Analysing Trading Strategies and Uses of Machine Algorithm to Predict Stock Prices

Pankaj Gupta[1*], Dr. Satyen M. Parikh[2], Dr. Meghna B. Patel[3]

[1*]FCA, Ganpat University, Gujarat, India, pangup_74@yahoo.com
[2]FCA, Ganpat University, Gujarat, India, satyen.parikh@ganpatuniversity.ac.in
[3]FCA, Ganpat University, Gujarat, India, meghna.patel@ganpatuniversity.ac.in

Abstract
The investment banking and financial sectors have undergone significant transformation from the 19th century to the present, driven by technological advancements and evolving trading strategies. Traditional methods of buying and selling stocks have been replaced or supplemented by automated algorithms and machine learning techniques, enabling more precise stock market predictions. This review paper consolidates historical and contemporary strategies employed by traders and investors to maximize profits in the stock market. It emphasizes the application of machine learning in predicting stock market trends and decision-making for buying and selling securities.
A systematic review of journal articles published between 2016 and 2022 was conducted to identify the primary markets, stock indices, and trading strategies utilized in stock market predictions. The paper contributes to the literature by providing (1) a detailed analysis of trading strategies and market indicators, and (2) a comprehensive review of machine learning techniques employed in stock market forecasting. Furthermore, a bibliometric analysis highlights the most influential studies in this domain.
Technical analysis tools, such as moving averages and candlestick patterns, are explored alongside modern methodologies. The study also identifies existing gaps in trading strategy research, offering insights for future advancements in this field. This review serves as a valuable resource for understanding the intersection of traditional trading methods, machine learning algorithms, and stock market prediction.

Keywords: Stock Market, Machine Learning, Prediction, Neural Networks, Deep Learning

Introduction
Over time, stock markets worldwide have experienced exponential growth. Consequently, it is unsurprising that millions and billions of dollars are traded daily in these markets as participants seek to profit from share trading. For traders, investment bankers, and fund managers, accurately forecasting stock price trends can lead to significant benefits and profits for both themselves and their clients.
Historically, traditional models were widely used for stock price prediction. However, with technological advancements and the rapid evolution of data science, artificial intelligence (AI) models, including machine learning (supervised, unsupervised, and reinforcement learning) and deep learning, have become increasingly popular for predicting stock prices. These AI-driven models have demonstrated superior performance compared to traditional methods. Numerous researchers have explored this domain, resulting in a vast body of work highlighting the effectiveness of these innovative models. Despite their promise, stock price prediction remains an extremely complex and challenging task.
Eugene Fama's seminal work in 1965, *"The Behavior of Stock Market Prices,"* introduced the random walk model, which posits that predicting stock prices based on historical data is impossible. According to this model, stock price changes are distributed randomly and are independent of past movements. Consequently, it assumes that historical trends or data cannot reliably predict future price movements.
Fama presented compelling evidence supporting the random walk hypothesis. However, in business and economic research, no hypothesis is ever beyond question. Additional tests and analyses may either confirm or challenge its validity. For investors to achieve high returns with low risk, the market would need to deviate from the efficient market hypothesis (EMH) or random walk theory. While these deviations may create opportunities, no strategy can consistently guarantee high returns.
Over time, there has been a growing interest in studying both traditional and technical models related to stock market pricing, with many claiming alignment with the efficient market hypothesis (EMH). This enduring debate over the pricing of stock market instruments continues to be a trending and highly discussed topic.
Research in this area spans from traditional models to modern machine learning and deep learning techniques. Machine learning and deep learning models are relatively recent developments, highlighting the contemporary nature of these studies.
Svetlana Borovkova and Ioannis Tsiamas (2019) proposed an ensemble of Long-Short Term Memory (LSTM) Neural Networks for intraday stock predictions. They incorporated a variety of technical analysis indicators as inputs to the network. The model's predictive capabilities were tested on several large-cap U.S. stocks and compared against Lasso

and Ridge logistic classifiers. Their findings revealed that the proposed LSTM model outperformed the benchmark models and equally weighted ensembles.

Efendi et al. (2018) explored various stock market forecasting models based on data types, forecasting methodologies, and performance measures. Efendi noted that many models lacked sufficient discussion of forecasting accuracy. To address this, Efendi focused on the triangular fuzzy number data preparation technique to enhance a fuzzy random autoregression model. Using real data from the Kuala Lumpur Stock Exchange (KLSE), the study demonstrated that variability and spread adjustments during data preparation significantly improve the accuracy of fuzzy random autoregression models.

Over time, increased analysis of time-series data has sought to prove the superiority of newer models over their predecessors. Financial time-series data analysis and prediction remain complex tasks, essential for making informed investment decisions (Das & Padhy, 2018). These challenges underscore the ongoing efforts to refine methodologies for better forecasting accuracy.

Das and Padhy (2018) introduced a hybrid model that combines Support Vector Machine (SVM) with Teaching-Learning-Based Optimization (TLBO) to forecast daily closing prices of the Multi Commodity Exchange of India Limited's COMDEX commodity futures index. This model demonstrated superior performance compared to the Particle Swarm Optimization (PSO)+SVM hybrid and standard SVM models, showcasing its effectiveness in price prediction.

While traditional time series-based models have been extensively used to predict stock market price movements, many other factors beyond historical data influence stock market behavior. These include human behavior, market sentiment, and incomplete or missing key information, which often drive market fluctuations. Relying solely on historical data may therefore be insufficient to predict stock prices accurately. To address these uncertainties, Che Lah and Efendi (2019) applied a fuzzy approach using triangular fuzzy numbers and a standard deviation methodology.

The fuzzy model proposed by Che Lah and Efendi (2019) was tested on real data sets from five ASEAN countries to forecast stock buying and selling prices. While this approach achieved accuracy comparable to traditional methods, it did not surpass them, highlighting the ongoing challenges in creating highly accurate predictive models.

Most studies in stock price forecasting have traditionally relied on historical data using time series models such as Autoregressive Moving Average (ARMA), Autoregressive Conditional Heteroskedasticity (ARCH), and Autoregressive Integrated Moving Average (ARIMA). These methods have provided foundational insights but often fall short in addressing the complex, multifactorial nature of stock market behavior.

The Autoregressive Moving Average (ARMA) model combines autoregression (AR) and moving average (MA) techniques to analyze well-behaved time-series data. It assumes that the time series is stationary and fluctuates uniformly around a particular mean over time (source: www.google.com). ARMA is widely applied in time-series forecasting due to its simplicity and effectiveness in stationary data analysis.

Autoregressive Conditional Heteroskedasticity (ARCH) is a statistical model used to assess and forecast volatility in time-series data, particularly in financial markets. ARCH modeling provides a closer approximation of real market volatility, making it a valuable tool for estimating financial risk (source: www.google.com).

The Autoregressive Integrated Moving Average (ARIMA) model is extensively used in demand forecasting, such as predicting future demand in industries like food manufacturing. ARIMA helps managers make informed supply chain decisions. Additionally, ARIMA models are applied to predict stock prices based on historical data. According to Namini, Tavakoli, and Namin, ARIMA has shown superior precision and accuracy in forecasting time-series lags.

However, advancements in machine learning have led to the adoption of more sophisticated models. Among these, the Recurrent Neural Network (RNN)-based Long Short-Term Memory (LSTM) approach has emerged as a standout method for stock market prediction.

Research by Namini et al. demonstrates that LSTM outperforms traditional models like ARIMA, achieving error accuracies between 84% and 87%. LSTM models exhibited consistent behavior and accuracy across multiple training sessions, indicating robustness and reduced sensitivity to random fluctuations.

Research by Mohapatra et al. focused on predicting stock returns for Indian banks using machine learning models and technical indicators, categorized into price, volume, and turnover metrics. The study evaluated model performance using statistical metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). Among the models tested—XGBoost, Gradient Boosting, AdaBoost, and Random Forest—XGBoost emerged as the most effective, with an MAE close to 35%.

These ensemble machine learning models demonstrated potential for forecasting stock prices with large deviations and relied significantly on macroeconomic and sectoral variables. Further research is needed to test the predictive capabilities of these models across various sectors and asset classes. Expanding their applicability could provide valuable insights for short-term investors and intraday traders, making these techniques widely useful across the financial industry.

Technical indicators, financial variables, and macroeconomic factors are widely recognized as crucial components in stock price prediction research (Tsai & Hsiao, 2010). However, there is no comprehensive study that conclusively identifies which specific variables or approaches are best suited for stock price forecasting. Tsai et al. (2010) compared three feature selection methods—Principal Component Analysis (PCA), Genetic Algorithms (GA), and decision trees (CART)—and combined them using union, intersection, and multiintersection approaches to assess prediction accuracy. The results demonstrated the best prediction accuracy when combining PCA and GA using the intersection method, and further improving performance with the multi-intersection of PCA, GA, and CART. This approach yielded the highest accuracy rate and the lowest error in predicting stock growth.

Traditional and technical models have long been employed for forecasting stock prices. Despite extensive research, accurately and efficiently predicting stock market prices remains a challenging issue. Data mining techniques have been used to filter out irrelevant variables to improve accuracy, but performance can vary depending on the feature selection method used. Tsai and Hsiao (2010) used a combination of PCA, GA, and CART to eliminate extraneous variables, achieving a 79% accuracy rate. They successfully reduced the original 85 variables by nearly 80%, leaving 1,417 key variables that were proven to be significant for stock prediction and could inform future investment decisions.

While Tsai and Hsiao (2010) focused on three popular feature selection methods (PCA, GA, and CART), there are other potential approaches available. Combining different feature selection techniques may result in a more accurate prediction of stock prices, though it is not always practical or feasible to test every available method. The type of analysis and research design often dictates which feature selection methods are used. The question remains: does using different feature selection approaches separately or in combination produce a more accurate predictive model?

Stock price prediction is generally categorized into three main approaches: fundamental analysis, technical analysis, and technological approaches (Sedighi et al., 2019). All these approaches play an important role in forecasting stock prices. Predicting stock prices accurately, particularly when accounting for market sentiment, is a complex challenge. Over time, various researchers have developed proportional models claiming to offer more accurate and efficient predictions for stock buying and selling.

Fundamental analysis involves evaluating a company's value based on economic, monetary, and financial variables, typically derived from financial statements. It attempts to estimate a company's worth by analyzing these key financial factors.

Technical analysis, on the other hand, focuses on forecasting price trends by examining patterns and fluctuations driven by a variety of factors, including economic, commercial, financial, political, and psychological influences. This approach uses technical indicators, which are mathematical algorithms based on historical price and volume data, to assess stock behavior and predict future movements.

In the financial world, a wide range of instruments, both simple and complex, are used to generate profit. Simple products like equity stocks are traded on exchanges, while more complex products include ETFs, unit trusts, derivatives, and even cryptocurrencies.

Understanding the market forces that drive stock prices is a major concern for both individual and institutional investors. No one wants to lose money, and identifying these driving factors is critical for making informed decisions. Stock prices are influenced by a mix of fundamental factors, market sentiment, and emotional or psychological influences. These various factors can be understood through a combination of fundamental and technical analysis, each providing valuable insights into stock price movements.

Fundamental analysis focuses on the economic factors influencing the price movements of securities and is considered a conservative approach primarily used for long-term investments. This method aims to determine the intrinsic value of a stock by examining a company's financial situation, balance sheet, and past performance. The goal is to assess the company's financial strength and its potential for future growth, ultimately determining whether its stock is a worthwhile investment. Investors typically look at a firm's earnings, dividend prospects, interest rate expectations, and risk assessment to establish the appropriate stock price. If a company's intrinsic value exceeds its current market price, it is deemed a good investment opportunity.

In fundamental analysis, investors first evaluate the company's financial performance by considering its earnings growth and comparing its price-to-earnings (P/E) ratio to that of similar companies. This process provides an analytical framework for making investment decisions, which is divided into three stages: Economic, Industry, and Company analysis.

The economic environment plays a significant role in investment decisions. A rapidly growing economy often leads to industry expansion, and consequently, higher stock prices. Conversely, slow economic growth typically results in lower stock prices. Analyzing macroeconomic factors such as GDP growth, interest rates, inflation, exchange rates, fiscal policies, and even factors like monsoon conditions is critical for understanding stock price behavior.

In addition to the broader economic landscape, investors must also assess industry-level factors that influence stock values. Analyzing the performance of different sectors helps identify industries that are thriving. For example, the technology sector has seen significant growth, particularly in US markets, where investors are reaping substantial profits.

The final stage of fundamental analysis is company analysis, where investors examine both quantitative and qualitative factors such as liquidity, market share, business type, and financial condition. This comprehensive analysis helps determine the company's potential for future success. Ratio analysis is a key method used in company analysis, providing insights into the company's financial health and prospects.

Technical analysis is focused on forecasting prices and studying the behavior of financial markets. The goal is to predict the future price trends of an instrument—such as stocks, commodities, currency pairs, or even cryptocurrencies—and assess whether it is profitable to trade the instrument. This approach relies on observing price changes over time, often displayed graphically in charts, with price and volume being the two critical components analyzed.

Forecasting market movements based on historical price and volume data is applicable across various market instruments. Technical analysis tools help identify statistical trends from past data, which is used to make informed decisions about potential investment opportunities.

Charles Dow, a prominent financial journalist, is credited with laying the foundation of technical analysis through the Dow Theory. Dow proposed two key assumptions: historical patterns and trends in the market. He argued that all markets are efficient and that even random price fluctuations could be identified using patterns that tend to repeat over time.

Dow's theory is based on three primary assumptions:

The Market Discounts Everything: This assumption states that past prices can be analyzed to identify the differences between a stock's fair price and its market value. Market prices reflect all available information, including the ups and downs of market conditions, and this has already been factored into the current price.

Prices Follow a Past Trend: According to this assumption, stock prices move in patterns over time. Whether analyzing short-term fluctuations or long-term trends, prices tend to follow recognizable patterns. Stock prices generally revert to previous levels unless a major crisis or extreme event causes a significant shift.

History Tends to Repeat Itself: The final assumption suggests that price movements are often repetitive. Market psychology, driven by emotions such as panic and greed, leads to recurring patterns of behavior, such as impulsive buying or selling. Technical analysis examines these emotional responses to predict future price trends. By studying historical price movements and volumes, analysts can identify patterns that are likely to recur.

Different types of charts, such as line charts and bar charts, are used in technical analysis. Line charts are simple and provide a long-term view of price trends, though they are less commonly used than bar charts, which offer a more detailed analysis by showing the open, close, high, and low prices for a given trading period. These charts help identify trends and patterns, forming the basis of technical trading strategies.

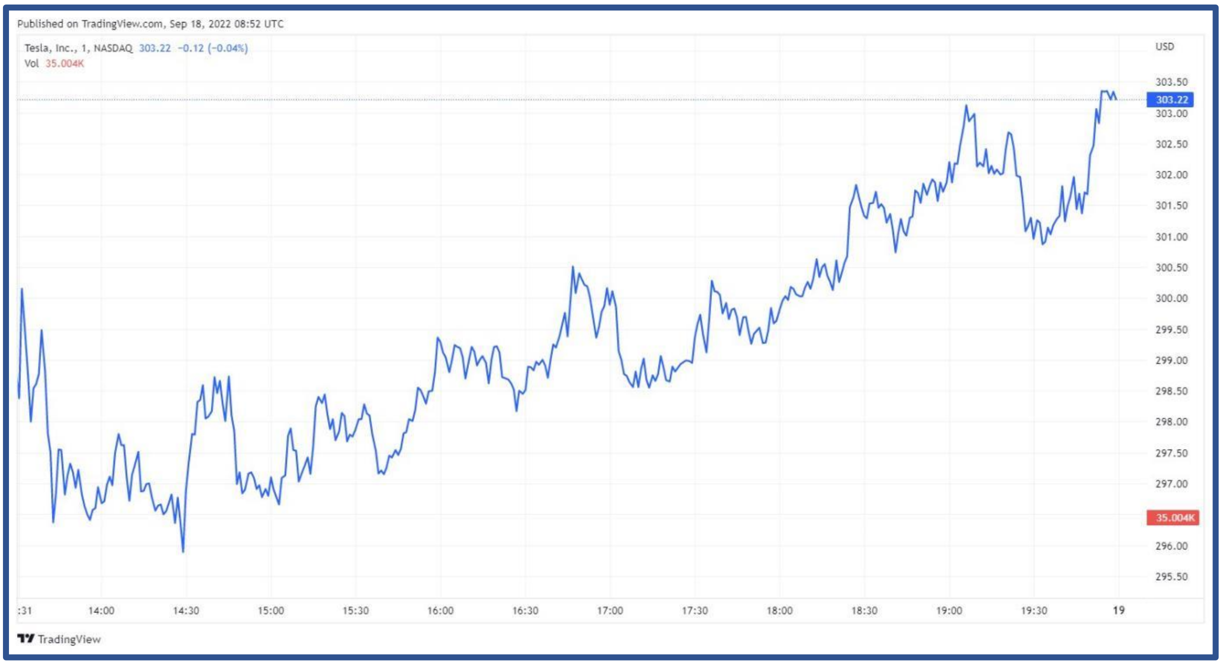

Figure 1. A line chart in technical analysis (Source: tradingview.com)

The line on the chart connects the market's closing prices for a selected period, such as weekly closings for a weekly chart or monthly closings for a monthly chart.

Bar charts are commonly used in technical analysis. These charts display the opening price, as well as the highest and lowest prices, during a specified trading period.

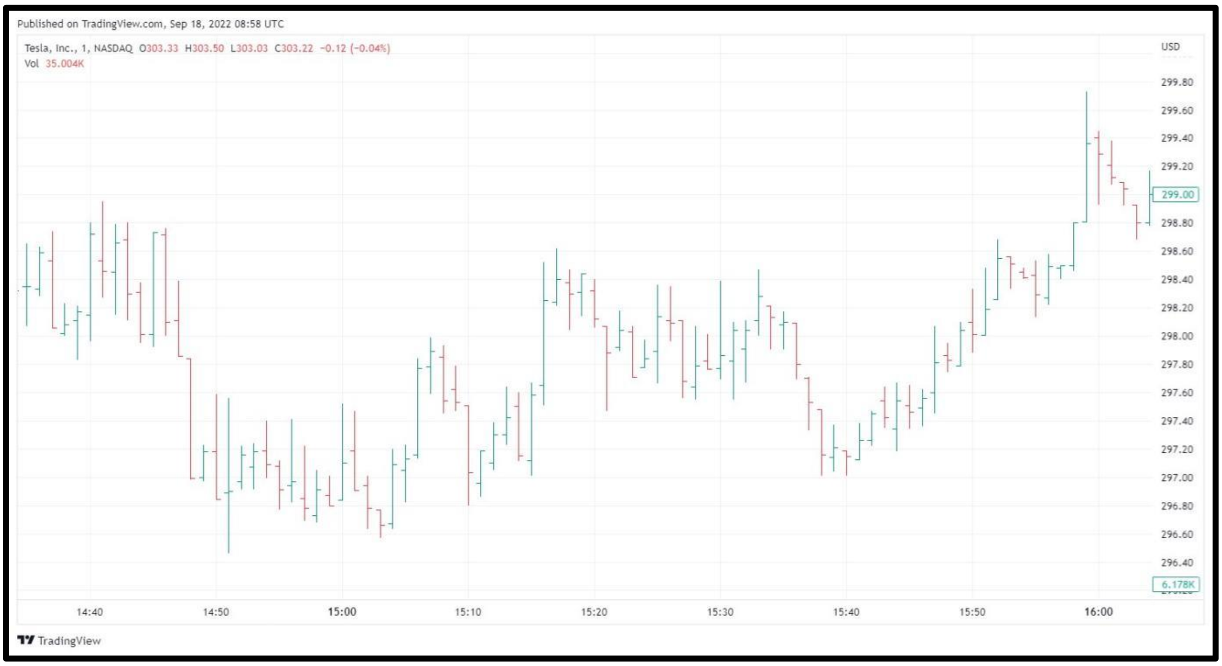

Figure 2. A bar chart used in technical analysis (Source: tradingview.com)

A candlestick chart, widely used by traders, displays market patterns reflecting the sentiment and emotions of buyers and sellers. It shows the same information as bar charts, including the open, high, low, and close prices, but with a visual emphasis on price movement.

Figure 3. Candlestick chart for technical analysis (Source: tradingview.com)

Artificial intelligence (AI) and soft computing have garnered significant attention in stock market prediction studies. Unlike traditional time series methods, these techniques can handle the stock market's nonlinear, chaotic, noisy, and complex data, which often leads to more accurate predictions (Chen & Hao, 2017). These methods offer a novel and advantageous approach, making them increasingly popular among researchers in the field of financial market forecasting. In recent years, the scope of AI-based forecasting research has expanded beyond finance, covering areas like medicine, agriculture, social media, and engineering, reflecting the growing interest in applying machine learning to diverse fields. A 2019 review of financial market prediction research by Henrique et al. (2019) analyzed over fifty studies, proposing a classification for markets, assets, methods, and variables. The review found that a significant portion of stock price prediction studies focused on North American markets, with many models utilizing support vector machines (SVMs) and neural networks. This highlights the continued importance of data-driven approaches in stock market prediction, offering ample opportunities for further research.

Accurate forecasting remains a challenging task, and continuous machine learning research is necessary to improve stock price predictions. The complexity of this field makes it one of the most demanding areas of study for researchers.

Shahet et al. (2019) proposed a taxonomy of computational approaches for stock market analysis and prediction, emphasizing that hybrid methods combining statistical and machine learning techniques are likely to be more effective. Similarly, Gandhmal & Kumar et al. (2019) reviewed over fifty publications, focusing on models such as Bayesian models, fuzzy classifiers, artificial neural networks (ANN), support vector machines (SVM), and other machine learning methods. Their findings suggest that stock market prediction is a complex task, requiring consideration of various factors for more accurate forecasting.

Bustos et al. (2020) provided a systematic review of stock market prediction models from 2014 to 2018, focusing on techniques such as deep learning, text mining, and ensemble methods. Jiang et al. (2020, 2021) conducted an extensive survey of models developed between 2017 and 2019, reviewing over 100 papers to assess the implementation and accuracy of deep learning-based models. These studies aim to lay the groundwork for future advancements in deep learning for stock market prediction.

Stock market forecasting remains a challenging task due to the volatile nature of the market. Noisy and irrelevant data can undermine forecasting accuracy, leading to poor decision-making. Kumar et al. (2022) reviewed more than 30 research papers based on machine learning and deep learning techniques, noting that most used ANN and NN for stock prediction. Despite this, many questions remain unanswered, highlighting the need for ongoing research to improve accuracy.

Sondo et al. (2020) explored the use of Effective Transfer Entropy (ETE) to predict US stock prices based on 3- and 6-month moving windows. This study examined 11 sectors and the impact of financial crises, demonstrating that ETE, when combined with machine learning models such as LR, MLP, RF, XGB, and LSTM, can improve prediction accuracy. However, the study's focus on a limited number of stocks suggests that further research is needed, particularly with larger datasets like the S&P 500.

Yaohu et al. (2021) improved stock market forecasting for the Chinese market by combining traditional candlestick charting with AI methods. Their findings suggest that momentum indicators outperform other variables like volume and volatility. However, they found that some models, such as SVM, are less effective for large datasets, and LSTM has yet to be fully explored in this context.

Suman et al. (2021) applied a graph-based approach to stock ranking prediction, showing that their model could achieve a 54.4% prediction accuracy. This study demonstrated that graph-based techniques could improve prediction performance, especially for denser graphs.

Research on stock market volatility has been a long-standing challenge. Huang et al. (2022) proposed a Multilevel Graph Attention Model, which outperforms traditional models by integrating financial market news into prediction models. This approach improves the accuracy of stock price forecasts, highlighting the need for more sophisticated tools to process news and current events.

Sentiment analysis has also become crucial in stock price prediction. Bouktif et al. (2020) demonstrated that sentiment analysis using social media data, such as Twitter, can significantly impact stock price predictions. Their research showed a 60% improvement in prediction accuracy compared to other sentiment-based algorithms. This highlights the importance of developing advanced sentiment analysis methods to predict stock market trends more accurately.

Xuan et al. (2021) used a deep learning-based model to incorporate both traditional financial index data and social media text features in stock prediction. Their findings suggest that social media-based models, particularly LSTM, outperform traditional models in predicting stock prices. However, the study's limitations, such as focusing on a single stock and one social media platform, suggest that future research should expand to include multiple stocks and social media sources.

In another study, Salvatore et al. (2021) predicted stock prices using an idiolects machine learning model based on global newswire data. Their model outperformed previous models but still faced challenges in extracting sentiment information, which is crucial for more accurate predictions. Future work in this area should integrate sentiment analysis to improve forecasting accuracy.

Xingqi et al. (2021) used a morphological similarity distance (MSD) and k-means clustering approach to predict stock prices from a similar set of stocks. The filtered data was processed using an online learning model, Hierarchical Temporal Memory, and showed superior performance for short-term stock price prediction. However, further research is needed to improve long-term prediction accuracy.

Mojtaba et al. (2021) examined stock market prediction across four sectors using nine machine learning approaches, including decision trees, random forests, and support vector classifiers. Their results showed that deep learning models like RNN and LSTM significantly outperformed other models, achieving up to 83% accuracy in binary data.

Yujie et al. (2019) developed a hybrid time-series predictive neural network (HTPNN) that incorporated social news headlines. This model improved stock price prediction accuracy by 5%, indicating that integrating social media and news data can provide valuable insights into stock market trends.

Despite the availability of commercial financial systems, accurately predicting stock prices remains a challenging task. Traditional models primarily use historical data, but they often fall short in accounting for the broader factors influencing stock market movements. The increasing availability of diverse data sources, such as social media and news platforms, offers new opportunities for improving prediction accuracy.

In this study, we provide an in-depth review of machine learning models used in stock market forecasting. We discuss the methods, their variants, and the sub-techniques used to enhance the predictive power of these models. Additionally, we analyze the variables and markets involved, identifying key areas for future research to bridge existing gaps and improve forecasting accuracy.

Research Methodology

Planning

To guide the literature review, we defined our research goals and questions, establishing the scope of our review. We focused on academic publications from 2019 to 2022 that explored machine learning techniques for forecasting stock prices. A well-defined review process was essential to minimize bias, particularly publication bias, which is critical for conducting a systematic literature review. Following the PRISMA 2020 guidelines (Page et al., 2021), we structured our review to track the number of records found, included, and excluded, and provided reasons for exclusions. The review process involved three key phases: planning, searching, and follow-up.

The planning phase involved setting up tools for organizing the search and selecting relevant search areas. We also formulated a comprehensive search query. This was followed by the search phase, where we tested and refined our search approach across multiple databases. The follow-up phase involved examining records, ensuring comprehensive screening after collecting the full texts of selected articles.

By adhering to this process, we ensured that we identified studies most pertinent to our research.

Translating our research questions into a usable search strategy was essential. The aim of our search method was to identify a comprehensive and relevant set of papers aligned with our research objectives. Our literature search was conducted in two stages: a "manual search" and an "automatic search."

Manual Search: In the first stage, we manually searched for literature on stock market prediction, starting with six journal papers that we had identified in preliminary searches (Svetlana et al., 2019; Efendi et al., 2018; Das & Padhy, 2018; Che Lah & Efendi, 2019; Chen & Hao, 2017; Shah et al., 2019). These articles served as the foundation for our review. After thoroughly reading these articles, we selected them as our primary studies.

We then employed snowballing techniques to find additional relevant articles:

Backward Snowballing: We examined the reference lists of the selected papers and included articles that met our inclusion and exclusion criteria. We looked at article titles and, when necessary, examined other parts of the papers to assess their relevance.

Forward Snowballing: We identified articles citing the initial studies using citation databases like Web of Science. Filtering options in the database allowed us to narrow down the results and remove duplicates. We then examined abstracts and other sections of the articles, removing irrelevant studies.

Automated Search: To ensure we had a comprehensive selection of papers and to minimize bias, we employed an automated search. We utilized the IEEEXplore database to uncover additional relevant papers that may have been missed in the manual search. Our search query was refined based on the inclusion and exclusion criteria, and duplicate articles were removed.

Despite extensive manual searching, we only found seven articles directly relevant to our focus. The automated search helped expand the search by identifying other pertinent studies that aligned with our research goals.

Search Queries: To define the search queries, we considered the specific focus of our research. We used a combination of keywords from the titles, abstracts, and keywords of the manually identified papers, along with our domain knowledge and expertise. The search queries aimed to capture publications that discussed stock price forecasting models using machine learning. Terms such as "Stock price," "Stock return," and "Machine Learning" were combined to identify relevant studies. These keywords were applied across various databases to ensure a broad search and identify studies that applied machine learning techniques for stock prediction. The search queries used were designed to capture articles that provided predictive models involving machine learning algorithms to forecast stock prices or returns. This approach was crucial in ensuring we gathered a wide range of relevant studies for our review.By following this structured methodology, we ensured the selection of the most pertinent and comprehensive studies, minimizing bias and maximizing the relevance of the literature included in our review.

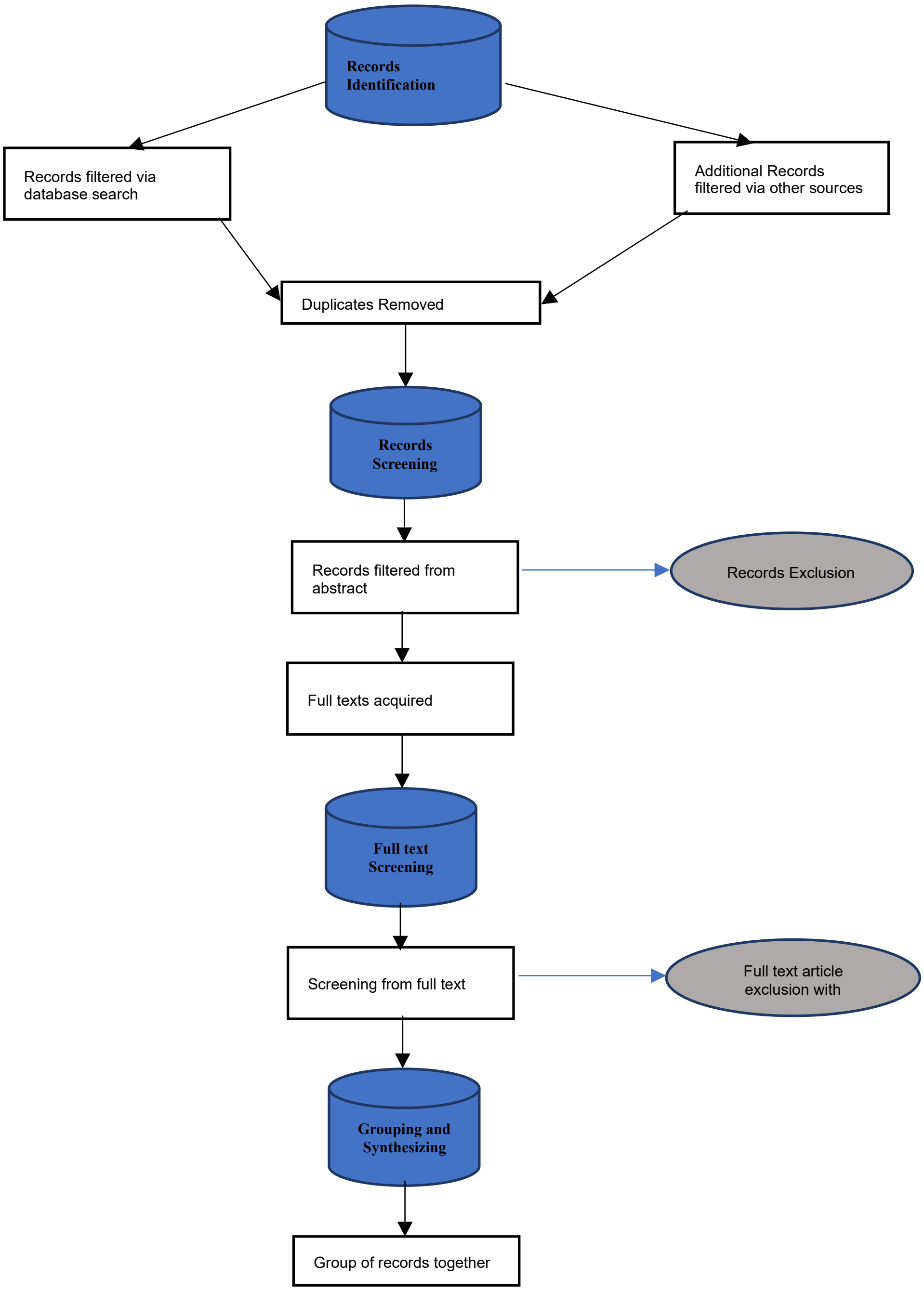

Figure 4. Research Methodology

Results & Discussion

Figure 5 shows the all search results relevant to the present study while Figure 6 shows search results which are only open access.

Figure 5. Screenshot showing complete search results

Figure 6. Screenshot showing only open access search results

Figure 7 shows search results with additional filters applied to the system.

Figure 7. Screenshot showing search results with additional filters

This study conducted an in-depth analysis of the literature on stock market forecasting using machine learning methods, focusing on publications from 2019 to 2022. A total of 29 journal articles were thoroughly reviewed and analyzed, with 22 ultimately included in the review after filtering. The analysis revealed several key trends and insights about the state of machine learning in stock market prediction.

The initial search yielded 2,000 articles from the years 2019 to 2022, which were subsequently filtered for relevance. Through a systematic process of refining the selection, we narrowed down the articles to 22 that met the criteria for inclusion in the review. These articles were examined for key attributes related to machine learning models and their application to stock market forecasting.

The majority of studies focused on stock markets in the Asian region. This trend is indicative of the prominence of Asian stock exchanges, which have gained significant attention in the context of stock price forecasting using machine learning techniques.

Industry Focus: From an industry perspective, the healthcare and information technology sectors were the most frequently studied. This reflects the growing importance of these sectors in the global economy and their potential for stock market prediction models.

Machine Learning Models: Among the machine learning techniques reviewed, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and fuzzy theory were identified as the most commonly used models for stock price prediction. These models have established a solid foundation in stock market forecasting, showing their capability to predict market trends with reasonable accuracy.

Deep Learning Techniques: Over the past three years, deep learning models, particularly Long Short-Term Memory (LSTM) networks, have gained substantial attention. LSTM models were found to outperform other deep learning approaches, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), in terms of prediction accuracy. This study highlights the growing importance of LSTM networks, which have demonstrated robust forecasting capabilities in the latest studies from 2021 and 2022.

Advancements in Fuzzy-Set Theory: Fuzzy-set theory, which had previously received limited attention, has seen a resurgence in the last few years, especially during 2021 and 2022. This approach, alongside deep learning, appears to be gaining traction as a viable method for stock market prediction.

This literature review underscores the rapid evolution of machine learning methods in stock market forecasting, with a clear shift towards deep learning, particularly LSTM models, over the past few years. The study identifies key trends in the geographical and industrial focus of the research, as well as the growing role of fuzzy-set theory in conjunction with traditional machine learning models.

Our findings suggest that LSTM networks are particularly effective in handling the complexities of stock market data, providing more reliable predictions compared to other machine learning models like CNN and RNN. This shift towards deep learning techniques represents a significant advancement in stock market prediction methodologies, paving the way for more accurate and efficient forecasting in the future.

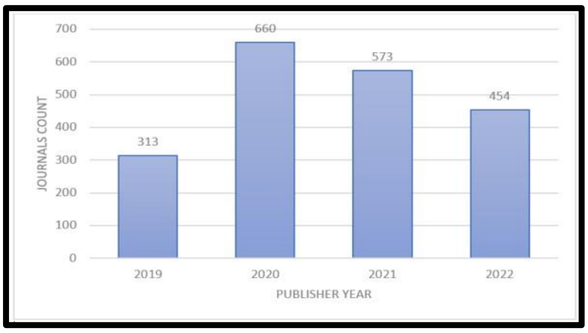

Figure 8. Articles by year

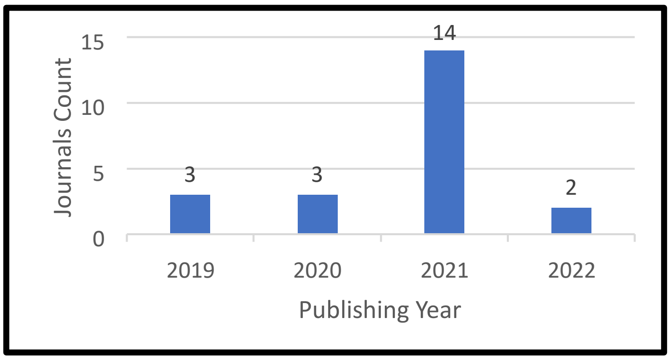

Figure 9. Most relevant Articles

Conclusion

This study provides a comprehensive overview of the recent advancements in stock market forecasting using machine learning techniques. It highlights the evolution of forecasting approaches, with a particular focus on the increasing use of deep learning methods, such as LSTM models, which have shown superior performance compared to traditional models like ANN, SVM, and fuzzy theory. The analysis of stocks, markets, and variables used for predictions reveals significant trends, including the dominance of Asian stock exchanges and the frequent use of sectors like healthcare and information technology. Furthermore, this research identifies current trends in stock market prediction and offers valuable insights into the growing reliance on machine learning for financial market analysis. These findings contribute to the ongoing development of more robust and accurate forecasting models, shaping the future direction of research in this field.

References

[1]. Badolia, L. (2016).How can i get started investing in the stock market. Edu creation Publishing, 2016

[2]. Lohrmann, C., & Luukka, P. (2019). Classification of intraday S&P500 returns with a random forest. International Journal Of Forecasting,35(1), 390–407. http://dx.doi.org/10.1016/j.ijforecast.2018.08.004.

[3]. Lohrmann, C., Luukka, P., Jablonska-Sabuka, M., & Kauranne, T. (2018). A combination of fuzzy similarity measures and fuzzy entropy measures for supervised feature selection. Expert Systems With Applications,110, 216–236

[4]. Fama, E. F. (1965). The behaviour of stock market prices. The Journal Of Business,38,34–105.

[5]. Fama, E. (1970). Efficient capital markets: A review of the theory. The Journal OfFinance,25, 383–417.

[6]. Fama, E. F. (1998). Market efficiency, long-term returns, and behavioural finance. Journal Of Financial Economics,49, 283–306.

[7]. Borovkova, S., & Tsiamas, I. (2019). An ensemble of LSTM neural networks for high-frequency stock market classification. Journal Of Forecasting,38(6), 600– 619.http://dx.doi.org/10.1002/for.2585.

[8]. Efendi, R., Arbaiy, N., & Deris, M. M. (2018). A new procedure in stock market forecasting based on fuzzy random auto-regression time series model. Information Sciences,441, 113–132

[9]. Das, S. P., & Padhy, S. (2018). A novel hybrid model using teaching–learning-based optimization and a support vector machine for commodity futures index forecasting. International Journal Of Machine Learning And Cybernetics,9(1), 97–111.http://dx.doi.org/10.1007/s13042-015-0359-0.

[10]. M Siddique, S Mohanty & D Panda (2019). A HYBRID MODEL FOR FORECASTING OF STOCK VALUE OF TATA STEEL USING ORTHOGONAL FORWARD SELECTION, SUPPORT VECTOR REGRESSION AND TEACHING LEARNING BASED OPTIMIZATION

[11]. Efendi, R., Arbaiy, N., & Lah Che, S. M. (2019). Stock Market Forecasting Model Based on AR(1) with Adjusted Triangular Fuzzy Number Using Standard Deviation Approach for ASEAN Countries

[12]. N. Tavakoli, D. Dai, Y. Chen, “Log-assisted straggler-aware I/O scheduler for high-end computing,” in 45th International Conference on Parallel Processing Workshops (ICPPW), 2016.

[13]. F. A. Gers, J. Schmidhuber, F. Cummins, “Learning to Forget: Continual Prediction with LSTM,” in Neural Computation 12(10): 2451- 2471, 2000.

[14]. S. S. Namini ,N. Tavakoli , A. S. Namin, “A Comparison of ARIMA and LSTM in Forecasting Time Series,” in 17th IEEE International Conference on Machine Learning and Applications, 2018.

[15]. Tsai, C. F., & Hsiao, Y. C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. Decision Support Systems,50(1), 258–269. http://dx.doi.org/10.1016/j.dss.2010.08.028.

[16]. Tsai, C. F., Lin, Y. C., Yen, D. C., & Chen, Y. M. (2011). Predicting stock returns by classifier ensembles. Applied Soft Computing,11(2), 2452–2459. http://dx.doi.org/10.1016/j.asoc.2010.10.001

[17]. Tsai, C. F., Lin, Y. C., Yen, D. C., & Chen, Y. M. (2011).

[18]. S. Mohapatra , R. Mukherjee, A. Roy, A. Sengupta and A. Puniyani (2022)
“Can Ensemble Machine Learning Methods Predict Stock Returns for Indian Banks Using Technical Indicators? “ J. Risk Financial Manag. 2022, 15, 350. https://doi.org/10.3390/jrfm15080350

[19]. Sedighi, M., Jahangirnia, H., Gharakhani, M., & Fard, S. F. (2019). A novel hybrid model for stock price forecasting based on metaheuristics and support vector machine.Data,4(2), 1–28. http://dx.doi.org/10.3390/data4020075.

[20]. N. Kwak, C. Kim, H. Kim, H. Dimensionality, Reduction based on ICA for regression problems, Neurocomputing 71 (2008) 2596–2603.
[21]. R.K. Lai, C.Y. Fan, W.H. Huang, P.C. Chang, Evolving and clustering fuzzy decision tree for financial time series data forecasting, Expert Systems with Applications 36 (2) (2009) 3761–3773.
[22]. M. Lam, Neural network techniques for financial performance prediction: integrating fundamental and technical analysis, Decision Support System 37 (4) (2004) 567–581.
[23]. D.T. Larose, Data Mining Method and Models, John Wiley & Sons, Inc., New Jersey, 2006.
[24]. C. Lee, G.G. Lee, Information gain and divergence-based feature selection for machine learning-based text categorization, Information Processing and Management 42 (1) (2006) 155–165.
[25]. X. Jiang, Asymmetric principal component and discriminant analyses for pattern classification, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (5) (2009) 931–937
[26]. Shah, H.; Tairan, N.; Garg, H.; Ghazali, R. A quick gbest guided artificial bee colony algorithm for stock market prices prediction. Symmetry 2018, 10, 292.
[27]. Chandar, S.K. Fusion model of wavelet transform and adaptive neuro fuzzy inference system for stock market prediction. J. Ambient Intell. Humaniz. Comput. 2019, 1–9.
[28]. Nguyen, N. An Analysis and Implementation of the Hidden Markov Model to Technology Stock Prediction. Risks 2017, 5, 62.
[29]. Ritchie, J.C. Fundamental Analysis: A Back-to-the-Basics Investment Guide to Selecting Quality Stocks; Irwin Professional Pub: Burr Ridge, IL, USA, 1996. 3.
[30]. Murphy, J.J. Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications; Penguin: London, UK, 1999.
[31]. Zhang, J., Shao, Y. H., Huang, L. W., Teng, J. Y., Zhao, Y. T., Yang, Z. K., & Li, X.Y. (2019). Can the exchange rate be used to predict the shanghai composite index? IEEE Access,8, 2188–2199.
[32]. Zhang, J., Teng, Y. F., & Chen, W. (2019). Support vector regression with modified firefly algorithm for stock price forecasting. Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-SolvingTechnologies,49(5), 1658–1674.
[33]. Zhang, X., Zhang, Y., Wang, S., Yao, Y., Fang, B., & Yu, P. S. (2018). Improving stock market prediction via heterogeneous information fusion. Knowledge-Based Systems,143, 236–247
[34]. Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. Expert Systems With Applications,124, 226–251
[35]. Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. Computer Science Review,34, Article 100190. http://dx.doi.org/10.1016/j.cosrev.2019.08.001.
[36]. Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A systematic review. Expert Systems with Applications,156, Article 113464.
[37]. Jiang, W. (2021). Applications of deep learning in stock market prediction: Recent progress. Expert Systems with Applications,184, Article 115537.
http://dx.doi.org/10.1016/j.eswa.2021.115537
[38]. Jiang, M., Liu, J., Zhang, L., & Liu, C. (2020). An improved stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. Physica A: Statistical Mechanics and Its Applications,541(258), Article122272. http://dx.doi.org/10.1016/j.physa.2019.122272
[39]. Kumar, D., Sarangi, P. K., & Verma, R. (2021). A systematic review of stock market prediction using machine learning and statistical techniques. Materials Today: Proceedings, http://dx.doi.org/10.1016/j.matpr.2020.11.399
[40]. Weng, B., Ahmed, M. A., & Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. Expert Systems with Applications,79, 153–163. http://dx.doi.org/10.1016/j.eswa.2017.02.041
[41]. Kitchenham, B. (2004).Procedures for performing systematic reviews. Keele University,UK and National ICT Australia.
[42]. Kitchenham, B., & Brereton, P. (2013). A systematic review of systematic review process research in software engineering. Information And Software Technology,55,2049–2075
[43]. S. Kim, S. Ku, W. Chang and J. W. Song, "Predicting the Direction of US Stock Prices Using Effective Transfer Entropy and Machine Learning Techniques," in
IEEE Access, vol. 8, pp. 111660-111682, 2020, doi:
10.1109/ACCESS.2020.3002174.
[44]. Y. Lin, S. Liu, H. Yang and H. Wu, "Stock Trend Prediction Using Candlestick Charting and Ensemble Machine Learning Techniques With a Novelty Feature Engineering Scheme," in *IEEE Access*, vol. 9, pp. 101433-101446, 2021, doi: 10.1109/ACCESS.2021.3096825.
[45]. S. Saha, J. Gao and R. Gerlach, "Stock Ranking Prediction Using List-Wise Approach and Node Embedding Technique," in IEEE Access, vol. 9, pp. 8898188996, 2021, doi: 10.1109/ACCESS.2021.3090834.
[46]. K. Huang, X. Li, F. Liu, X. Yang and W. Yu, "ML-GAT:A Multilevel Graph Attention Model for Stock Prediction," in IEEE Access, vol. 10, pp. 86408-86422, 2022, doi: 10.1109/ACCESS.2022.3199008.
[47]. S. Bouktif, A. Fiaz and M. Awad, "Augmented Textual Features-Based Stock Market Prediction," in IEEE Access, vol. 8, pp. 40269-40282, 2020, doi:
10.1109/ACCESS.2020.2976725.

[48]. X. Ji, J. Wang and Z. Yan, "A stock price prediction method based on deep learning technology," in International Journal of Crowd Science, vol. 5, no. 1, pp. 5572, April 2021, doi: 10.1108/IJCS-05-2020-0012.
[49]. S. M. Carta, S. Consoli, L. Piras, A. S. Podda and D. R. Recupero, "Explainable Machine Learning Exploiting News and Domain-Specific Lexicon for Stock Market Forecasting," in IEEE Access, vol. 9, pp. 30193-30205, 2021, doi: 10.1109/ACCESS.2021.3059960.
[50]. X. Wang, K. Yang and T. Liu, "Stock Price Prediction Based on Morphological Similarity Clustering and Hierarchical Temporal Memory," in IEEE Access, vol. 9, pp. 67241-67248, 2021, doi: 10.1109/ACCESS.2021.3077004.
[51]. M. Nabipour, P. Nayyeri, H. Jabani, S. S. and A. Mosavi, "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis," in IEEE Access, vol. 8, pp. 150199-150212, 2020, doi: 10.1109/ACCESS.2020.3015966.
[52]. Y. Wang, H. Liu, Q. Guo, S. Xie and X. Zhang, "Stock Volatility Prediction by Hybrid Neural Network," in IEEE Access, vol. 7, pp. 154524-154534, 2019, doi: 10.1109/ACCESS.2019.2949074.
[53]. Mahinda Mailagaha Kumbure, Christoph Lohrmann, Pasi Luukka, Jari Porras," Machine learning techniques and data for stock market forecasting: A literature review,Expert Systems with Applications," vol. 197,2022,116659,ISSN 0957-4174,https://doi.org/10.1016/j.eswa.2022.116659.