ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

A Scalable Analytics System for Live Stock Market

[1]T. Surya Naga Durga Anvitha, [2]D. Naveen Raju,[3]R. kusuma,[4]V. Uday Krishna, [5]Mr.N.S. V Prasad

[1,2,3,4]U. G Student, Department of Artificial Intelligence & Data Science,

D.N.R. COLLEGE OF ENGINEERING & TECHNOLOGY (AUTONOMOUS)

Balusumudi, Bhimavaram, West Godavari District, Andhra Pradesh -534202

[5]Assistant Professor, Department of Artificial Intelligence & Data Science,

D.N.R. COLLEGE OF ENGINEERING & TECHNOLOGY (AUTONOMOUS)

Balusumudi, Bhimavaram, West Godavari District, Andhra Pradesh -534202

ABSTRACT

In the fast-paced world of financial markets, the ability to analyze stock market trends in real-time is crucial for making informed investment decisions. This study explores the application of machine learning algorithms, including XGBoost, Decision Trees (DT), and K-Nearest Neighbors (KNN), for predicting stock market trends based on historical data and real-time market indicators. By leveraging these machine learning models, we aim to predict stock price movements, identify trends, and detect anomalies that could indicate potential market shifts. The integration of these algorithms with Power BI, a powerful data visualization and business analytics tool, allows for real-time analysis and dynamic dashboard reporting. This system processes large datasets from various market sources and presents actionable insights through interactive Power BI dashboards, enabling investors to make data-driven decisions. The effectiveness of the proposed approach is evaluated based on performance metrics such as accuracy, precision, and recall. Results demonstrate the potential of machine learning in financial forecasting, improving decision-making processes and offering a competitive edge in the stock market.

KEYWORDS: - XGBoost, Decision Trees (DT), and K-Nearest Neighbours (KNN), Power BI, Stock Market

INTRODUCTION

Social The stock market is a dynamic and complex system influenced by numerous factors such as economic data, political events, and investor sentiment. As a result, predicting market trends and stock price movements has always been a significant challenge for investors and analysts. Traditional methods often fail to keep up with the rapid changes and large volumes of data in real-time. Machine learning (ML) techniques, such as XGBoost, Decision Trees (DT), and K-Nearest Neighbors (KNN), have emerged as powerful tools for forecasting stock market trends by analyzing historical price data and real-time market signals. These models can identify hidden patterns and predict future trends with a high degree of accuracy. Coupling machine learning with data visualization platforms like Power BI enables users to interactively explore trends, monitor market movements, and make more informed investment decisions. Power BI's real-time dashboard capabilities enhance the process by providing visual insights, which are essential for responding quickly to changing market conditions. By integrating machine learning with advanced visualization tools, this approach aims to provide investors with actionable insights and a competitive edge in stock market trading.

RELATED WORK

Several studies have explored the application of machine learning techniques for stock market prediction and trend analysis. Early approaches relied on statistical models such as linear regression and time series analysis, which had limitations in capturing complex market patterns. With advancements in machine learning, algorithms like Decision Trees and K-Nearest Neighbors (KNN) have been widely used for classification and trend prediction tasks. More recently, ensemble methods such as XGBoost have gained popularity due to their ability to handle large datasets and improve prediction accuracy. Many researchers have utilized historical stock price data along with technical indicators to train predictive models. Some studies have also integrated sentiment analysis from news and social media to enhance prediction performance. Visualization tools like Power BI and Tableau have been increasingly used to present analytical results through interactive dashboards. Comparative analyses show that ensemble and boosting techniques often outperform traditional models in financial forecasting. However, challenges such as market volatility, noise in data, and overfitting remain significant concerns. Overall, existing research highlights the potential of combining machine learning with real-time visualization tools for effective stock market analysis.

LITERATURE SURVEY

The literature Various studies have explored different approaches for stock market prediction and financial analysis using machine learning and data mining techniques. Research on Twitter sentiment analysis demonstrated that public emotions extracted from social media can influence stock price movements, achieving prediction accuracy of over 76% for certain companies. Network analysis using datasets like the Enron email corpus has shown that organizational behavior and communication patterns can indirectly reflect financial stability and market trends. Neural network models such as Back Propagation Neural Networks (BPNN), GRNN, and RBNN have been applied in predictive tasks, highlighting the effectiveness of deep learning in forecasting complex patterns. Analytical techniques using indicators like MACD, RSI, MFI, and ATR have also been widely used for modeling stock price behavior. Studies on investor relations emphasize the importance of communication and stakeholder interaction in influencing financial performance. Machine learning approaches such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN) have shown better performance compared to traditional statistical methods. Time series models like ARIMA provide baseline forecasting but struggle with nonlinear patterns in stock data. Big data frameworks such as Hadoop and Spark enable real-time processing of large-scale financial datasets. Additionally, anomaly detection techniques and visualization tools like dashboards play a crucial role in identifying unusual market behavior and supporting data-driven decision-making.

EXISTING METHOD

The existing stock market analysis system primarily relies on traditional machine learning and statistical models such as Linear Regression, Support Vector Machines (SVM), and basic Decision Tree algorithms for predicting stock price movements using historical market data. These systems focus mainly on trend identification through technical indicators and rule-based analysis rather than real-time intelligent forecasting. Visualization is usually performed using static charts and reports, which do not support interactive or dynamic dashboards. Anomaly detection and market shift identification are limited and often require manual interpretation by analysts. Moreover, these models struggle with large-scale and real-time data processing, resulting in lower prediction accuracy and delayed decision-making. Due to the limited use of advanced ensemble models and real-time visualization tools, the existing system does not fully support investors in making fast, accurate, and data-driven investment decisions.

PROPOSED METHOD

The proposed method integrates machine learning algorithms, including XGBoost, Decision Trees (DT), and K-Nearest Neighbors (KNN), with Power BI for real-time stock market trend analysis and prediction. First, historical stock data, along with real-time market indicators such as volume, volatility, and economic signals, are collected and preprocessed. Feature engineering is applied to extract relevant patterns and trends that could influence stock price movements. XGBoost is used for its ability to handle complex data relationships and improve predictive accuracy through boosting, while Decision Trees provide interpretability for decision-making processes, and KNN is used for its simplicity in classification. These models are trained on the prepared dataset and validated using performance metrics like accuracy, precision, and recall. The predictions generated by the models are then visualized in real-time using Power BI, which creates interactive dashboards that allow users to track stock price movements, identify trends, and monitor market anomalies. This integration helps investors quickly identify market shifts and make data-driven decisions. The system is designed to continuously update predictions as new data becomes available, ensuring up-to-date insights for market participants.

SYSTEM ARCHITECTURE

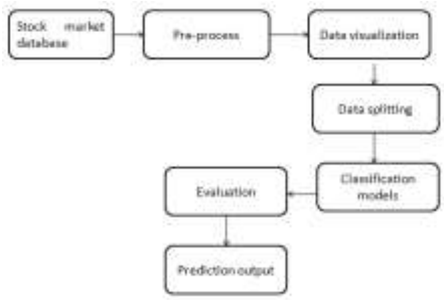

Figure 1: Architecture of the Project

METHODOLOGY DESCRIPTION

Data Collection: The Stock Market Database module is responsible for collecting and storing historical and real-time stock market data used for analysis and prediction. It gathers information such as stock prices, trading volume, opening price, closing price, high and low values, and timestamps from reliable financial sources or APIs. The module ensures structured storage using databases or datasets for easy retrieval. Data consistency and accuracy are maintained through validation mechanisms. Historical records help the system learn past market behavior and trends. The database supports continuous updates to reflect changing market conditions. It also manages large datasets efficiently for machine learning operations. Proper indexing improves data access speed during processing. Secure storage mechanisms protect financial information from unauthorized access. This module serves as

the foundation for all further processing and prediction activities.

Pre-processing Module: The Pre-processing module prepares raw stock market data for machine learning analysis. It removes missing values, duplicate entries, and noisy or inconsistent records that may affect model performance. Data normalization and scaling are applied to ensure uniform feature distribution. The module converts categorical information into numerical formats when required. Feature engineering techniques are used to extract meaningful indicators such as moving averages or price changes. Time-series formatting is maintained for sequential learning models. Outliers are detected and handled to avoid misleading predictions. Data transformation improves learning efficiency and algorithm stability. Cleaned datasets reduce computational complexity and increase prediction accuracy. This module ensures that only high-quality data moves to the next stage.

Data Visualization Module: The Data Visualization module presents stock market data in graphical and analytical formats. It generates charts such as line graphs, candlestick charts, histograms, and trend plots to understand market behavior. Visualization helps identify patterns, seasonal trends, and sudden fluctuations in stock prices. Users can visually inspect correlations between different financial indicators. This module improves decision-making by simplifying complex numerical data. Interactive dashboards may be used to enhance user understanding. Visualization also helps detect anomalies before model training begins. Comparative analysis between different stocks becomes easier through graphical representation. It assists researchers in validating preprocessing results visually. Overall, this module transforms raw numerical information into meaningful visual insights.

Data Splitting Module: The Data Splitting module divides the prepared dataset into training and testing subsets. The training dataset is used to teach machine learning models, while the testing dataset evaluates performance. Proper splitting prevents data leakage and ensures unbiased prediction results. Typically, datasets are divided using ratios such as

70:30 or 80:20. Time-based splitting may be applied for stock market time-series data. Randomization techniques maintain balanced class distribution. This module ensures fair performance evaluation of algorithms. Validation datasets can also be created for hyperparameter tuning. Data splitting improves generalization capability of models. It plays a critical role in achieving reliable and realistic prediction outcomes.

Classification Models Module: The Classification Models module applies machine learning algorithms to predict stock market trends or categories. Models such as Random Forest, Support Vector Machine, Decision Tree, and Neural Networks may be implemented. These algorithms learn patterns from historical data during training. Feature relationships are analyzed to classify future market movements such as price rise or fall. Model parameters are optimized to improve learning efficiency. Multiple algorithms may be compared to identify the best performer. The module handles model training, validation, and prediction processes. Advanced techniques help capture nonlinear relationships in financial data. Automation enables continuous learning with updated datasets. This module forms the intelligence core of the prediction system.

Evaluation Module: The Evaluation module measures the performance of trained classification models. It uses metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis. Evaluation ensures that the model performs reliably on unseen data. Overfitting and underfitting issues are identified through performance comparison. Cross-validation techniques may be applied to improve reliability. Graphical evaluation tools help visualize prediction performance. The module compares multiple algorithms to select the most accurate model. Error analysis helps refine preprocessing and feature selection steps. Reliable evaluation increases confidence in prediction results. This module guarantees that only optimized models are used for final forecasting.

Prediction Output Module: The Prediction Output module generates the final stock market prediction results. It displays whether stock prices are expected to increase, decrease, or remain stable. Predictions may be presented through dashboards, reports, or graphical outputs. Real-time prediction results assist investors in decision-making. The module converts model outputs into user-friendly formats. Alerts or notifications can be generated for significant market changes. Historical predictions may be stored for performance tracking. Visualization of predicted trends enhances interpretability. The system ensures fast response time for real-time forecasting. This module delivers actionable insights derived from machine learning analysis.

RESULTS AND DISCUSSION

This project shows the details of profile how we can detect easily.

Figure 2.1: Index Page

In this picture we showed index page of the project in these basic details we can get.

Figure 2.2: Create Account Page

If we clicked Create account button directly open this page user give details and create account

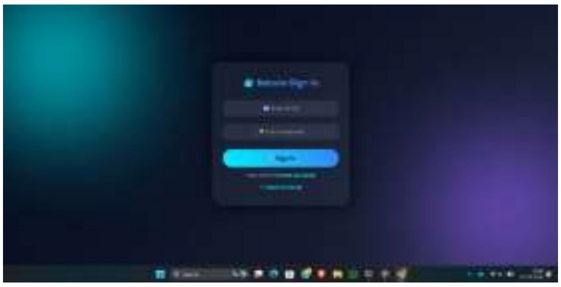

Figure 2.3 Sign in page

Then directly open this page in this page user will give credentials here and click sign in button.

Figure 2.4: Home page

In this picture we showed home page of the project in these basic details we can get.

Figure 2.5: Stock input details page

In this we give inputs of stock according with that details we predict the price of the stock

Figure 2.6: Result page

With input details we predict the cost it showed as a dashboard

CONCLSUION

In conclusion, the study demonstrates that organizational communication patterns, particularly email interactions, can provide valuable insights into company behavior and stock market performance. The analysis of the Enron email dataset reveals a strong relationship between communication frequency and stock price movements. The proposed data mining approach effectively captures hidden patterns that traditional statistical methods fail to identify. These findings highlight the potential of using communication networks as predictive indicators for financial performance. Overall, the research establishes that data-driven techniques can enhance understanding and forecasting of stock market trends.

FUTURE SCOPE

In the future, the system can be enhanced by incorporating text mining techniques to analyze the content of emails and social media data for more accurate predictions. Advanced methods such as fuzzy logic can be applied to model varying levels of communication strength more effectively. Integration of real-time data from platforms like Twitter and financial news sources can further improve prediction performance. Machine learning and deep learning models can be explored to capture more complex relationships between communication patterns and stock prices. Additionally, extending the approach to multiple organizations can improve generalization and applicability in real-world financial analysis.

REFERENCES

1.M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques", 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), pp. 322-327, 2016.

2.K. Raza, "Prediction of Stock Market performance by using machine learning techniques", 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), pp. 1-1, 2017.

3.H. Gunduz and Z. Cataltepe and Y. Yaslan, "Stock market direction prediction using deep neural networks", 2017 25th Signal Processing and Communications Applications Conference (SIU), pp. 1-4, 2017.

4.M. Billah and S. Waheed and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network", 2016 2nd International Conference on Electrical Computer & Telecommunication Engineering (ICECTE), pp. 1-4, 2016.

5.H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend", 2012 International Conference on Statistics in Science Business and Engineering (ICSSBE), pp. 1-5, 2012.

6.K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions", INTERACT-2010, pp. 59-63, 2010.

7.S. Liu and G. Liao and Y. Ding, "Stock transaction prediction modelling and analysis based on LSTM", 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), pp. 2787-2790, 2018.

8.T. Gao and Y. Chai and Y. Liu, "Applying long short term memory neural networks for predicting stock closing price", 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), pp. 575-578, 2017.

9.K. A. Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction", 2018 9th International Conference on Information and Communication Systems (ICICS), pp. 151-156, 2018.

10. J. Patel, S. Shah, P. Thakkar and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 1, pp. 259-268, 2015.

11. Y. Kim, "Predicting stock market movements using neural networks and reinforcement learning," *Applied Soft Computing*, vol. 55, pp. 132-143, 2017.

12. A. Adebiyi, A. Adewumi and C. Ayo, "Stock price prediction using the ARIMA model," *UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pp. 106-112, 2014.

13. W. Bao, J. Yue and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and LSTM," *PLoS ONE*, vol. 12, no. 7, pp. 1-24, 2017.

14. Q. Song, A. Liu and S. Yang, "Stock portfolio selection using learning-to-rank algorithms," *IEEE Access*, vol. 5, pp. 11867-11876, 2017.

15. D. Zhang, L. Wang and B. Huang, "Forecasting stock market trends using hybrid machine learning models," *IEEE International Conference on Data Mining Workshops*, pp. 695-702, 2018.

16. X. Li, P. Wu and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction," *Knowledge-Based Systems*, vol. 69, pp. 1-10, 2014.

17. S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," *International Conference on Advances in Computing Communications and Informatics*, pp. 1643-1647, 2017.

18. M. Dixon, D. Klabjan and J. Bang, "Classification-based financial markets prediction using deep neural networks," *Algorithmic Finance*, vol. 6, no. 3–4, pp. 67-77, 2017.

19. J. Nelson, A. Pereira and R. Oliveira, "Stock market's price movement prediction with LSTM neural networks," *International Joint Conference on Neural Networks (IJCNN)*, pp. 1419-1426, 2017.

20. H. Mo, "Stock market prediction using machine learning approaches," *International Conference on Big Data Analysis*, pp. 1-5, 2016.

21. Y. Zhang and L. Wu, "Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network," *Expert Systems with Applications*, vol. 36, no. 5, pp. 8849-8854, 2009.

22. M. Ballings, D. Van den Poel, N. Hespeels and R. Gryp, "Evaluating multiple classifiers for stock price direction prediction," *Expert Systems with Applications*, vol. 42, no. 20, pp. 7046-7056, 2015.

23. J. Chong, S. Han and A. Park, "Deep learning networks for stock market analysis and prediction," *Expert Systems with Applications*, vol. 83, pp. 187-205, 2017.

24. R. Akita, A. Yoshihara, T. Matsubara and K. Uehara, "Deep learning for stock prediction using numerical and textual information," *IEEE/ACIS International Conference on Computer and Information Science*, pp. 1-6, 2016.

25. T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654-669, 2018.

26. S. Basak, S. Kar, S. Saha, L. Khaidem and S. R. Dey, "Predicting stock market price using artificial neural network," *International Conference on Machine Learning and Data Science*, pp. 1-6, 2019.

27. Z. Hu, Y. Zhao and M. Khushi, "A survey of forex and stock price prediction using deep learning," *Applied System Innovation*, vol. 4, no. 1, pp. 1-19, 2021.

28. A. Sezer, M. Ozbayoglu and E. Dogdu, "Financial time series forecasting with deep learning: A systematic literature review," *Applied Soft Computing*, vol. 90, pp. 106181, 2020.

29. P. Henrique, V. Sobreiro and H. Kimura, "Literature review: Machine learning techniques applied to financial market prediction," *Expert Systems with Applications*, vol. 124, pp. 226-251, 2019.

30. S. Tripathy, A. Rahman and A. K. Laha, "Stock market prediction using ensemble machine learning approaches," *IEEE International Conference on Intelligent Systems Design and Applications*, pp. 789-794, 2020.