predicts the closing price of five companies across different sectors the next day using support vector machines and machine learning techniques. For this reason, financial data such as the stock's open, high, low, and closing prices are used; these values are transformed into new variables and then input into the model. Two popular strategic metrics, RMSE and MAPE are evaluated using the models. The modelling ability to accurately predict the stock's closing price is supported by the low levels of these two indicators.

The authors Farman Ali and Pradeep Suri present a neural network approach using a neuro-fuzzy system and support vector machine architecture. This study's main goal is to perform a scientometric analysis of artificial intelligence-based stock market projections [22].

It was in a framework that incorporates Random Forests, Long Short-Term Memory (LSTM) networks, and Support Vector Machines that machine learning was first proven by Pushpendu Ghosh et al. They have attempted to predict the intraday out-of-sample directional movements of the S&P 500 equities from January 1993 to December 2018 using random forests and LSTM networks as training methods. Examples have included random forests and cuDNN LSTM, which stands for CUDA Deep Neural Network Long Short-Term. In addition to starting prices and intraday returns, their multi-feature design also compares returns to closing prices [23].

Isaac Kofi Nti et al. have developed a novel framework named Ensemble Support Vector Machine (ESVM) that is based on the GASVM method. Here, we present GASVM, a new homogeneous ensemble classifier. It incorporates genetic algorithms (GA) into its architecture, which is based on support vector machines. Stock market forecasting makes use of this classifier to refine SVM kernel parameters and select features. In this study, GA is used to optimize all the various SVM design components at the same time. In their trials, the team used stock data from the Ghana Stock Exchange (GSE) spanning over eleven years [24].

3.5 Artificial Intelligence Framework

The paper [25] two artificial neural networks (ANNs) that use Candlestick Patterns and Technical Indicators to build stock market prediction systems were presented by Ashraf S. Hussein et al. These networks are the Multi-Layer Perceptron (MLP) ANN and the Radial Basis Function Neural Network (RBFNN). The use of artificial neural networks (ANNs) in this system aims for accuracy, self-confidence, universal application, and responsiveness. Because of its speed and capacity to deliver learning mechanism generalizations, they employed Multi-Layer Perceptron (MLP) ANN trained using the Kullback Leibler Divergence (KLD) learning technique for technical indicators. For candlestick patterns, trained Radial Basis Function Neural Networks (RBFNNs) with Localized Generalization Error (L-GEM) were used.

The authors of a research paper [26], among the many AI methods proposed by R. Dileep Kumar et al. are neural networks (NN), support vector machines, and (neuro-fuzzy) k-Nearest Neighbour. Additionally, they provide a Python Pandas module that may be used for various data processing applications, including classification and regression. In order to use AI for stock market analysis and trend prediction, this study set out to construct complex nonlinear connections between input and output data. The research paper is divided into four sections: financial sentiment analysis, portfolio optimization, stock market prediction using AI, and combinations of two or more methodologies. Preliminary results of state-of-the-art applications are shown in the article for each domain. More and more people are paying attention to this area of study, and the literature is becoming deeper and more specialized, according to the survey's overall results.

The author, Shreya Pawaskar, introduces a methodology for artificial intelligence (AI) stock market return prediction that uses a mix of deep learning, reinforcement learning. It is difficult to forecast stock market returns due to the market's notoriously volatile character. Nevertheless, the author's proposed framework strives to provide a great deal of clarity and accuracy. The framework may improve its efficiency in forecasting stock market returns with the use of AI and high computing power [27].

An AI architecture that combines SVM, RNN, DNN, and SVR was given in the publication of Mariam Moukalled et al. In order to increase the precision of stock forecasts and execute lucrative trades, they want to create an automated trading system that integrates machine learning algorithms, mathematical operations, and external factors such as news sentiment. By the end of the trading day, they hope to have determined the price or trend of a particular stock, which they set out to do in the morning [28].

Priyanka Tupe-Waghmare, the author, has introduced artificial intelligence and a machine learning framework that uses sentimental analysis to predict future stock prices. The goal of this study is to predict stocks and stock prices using artificial intelligence, and in their bibliometric study, they have found that there is significant potential for further research in the financial market domain [29].

3.6 Business Cycle and Computer Framework

In the paper [30], Lufuno Ronald Marwala, a writer, suggests an artificial intelligence approach for index prediction that is based on computer architecture and the economic cycle. The data pattern and its effects on each predicting method, the effect of a time horizon on forecasting techniques, and accuracy, as measured by the Mean Absolute Percent Error, RMSE, MSRE, and Anderson-Darlington calculation methods, are the main factors to be considered when choosing and comparing forecasting methods. With the use of AI models, this study aimed to forecast how much a stock market index will be worth in the future. Based on previous price data, three artificial intelligence techniques—neural networks (NN), support vector machines (SVMs), and neuro-fuzzy systems—are used to forecast the upcoming price of a stock market index.

3.7 SMP Framework

The SMP Framework, developed by Nusrat et al., compiles a literature review on machine learning methods used to forecast stock market movements. In their work, the authors detailed the process of gathering literature on machine learning for stock market prediction. A series of web searches including "stock market prediction using machine learning" in several databases (e.g., "Google Scholar," "Research Gate," "ACM Digital Library," "IEEE Explore," "Scopus," etc.) set the ball rolling. The operators OR and AND were used for searches in single and multiple classes that contained keywords, respectively, during the literature review. Terms such as "impact of sentiments on stock market prediction," "stock market prediction methods," and "machine learning-based approach for stock market prediction" were searched via. This resulted in the identification of foundational literature about the forecasting of stock market fluctuations. The writers acquired a foundational understanding of the issue by reading only a small number of introductory publications. A more recent attempt to strengthen and advance the subject included further refining the search criteria to gather content from the last decade. Indexing, impact factors, quartiles, and publishers were used to further narrow the chosen literature based on quality measures. [31].

Li et al. propose a technique named SMeDA-SA for mining Twitter data to perform sentiment analysis aimed at predicting stock movements of specific publicly listed companies. This framework incorporates the extraction of sentiment words from tweets, association rule discovery, and classification of sentiments to predict stock price movements accurately [32].

Oscar Alsing and Oktay Bahceci developed a framework to explore the impact of social media sentiment on stock market predictions. They hypothesized that public opinion on Twitter could be an indicator of stock price movements. Their approach integrated sentiment analysis with machine learning techniques to analyse tweets. The framework aimed to identify correlations between the sentiment expressed about companies on Twitter and their stock performance. This innovative strategy sought to bridge the gap between financial market analysis and the burgeoning field of social media analytics [33].

3.8 Double Deep Q-Network (DDQN) (Logistic Regressions)

Financial market AI and DL research was pioneered by author Andrew W. Brim with the introduction of the Double Deep Q-Network (DDQN) architecture. For prediction purposes, the framework employs reinforcement learning, Q-functions, logistic regressions, and a double Q-learning algorithm. This dissertation contains three research articles. The first article uses reinforcement learning with a Double Deep Q-Network (DDQN) for candlestick pattern trading in an effort to outperform the returns on the S&P 500 Index. To further examine the DDQN's trading patterns, it employs feature map visualizations. By using fuzzy logic to arbitrage the causal link in a pair-trading strategy, the second research finds larger returns by identifying the relationship between two co-integrated assets as a fuzzy relationship. The third research project uses a Double Deep Q-Network (DDQN) to forecast the spread of a pair trading strategy using two co-integrated assets [34].

3.9 Design Science Research Methodology (DSRM)

By combining sentiment analysis of social media with monetary stock data and machine learning, author Mohammad Al Ridhawi's Design Science Research Methodology (DSRM) is able to forecast market movements. Research Contributions, Designing as a Search The procedure, Design Evaluation, Problem Relevance, Design as an Artifact, Research Rigor, and Communication of Research are the several components that make up the method. We suggest an ensemble-based approach that uses convolutional deep neural networks (CNNs), multi-layer perceptron's (MLPs), and long short-term memory (LSTMs) to evaluate the sentiment in social media posts. Financial stock prediction is another use of an LSTM model. In order to train the models, we use financial market data from 2015–2019 as well as sentiment extracted from Twitter posts. Microsoft, Apple, Cisco, and IBM equities are the specific targets. According to the experiment's findings, combining emotional intelligence with financial data might lead to more precise stock market forecasts [35].

3.10 Node2vec Framework

In paper [36], to forecast short-term stock price movements, Huihui Ni et al. provide the Node2vec framework along with graph techniques including DFS, BFS, and tweet vectors. They suggest a hybrid strategy for stock market forecasting that makes use of historical pricing and tweet embedding. In contrast to conventional text embedding techniques, this method takes into account the outward structural properties and internal semantic elements of Twitter data, producing tweet vectors with more useful information. By building the Tweet Node network, the authors specifically create a Tweet Node algorithm to characterize possible relationships in Twitter data.

Table 1. Summary of different framework solution

Framework	Advantage	Disadvantage
SMP [31], [32], [33]	Taps into the pulse of public opinion on social media to predict stock movements with surprising accuracy	It can be tricked by false information on social media.
Node2vec [36]	Finds hidden stock connections, showing new insights.	Needs a lot of data and power, making it heavy-duty work.
Deep Learning Neural Network [1], [5], [6], [7], [8]	Brainy networks can digest vast amounts of data to uncover complex stock market trends.	Needs lots of data and power, which can be demanding.
Double Deep Q-Network (DDQN) [34]	Learns from past actions to make better future stock predictions.	May not always work well in the unpredictable stock market.
Machine Learning [2], [3], [4], [9], [10], [11], [12], [13], [14], [15], [16], [17]		
SVM Machine Learning [21], [22], [23], [24]	Good at predicting stock direction using past data.	Struggles with the stock market's complex nature.
Artificial Intelligence [25], [26], [27], [28], [29]	Analyzes and predicts stock trends very well.	Can get overwhelmed with too much data, leading to mistakes.
LSTM Neural Network [18], [19], [20]	They remember long-term patterns, making them great for predicting stock prices over time.	Yet, their deep thoughts need lots of training data and time, making them slow learners.
Business cycle and computer [30]	It provides a macroeconomic perspective, essential for long-term stock predictions.	It's like predicting the weather long-term, often too broad and missing the micro-movements.
Design Science Research Methodology [35]	It's a structured approach to creating and evaluating AI models, ensuring they meet real-world stock market challenges.	Can be too strict, limiting how models are developed.

4. Comparative Review Based on Methodology

Figure 2 shows the methodologies of the stock market prediction using Artificial Intelligence.

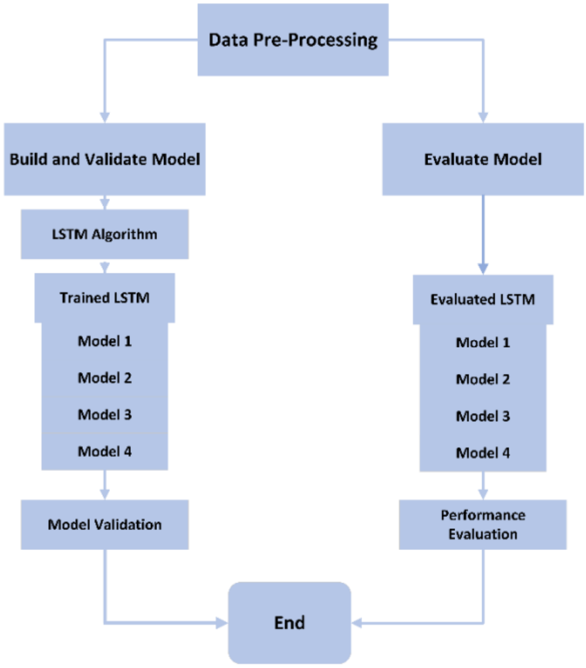

Fig. 2. Methodologies of stock market prediction using Artificial Intelligence

4.1 Studies Using Artificial Neural Networks to Predict Stock Market Values:

Xiongwen Pang et al. represented a word's location as 1 and the rest as 0 using the one-hot vector approach. Based on these two tests, they visually analyzed the stock vector in the ELSTM model and noted the morphological properties of the stock data [1]. Victor Rubio Jornet introduced ANN to demonstrate its capability in financial modelling and predicting stock price index movement using a three-layered feedforward ANN model [9].

Neural networks are highly parallel distributed processors made up of simple processing units, according to Yakup Kara et al. It is the natural tendency of this kind of network to preserve experimental knowledge and make it accessible when required. The network and the human brain are similar in two aspects: first, both require learning from the environment to acquire knowledge, and second, retain that knowledge in interneuron connection strengths, or synaptic weights [2].

Convolutional neural networks, or CNNs, are made to extract topology-capturing characteristics. Filters are applied to sets of input data in order to achieve this. CNNs are utilized in many different fields, including speech recognition, image processing, and time series analysis because of their capacity to record both sequential and spatial input. The authors Theyazn H. and Ali Alzahrani defined the layers of a convolutional neural network as input, convolution, pooling, and output [18].

The author of this article [19], author Yixin Guo, introduces an alternative to the traditional ARIMA model. In addition to ARIMA, the article will also use the LSTM model to build our model. The training data will consist of 70% of the data, while the remaining 30% will be used for testing. To optimize the model for training, we will be using the Root Mean Square Error and Adam algorithm. The ARIMA and GARCH model will be calculated using Stata12, while Mat lab will be used for the training.

In his discussion of a numerical analysis, author Abhinav Tipirisetty focuses on developing a model based on recurrent neural networks (RNNs) to forecast S&P 500 index stock prices. Given that stock market data is fundamentally a time series and that RNNs are very good at learning and predicting time series data, they are an excellent choice for this purpose. Tipirisetty employs a specific kind of RNN called Long Short-Term Memory for this investigation [6].

Artificial Neural Networks (ANN), a clever data mining technology that finds underlying trends in data and makes generalizations from it, were introduced by Mehar Vijha et al. Because ANN can simulate and analyse complicated patterns in unstructured data, it has a considerable edge over most conventional approaches. The model uses the fundamental architecture of a neural network, with neurons arranged in three layers: input, hidden, and output [21].

According to Pushpendu Ghosha et al., LSTM was first presented in 1997 by Schmidhuber & Hochreiter and is a kind of recurrent neural network. They cited Fischer & Krauss (2018) for more information. Experiments were carried out utilizing CuDNN LSTMs since LSTM training is time-consuming and because GPU power is best used in this way. One such framework that makes use of graphics processing units to speed up deep neural networks is CUDA [23].

The authors Ramit Sawhney et al. utilized the Stock-Net dataset to develop a process for training and evaluating MAN-SF. The dataset consists of data on high-trade-volume stocks in the S&P 500 index in the NYSE and NASDAQ markets. To extract stock-specific tweets, regex queries are used from NASDAQ ticker symbols such as $AMZN for Amazon [37].

The study by author Hari Kiran Sai Surayagari makes use of two networks. We go over each network in detail to help you better grasp the process. The first one is the LSTM network, which is employed in the news article classification process. Any machine learning algorithm must include pre-processing, and LSTM is no exception. Pre-processing needs to be done carefully. Although the quantity of data was covered in detail in the preceding chapters, the significance of data quality was not sufficiently stressed. Data quality is just as important as data quantity, thus it's critical to have the right information in the proper format [3].

LSTM is a type of recurrent neural network, also known as RNN. Unlike the regular RNN module which uses a single TANH function to process the output of the last layer, LSTM uses a feedback loop and gates to remember important information. According to Iyyappan. M. et al., LSTM consists of four interacting RNN layers in each module [4].

In order to comprehend the contents and processes of a convolutional neural networks (CNN) learning, authors Johnson and Jaya use feature map representations. The feature visualization work conducted by the Google Brain Team, including that of Olah, Mordvintsev, and Schubert (2017), makes use of neural network weight reconstruction to generate visuals [38].

In the pursuit of better interpretable neural networks, feature visualization has emerged as a key area of study. Author Andrew W. Brim proposed this method, which uses neural network weight reconstruction to produce pictures. In their quest to make neural networks interpretable, they found that feature visualization was among the most advanced and promising research topics [34].

Authors R. Dileep Kumar et al. have explored the use of the k-Nearest Neighbour classification algorithm for economic forecasting. Predictive models for financial distress have been an area of great interest in financial research due to the impact of an organization's financial difficulties on its stakeholders [26].

4.2 Studies Using Support Vector Machines (SVM) to Analyse Stock Markets

Support Vector Machine, developed by Ashraf S. Hussein et al., seeks to locate a hyperplane in an N-dimensional space that can reliably classify data points. Ritika Chopra and Gagan Deep Sharma demonstrated in their research how this hyperplane may segregate the N-dimensional space into the 0s and 1s. Whether a Linear SVM, RBF SVM (Gaussian), or Polynomial SVM is used depends on the kind of issue being addressed [25].

Isaac Kofi Nti and his team introduced the Support Vector Machine (SVM) as a supervised Machine Learning algorithm that is used for both regression and classification tasks. The SVM acts as a linear separator that is placed between two data nodes and used to detect two different classes in a multidimensional environment [24].

According to author Prince Vipulbhai Patel, LSTM cells have a beneficial mechanism that acts as a gate. In an average RNN cell, there is a tanh activation function that collects output and a new hidden state. However, LSTM has a more complex arrangement compared to RNN. This is because the LSTM cells require three pieces of information: the current input data, the short-term memory from the previous cell, and the long-term memory cell [11].

The authors, Alice Zheng et al., began by simplifying the difficult task of forecasting, based on historical stock prices and volumes, whether stock prices will rise or fall over the next n days. They used the rbf kernel from the sklearn library along with Bayesian networks, SVMs, Logistic Regression, and Simple Neural Networks to solve this classification challenge. These models were evaluated using "MSFT" stock prices [13].

4.3 Studies Using Hybrid and Ot`her AI Techniques to Analyse Stock Markets:

The paper's methodology describes the categorization system, acceptance criteria, and keywords that were searched. Model attributes fall under the category of data pre-processing.

The author Rafael Kostantinou employed artificial intelligence techniques, which included training algorithms and performance measures. The study provides information on the stock markets that were covered, the types of input variables used, and whether the author conducted a comparative analysis of various AI techniques [5].

News reports from Reuters.com that reveal the opinions of writers Jaya and Johnson serve as the source of data for the analysis. In addition, Yahoo.com is used to gather historical price data for the stock tickers of the firms that are mentioned in the articles. Sentiment scores and classes are extracted from news article tones using natural language processing algorithms. Machine learning algorithms then employ technical indicators, like moving averages of the closing prices, as input variables to forecast the direction of the stock prices. [38].

Standard supervised learning based on SRLP training loss and a self-supervised training task are proposed by Jinan Zou et al. for stock movement classification as an additional activity to train the network [39].

To capture high-level abstractions in data, deep learning models apply non-linear modifications layer by layer. Unlike previous algorithms, this supervised learning paradigm may make use of unlabelled data. The basis of deep learning is a multi-layer feed-forward artificial neural network trained via back-propagation and stochastic gradient descent. The network might consist of several hidden layers, neurons with Tanh, rectifier, and max-out activation functions, or neurons. Sachin Deshmukh and Vaishali Ingle authored this statement [40].

Algorithms for machine learning are fundamental to artificial intelligence. These algorithms are trained to find patterns in data and approximate a function that can compute outputs for certain issues based on inputs. Three stages are involved in the operation of machine learning models: algorithm testing, results validation, and algorithm training. Data sets are needed at every stage of the testing process, which evaluates the algorithm's performance. The authors Huihui Ni et al.'s study served as the basis for this material [36].

The study by authors Sohrab Mokhtari et al. outlines the use of machine learning tools for predicting the stock market. The process involves four main steps: building the dataset, engineering the data, training the model, and making the prediction. The authors have provided a detailed explanation of each of these steps [41].

After collecting the relevant data, author Nousi Christina takes the next step of determining people's views on Microsoft and its products and services. The author intends to make sense of the data gathered from Twitter and Stock Twits by performing sentiment analysis, using two sentiment analysis tools, Text Blob and VADER, for both platforms. The goal is to determine which of the two tools provides the best result. Then, the author prepares the data for building the ML models by pre-processing it. The final aim is to merge the sentiment score and stock prices into a single table to forecast Microsoft's stock prices [12].

The methodology proposed by Long et al. involves an end-to-end deep learning model that performs feature engineering and prediction in a single step, contrary to traditional two-stage models. They utilize a combination of convolutional and recurrent units within their MFNN to handle sequential market data efficiently. This methodology emphasizes the direct transformation of raw market data into a structured format that enhances the predictability of stock price movements, thereby refining the process of identifying trading signals within extreme market conditions [42].

The methodology introduced by the Abdelfattah et al. involves a multi-step process. Initially, social media data are collected and pre-processed to remove noise. Sentiment analysis is then conducted using a neutrosophic-logic-based approach, which allows for a nuanced interpretation of sentiment by accounting for truth, indeterminacy, and falsity. This sentiment analysis is combined with historical stock market data to predict future stock movements. The inclusion of neutrosophic logic aims to address the limitations of traditional sentiment analysis techniques by effectively managing the ambiguity and uncertainty present in social media data [43].

The methodology outlined by Khan and colleagues involves collecting data from Twitter and Business Insider, followed by rigorous pre-processing steps. They employ sentiment analysis to gauge the impact of social media and news on stock market trends. The approach includes feature extraction from stock data to identify predictive indicators. They create final datasets that merge sentiment scores with historical stock market data, setting the stage for applying machine learning algorithms. This methodical process is designed to systematically analyse the influence of external data on stock market predictions [44].

The authors employ a sophisticated methodology that combines traditional econometric models with advanced sentiment analysis techniques. They extract sentiment data from social media platforms and analyse its correlation with stock returns over time, using both linear and nonlinear models to capture the complex dynamics between public sentiment and market outcomes [45].

This research employed a Systematic Literature Review (SLR) methodology. We began by formulating specific research questions. Next, relevant data was gathered from academic sources such as journal articles and existing literature reviews. To ensure the quality and relevance of the included studies, a modified PRISMA checklist was utilized for evaluation [46].

Our research methodology for constructing the stock prediction system involved a three-step process. First, data was collected from various sources and rigorously evaluated for relevance. Second, the data underwent a comprehensive analysis that considered current market trends, industry regulations, and individual company performance. Finally, based on the analyzed data, an Artificial Neural Network (ANN) was developed, and the most accurate method for stock price prediction was selected [47].

In this study, we employed a feature extraction approach to generate daily data points suitable for stock price prediction using a Support Vector Machine (SVM). The features were categorized into two classes: "up" and "down," signifying potential price increases and decreases, respectively. To ensure a fair comparison with the methodology presented by Nguyen et al. (2015) (Section 3), we adopted a similar linear kernel for the SVM and maintained consistency in both the data and model parameters [48].

This research explores several methodologies for stock price prediction. Big data approaches leverage vast amounts of publicly available information, often analyzed using platforms like Hadoop. Deep learning techniques, inspired by neural networks, offer another avenue for prediction. Long Short-Term Memory (LSTM) networks, a specific type of Recurrent Neural Network (RNN), address challenges associated with long-term dependencies in data. Additionally, sentiment analysis of social media data and news articles can provide insights into broader market trends and potential shifts in stock prices. Finally, time series analysis, recognizing the time-dependent nature of stock prices, serves as another popular method for forecasting [49].

This research employs a literature review and meta-analysis methodology to investigate the efficacy of machine learning in enhancing existing stock pricing strategies. By systematically analyzing findings from prior empirical studies, this research aims to address the question: To what extent does machine learning provide value compared to traditional stock pricing methods? Given the presence of conflicting research regarding machine learning's effectiveness in investment, a meta-analysis is particularly valuable for synthesizing past results. To locate relevant studies, we will utilize resources from both the Utwente library and online databases [50].

This section is unlikely to belong to the methodology part of a research paper focusing on image detection and recognition with neural networks. The methodology describes the specific steps taken in your research, while this paragraph talks about the general capabilities of neural networks (NN) and convolutional neural networks (CNN) in the field [51].

This research focuses on traffic infractions and their impact on road safety. To investigate this issue, we will employ [insert your chosen methodology, e.g., a survey, interviews with traffic authorities, analysis of accident data]. This methodology will allow us to gather data on the prevalence of traffic infractions and their contributing factors [52].

Table 2. Summary of different methodology

Paper Number	ANN	SVM	Hybrid & Other Ai Techniques
[1]	✓		
[5]			✓
[6]	✓		
[9]	✓		
[2]	✓		
[3]	✓		
[4]	✓		
[11]		✓	
[12]			✓
[13]		✓	
[18]	✓		
[19]	✓		
[21]	✓		
[23]	✓		
[24]		✓	
[25]		✓	
[26]	✓		
[34]	✓		
[36]			✓
[37]	✓		
[38]	✓		✓
[39]			✓
[40]			✓
[41]			✓

[42]	✓
[43]	✓
[44]	✓
[45]	✓

5. COMPARATIVE REVIEW BASED ON ALGORITHM

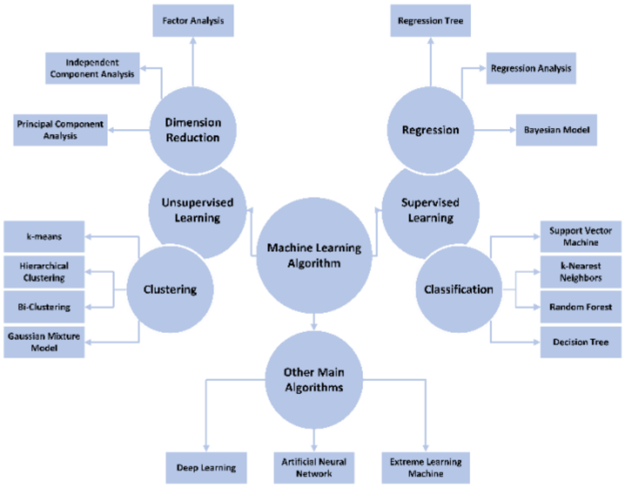

Figure 3. Algorithms for Stock Market Prediction using Artificial Intelligence.

5.1 Long Short-Term Memory

In paper [1], authored by Xiongwen Pang et al. propose a deep LSTM algorithm that outperforms bedded layers in forecasting immature stock requests froms stock time series by affecting the input of each LSTM.

Built on top of RNNs, the authors Theyazn H. H. Aldhyani et al. present the Long Short-Term Memory (LSTM) model. When dependencies over the long term are crucial, the LSTM model comes in handy. To do this, it uses input gates to remember information about dependencies for a longer time and forget gates to erase it from the input. When it comes to predicting outcomes from unknown time series, LSTM shines [18].

The author Johnson Jaya stated that LSTMs maintain a memory-like structure, which transfers information between cells and preserves long-term dependencies. This helps in training hyperactive parameters like model and algorithm parameters [38].

Author Abhinav Tipirisetty describes LSTM as a cellular model specifically designed to facilitate numerical evaluation, helping networks learn long-term dependencies, analysing stock prices over time and improving data records with textual data for future analysis [6].

The researchers Pushpendu Ghosh et al. have proposed the use of long-short-term memory networks (LSTM) and random forests to predict stock prices. These methods have been found to outperform the traditional ARIMA model in terms of lower forecast errors and greater accuracy. In addition, the use of attention mechanisms has improved the predictions, while sentiment analysis has significantly enhanced the accuracy of both LSTM and ARIMAX models [23].

In paper [3], author Hari Kiran Sai Surayagari explains that LSTM (Long Short-Term Memory) is a type of neural network that uses gating and issue-smart addition. LSTM has four types of gates: cellular, input, output, and forget gates. These gates act as switches, determining whether input should be written to the memory cell, whether the network should pay attention to certain inputs, or whether the values in the memory cell should be reset.

In paper [11], Prince Vipulbhai Patel introduces LSTM, a supervised learning model that converts time series data into supervised learning problems by splitting it into input and output sets. This approach enhances the predictive accuracy of the model. LSTM works on data within the activation function's scale, similar to hyperbolic functions, and requires three inputs for forecasting.

The authors, Jingyi Shen et al. introduce PCA, a technique that reduces the constraints of input data. However, it is necessary to pre-process the data before feeding it into the LSTM layer. The LSTM structure is composed of two layers: the input LSTM layer and the affair layer. These layers predict stock price trends based on time series data [7].

Standard Intermittent Neural Networks such as LSTM and GRU have been enhanced to overcome memory issues and improve predictions. According to author Debalina Barua, Bidirectional LSTMs utilize two separate intermittent nets for tasks including speech recognition and textbook categorization. Multi-head attention schemes have been introduced to optimize queries, values, and keys. VGRUs and LSTM have simpler infrastructures and improved control over retired states [15].

Chen et al.'s RNN-Boost algorithm utilizes a hybrid approach combining Recurrent Neural Networks (RNN) with AdaBoost. RNNs are chosen for their ability to process sequential data and capture temporal dependencies, while AdaBoost is used to enhance the model's prediction accuracy by focusing on instances that are hard to predict. The RNN employed utilizes Gated Recurrent Units (GRU) to mitigate the vanishing gradient problem, ensuring more effective learning of dependencies over time [53].

The author Akash et al. discussed a algorithmic approach incorporates a blend of ANN, Random Forest, SVM, and LSTM algorithms. Their selection is strategic, aiming to leverage the strengths of each algorithm to predict stock market movements more accurately. They underscore the importance of algorithmic diversity, allowing for a more nuanced analysis of data points from different dimensions, including historical prices and news sentiment. This multi-algorithm strategy is a key aspect of their research, distinguishing their work by combining outcomes from different algorithms to achieve higher accuracy in stock market predictions [54].

5.2 Learning and Double Deep-Q Learning Algorithms

In the paper [5], author Rafael Konstantinou explained that supervised learning is a system where a network is trained to produce a desired response to a training vector. The network's parameters are adjusted based on the training vector and an error signal until the output is close enough to the desired result. Supervised learning is generally used for classification and regression tasks. Regression involves estimating connections between data points, similar to function approximation, by finding a function that can describe these points. Unsupervised learning optimizes network parameters based on task-independent measures of input features. This enables the network to automatically encode input features and create new classes. It is commonly used to create auto-encoders and auto-decoders, which capture important features and discard non-important ones. Reinforcement learning is a machine learning field inspired by behaviourist psychology. Mapping scenarios to behaviours that optimize numerical reward signals is its goal. The four sub-elements of reinforcement learning are the environment, the agent, the reward function, the value function, and an environment model. The value function establishes the states' long-term desirability, the reward function links perceived states to actions, and an environment model is utilized for planning. The policy establishes the agent's behaviour. The long-term objective is to optimize the total reward earned.

The Double Deep Q-Network (DDQN) is a reinforcement learning system that Andrew W. Brim has described. The architecture of the RL landscape is generated by analyzing candlestick photos using a multi-layer Convolutional Neural Network (CNN). The assessment network replicates the target network's weights one hundred times, serving as an open AI gym. DDQN learns from input states using spread mean regression and employs a vector of action values. Input characteristics are used to optimize a Q-function that maximizes pricing. The DDQN architecture makes use of a Pytorch NN that has a triple-layer output, a 50-node fully-connected layer with a RELU non-linear activation function, and ten-feature input[34].

5.3 Genetic Algorithm

Isaac Kofi et al. introduce Genetic Algorithm (GA) as an evolutionary algorithm that's based on Darwinian natural selection and genetics in biological systems. GA operates by working with a population of feasible solutions, where each chromosome is selected through fitness functions. The GA cycle evolves from the first generation to the last generation, with acceptable results obtained when stopping criteria are met. This study implements the GA feature selection process [24].

Authors Troy J. Strader et al. propose an algorithmic approach to determine point discretization and connection weight for artificial neural networks that can predict stock price indicators. This approach combines inheritable algorithms with either artificial neural networks or support vector machines to overcome the limitations of single-fashion approaches and improve the algorithm's performance. Previous research has used GA for training networks, point subset selection, and topology optimization, but this study uses GA to reduce the point space complexity [55].

According to author Nusrat Rouf et al. genetic Algorithm (GA) is a heuristic approach that imitates natural evolution to solve problems. In SMP, it is utilized to refine trading rule parameters to improve accuracy. For example, a smart decision-support system for stock trading employs GA and rough sets to establish optimal trading rules [31]. Author Erhan Beyaz, proposed developing a fuzzy system for stock forecasting. The system combined SOM for clustering, fuzzy sense for rule birth, and an inheritable algorithm for optimization, and it was performed in a database of rules [14].

5.4 Back-propagation learning Algorithms

The authors, Yakup Kara et al., used a back-propagation algorithm to train a three-layered feedforward artificial neural network (ANN). The neurons in each layer were connected to the neighbouring layers using weights. The initial weights were assigned randomly, and a range of input patterns were created for a given set of relationship pairs. The model uses the gradient descent method to update the weights to minimize the root mean square error (RMS). The retired layer uses a tangent sigmoid transfer function, while the affair layer uses a logistic sigmoid transfer function. The model's outputs range from 0 to 1, where a value below 0.5 indicates a downscaling direction or an adding direction [2].

5.5 Levenberg Marquardt and Adam Algorithms

Ritika Chopra et al. introduce the Levenberg-Marquardt Artificial Neural Network (ANN) which is a mechanism for learning that utilizes back-propagation of previous errors to adjust weights and reduce errors. This process involves communicating the error of one neuron to all other neurons in the network, leading to optimal network weights. The ANN's structure, including the number of layers, neurons, input and output neurons, is crucial for effective learning. In the hidden layer, the activation function performs a non-linear transformation of the input data using a standard logistic function [56].

According to author Yixin Guo, the Adam algorithm is utilized as the optimizer for a model where the literacy rate is 0.001 and training data is randomly chosen. The loss function is measured using the MSE index, which decreases rapidly and ultimately converges. To make the prediction model more accurate, the root mean square error is taken into account [19].

5.6 Random Walk Theory (RWT) and High-Performance Computing (HPC) Algorithms

The authors Mehar Vijha et al. present an ensemble machine learning technique called Random Forest (RF) in this paper. RF is used for regression and classification tasks by combining multiple decision trees to reduce model friction. It creates new variables for each tree to train, which determines decision-making at bumps. The main objective of RF is to minimize forecasting errors by treating stock request analysis as a bracket problem. It predicts the coming-day closing price of a company based on training variables. However, high noise in stock request data can cause tree growth to differ from the anticipated outcome [21].

Ashraf S. Hussein et al. have developed a system for predicting stock requests using high-performance computing methods. The system focuses on accuracy, confidence, and understanding while keeping the prediction time in mind. However, the advanced computational resources used in the system, such as computational grids, are not easily accessible to most users and financial experts [25].

5.7 Term Frequency-Inverse Document Frequency (TF-IDF), k-Means and Convolutional Neural Networks (CNNs) Algorithms

The paper [40], two algorithms are shown in the work of Vaishali Ingle et al. To find the word scores in a plain text corpus, TF-IDF uses text features. It does this by comparing the document's frequency of a phrase to its overall dataset frequency, and then using that comparison to find the term's relevance. The term frequency measures the importance of a word based on how often it occurs in a text. But TF-IDF evens out this value by taking the term's frequency in the whole dataset into account, which is particularly useful when several documents share a single phrase. However, in order to partition data into groups, deep literacy frameworks use the k-means algorithm, an unsupervised learning method. Based on characteristics that are expected to vary within the population, it groups compliances into k-figures with the closest mean. Among the columns used by the method for clustering are "TF-IDF," "open price," "high price," "low price," and "close price" in the dataset.

Image processing, voice recognition, and time series analysis are just a few of the many applications of the Convolutional Neural Network (CNN) introduced in the paper by Theyazn H. H. Aldhyani and colleagues. CNN uses filters to extract features from input data, which is how it works. The output, pooling, input, and convolutional layers make up its architecture. The input matrix is multiplied by a filter with weight and form characteristics to yield the output matrix [18].

The Author Mohammad Al Ridhawi, presents the CNN model that harnesses Google's Word2Vec8 model to capture the original textbook behaviour. The model uses Gaussian noise, complications, and Tanh activation layers to mitigate the issue of overfitting and produce output vector sizes that vary. The model remains active during the training process [35].

5.8 Tweet Node, Bert model and Regression machine learning (ML), classification ML Algorithms

The paper [36] authored by Huihui Ni et al. focuses on the creation of a tweet node model that uses tweet data and a modified term frequency-inverse document frequency (TF-IDF) algorithm. The model extracts important elements, such as tweet replies and emotional factors, to create a network of tweet nodes. The improved node2vec is used to encode tweet bumps, which generates representations for Twitter data. Additionally, BERT is used to analyse emotional factors in tweet text and supplement the representation using multilayer transformers.

In their discussion of stock market analysis methods, Sohrab Mokhtari et al. include both technical and abecedarian approaches. Abecedarian analysis uses classification algorithms to categorize public sentiment based on news and social media, while technical analysis uses regression machine learning techniques to forecast stock price movements. This study's findings suggest that present AI technology isn't sophisticated enough to defeat the stock market, as the median performance of these algorithms still can't do better than the market [41].

In the paper [31], In introduction, Nusrat Rouf et al. provide regression algorithms, a class of prediction approaches for modelling the interplay between independent and dependent variables. These algorithms have been employed in various studies, including predicting a company's stock prices through regression analysis and candlestick pattern detection.

5.9 Neuro-fuzzy, Gustafson-Kessel and Clustering, Boruta Algorithms

The paper [30] authored by Lufuno Ronald Marwala, introduces the concept of neuro-fuzzy models that combine both fuzzy sense and neural network models. These models produce data-driven models that interpret mortal logic using information and query. The Adaptive Neuro-Fuzzy Conclusion System (ANFIS) serves as an alternative to the Sugeno-type fuzzy rule base and radial base function network. These models can be fine-tuned structurally and parametrically to achieve optimal performance. With the use of automated styles, computational efficiency may be improved while colourful literacy styles are used. An expansion of the conventional fuzzy c-means method, the Gustafson-Kessel (GK) algorithm is also covered in the article. GK algorithm uses an adaptive distance norm to describe clusters of different geometric shapes in a data set. The algorithm comprises four-way calculations of cluster prototypes, computation of covariance matrices, computation of distances, and streamlining of the partition matrix. Advantages include compact, asleep fuzzy sets, lower perceptivity to data normalization, and the capability to descry different shapes.

The paper [31], authored by Rouf et al. presents fuzzy algorithms (FA), which are decision-making methods based on human reasoning. The authors discuss the use of the adaptive neuro-fuzzy inference system (ANFIS) for sentiment analysis on social media. Additionally, hybrid fuzzy approaches have been proposed by researchers such as Sedighi et al. (2019), who developed a model for prediction using Artificial Bee Colony (ABC), Support Vector Machine (SVM), and ANFIS. The study focused on achieving accuracy and quality in forecasting, and the results showed that this approach is more accurate.

The author Erhan Beyaz, has developed a forecasting process that incorporates the market state by using a clustering algorithm to identify different request states. The process involves developing forecasting models for the target companies for each state, accommodating stocks not affected by different request states, and using request sentiment index values to generate moods for the overall stock request. The "k-means++" clustering algorithm is used for this purpose. Additionally, the Boruta algorithm is a feature selection system that uses arbitrary forests as a forecaster. It determines if a point is significant by using the variable significance statistic. The method sorts the original characteristics as either pertinent or unimportant based on the mean loss in accuracy after adding random features dubbed "shadows" to the dataset. A 252-day forecasting horizon was used for the study in the simulations [14].

5.10 Multiple Linear Regression, Polynomial Regression, Decision Tree Regression, and Random Forest Regression Algorithm

Author Shreya Pawaskar describes three different regression algorithms that are commonly used in data analysis. Multiple linear regression is a more advanced version of simple linear regression, which is used to predict a variable's value based on two or more independent variables. It is useful for accurate data forecasting, time series analysis, and identifying causal effect dependencies. This algorithm has practical applications in many fields, including college CGPA prediction, weather forecasting, and stock market analysis.

A second statistical technique that makes use of nth-degree polynomials to describe the connection between variables that are independent and dependent is polynomial regression. The use of higher-order polynomials, such square, cubic, or quadratic, enables the fitting of non-linear lines to data sets. Health outcomes and sediment isotopes are two areas where it finds widespread usage. The polynomial regression approach is well-suited for prediction since the dataset used is non-linear.

One supervised learning approach that finds usage in both classification and regression is Decision Tree Regression. It breaks down nodes into sub-nodes using mean squared error, and it uses strategic splits to forecast both continuous and categorical output variables; the level of delicacy depends on the splits.

Lastly, Random Forest Regression employs ensemble learning to provide precise predictions; it is a supervised learning technique. It finds application in tasks like diabetes prediction and product suggestion, and it shines with big data sets. Multiple regression decision trees work together to provide output predictions [27].

5.11 Other Algorithms

Author Alice Zheng et al. discuss binary Logistic Regression as the first algorithm used to classify outcomes between two possibilities. The weight matrix (W) is obtained by multiplying the input variables (X) which have m rows and n columns. The sigmoid6 function is used to transform real numbers into a 0 to 1 space, calculate the cost function for each iteration, obtain the gradient of the cost function and update the weight matrix W [13]. The study incorporated technical indicators and stock prices into a model to predict the changes in stock prices for the next n days. The model achieved a high prediction rate of 61.5% for n = 1 and 55% for n greater than or equal to 10, with improvements in prediction for larger n values [8].

In paper [9], authored by Víctor Rubio Jornet explains that a multi-layer perceptron (MLP) is a type of neural network that combines multiple perceptron's to create a fully-connected network. The network is used for binary classification, and an activation function is added at the end of the process. The K-NN algorithm is a classification method that predicts the category of a data point based on its distance from existing data points. It can use different prediction types, such as Euclidean distance. The number of data points used is critical, and a high value of k results in lower bias. The Random Forest algorithm is a method that uses a large ensemble of decision trees to classify data entries between 1 and 0. To ensure low correlation between models, the algorithm uses two methods: bagging, where each tree uses the same amount of data, and fine randomness, where each tree has a random subset of features selected from the original ones. The project employs the Support Vector Machine (SVM) algorithm to classify data points into binary categories of 0 and 1. The SVM algorithm can be varied between Linear, RBF, or Polynomial SVMs, with the kernel determining the shape of the classification function.

Isaac Kofi Nti et al. explain that Support Vector Machine (SVM) is a type of supervised machine learning algorithm that is used for both regression and classification tasks. It works by separating data nodes using a linear separator to identify different classes in multidimensional environments. SVM applies a hyperplane to each expected output and input feature in the training dataset, and this separates binary-decision classes in 2-attributes. The algorithm transforms input features into a high-dimensional feature space [24].

In paper [4], author Iyyappan. M et al. introduce the Holt-Winters triple exponential algorithm. This algorithm is used to capture demand and trend levels over time by breaking down demand observation into three components: base level, trend level, and seasonal factor. The algorithm applies smoothing factors α, β, and c, and then recomposes demand expectations.

The author, R. Dileep Kumar, et al. discuss machine learning and deep learning are two techniques used to create predictive models from historical data. These techniques use algorithms such as supervised, semi-supervised, unsupervised, and reinforcement to transform data into models. The algorithms function by predicting output values from input data, and classification and regression are the primary processes involved in this. The four categories of machine learning algorithms are Supervised, semi-supervised, unsupervised, and reinforcement learning algorithms. Over time, supervised models can improve in accuracy since they are trained on labelled data sets. These algorithms are essential for problem solving and outcome prediction based on past data [26].

In their publication, Mariam Moukalled et al. addressed the application of RNNs. Various RNN cell types, including Vanilla RNN, LSTM, and GRU, are used in sequence data machine learning algorithms. After training the RNN model using normalized records, the researchers were able to determine the final price of the day. When dealing with problems with regression and classification, another supervised machine learning approach is Support Vector Machine (SVM). To optimize the border between outputs and alter the data, SVM employs a kernel trick approach. Stock market forecast makes extensive use of it. The Support Vector Regression (SVR) model is used for regression, while the Support Vector Machine (SVM) model is utilized for classification. With the end-of-the-day price serving as the output, both models use the same vocabulary and features to forecast continuous value [28].

Author Jingyi Shen et al. explain the process of extending features (FE), which involves three methods of processing: normalization, polarization, max-min scaling, and fluctuation percentage. To select features, they use recursive feature elimination (RFE), which has three parts: ranking, resampling, and external validation. To eliminate overfitting and ensure high-weighted features, they use data resampling. By using randomized principal component analysis (PCA), they maximize the decrease of the training data matrix's scale. Classification accuracy may be negatively impacted by characteristics that are either irrelevant or of low relevance [7].

In their paper [31], an investigation of several algorithmic strategies for SMP is presented by Nusrat Rouf et al. Utilizing sentiment and historical data from the banking, mining, and oil industries, they evaluate the efficacy of Logistic Regression, Decision-Boosted Tree, and Support Vector Machine algorithms. Decision-Boosted Tree is shown to be the most effective algorithm among the three algorithms tested. Taking intra-day price fluctuation into account is something the authors recommend for better accuracy. A classification approach that relies on the Bayesian Theorem of probability is Naïve Bayes (NB). Textual data from many sources, such as traditional and social media data, may be classified and compared using it, and sentiment analysis makes extensive use of it for this purpose. To improve prediction performance, hybrid methods mix non-linear and linear models. According to research, a smart hybrid model combining linear and non-linear components, such as an autoregressive moving reference neural network and an exponential smoothing model, may accurately forecast stock prices. The model derived predictions from prediction mistakes while minimizing errors caused by non-linear processing. More complex neural networks, known as Deep Neural Networks (DNNs), have more hidden layers and neurons, making them better at learning. Algorithms like Convolutional Neural Networks, Deep Belief Networks, and Long-Short Term Memory rely on them often when making financial forecasts.

LSTM networks excel at learning long-term dependencies and achieve higher accuracy than other methods.

Central to the paper's exploration is the employment of several machine learning algorithms, each chosen for its proven efficacy in predictive analytics. The authors delve into the specifics of algorithms such as linear regression, decision trees, and neural networks, discussing their application to the stock market data. Notably, they highlight the customization and tuning of these algorithms to suit the nuances of financial data, optimizing them for better prediction accuracy. This detailed exploration into algorithms demonstrates the authors' deep understanding of machine learning's capabilities and limitations in the context of stock market prediction [57].

The study employs an elastic net regression-based training method, specifically highlighting the use of a multi-step adaptive elastic-net (MSAENet) algorithm. This algorithm is pivotal for feature selection in the context of high-dimensional regression, optimizing the prediction model by identifying the most relevant features from the extensive dataset [58].

Table 3. Summary of different algorithm solution

Algorithm	Advantage	Disadvantage
Long Short-Term Memory [1], [3], [6], [7], [11], [15], [18], [23], [38], [53], [54], [58]	Excellently retains information over long periods, making it ideal for sequential data like speech and text.	Computationally intensive, requiring significant resources and time to train.
Learning and Double Deep-Q Learning [5], [34]	Autonomously over time through data exposure, enhancing accuracy and efficiency.	Complex to implement and tune, with higher computational costs than simpler methods..
Genetic [14], [24], [31], [55]	Offers robust solutions to optimization and search problems by mimicking natural selection processes.	Can be slow to converge and may get stuck in local optima, missing the global best solution.
Back Propagation Learning [2]	Fundamental for efficiently training neural networks, enabling them to learn from their errors and improve.	Gradient vanishing or exploding problems can occur, making it challenging to train deep networks without careful initialization and adjustments.
Levenberg-Marquardt and Adam [19], [56]	Highly efficient for solving nonlinear least squares problems. Combines the benefits of adaptive gradient algorithm and RMSprop,	May lead to convergence to suboptimal solutions due to its bias-correction mechanism which can be misleading in certain conditions.
Random Walk Theory (RWT) and High-Performance Computing (HPC) [21], [25]	Provides a simple model for stock price movements, emphasizing. unprecedented speeds, significantly advancing research and applications in various fields.	Oversimplifies market behavior, ignoring factors like trends, patterns, and investor psychology. Requires significant investment in infrastructure, expertise, and energy resources.
Term Frequency-Inverse Document Frequency (TF-IDF), k-Means and Convolutional Neural Networks (CNNs) [18], [35], [40]	Effectively measures word importance in documents within a corpus, enhancing the accuracy of information retrieval and text mining.	Requires significant computational power and data for training, limiting accessibility for some applications.
Tweet Node, Bert model and Regression machine learning (ML), classification ML [31]	Provides valuable insights into relationships between variables, aiding in predictions and decision-making.	Can be prone to overfitting, especially with complex models or insufficient regularization.
Neuro-fuzzy, Gustafson-Kessel and Clustering, Boruta [14], [30], [31]	Combines neural networks' learning capabilities with fuzzy logic's handling of uncertain or imprecise information, offering robust models.	Complexity and interpretability can be challenging, requiring expertise to design and tune. Determining the optimal number of clusters a priori can be difficult, affecting its effectiveness.
Multiple Linear Regression, Polynomial Regression, Decision Tree Regression, Random Forest Regression [27]	Predicts outcomes based on multiple predictors, providing insights into the relationships between variables. Models nonlinear relationships between variables, offering more flexibility than linear regression.	Assumes a linear relationship, potentially oversimplifying complex real-world interactions. Risk of overfitting increases with the degree of the polynomial, making model selection critical.
Multiple Machine Learning [4], [7], [8], [9], [24], [26], [28], [31], [57], [58]		

6. COMPARATIVE REVIEW BASED ON DATASET

All the work done here is done on stock market prediction using machine learning. Here uses different kind of datasets based on various kind of algorithms to a better prediction. The main goal here is to solve stock market predictions through machine learning and using different datasets.

6.1 Stock Exchanges by countries

In the paper [1], Xiongwen Pang et al. utilize the machine learning models RNN and LSTM on two different datasets. Due to the large size of the datasets, these two algorithms are the best fit for implementation. Additionally, the lack of historical data on paper has been noted.

Data pre-processing is crucial for neural network stock market return prediction, according to authors Ritika Chopra and Gagan Deep Sharma. Data standardization, principal component analysis, and Z score are used by around 55% of the publications in the dataset to decrease mistakes and reduce overfitting. The paper recommends that feature selection algorithms and parameter optimization approaches be the subject of future research, highlighting the significance of input data in AI models. Additionally, the authors recommend that researchers look into financial series forecasting, compare and contrast AI models in existing and emerging economies, and prioritize data pre-processing methods to improve prediction accuracy in future studies [56].

Author, Jaya Johnson used SVM regressor and LSTM models to predict stock prices. The SVM regressor was employed for stock price prediction, while the LSTM model was used for sentiment analysis. The sentiment analysis was performed on merged datasets from Kaggle and Yahoo Finance, and a linear kernel from NLTK's Sentiment Analyser was used for this purpose. The prediction was based on categorized news data from the UCI repository, and the accuracy varied based on news alignment [38].

Author, Yixin Guo evaluates the dataset using ARIMA, GARCH, LSTM, and a mixed model for stock price prediction. The focus is on the MSE and Average Error Rate metrics. The study highlights the potential of deep learning, specifically LSTM, in forecasting financial time series. It offers potential investment guidance [19].

The research paper [5] authored by Rafael Konstantinou. The paper focuses on a dataset of daily stock transactions, particularly the OMSX30 index that represents the 30 most significant companies in Sweden's stock market. The dataset was gathered from November 2016 to May 2017. To evaluate the model's generalization, it was extended to include additional companies. The data is processed using temporal representation and additional financial indicators and oscillators are added. Normalization of the data is carried out using four methods: Along Channel, Across Channel, Mixed Channel, and External Normalization.

The goal of Yakup Kara et al.'s research is to forecast the future movement of stock price indices in the Turkish stock market using ANN and SVM. In order to determine the optimal parameter combinations for each model, the research carried out parameter setting experiments. The results show that with an average holdout performance of 75.74 percent, the ANN model achieves better results than the SVM model. Especially in developing economies like Turkey's, this study adds to our knowledge of how well machine learning methods can forecast stock market fluctuations. Keep in mind that owing to differences in data, characteristics, and technique, it may not be totally fair to compare various research [2].

Isaac Kofi Nti et al. highlights the significance of comprehending financial markets, particularly stock markets, for the benefit of the economy and society. It covers three main approaches to stock market forecasting: fundamental analysis, technical analysis, and technology (machine learning) methods. Fundamental analysis concentrates on a company's financial position and economic indicators, while technical analysis predicts future stock prices by analysing historical market trends. The research aims to systematically review previous studies on stock market predictions, taking into account differences among existing works and categorizing decision-making techniques. It also discusses the unpredictability of the stock market and the role of machine learning techniques in stock market analysis [10].

In this paper [8], authored by Jinan Zou et al. discuss a comprehensive overview of the challenges faced in stock market prediction and the future directions of research is provided. The paper focuses on areas of improvement and potential research avenues. It highlights the need for better evaluation metrics, improvements in generalization ability, and addressing anomalies. The text also integrates diverse techniques such as continual learning and distributional RL. However, it lacks specific details on dataset usage, concrete examples, and connecting challenges with solutions.

Islam et al.'s study utilizes a rich dataset collected from the Bangladesh stock market, covering financial statements, macroeconomic indicators, news articles, , historical stock prices, and social media data for the top organizations listed on the Dhaka Stock Exchange (DSE) from January 2022 to December 2022. This dataset includes both numerical and textual data, allowing for a detailed analysis of stock market trends and sentiments. The selection of predictive features is based on a literature review, encompassing both fundamental and technical indicators [59].

In paper [4], authored by Iyyappan. M et al. discuss a project that uses machine learning algorithms like the Holt-Winters algorithm, recurrent neural network (RNN), and recommendation system to predict stock market prices. The system utilizes historical stock data from the National Stock Exchange (NSE) and operates in real-time, allowing users to make informed investment decisions. However, there are some areas for improvement, such as enhancing evaluation metrics, improving data quality, comparing the system with existing models, incorporating user feedback, and addressing ethical considerations such as transparency and potential biases. The system's strengths include diverse algorithms, a rich dataset, and real-time execution.

Authors, Jinan Zou et al. demonstrate the immense value of the A-Stock dataset in the field of Natural Language Processing (NLP) and stock prediction. The dataset contains a broad range of stocks in the China A-shares market, providing a diverse set of data for analysis. Each stock is accompanied by financial news, which is crucial for capturing the dynamic nature of financial markets. Additionally, various stock factors such as free float share, dividend yield, and turnover rate are included, providing additional dimensions for algorithmic trading research. The dataset also includes minute-level historical prices for the news, enabling researchers to analyse stock movements with higher temporal granularity. The task of predicting stock movement is defined as a text-based classification problem, and the annotated dataset includes information on trading actions in news articles. The AStock dataset is a valuable resource for researchers and practitioners in the field of NLP-based stock prediction, as it offers rich and diverse content and a well-designed task formulation [39].

The authors Theyazn H. H. Aldhyani and Ali Alzahrani have conducted research using historical stock market data for Tesla and Apple, with a focus on their electric vehicles and sustainable energy sectors. The Tesla dataset spans from 2014 to 2017 and provides insights into Tesla’s stock performance, while the Apple dataset covers a broader period up to 2020, offering a more comprehensive view of Apple’s performance. The researchers have employed data normalization techniques to enhance gradient descent efficiency. The datasets have been divided into training and testing sets, with 70% of the data reserved for Tesla and 30% for Apple. A validation split has also been set aside during the training phase to prevent overfitting. These datasets are of great importance for training and evaluating deep learning models, particularly LSTM and CNN-LSTM. Since stock market data is temporal in nature, these models can learn patterns and trends, and their performance is evaluated based on their ability to predict stock prices accurately [18].

6.2 Government, Financial, Health and Marketing

Isaac Kofi Nti et al. presents a study on predicting stock prices using a novel ensemble classifier called GASVM, which is based on Support Vector Machines (SVM). The study utilized data from the Ghana Stock Exchange (GSE) over a span of eleven years. The paper highlights the challenges faced by SVM in handling high-noise and high-dimensional datasets, particularly overfitting. The GASVM model outperforms classical machine learning algorithms with an impressive prediction accuracy of 93.7%. The paper emphasizes the importance of using ensemble methods for feature selection and parameter optimization, which eliminates the need for manual parameter tuning. The study provides a comprehensive evaluation of the model's performance using various metrics [24].

6.3 Giant Companies

This paper [3], by Hari Kiran and Sai Surayagari delves into the topic of using machine learning to forecast stock market movements. Stock price predictions from news items are the main focus of fundamental and technical analysis, data collecting, and the use of long short-term memory (LSTM) neural networks. Google Finance and Event Registry are introduced as data sources in the study, with an emphasis on how reliable and accurate they are. It also covers the basics of machine learning, including supervised, unsupervised, and semi-supervised learning, as well as how to employ neural networks—specifically Long Short-Term Memory (LSTM) networks—to forecast stock values. The study does note certain caveats, however, including the fact that massive volumes of data are required, that news may take some time to spread, and that unpredictable stock prices may have an effect. Study findings and interpretations give light on real-world issues and pave the way for further investigation.

6.4 News and social media

Authors, R. Dileep Kumar et al. discuss the dataset they used, which includes its size, sources, and types. They also delve into the data pre-processing, training and testing splits, feature extraction, visualization, and model performance. There is a training set that contains 80% of the data and a testing set that contains 20% of the data.. The visualization helps to reveal patterns and trends, while the model performance provides insights into the performance of the model [26].

Derakhshan and Beigy's research is anchored in a meticulously compiled dataset, encompassing a diverse array of stock-related comments sourced from social media platforms. This dataset is bifurcated into English and Persian segments, showcasing the authors' commitment to exploring sentiment analysis across different linguistic and cultural contexts. For the English dataset, comments pertaining to 18 different stocks were collected over a year from Yahoo Finance Message Boards, providing a rich tapestry of investor sentiment and opinions. In contrast, the Persian dataset was curated from the Iranian stock market social network, Sahamyab, focusing on comments related to five specific stocks over a six-month period. Both datasets include not only textual comments but also explicit sentiment labels provided by users, offering a valuable resource for training and testing the sentiment analysis model. The authors took great care in balancing these datasets, ensuring a representative mix of positive and negative sentiments to refine the accuracy of their predictive models. By leveraging these comprehensive datasets, Derakhshan and Beigy aim to capture a wide spectrum of investor sentiments, providing a solid foundation for their analysis of stock price movements [60].

6.5 Historical Data of the last 5 years Tata Consultancy Services.

The research paper [27], written by Shreya Pawaskar, investigates the use of machine learning to forecast stock market movements using data collected by Tata Consultancy Services (TCS) from 2016 to 2021. The dataset includes a number of parameters, including the following: date, volume, high price, low price, adjusted close price, closing price, and open price. 'Day Perc Change' and 'HL_PCT' are two more columns added to the research, which uses machine learning methods such Decision Tree Regressor, Random Forest Regressor, Multiple Linear Regression, and Polynomial Regression. Decision Tree Regressor was shown to be the most accurate model in the research. The study also notes that there are advantages to employing machine learning to forecast the stock market, such as less human mistake and more consumer profits. Additional data integration, hyper parameter optimization, deep learning algorithm utilization, and the creation of intuitive machine learning web apps are all potential areas for future development.

6.6 Alpha Vantage

In the paper [13], authors, Alice Zheng and Jack Jin explore the use of machine learning techniques to predict short-term stock prices using time series data. The paper emphasizes the significance of stock trading in finance and the various analysis methods employed by professional traders. To access time series data from 82 randomly chosen stocks traded at NYSE, the authors use the Alpha Vantage API. The methods section describes their approach, which includes classification models, Linear Regression, and Support Vector Regression. Initial experiments reveal that models based solely on past prices have limited predictive capability. However, the integration of technical indicators, particularly with Support Vector Regression, displays significant improvement, achieving correct prediction rates of nearly 70%.

6.7 S&P 500 Index

The authors Andrew Brim and Nicholas S. Flann investigate the use of deep learning techniques, namely CNNs and DDQNs, to forecast financial market movements. The 2020 COVID-19 pandemic stock market catastrophe is their primary focus. Using feature map visualizations, the research compares the DDQN's performance to that of the S&P 500 Index and aims to understand how the neural network generates predictions. Apple, Amazon, Microsoft, Google, and Facebook are among the top 30 stocks in the S&P 500 Index that are included in the dataset, which covers stock prices from 2013 to 2019. CNN uses candlestick charts that display stock prices over the last 28 days as an input. The next day, the DDQN decides whether to go long, short, or do nothing at all with the stock. The training dataset consists of 52,920 observations over seven years (2013-2019). There are 3,780 observations in the testing dataset, covering the six months from January 2, 2020 to June 30, 2020. When comparing the DDQN's performance throughout the testing period to that of the S&P 500 Index, geometric returns are the primary criterion to consider [34].

The study utilizes a dataset comprising tweets related to eight S&P 500 firms and corresponding stock market data from Yahoo! Finance. Tweets were collected based on mentions of these firms using a dedicated application that leverages Twitter's API, focusing on interactive tweets to build the trust network. The dataset covers a period from January 1, 2015, to August 31, 2015, including 167 trading days and yielding significant insights into the correlation between Twitter sentiment and stock market behaviour [60].

Table 4. Summary of different dataset

Dataset	Paper Number
Government, Financial, Health and Marketing	[24]
Giant Companies	[3]
News and social media	[26]
Tata Consultancy Services	[27]
Vantage, Alpha, Alpha Vantage	[13]
S&P 500 Stocks	[34]
Stock Exchanges by Countries	[1], [2], [5], [10], [19], [38], [39], [56], [59]

Table 5. Comparisons among datasets

Dataset	Comparisons
Government, Financial, Health and Marketing	The GASVM model, based on SVM, was utilized with data from the Ghana Stock Exchange over eleven years, achieving a prediction accuracy of 93.7%.
Giant Companies	Google Finance and Event Registry were used as data sources for stock price predictions from news items using LSTM networks.
News and social media	The dataset included 80% training and 20% testing sets, with visualizations revealing patterns and trends to evaluate model performance.
Tata Consultancy Services	The dataset from Tata Consultancy Services, spanning 2016 to 2021, was analyzed using various regression models, with Decision Tree Regressor being the most accurate.
Vantage, Alpha, Alpha Vantage	Time series data from 82 NYSE stocks accessed via Alpha Vantage API, with technical indicators improving prediction rates to nearly 70% using SVR.
S&P 500 Stocks	The dataset covering 2013-2019 for the top 30 stocks in the S&P 500 Index was used to compare CNN and DDQN performance during the 2020 COVID-19 pandemic.

7. COMPARATIVE REVIEW BASED ON RESULT

7.1 Stock Exchanges by countries

In paper [5], the main goal was gathering information for the Stockholm stock market's OMSX30 index. The 30 most significant Swedish corporations are included in this index. Each trading day's transactions were kept in an ASCII text file that was arranged according to the stock name. Nevertheless, the neural network could not be trained efficiently, and the outcomes were unsatisfactory. It was discovered that the neural network was diverging in every instance, and the accuracy of the correctly classified data was marginally more than 50%. 51.2% was the lowest accuracy attained, and 62.5% was the highest.

In their research paper [2], in an effort to foretell the future of the daily ISE National 100 Index, the authors developed two practical models. We employed two classification methods, support vector machines (SVMs) and artificial neural networks (ANNs), and ten technical indicators as inputs to our model. In order to enhance the accuracy of their predictions, the scientists conducted two thorough trials of setting the parameters for both models. With an average accuracy rate of 75.74% and a standard deviation of 71.52%, the ANN model clearly beat the SVM model in the experiments.

Artificial neural networks, decision trees, and support vector machines were among the models tested for their ability to forecast the Ghana Stock Exchange stock market. Using error measures like RMSE, MAE, and MSE, the results demonstrated that artificial neural networks outperformed the other two techniques. Predictions of future performance varied between 36.55% and 97.58%, with testing data making up 20% and training data 80% [10].

One specific study uses a real-time dataset with fifteen equities as the system's input. The data is used by the algorithm to estimate or predict future stock values of different firms from different industries. The National Stock Exchange (NSE), the largest stock exchange in India, was the subject of the research despite the availability of several stock exchange marketplaces throughout the world. With the largest online digital exchange that makes it simple for customers to buy or sell stocks online, NSE is the biggest stock exchange in India for stock traders. The investigation concluded that the findings were satisfactory, with some RMSE values of less than 50 [4].

In the paper referenced as [15], historical data used to train the authors' models was sourced from the official website of the Dhaka Stock Exchange (DSE). Date, volume, high, low, close, and adj close prices are some of the financial data elements used in the models. For this purpose, the researchers used R-Squared, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE), among other standard strategic analytics. An accuracy rating of more than 90% was claimed for this model. The proposed LSTM model was implemented using Python. Using past data, it predicts how much BPML, ASIAINS, SAIHAMCOT, ARAMITCEM, and BEXIMCO will be worth in the future.

The paper [18], a methodology for forecasting the closing prices of two American corporations, Tesla, Inc. and Apple, Inc., using a mix of CNN-LSTM and LSTM, is presented by the authors. Both companies' shares may be found on the New York shares Exchange. Researchers in this study compared the accuracy of two deep learning algorithms—LSTM and a hybrid

CNN-LSTM—in predicting the future values of Apple and Tesla shares. It was found that CNN-LSTM performed better than LSTM in both the training and testing phases, with CNN-LSTM achieving 99.58% and LSTM 97.20%, respectively. This meant that CNN-LSTM was more accurate than LSTM.

The study described in paper [24] used data from the Ghana Stock Exchange (GSE) over an eleven-year period. The paper discusses the difficulties faced by SVM when handling high-noise and high-dimensional datasets, particularly overfitting. The GASVM model showed better performance compared to classical machine learning algorithms, with an impressive prediction accuracy of 93.7%, which is a remarkable achievement.

Astock is a dataset from China's A shares market that includes news and stock factors. The dataset consists of 40,963 news articles that are annotated with all instances of three trading actions (long, preserve, short). These news articles have a valid official license and originate from Tushare. To make the dataset realistic, it also includes various stock factors. Astock was used to create a semantic role labelling pooling (SRLP) representation for stock-specific news and predict stock movement. The self-supervised SRLP approach achieved a prediction accuracy of 64.09%, which is higher than the accuracy of StockNet and HAN Stock [39].

The results section of Sable et al.'s review encapsulates the findings from a comprehensive analysis of stock market prediction techniques. It underscores the efficacy of ARIMA for statistical predictions, while highlighting the advanced capabilities of LSTM and SVM in ML and DL frameworks, respectively. The review reveals a nuanced understanding of the advantages and limitations of these methods, offering insights into their application across different market conditions. Furthermore, the emphasis on Indian stock market datasets in the literature points to a focused area of research, suggesting a need for broader geographic diversity in future studies [61].

7.2 Financial Data Providers

Alice Zheng and Jack Jin have conducted research on using machine learning techniques and time series data for short-term stock price prediction. Experimented with several models such as logistic regression, Bayesian networks, simple neural networks, and support vector regression to forecast stock prices for the next n days. The initial models had error rates ranging from 40% to 50%, but the SVMs with a radial basis kernel demonstrated the best performance. The researchers improved the results by combining linear regression with technical indicators, which led to a correct prediction rate of 61.5% for $n = 1$ and 55% for $n \geq 10$. The Support Vector Regression (SVR) with a Gaussian kernel outperformed linear regression, achieving correct rates of 68.5% to 69.5%. [13].

In paper [38], various machine learning models and sentiment analysis techniques were tested to predict stock prices and the S&P 500 index over six months and one year. The results showed different accuracy levels for each model. Logistic Regression (LG) had a 61% accuracy rate, while SVM had a 95% accuracy rate, LSTM had a 63% accuracy rate, and SVM had a 99.5% accuracy rate. Although LSTM showed potential for improvement with larger datasets, it still remained lower at 34%.

The paper [40], describes a way to forecast stock values by combining deep learning, gradient-boosted models (GBM), and generalized linear models (GLM). The study states that over all firms and dates, GLM's average mistake rate is 0.70%, but GBM's average error rates are 0.00% for ITC and 7.52% for SUN. Although the article makes reference to a multi-layer feed-forward artificial neural network—the model used in deep learning—no details on its performance metrics or error rates are provided.

7.3 Giant Companies

Regression model performance needs to be expressed as an error. The performance could alter if the datasets are updated and produce different outcomes. There is no accuracy score available. In this situation, the coefficient of determination (R^2 error) and the root mean squared error (RMSE) can be applied.

1) R^2 error: It is the percentage of the dependent variable's variance that the independent variable can explain. It is employed to gauge the fit's goodness. A higher number indicates a superior model.

2) RMSE: It is the residuals, or standard deviation of the prediction errors. It assigns a comparatively high weight to the significant errors. A lower number indicates a superior model [27].

7.4 News and Social Media

The results for the models on the different stocks that were defined in section 3.6 are displayed in this part. The assessment criteria are as follows: directional accuracy, which examines the predicted value's direction in relation to yesterday's closing price; precision, which gauges the result's relevance; recall, which counts the number of genuine, pertinent results that were returned; and F-measure, which calculates the weighted average of precision and recall. For the several examined equities, SVM performs better than RNN, SVR, and DNN according to the directional accuracy metric (table 2). We describe the input data in Table 1 [28].

The authors were able to obtain a 68.5% prediction vs. real pricing performance just using the FSD model. However, the paper were attains a 74.3% prediction vs real pricing performance when the paper coupled the FSD model and the ESE model. Therefore, sentiment analysis from social media communications used for stock price assessment enhances stock price estimation. The paper compared our suggested SFD model to the most advanced models currently in use that have been published in the literature. Compared to current state-of-the-art models, our SFD model results boosted price prediction performance by as much as 17% [35].

The majority of stocks can achieve more than 50% accuracy according to the experimental results of two widely used embedding models, doc2vec and average bag of words. This indicates that the combination of historical stock price and tweet text information has a positive impact on the task of stock forecasting. An improved benchmark can be achieved by comparing the first two rows of the table with the third and fourth rows. This is demonstrated by the average bag-of-words embedding model. Compared to doc2vec, which encodes using the semantic link between sentences, it is superior. The meaning of each tweet can be better indicated by the average bag-of-words model because Twitter (now X) data cannot be compiled into a contextual whole like an article can. Second, the authors established whether the model outperforms the baselines for each stock by comparing the outcomes of models with the baselines, which demonstrates the superiority of the model's performance. In particular, the fully-equipped model outperforms the best baseline result on $Apple, $SPX, $PepsiCo, and $Amazon by 9.5%, 6.97%, 4.65%, and 6.98%, respectively. Furthermore, the precision of the model aligns with Doc2vec-CNN's on $WMT [36].

This paper's use of ML algorithms to predict public sentiment does not yield encouraging findings. With an accuracy of 76%, the SVM algorithm is the most accurate. Additionally, the effectiveness of these methods displays the AUC for each algorithm and compares the ROC curves. The best AUC score belongs to the SVM algorithm [41].

Nguyen and Shirai's results demonstrate the effectiveness of their proposed model. They report that their method outperforms traditional models based solely on historical prices by approximately 6.07% in accuracy. Furthermore, when compared to other sentiment analysis methods, their approach shows better accuracy than both LDA and JST (Joint Sentiment/Topic) models by 6.43% and 6.07%, respectively. These findings underscore the potential of integrating sentiment analysis with topic modelling for stock market prediction, highlighting the value of social media as a predictive tool in financial markets [62].

7.5 S&P 500 Index

In the paper [19]'s mix model process, As we can see, the closing index sequence was used to create an ARIMA model initially. The prior day's closing index served as the input feature. The MSE was 2.185 and the average error rate was 4.48% according to the results of the first test. Following hyper parameter adjustment, the prediction effect became much more accurate, and the average error rate decreased to 3.19% while the mean square error (MSE) fell to 1.213. The closing index's average yield fluctuation during the test period was 0.62%, which was substantially less than the model's ideal predicting effect. Second, following hyper parameter adjustment, the GARCH model demonstrated the best prediction effect, with an MSE of 1.923 and an average error rate of roughly 1.65% during the test period. They provide additional details about the S&P 500, and the three-factor forecast model is highly significant in real-world scenarios. Nevertheless, the average error rate at this point remains greater than the average yield volatility of the closing index, which is 0.62%, and more research is still needed to determine whether the GARCH method can be used to forecast the Shanghai Composite Index. Ultimately, the LSTM model's hyper parameters, which were set using 13 initial indications, were adjusted in order to test its predictive power. It was discovered that the average error rate was further decreased to 0.40%, with noticeably improved predictive capacity, and that the mean square error (MSE) between the model's predicted value and the actual value throughout the test period was 0.876. The Mixed Model's average error rate is 0.27%, and its mean square error is 0.412.

The findings of paper [23] show that the multi-feature method, which uses intraday returns in addition to closing price returns and opening price returns, performs better than the single-feature methods used by Krauss et al. (2018) and Fischer & Krauss (2017) for LSTM and random forests, respectively. The LSTM method generates a better return of 0.64% before transaction fees, as opposed to the 0.41% daily return achieved by Fischer & Krauss (2018). The use of random forests also improves daily return, going from 0.39% in Krauss et al. (2017) to 0.54%. The percentage of successful outcomes is 69.55% for LSTM and 69.57% for random forests. Also, when looking at annualized risk-return measures, the strategy beats both Krauss et al. (2017) and Fischer & Krauss (2018). This is because the approach has a smaller standard deviation and a better Sharpe ratio. Both the maximum drawdown and the daily value at risk are reduced by the technique.

The paper [34] uses a technique that is tested on 2018 stock prices and applied to all 38 stock pairings. For each of the 38 stock pairs, the results display the total spread cumulative returns for a range of NRM values from 1 to 1000. For all 38 stock pairs, the cumulative total returns are 131.33. For 249 out of the 252 test days, the DDQN did not produce any position actions while the NRM was set at 50.0. Nonetheless, returns of 1.42 were obtained throughout the three days that a lengthy activity was performed. This suggests that the system is capable of becoming more conservative while maintaining high performance. In the testing data, the DDQN generated a Q-function and successfully executed actions to long, short, and have no position on the pair's spread during training. The ADBE/RHT Spread mean, for example, has reverted at least three times in the last fifty trading days. The four pairs with the highest spread returns were CTWS/AWR 71.28, HBI/MRO 27.41, FCX/HBI 25.67, and CNX/HBI 7.52. The two lowest performing groups were ESV/GNW -9.64 and ESV/RRC -0.78. Overall, the findings imply that trading stock pairs using the DDQN strategy might be profitable while maintaining caution.

In the paper [37], The experimental results and several conclusions having financial consequences are presented by the writers. On StockNet's test data split from January 10, 2015, to December 31, 2015, on the S&P 500 index, they compared various approaches. Ten separate runs were used to average the comparison. MANSF outperformed the strongest baselines, StockNet, and Adversarial LSTM, by utilizing a learnt blend of historical prices and tweets using corporate ties. The price and text models' partial capacity to forecast market trends using unimodal features was noted by the authors. Additionally, they pointed out that the incorporation of a graph-based learning model, such GCN or GAT, can lead to improvements over individual modalities and verify the idea of exploiting inter-stock interactions for improved forecasting. To assess the practicality of MAN-SF, the authors also performed a profitability analysis on actual stock data. They discovered that MAN-SF generated an overall profit and greater risk-adjusted returns. Using S&P 500 index stock data, MAN-SF outperformed various baselines during the three-month common testing period.

Table 6. Summary of different results

Result based on	DataSet	Measures
Stock Exchanges by Countries [5], [2], [10], [4], [15], [18], [24], [39]	New york, Dhaka Stock exchange, Indain Stock, Stockholm stock market, China's A shares market	Average Accuracy=81.011%
Financial Data Providers [13], [38], [40]	Alpha Vendor	Average Accuracy=85.00%
Giant Companies [27]	Tata Company	Average Accuracy=82.00%

News and Social Media [28], [35], [36], [41]	X,(formerly named twitter), Social media, Yahoo finance.	Average Accuracy=87.08%
S&P 500 Index [19], [23], [34], [37]	S&P 500 Index	Average Accuracy=84.09%

Table 7. Comparison among results.

Result based on	Comparison
Stock Exchanges by Countries	The New York, Dhaka, Indian, Stockholm, and China's A shares stock exchanges achieved an average accuracy of 81.011%.
Financial Data Providers	Financial data providers like Alpha Vendor showed an average accuracy of 85.00%.
Giant Companies	Tata Company's predictions showed an average accuracy of 82.00%.
News and Social Media	News and social media platforms like X (formerly Twitter), Yahoo Finance, and others had an average accuracy of 87.08%.
S&P 500 Index	Predictions for the S&P 500 Index achieved an average accuracy of 84.09%.

8. DISCUSSIONS

Artificial Intelligence (AI) has the ability to change the ecosystem financial professionals and investors assess and make decisions regarding stock market predictions. It is important to include AI in stock market trading tools as it improves their ability to analyse market data and execute deals efficiently and rapidly.

8.1 Model Robustness Across Market Conditions

Artificial intelligence (AI) performs better when it has access to a wide range of diverse data. However, it is not extensively studied how well AI models can perform in different market conditions, including various economic cycles. As a result, it is crucial to conduct research to explore how well these models can generalize and perform in various market environments to ensure their reliability in dynamic and changing scenarios.

8.2 Handling Market Anomalies

Stock markets can experience exceptions, such as extreme price movements or sudden changes in trading volumes, sometimes due to insider trading. Therefore, research is needed to investigate how AI models can effectively handle and adapt to these exceptions, ensuring that predictions are not overly influenced by abnormal market behaviour.

8.3 Incorporation of Macro-Economic Indicators

While some studies consider historical stock data, the integration of macro-economic indicators and global events is often limited. Therefore, research should explore effective ways to incorporate external factors such as economic indicators, geopolitical events, and regulatory changes into predictive models for a more holistic approach to predictions.

8.4 Integrating Historical Data with Real-Time News, Social Media, and Sentiment Analysis:

The current research lacks comprehensive integration of historical data with real-time information from news, social media, and sentiment analysis. Merging these sources is crucial for a holistic understanding of market dynamics. Historical data provides context, while real-time news and sentiment analysis offer insights into dynamic market sentiments. The synergy of these sources can enhance predictive models, capturing the impact of external events on stock prices. Addressing this gap is essential for developing more accurate and adaptive AI models that can navigate both historical trends and rapidly changing market sentiments. This integration would provide investors with a more complete and timely picture for informed decision-making.

8.5 Ethical Considerations and Regulation:

The ethical considerations in stock market prediction using AI lack comprehensive exploration. There is a need for research that critically examines the potential ethical implications, including issues of transparency, fairness, and accountability in AI models. Current literature often overlooks the impact of AI-driven predictions on market dynamics, potentially leading to market manipulation or unintended consequences. In order to build ethical standards and legal frameworks that guarantee the responsible development and use of AI in the financial markets, it is essential to fill this gap. In order to preserve the integrity of financial institutions and develop confidence among investors, ethical concerns in stock market prediction demand specific attention.

Table 8. Comparison among discussions

Result based on	Comparison
Model Robustness Across Market Conditions	AI performs better with diverse data, but its performance across different market conditions needs further study to ensure reliability in dynamic scenarios.
Handling Market Anomalies	Research is needed to investigate how AI models can adapt to extreme price movements and sudden changes in trading volumes due to insider trading.
Incorporation of Macro-Economic Indicators	Effective ways to integrate economic indicators, geopolitical events, and regulatory changes into predictive models should be explored for a holistic approach.
Integrating Historical Data with Real-Time News, Social Media, and Sentiment Analysis	Current research lacks comprehensive integration of historical data with real-time news and social media, which is crucial for a holistic understanding of market dynamics.
Ethical Considerations and Regulation	Ethical considerations in AI-driven stock market prediction need more exploration, including issues of transparency, fairness, and accountability.

9. CONCLUSION

Predicting the stock market is a frequently discussed topic nowadays, and many researchers are striving to build models that can accurately forecast stock prices. The stock market is influenced by various factors, hence most researchers analyse different aspects to make their predictions. However, only a handful of them try to consider more than one factor. One common method for predicting stock prices is by analysing trends. This means scrutinizing a company's past stock prices to speculate on what its future value might be. It's a popular approach because, in the world of investing and trading, the actions of investors and traders have a significant impact on a company's value. If everyone follows the same trend-based approach and obtains similar results, more people are likely to use that approach, leading to the expected outcome [8].

During our study, we have identified three major frameworks that are mostly utilized for predicting the stock market and prices. These frameworks are Machine Learning, which is the most commonly used among the papers, followed by Deep Learning Neural Network and Artificial Intelligence (AI). Utilizing these frameworks can prove to be helpful in accurately forecasting the stock market and stock prices. There are certain algorithms, such as Long Short-Term Memory (LSTM) and Back-propagation that have proven to be effective in accurately predicting stock prices. LSTM has an average accuracy rate of 80 to 85%, which is considered to be efficient, according to some research papers. Additionally, Back-Propagation algorithm has an accuracy rate of up to 71%. Therefore, these algorithms are useful tools for predicting stock prices and trends.

It is important to recognize that there are certain challenges and limitations associated with using AI to predict stock market trends. This is because market dynamics are influenced by a multitude of unpredictable factors, and past performance does not always guarantee future results. Going forward, it is likely that the relationship between human expertise and AI capabilities will become even more crucial for successful stock market prediction. Striking a balance between leveraging the power of AI and maintaining a human touch in decision-making processes will be key to unlocking the full potential of AI in navigating the complexities of the stock market.

ACKNOWLEDGEMENT

We sincerely thank the university and department that provided essential materials and resources necessary for this research. Their contributions have been invaluable in the completion of this work.

REFERENCES

[1] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, “An innovative neural network approach for stock market prediction,” Journal of Supercomputing, vol. 76, no. 3, pp. 2098–2118, Mar. 2020, doi: 10.1007/s11227-017-2228-y.

[2] Y. Kara, M. Acar Boyacioglu, and Ö. K. Baykan, “Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange,” Expert Syst Appl, vol. 38, no. 5, pp. 5311–5319, May 2011, doi: 10.1016/j.eswa.2010.10.027.

[3] H. Kiran, S. Surayagari, A. Ben-Hur, and C. Stein, "Stock market predictions using machine learning," in Proceedings of the IEEE International Conference on Data Mining (ICDM), 2021.

[4] M. Iyyappan, S. Ahmad, S. Jha, A. Alam, M. Yaseen, and H. A. M. Abdeljaber, “A Novel AI-Based Stock Market Prediction Using Machine Learning Algorithm,” Sci Program, vol. 2022, 2022, doi: 10.1155/2022/4808088.

[5] R. Konstantinou, "Stock Market prediction using Artificial Neural Networks," in Proceedings of the International Joint Conference on Neural Networks, vol. 4, pp. 1847-1852, 2017.

[6] A. Tipirisetty, “Stock Price Prediction using Deep Learning,” in San Jose State University, San Jose, CA, USA, 2018. doi: 10.31979/etd.bzmm-36m7.

[7] J. Shen and M. O. Shafiq, “Short-term stock market price trend prediction using a comprehensive deep learning system,” J Big Data, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40537-020-00333-6.

[8] J. Zou et al., "Stock Market Prediction via Deep Learning Techniques: A Survey," in arXiv preprint arXiv:2212.12717, 2022.

[9] V. Rubio Jornet, V., "Stock Prediction with Machine Learning," Technical Report, Universitat Politècnica de Catalunya, Barcelona, Spain, 2019.

[10] 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: 10.1007/s10462-019-09754-z.

[11] B. Prince Vipulbhai Patel, J. Weigley, K. Mkrtchyan, and R. Mcilhenny, “Stock Market Prediction using Machine Learning,” 2021.

[12] N. Christina, “Stock Market Prediction using Sentiment Analysis,” in Master’s thesis, School of Science & Technology, Thessaloniki, Greece, pp. 9-11, 2021.

[13] A. Zheng and J. Jin, "Using AI to Make Predictions on Stock Market," Stanford University, Tech. Rep., 2017.

[14] E. Beyaz, "Effective Stock Price Forecasting Using Machine Learning Technique Whilst Accounting for the State of the Market," M.S. thesis, Dept. Comput. Sci., Univ. Manchester, Manchester, UK, 2019.

[15] D. Barua, "Dhaka Stock Exchange Stock Price Prediction using Machine Learning and Deep Learning Models," in Brac University, Dhaka, Bangladesh, 2022.

[16] S. Dixit and N. Soni, “Enhancing stock market prediction using three-phase classifier and EM-EPO optimization with news feeds and historical data,” Multimed Tools Appl, 2023, doi: 10.1007/s11042-023-17184-x.

[17] G. Sonkavde, D. S. Dharrao, A. M. Bongale, S. T. Deokate, D. Doreswamy, and S. K. Bhat, “Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications,” International Journal of Financial Studies, vol. 11, no. 3. Multidisciplinary Digital Publishing Institute (MDPI), Sep. 01, 2023. doi: 10.3390/ijfs11030094.

[18] T. H. H. Aldhyani and A. Alzahrani, “Framework for Predicting and Modeling Stock Market Prices Based on Deep Learning Algorithms,” Electronics (Switzerland), vol. 11, no. 19, Oct. 2022, doi: 10.3390/electronics11193149.

[19] Y. Guo, "Stock Price Prediction using Machine Learning," in Proceedings of International Conference on Machine Learning, Stockholm, Sweden, 2022.

[20] K. Bin Saboor, Q. Ul, A. Saboor, L. Han, and A. S. Zahid, “Predicting the Stock Market using Machine Learning: Long short-term Memory,” 2020. [Online]. Available: https://ssrn.com/abstract=3810128

[21] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, “Stock Closing Price Prediction using Machine Learning Techniques,” in Procedia Computer Science, Elsevier B.V., 2020, pp. 599–606. doi: 10.1016/j.procs.2020.03.326.

[22] F. Ali and P. Suri, “The Eurasia Proceedings of Educational & Social Sciences (EPESS) The Eurasia Proceedings of Educational A Bibliometric Analysis of Artificial Intelligence-Based Stock Market Prediction,” & Social Sciences (EPESS), vol. 27, 2022, [Online]. Available: www.isres.org

[23] P. Ghosh, A. Neufeld, and J. K. Sahoo, “Forecasting directional movements of stock prices for intraday trading using LSTM and random forests,” Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.10178

[24] I. K. Nti, A. F. Adekoya, and B. A. Weyori, “Efficient Stock-Market Prediction Using Ensemble Support Vector Machine,” Open Computer

Science, vol. 10, no. 1, pp. 153–163, Jan. 2020, doi: 10.1515/comp-2020-0199.

[25] A. S. Hussein, I. M. Hamed, and M. F. Tolba, "An Efficient System for Stock Market Prediction," Advances in Intelligent Systems and Computing, vol. 323, pp. 871–882, 2015, doi: 10.1007/978-3-319-11310-4_76.

[26] R. Dileep Kumar, "Stock Market Prediction using Machine Learning," 2022. [Online]. Available: www.jespublication.com

[27] S. Pawaskar, "Stock Price Prediction using Machine Learning Algorithms," Int J Res Appl Sci Eng Technol, vol. 10, no. 1, pp. 667–673, Jan. 2022, doi: 10.22214/ijraset.2022.39891.

[28] M. Moukalled, W. El-Hajj, and M. Jaber, "Automated Stock Price Prediction Using Machine Learning," in M.S. thesis, Dept. Comput. Sci., Amer. Univ. Beirut, Beirut, Lebanon, 2019.

[29] P. Tupe-Waghmare, "Prediction of Stocks and Stock Price using Artificial Intelligence : A Bibliometric Study using Scopus Database," 2021.

[30] L. Ronald Marwala, "Forecasting the Stock Market Index Using Artificial Intelligence Techniques," 2010.

[31] N. Rouf et al., "Stock Market Prediction using Machine Learning Techniques: A Decade Survey on Methodologies, recent Developments, and Future Directions," Electronics (Switzerland), vol. 10, no. 21. MDPI, Nov. 01, 2021. doi: 10.3390/electronics10212717.

[32] B. Li, K. C. C. Chan, C. Ou, and S. Ruifeng, "Discovering public sentiment in social media for predicting stock movement of publicly listed companies," Inf Syst, vol. 69, pp. 81–92, Sep. 2017, doi: 10.1016/j.is.2016.10.001.

[33] O. Alsing and O. Bahceci, "Stock Market Prediction using Social Media Analysis," in B.S. thesis, Dept. Comput. Sci., KTH Royal Inst. Technol., Stockholm, Sweden, 2015. [Online] Available: https://www.diva-portal.org/smash/get/diva2:811087/FULLTEXT01.pdf

[34] Andrew W. Brim, "Artificial Intelligence and Deep Reinforcement Learning Stock Market Predictions." [Online]. Available: https://digitalcommons.usu.edu/etd/8393

[35] M. Al Ridhawi, "Stock Market Prediction Through Sentiment Analysis of Social-Media and Financial Stock Data Using Machine Learning," in M.S. thesis, Dept. Elect. Eng. Comput. Sci., Univ. Ottawa, Ottawa, ON, Canada, 2021. [Online] Available: https://ruor.uottawa.ca/items/f37187bc-6b98-4a74-b7ba-3893f806bc1a

[36] H. Ni, S. Wang, and P. Cheng, "A Hybrid Approach for Stock Trend Prediction Based on Tweets Embedding and Historical Prices," World Wide Web, vol. 24, no. 3, pp. 849–868, May 2021, doi: 10.1007/s11280-021-00880-9.

[37] R. Sawhney, S. Agarwal, A. Wadhwa, and R. Ratn Shah, "Deep Attentive Learning for Stock Movement Prediction From Social Media Text and Company Correlations." [Online]. Available: https://www.investopedia.com/

[38] J. Johnson, "Machine Learning for Financial Market Forecasting," in M.S. thesis, Harvard Univ. Division of Continuing Education, Cambridge, MA, USA, 2023.

[39] J. Zou, H. Cao, L. Liu, Y. Lin, E. Abbasnejad, and J. Q. Shi, "Astock: A New Dataset and Automated Stock Trading based on Stock-specific News Analyzing Model," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.06606

[40] V. Ingle and S. Deshmukh, "Ensemble deep learning framework for stock market data prediction (EDLF-DP)," Global Transitions Proceedings, vol. 2, no. 1, pp. 47–66, Jun. 2021, doi: 10.1016/j.gltp.2021.01.008.

[41] S. Mokhtari, K. K. Yen, and J. Liu, "Effectiveness of Artificial Intelligence in Stock Market Prediction based on Machine Learning," Int J Comput Appl, vol. 183, no. 7, pp. 1–8, Jun. 2021, doi: 10.5120/ijca2021921347.

[42] W. Long, Z. Lu, and L. Cui, "Deep learning-based feature engineering for stock price movement prediction," Knowl Based Syst, vol. 164, pp. 163–173, Jan. 2019, doi: 10.1016/j.knosys.2018.10.034.

[43] B. A. Abdelfattah, S. M. Darwish, and S. M. Elkaffas, "Enhancing the Prediction of Stock Market Movement Using Neutrosophic-Logic-Based Sentiment Analysis," Journal of Theoretical and Applied Electronic Commerce Research, vol. 19, no. 1, pp. 116–134, Jan. 2024, doi: 10.3390/jtaer19010007.

[44] W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi, and A. S. Alfakeeh, "Stock market prediction using machine learning classifiers and social media, news," J Ambient Intell Humaniz Comput, vol. 13, no. 7, pp. 3433–3456, Jul. 2022, doi: 10.1007/s12652-020-01839-w.

[45] C. S. Ho, P. Damien, B. Gu, and P. Konana, "The time-varying nature of social media sentiments in modeling stock returns," Decis Support Syst, vol. 101, pp. 69–81, Sep. 2017, doi: 10.1016/j.dss.2017.06.001.

[46] L. N. Mintarya, J. N. M. Halim, C. Angie, S. Achmad, and A. Kurniawan, "Machine learning approaches in stock market prediction: A systematic literature review," in Procedia Computer Science, Elsevier B.V., 2022, pp. 96–102. doi: 10.1016/j.procs.2022.12.115.

[47] A. R. Khan et al., "Stock market prediction in bangladesh perspective using artificial neural network," International Journal of Advanced Technology and Engineering Exploration, vol. 9, no. 95, pp. 1397–1427, Oct. 2022, doi: 10.19101/IJATEE.2021.875852.

[48] A. Derakhshan and H. Beigy, "Sentiment analysis on stock social media for stock price movement prediction," Eng Appl Artif Intell, vol. 85, pp. 569–578, Oct. 2019, doi: 10.1016/j.engappai.2019.07.002.

[49] P. Soni, Y. Tewari, and D. Krishnan, "Machine Learning Approaches in Stock Price Prediction: A Systematic Review," in Journal of Physics: Conference Series, IOP Publishing Ltd, Jan. 2022. doi: 10.1088/1742-6596/2161/1/012065.

[50] V. Van Den Beukel, R. Gutsche, and J. Osterrieder, "Influence of machine learning on stock pricing: a meta-analysis CC-BY-NC 2," in B.S. thesis, School of Behavioral, Management and Social Sciences, Univ. Twente, Enschede, Netherlands, 2022.

[51] R. Mahmud, A. F. M. S. Saif, and D. Gomes, "A comprehensive study of real-time vacant parking space detection towards the need of a robust model," AIUB Journal of Science and Engineering, vol. 19, no. 3, pp. 99–106, Mar. 2021, doi: 10.53799/AJSE.V19I3.80.

[52] K. Biswas, N. K. Paul, D. Saha, T. Ahmed, and R. Mahmud, "Detection of Traffic Rule Violations Using Machine Learning: An Analytical Review," Malaysian Journal of Science and Advanced Technology, pp. 37–47, Mar. 2023, doi: 10.56532/mjsat.v3i1.146.

[53] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, "Leveraging social media news to predict stock index movement using RNN-boost," Data Knowl Eng, vol. 118, pp. 14–24, Nov. 2018, doi: 10.1016/j.datak.2018.08.003.

[54] A. Patel, D. Patel, and S. Yadav, "Prediction of stock market using Artificial Intelligence," 2021. [Online]. Available: https://ssrn.com/abstract=3871022

[55] T. J. Strader, J. J. Rozycki, T. H. ROOT, and Y.-H. J. Huang, "Machine Learning Stock Market Prediction Studies: Review and Research Directions," Journal of International Technology and Information Management, vol. 28, no. 4, pp. 63–83, Jan. 2020, doi: 10.58729/1941-6679.1435.

[56] R. Chopra and G. Deep Sharma, "Risk and Financial Management Application of Artificial Intelligence in Stock Market Forecasting: A Critique, Review, and Research Agenda," 2021, doi: 10.3390/jrfm.

[57] M. U. Ghani, M. Awais, and M. Muzammul, "Stock Market Prediction Using Machine Learning (ML) Algorithms ," Advances in Distributed Computing and Artificial Intelligence Jourmal, vol. 20, pp. 1–20, 2019.

[58] O. C. Sert, S. D. Şahin, T. Özyer, and R. Alhajj, "Analysis and prediction in sparse and high dimensional text data: The case of Dow Jones stock market," Physica A: Statistical Mechanics and its Applications, vol. 545, May 2020, doi: 10.1016/j.physa.2019.123752.

[59] M. Z. Islam, Md Maruful Hoque Chowdhury, and M. M. Sarker, "The Impact of Big Data Analytics on Stock Price Prediction in the Bangladesh Stock Market: A Machine Learning Approach," International Journal of Science and Business, vol. 28, no. 1, pp. 219–228, 2023, doi: 10.58970/IJSB.2216.

[60] Y. Ruan, A. Durresi, and L. Alfantoukh, "Using Twitter trust network for stock market analysis," Knowl Based Syst, vol. 145, pp. 207–218, Apr. 2018, doi: 10.1016/j.knosys.2018.01.016.

[61] R. Sable, S. Goel, and P. Chatterjee, "Techniques for Stock Market Prediction: A Review," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11. Auricle Global Society of Education and Research, pp. 381–402, May 01, 2023. doi: 10.17762/ijritcc.v11i5s.7056.

[62] T. H. Nguyen and K. Shirai, "Topic modeling based sentiment analysis on social media for stock market prediction," in ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference, Association for Computational Linguistics (ACL), 2015, pp. 1354–1364. doi: 10.3115/v1/p15-1131.