A Survey: Predicting Stock Prices with Machine Learning

Aditya Karale[1], Pranav Waghmore[2], Vaibhav Bhangre[3], Vaibhav Kokate[4], Ganesh Jadhav[5]
[1,2,3,4,5] *Department of Computer Science and Engineering Smt. Kashibai Navale College Engineering, Pune Maharashtra India*

Abstract—The stock exchange has emerged as a crucial component in today's financial landscape, significantly impacting the global economy. The stock market appeals to individuals from diverse educational and professional backgrounds due to its potential for financial gain. Given the complex and unpredictable nature of the stock market, its study has become increasingly important. Investors often base their decisions on research and forecasts to mitigate risks and enhance returns. Traditional prediction methods, including fundamental and technical analysis, often fall short in terms of reliability and precision. Consequently, machine learning has gained prominence in stock price prediction, leveraging historical data to forecast future trends. This paper examines the use of Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks for forecasting trends in the stock market.

Keywords: Stock, Stock Market, Shares, Long Short-Term Memory

I. INTRODUCTION

In the rapidly advancing field of stock market prediction, a stock or share—representing a company's equity—is a financial instrument that denotes ownership in a company. This ownership provides shareholders with a proportional claim on the company's assets and earnings. Holding stock means that an individual owns a portion of the company, relative to the number of shares they possess compared to the total outstanding shares. For example, if someone holds 100,000 shares in a company with a total of 1 million shares, they effectively own a 10% stake in the company. Many companies have millions or even billions of outstanding shares.

Stock exchanges operate as secondary markets where existing shareholders can sell shares to potential buyers. Typically, companies listed on exchanges do not directly buy or sell their shares once issued, so when a stock is purchased, it is bought from another shareholder—not from the company itself.
Similarly, when a stock is sold, it is sold to another investor rather than directly to the company.

In economies like India, stock market performance is often closely linked with economic growth. An increase in the stock market generally indicates strong economic growth, while a decline may signal slower growth. Despite the potential for gains, only about 10% of the population invests in the stock market, as many view it as unpredictable or akin to gambling. This misconception could be reduced through better public education and awareness.

Recently, Machine learning has gained significant momentum in the field of stock market analysis.prediction, valued for its accuracy and efficiency. Techniques such as sentiment analysis, analysis of historical stock prices, and tracking of sales and dividend growth are increasingly applied in stock prediction models. For reliable predictions, it is essential to have both relevant data and robust analytical techniques.

II. RELATED WORK

A literature review encompasses various learning techniques for extracting ontology from data, including the following methods:
The paper "Stock Market Prediction using Machine Learning

Algorithms: A Classification Study" by Misra, Meghana. This paper aims to classify various machine learning algorithms. It introduces the concept of economic derivatives, which aligns with the "no arbitrage" principle, and discusses key predictive models such as stochastic process theory and the Efficient Market Hypothesis (EMH). News articles were analysed, and predictions were generated based on their content. The prediction process involved steps like data preparation, analysis,

aggregation, and visualization. Prediction results were categorised as neutral, positive, or negative, with data collected daily. Logistic regression was used for making predictions and achieved an accuracy of 70%. A generalised linear model (GLM) from the binomial family is commonly employed as a logistic regression model.

In "Stock Price Prediction Using Machine Learning Techniques," Sirimavo and Naadun predict the performance of the Karachi Stock Exchange (KSE) using various machine learning techniques. The paper discusses methodologies like single-layer perceptron, multi-layer perceptron, radial basis function, and support vector machines. The proposed approach involves using multiple market-impacting factors as input attributes for the machine learning model. Numerous factors were identified as having a substantial influence on market performance.

In the paper *"Survey of Stock Market Prediction Using Machine Learning Approach,"* Ashish Sharma provides a survey of effective regression techniques for predicting stock prices using data from stock exchange databases. The paper explores several regression methods, including polynomial regression, linear regression, RBF regression, and sigmoid regression. Notably, polynomial regression is emphasised as particularly effective for stock market prediction because of its ability to model the nonlinear relationships between dependent and independent variables.

In the paper "Stock Market Prediction using Data Mining Techniques" by Maini and K. Govinda, exchange trends were predicted using machine learning models such as the Random Forest model and Support Vector Machine. Both historical and sentiment data were analyzed, as they play a crucial role in influencing market performance. Researchers performed sentiment analysis on Twitter feeds to identify the relationship between "public sentiment" and "market sentiment." Data from Twitter was used to gauge public mood, and a self-organizing fuzzy neural network was applied to the predicted mood from Twitter feeds along with the previous day's Dow Jones Industrial Average values to forecast future stock market movements.

The paper "Stock Closing Price Prediction Using Machine Learning" by Vijh and Mehar focused on predicting stock prices in the Thailand Stock Exchange (SET). The authors employed various methods, including the multi-layer perceptron model, support vector machine model, and partial least squares classifier. Their experiments showed that the Partial Least Squares Classifier performed better than the other two algorithms in forecasting stock prices.

In "Empirical Study on Stock Market Prediction Using Machine Learning" by Sable and Rachana, the study reviews stock market prediction using a variety of traditional, machine learning, and deep learning algorithms. The survey also considered multiple datasets used for stock market prediction, selected features of these datasets as input parameters, and evaluation metrics to compare prediction results. All of these factors were analyzed to develop an effective prediction approach.

The paper titled "DP-LSTM: Differential Privacy-inspired LSTM for Stock Prediction Using Financial News" by Xinyi Li introduces an innovative deep neural network model called DP- LSTM for stock price prediction. This model incorporates financial news as restricted data and integrates various news sources using a differential privacy mechanism. Based on the autoregressive integrated moving average (ARIMA) model, an estimation is made by considering the financial news data within the model. Subsequently, a deep neural network utilizing Long Short-Term Memory (LSTM) is constructed, comprising three components: the LSTM, the VADER model, and a differential privacy (DP) mechanism. The DP-LSTM design significantly reduces prediction errors and enhances robustness. The paper combines the deep neural network with the popular natural language processing (NLP) model, VADER, to identify and extract sentiments from text, integrating stock price fluctuations and compound scores to mitigate investment risks. Relying solely on news-derived data for predictions may introduce bias due to some subjective reports. Therefore, the DP-LSTM enhances the robustness of the prediction model.

This paper "Prediction of Stock Market using Machine Learning Algorithms" of Narote is introduced an examination of AI-supported calculations to assess the stock costs later on to break down market conduct. R factual programming language has been utilized to make the examination. It gathered that Logistic- regression gives exactness in examining market development course and market expectations contrasted with different models like

Random Forest, ARIMA, and K-NN. Clarification of these models with their test concentrate on the datasets for various stocks was arranged understanding rate exactness, merits, and demerits

The paper *"Social Media Monitoring using Sentiment Analysis of Twitter Data"* by Bhardwaj and Akhil Ansari examines the prediction of future stock prices using sentiment values for each stock. It focuses on two techniques, Word2Vec and N-gram, to extract insights from sentiments in tweets. The study utilizes sentiment analysis alongside machine learning algorithms on tweets collected from Twitter, analyzing the correlation between a company's stock market movement and the sentiments expressed in tweets. It proposes a hybrid approach that combines unsupervised learning to cluster the tweets and then applies supervised learning methods for classification.

The paper *"Social Media Monitoring using Sentiment Analysis of Twitter Data"* by Bhardwaj and Akhil Ansari explores the prediction of future stock prices by analyzing sentiment values associated with each stock. The study focuses on two techniques—Word2vec and N-gram—to extract insights from sentiments in tweets. It combines sentiment analysis with machine learning algorithms applied to tweets from Twitter, analyzing the correlation between a company's stock market movement and the sentiments expressed in tweets. The paper proposes a hybrid approach that uses unsupervised learning to cluster the tweets, followed by supervised learning methods for classification. The authors applied various machine learning techniques, including Naive Bayes (NB), Maximum Entropy, and Support Vector Machine (SVM), concluding that NB and SVM with 89.4% accuracy outperformed other methods in sentiment classification. The study further investigated the correlation between Twitter sentiments and stock prices, identifying which words in tweets are associated with stock price changes by performing post-analysis of price fluctuations and tweet content.

This paper, "Stock Price Prediction Using ARIMA Model," by Jadhav, Rupwari The creators utilized the ARIMA model to anticipate the stock cost based on the information they got from the New York Stock Exchange (NYSE) and the Nigeria Stock Exchange (NSE). They have utilized data sets comprising four features: open, low, close, and high cost. They have accepted the end cost as the objective viewpoint to be anticipated. The explanation for this was that the closing price is the most relevant cost by the day's end. They have exhibited that there is no connection between the autocorrelation capacities (ACFs) and partial autocorrelation capacities (PACFs) utilizing Q-insights and correlation plots. For the non-fixed information, it was 207 fixed with the assistance of differencing methods. It was reasoned towards the end of the examination that the ARIMA model is exceptionally helpful for the transient forecast.

In this paper, "A machine learning model for stock market prediction "by Adebiyi, the proposed model is based on the study of historical data and technical data. It optimizes LS- SVM with a PSO algorithm for daily stock prediction. The Levenberg-Marquardt (LM) algorithm is employed as a benchmark for comparison with LS-SVM and LS-SVM- PSO models. This proposed algorithm was tested for many companies belonging to the sectors of information & technology, financials, health care, energy, communications, materials, and industries. Five indicators were calculated from the raw datasets, such as the Relative Strength Index (RSI), Money Flow Index (MFI), Exponential Moving Average (EMA), Stochastic Oscillator (SO), and Moving Average Convergence/Divergence (MACD). LS-SVM-PSO has better performance compared to other algorithms like LS-SVM and ANN, while the ANN-BP algorithm has the highest error value.

In the paper *"Stock Market Prediction Using Machine*

Learning" by Hegazy and Osman, the proposed model is based on the analysis of historical and technical data. It optimizes LS- SVM (Least Squares Support Vector Machine) using a PSO (Particle Swarm Optimization) algorithm for daily stock forecasts. The Levenberg-Marquardt (LM) algorithm is used as a benchmark for comparison with the LS-SVM and LS-SVM- PSO models. This proposed algorithm was tested on various companies across sectors such as information technology, financials, healthcare, energy, communications, materials, and industries. Five indicators were calculated from the raw datasets: Relative Strength Index (RSI), Money Flow Index (MFI), Exponential Moving Average (EMA), Stochastic Oscillator (SO), and Moving Average Convergence/Divergence (MACD). The

LS-SVM-PSO model demonstrated superior performance compared to other algorithms like LS-SVM and Artificial Neural Networks (ANN), while the ANN-BP algorithm showed the highest error rate.

The paper "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms via Continuous and Binary Data: A Comparative Analysis" presents a comparative study of nine machine learning models: random forest, decision tree, extreme gradient boosting, adaptive boosting, K-nearest neighbors, support vector classifier, Naïve Bayes, artificial neural network, and logistic regression. In addition, deep learning methods, including Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN), were also explored. For all these models, a total of ten technical indicators were provided as input values. Two distinct approaches for input data were considered: continuous data and binary data. The study examined the effect of preprocessing, where the continuous data inputs included stock trading data such as open price, close price, high value, and low value, while the binary data inputs involved a preprocessing step that converted the continuous data into binary form.

In the paper "NSE Stock Market Prediction Using Deep- Learning Models" by Nabi, Mojtaba, the authors explore the application of various deep learning architectures for predicting the future stock prices of companies listed on the National Stock Exchange (NSE). The study focuses on four prominent deep learning models: recurrent neural networks (RNN), multilayer perceptrons (MLP), convolutional neural networks (CNN), and long short-term memory (LSTM). These models are trained on historical stock price data to forecast future price movements. The paper compares the performance of these deep learning models with the ARIMA (AutoRegressive Integrated Moving Average) model, a traditional time-series forecasting method. The results of the study reveal that the deep learning models, particularly the LSTM and RNN, significantly outperform the ARIMA model in terms of prediction accuracy. The authors observe that the ARIMA model struggles to capture the complex, non-linear patterns present in stock market data, which are better addressed by the neural network models.

Through this comparative analysis, the paper highlights the superiority of deep learning approaches over traditional methods like ARIMA in stock market prediction, emphasizing the potential of neural networks in capturing the intricate relationships in financial data for more accurate forecasting.

Ref No.	Parameter	Algorithms	Limitation and Future work
1.	1)Accuracy	1) Logistic Regression 2) Linear Model	1) Low Accuracy 70% 2) Require more complex model for better prediction
2.	1) Market performance prediction 2) impact of input factor	1)Single Layer Perceptron 2) Multi- Layer Perceptron	1) Difficulty in identifying the most relevant factors impacting the market performance 2) Complex computation and longer training time
3.	1) Accuracy in Stock Price Prediction 2) Regression Techniques	1)Rectilinear Regression, RBF Regression, Sigmoid Regression, Polynomial Regression	1) Complexity in handling large stock exchange databases 2) Non-linear relationships are difficult to model accurately
4.	1)Sentiment Analysis 2) Accuracy in Market Trend Prediction	1) Random Forest, Support Vector Machine 2) Self Organizing Fuzzy Neural	1) Requires large historical and sentimental data

		Network	2) Complexity in analyzing Twitter feeds and public sentiment

5.	1) Accuracy in Stock Price Prediction 2) Experimental Comparison	1) PartialLeast Square Classifier 2) Multi-Layer Perceptron, Support Vector Machine	1) Limited accuracy for complex stock patterns 2) Partial Least Square performance depends on data pre-processing
6.	1) Accuracy in Stock Market Prediction2) Dataset and Feature Selection	1) Traditional Algorithms, Machine Learning, Deep Learning	1) Difficulty in selecting optimal datasets and input parameters 2) Evaluation metrics may vary, leading to inconsistent comparisons
7.	1) Accuracy 2) Recall Rate	1) Single Shot Multi Box Detector (SSD) 2) Adaptive Image Enhancement Algorithm	1) Small Dataset Size 2) Time-Consuming Image Enhancement
8.	1) Accuracy 2) Compatibility 3) Efficiency	1) ARIMA and LSTM are more compatible for NASDAQ stocks. 2) LSTM and Linear Regression are more efficient for NSE stocks.	1) Logistic Regression provides higher accuracy in analyzing market movement direction and market predictions compared to other models.
9.	1) Sentiment Analysis 2) Stock Price Prediction 3) Correlation Analysis	1) Word2vec 2) N-gram 3) Naive Bayes (NB)	1) Naive Bayes (NB) and SVM with 89.4% accuracy outperform other methods in sentiment classification. 2) The hybrid approach combines unsupervised and supervised learning methods.

10.	1) Accuracy	1) Word2vec 2) N-gram 3) Sentiment Analysis with Machine Learning (Naive Bayes, SVM, Maximum Entropy)	1) Correlation with stock market movement requires improvement 2) Need for more accurate unsupervised and supervised learning integration
11.	1) Accuracy 2) Forecasting	ARIMA Model	1) No significant correlation between ACFs and PACFs 2) Limited to short-term forecasting
12.	1) Accuracy 2) Prediction Performance	1) LS-SVM 2) LS-SVM-PSO 3) Levenberg-Marquardt	1) LS-SVM-PSO outperforms LS SVM and ANN 2) ANN BP has the highest

		(LM) 4) ANN-BP	error value
13.	1) Accuracy 2) Trend Prediction	1) Random Forest 2) Decision Tree 3) extreme Gradient Boosting	1) Performance comparison needed between continuous and binary data inputs 2) Impact of preprocessing on model performance

III. OBSERVATIONS AND FINDING

Stock Market Prediction

Stock market prediction has become a significant area of focus, leveraging advanced machine learning techniques to forecast stock prices and market trends. This survey provides an in-depth analysis of current methods, challenges, and advancements in this domain.

A. Key Issues and Insights

The stock market is inherently volatile and complex, influenced by a wide range of economic, social, and political factors that interact in unpredictable ways. This complexity poses significant challenges for developing accurate predictive models, as they must account for rapid market fluctuations and the diverse set of variables that impact stock prices.

In addition to the challenges posed by market complexity, external events such as economic policy changes, geopolitical tensions, natural disasters, and shifts in public sentiment can have profound effects on stock prices. Effective predictive models must be adaptable and capable of adjusting to these external factors in order to provide accurate forecasts under a variety of circumstances.

Furthermore, the success of stock market prediction models heavily depends on the availability of high-quality, diverse datasets that include historical stock prices, trading volumes, and relevant financial news. Ensuring the quality, consistency, and accuracy of this data is crucial for building reliable and robust predictive models that can withstand the unpredictable nature of the stock market.

Recent advancements in machine learning, especially in Long Short-Term Memory (LSTM) networks, have enhanced stock market prediction accuracy. These models capture temporal dependencies in data, which is essential for forecasting time series like stock prices.

Real-Time Processing Needs: For practical application, stock prediction systems often require real-time processing to make timely decisions. Advanced hardware solutions, such as GPUs and cloud-based services, facilitate real-time predictions. However, achieving low latency while maintaining high accuracy remains a substantial challenge.

Robustness and Adaptability: Developing models that can adapt to various market conditions and respond to previously unseen trends is a core area of research. Techniques such as data augmentation and ensemble methods are applied to enhance adaptability and performance.

Integration with Financial Decision-Making: Integrating stock market prediction models into broader financial decision-making frameworks enables more informed investment strategies. Such integration requires seamless collaboration between predictive models and portfolio management systems.

Hybrid and Transfer Learning Approaches: Combining traditional statistical methods with modern deep learning techniques, as well as employing transfer learning, has proven beneficial in improving model performance, especially when dealing with limited data. These approaches enhance both accuracy and generalization capabilities.

B. Findings

Deep Learning Dominance: Deep learning models, particularly LSTM and transformers, have achieved state-of-the-art performance in stock market prediction due to their capability to model sequential data and capture intricate patterns.

Importance of Feature Engineering: Incorporating additional financial indicators and external factors into predictive models has been found to improve

model robustness and accuracy.

Hybrid Approaches: Combining statistical methods with deep learning models can improve performance, especially in scenarios with limited computational resources or when addressing specific market sectors.

Advancements in Hardware: The development of powerful hardware solutions, such as GPUs and cloud- based infrastructures, has enabled faster and more efficient stock market predictions, supporting real- time analysis for practical applications.

Ongoing Research Directions: Future research focuses on enhancing model adaptability to unexpected market events, improving robustness under volatile conditions, and integrating prediction models with other data sources, such as social media sentiment and economic indicators, to strengthen overall predictive performance.

VI. CONCLUSION AND FUTURE WORK

This paper provides a comprehensive examination of stock market price prediction technologies, detailing the current research landscape and areas of application. It explores advancements in financial dataanalysis and the application of Long Short-Term

Memory (LSTM) networks and other deep learning models for sequential data processing and trend prediction. The paper outlines the framework and environment used for implementing the LSTM model, including key code implementations for various layers. The introduction of techniques such as Batch Normalization significantly addressed issues related to overfitting and improved model stability. A comparative analysis of different architectures determined that a two-layer LSTM with dropout provided optimal performance, balancing model complexity and accuracy. The experimental results demonstrate that the enhanced model achieves an impressive prediction accuracy, maintaining reliable results under varied market conditions. These findings validate the effectiveness of the proposed improvements and underscore the model's potential for real-world stock market forecasting applications.

In future work, further enhancements could include the integration of additional data sources, such as sentiment analysis from social media or financial news, to capture external market influencers more effectively. Additionally, exploring hybrid models that combine deep learning with traditional time series forecasting methods may improve accuracy in highly volatile markets. Overall, the study confirms that advancements in LSTM-based models offer a promising foundation for robust and adaptable stock market price prediction systems.

V.REFERANCES

[1] Misra, Meghna, et al. “Stock Market Prediction Using Machine Learning Algorithms: A Classification Study.” 2018 International Conference on Recent Innovations in Electrical, Electronics and Communication Engineering, ICRIEECE 2018, IEEE, 2018, pp. 2475–78.

[2] Sirimevan, Naadun, et al. “Stock Market PredictionUsing Machine Learning Techniques.” 2019 International Conference on Advancements in Computing, ICAC 2019, no. 4, 2019, pp. 192–97.

[3] Sharma, Ashish, et al. “Survey of Stock MarketPrediction Using Machine Learning Approach.” Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, vol. 2017- Janua, 2017, pp. 506–09.

[4] Maini, Sahaj Singh, and K. Govinda. “Stock Market Prediction Using Data Mining Techniques.”Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2017, no. Iciss, IEEE, 2018, pp. 654–61.

[5] Vijh, Mehar, et al. “Stock Closing Price Prediction Using Machine Learning Techniques.” Procedia Computer Science, vol. 167, 2020, pp. 599–606.

[6] Sable, Rachna, et al. “Empirical Study on Stock Market Prediction Using Machine Learning.” 2019 6th IEEE International Conference on Advances in Computing, Communication and Control, ICAC32019, IEEE, 2019,
pp. 1–5.

[7] Li, Xinyi, et al. “DP-LSTM: Differential Privacy- Inspired LSTM for Stock Prediction Using Financial News.” ArXiv, no. NeurIPS, 2019, pp. 1–9.

[8] Narote, Prof Amit, et al. Predicting NASDAQ and NSE Stocks Using Machine Learning Algorithms : ARIMA , LSTM &. no. Iii, 2020,

pp. 39–46.

[9] Bhardwaj, Neha, and Akil Ansari. “Prediction of Stock Market Using Machine Learning Algorithms.” International Research Journal of Engineering and Technology, vol. 5994, no. May, 2019, pp. 5994–6005.

[10] Jadhav, Rupawari, and Wakode M. S. “Survey: Sentiment Analysis of Twitter Data for Stock Market Prediction.” Ijarcce, vol. 6, no. 3, 2017, pp. 558–62.

[11] Adebiyi, Ayodele A., et al. “Stock Price Prediction Using the ARIMA Model.” Proceedings - UKSim- AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014, 2014, pp. 106–12.

[12] Hegazy, Osman, et al. "A Machine Learning Modelfor Stock Market Prediction" International Journal of Computer Science and Telecommunications, Volume 4, Issue 12. [Online]. Available: http://arxiv.org/abs/1402.7351

[13] Parmar, Ishita, et al. “Stock Market Prediction Using Machine Learning.” ICSCCC 2018 - 1st International Conference on Secure Cyber Computing and Communications, IEEE, 2018, pp. 574–76.

[14] Nabipour, Mojtaba, et al. “Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis.” IEEE Access, vol. 8, 2020, pp. 150199–212.

[15] Hiransha, M., et al. “NSE Stock Market Prediction Using Deep-Learning Models.” Procedia Computer Science, vol. 132, no. Iccids, Elsevier B.V., 2018,pp. 1351–62.

[16] Singh, Sukhman, et al. “Stock Market Forecasting Using Machine Learning: Today and Tomorrow.” 2019 2nd International Conference on Intelligent Computing, Instrumentation.