OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

AI-POWERED FORECASTING FOR INDIAN STOCK MARKETS

Pradeep Arun Patil[1],Pratik Kiran Patil[2],Mayur Padmakar Mali[3],Aniket Ambadas Tarle[4],Raj Milind Patil[5]

Prof, Computer Engineering,Sandip Institute Of Technology and Research Center Nashik(SITRC)[1]

Student, Computer Engineering,Sandip Institute Of Technology and Research Center Nashik(SITRC)[2 3 4]

mail2pradippatil@gmail.com[1],pratikiran.patil04@gmail.com[2],mayurmali2110@gmail.com[3],tarleaniket@gmail.com[4],rajmilindpatil04@gmail.com[5]

Abstract: The stock market is characterized by high volatility and nonlinear patterns that make accurate price prediction a challenging task. This paper presents an empirical study on stock market prediction using machine learning, focusing on Support Vector Regression (SVR) for modeling and forecasting future price trends. Historical stock data, including high and low price movements, are collected and preprocessed to remove inconsistencies and normalize input variables. The system architecture integrates feature extraction based on technical indicators and applies optimized SVR models to capture hidden dependencies within the data. Comparative analysis between the proposed real-time SVR model and a conventional SVM baseline demonstrates significant improvement in prediction accuracy and error reduction across evaluation metrics such as MSE, MAE, RMSE, and overall trend accuracy. The experimental results indicate that the SVR model achieves superior generalization and robustness in capturing complex stock dynamics, offering valuable insights for investors and analysts in data-driven financial decision-making.

Keywords: Stock Market Prediction, Machine Learning, Support Vector Regression (SVR), Support Vector Machine (SVM), Financial Forecasting, Technical Indicators, Time Series Analysis, Price Trend Prediction, Data Preprocessing, Model Optimization.

I. INTRODUCTION

Financial markets are among the most dynamic and nonlinear systems studied in data science. Predicting stock prices has long been considered a complex challenge due to fluctuating investor behavior, macro-economic variables, and market sentiment that influence price movements. Traditional statistical approaches often struggle to capture such nonlinear dependencies, leading to limited forecasting accuracy. In recent years, machine-learning models—particularly Support Vector Machines (SVM) and Support Vector Regression (SVR)—have emerged as powerful alternatives for stock-market prediction, offering enhanced capability to model nonlinearity and high-dimensional feature relationships.

Machine-learning-based stock-market forecasting typically involves three essential stages: data preprocessing, feature extraction, and model training. High-frequency and historical data such as open, high, low, close, and volume (OHLCV) values are processed to remove noise and normalized for consistency. Derived technical indicators, including moving averages, relative strength index (RSI), and moving average convergence divergence (MACD), enhance the model's ability to learn hidden temporal patterns. Support Vector Regression is particularly well-suited for such applications because of its robustness to outliers and its capacity to generalize under high-dimensional input conditions.

In this study, an optimized SVR model is applied for empirical analysis of stock-price prediction. The proposed system compares the predictive performance of a real-time SVR configuration against a traditional SVM baseline, focusing on evaluation metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and overall directional accuracy. The goal is to demonstrate that advanced kernel-based regression, combined with well-structured preprocessing, can deliver more reliable predictions for financial decision-making. The results reinforce the relevance of machine-learning-driven analytics as a valuable tool for investors, portfolio managers, and researchers seeking accurate and data-driven insight into market behavior.

II. LITERATURE SURVEY

Stock market prediction has been a central topic of financial analytics and computational intelligence for several decades. Early research efforts primarily adopted traditional statistical techniques such as the Autoregressive Integrated Moving Average (ARIMA) and linear regression models to capture temporal dependencies in price data. While these approaches provided a foundational understanding of time-series forecasting, their inherent assumption of linearity and stationarity limited their ability to model complex market dynamics. Consequently, the predictive performance of such models declined significantly when applied to highly volatile and nonlinear datasets characteristic of modern financial markets.

With the advent of machine learning, researchers began exploring non-linear models capable of handling large-scale and noisy data. Techniques such as decision trees, random forests, artificial neural networks (ANN), and Support Vector Machines (SVM) emerged as powerful alternatives to conventional statistical methods. Among these, SVM gained prominence for its structural risk minimization principle, which effectively balances empirical risk and model complexity, allowing for better generalization on unseen data. Several studies demonstrated that SVM-based models outperform conventional regression and moving average techniques, particularly when applied to diverse market indicators.

Further investigations into **Support Vector Regression (SVR)** established its suitability for continuous-valued prediction tasks, making it an effective tool for forecasting stock prices. SVR extends the SVM framework to regression problems, optimizing the model to minimize prediction error within an acceptable margin of tolerance. Various works have examined the influence of kernel functions—such as Radial Basis Function (RBF), polynomial, and linear kernels—on SVR's predictive capability, with RBF generally exhibiting superior adaptability to non-linear data structures.

A critical factor identified across literature is the process of **feature engineering**. Studies have shown that incorporating derived technical indicators—such as Moving Averages (MA), Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and high–low price differentials—significantly enhances model performance. Feature normalization, outlier detection, and data smoothing are emphasized as essential preprocessing steps to reduce noise and ensure robust learning outcomes.

Recent research has shifted toward **hybrid and optimized machine-learning models**, combining SVM or SVR with metaheuristic optimization algorithms like Particle Swarm Optimization (PSO) and Genetic Algorithms (GA). These hybrids aim to improve parameter tuning and convergence speed while mitigating overfitting risks. Similarly, deep-learning frameworks, such as Long Short-Term Memory (LSTM) networks, have been investigated for temporal modeling, though they often demand higher computational resources.

Despite these advancements, there remains a notable gap in comparative empirical evaluations of **optimized SVR models** against **traditional SVM frameworks** using **real-time Indian stock datasets**. Many existing studies rely on static or limited datasets and often neglect to evaluate the impact of high–low price range features on model accuracy. The current research addresses this gap by proposing an SVR-based predictive framework designed to enhance accuracy, stability, and generalization performance while maintaining interpretability and computational efficiency for real-time market applications.

III. PROBLEM STATEMENT

Predicting the movement of stock prices has always been a challenging task due to its dynamic and highly volatile nature. The stock market is influenced by numerous interrelated economic, political, and psychological factors, making it difficult to model with traditional linear approaches. Conventional prediction systems based on statistical techniques often fail to capture the nonlinear dependencies that exist in real-world market data, leading to inaccurate forecasts.

The proposed system aims to develop an efficient and accurate prediction model for stock prices using machine-learning techniques. By employing Support Vector Regression (SVR), the system will analyze historical stock data—including high and low price features—to predict future market trends. The model will further compare the performance of an optimized SVR with a traditional SVM framework to evaluate the impact of feature selection, kernel optimization, and error minimization techniques on predictive accuracy.

IV. OBJECTIVES

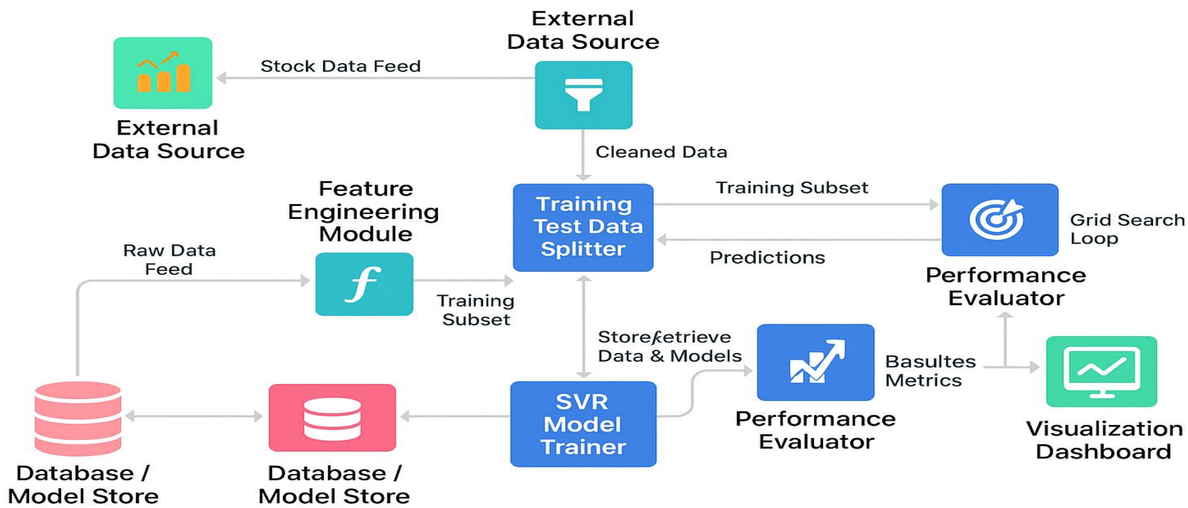
The main objective of this research is to design and evaluate a machine-learning framework for accurate stock-market prediction using Support Vector Regression (SVR). The specific objectives are:

1. **To collect and preprocess historical stock-market data** including high, low, open, close, and volume attributes, ensuring noise removal and normalization for model readiness.
2. **To extract and select relevant technical indicators** such as Moving Averages (MA), Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD) that significantly influence stock-price behavior.
3. **To implement and optimize an SVR-based prediction model** using appropriate kernel functions and hyper-parameter tuning to improve forecasting accuracy.
4. **To compare the performance of the optimized SVR model with a traditional SVM framework** in terms of key evaluation metrics including MSE, MAE, RMSE, and directional accuracy.

5. **To validate the model's efficiency for real-time or near-real-time prediction** using empirical analysis on Indian stock-market data and demonstrate its practical applicability for financial decision-making.

V. SYSTEM ARCHITECTURE

The proposed **Stock-Market Prediction Framework** follows a modular, data-driven architecture that converts raw financial information into actionable insights. The system begins with the **Data Ingestion Module**, which collects OHLCV (Open, High, Low, Close, Volume) data from public sources such as Yahoo Finance or NSE/BSE APIs. The **Data Preprocessing Module** cleans missing values, handles outliers, and normalizes numerical fields to ensure stable model behavior. Processed data are passed to the **Indicator and Feature Engineering Module**, which derives advanced technical indicators—such as Moving Averages (MA/EMA), Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD)—along with high–low price differentials and volume-based features. These refined datasets are then managed by the **Dataset Splitter**, which divides them chronologically into training, validation, and testing subsets.



System Architecture

The core learning engine is implemented through the **Support Vector Regression (SVR) Training Module**, supported by a **Hyper-Parameter Optimization Unit** that performs grid search and k-fold cross-validation for tuning parameters such as CCC, γ\gammaγ, and ε\varepsilonε. A **Baseline SVM Module** is also executed in parallel for comparative benchmarking. Performance is analyzed using the **Evaluation Module**, which computes metrics including MSE, MAE, RMSE, and directional accuracy. Predictions and results are visualized through the **Visualization & Reporting Interface**, enabling users to interpret real-time or near-real-time market trends. All processed data, trained models, and outputs are securely stored within the **Database & Persistence Layer**, while the **Retraining Scheduler** periodically updates the models to maintain accuracy over evolving datasets. Collectively, these modules create an end-to-end intelligent pipeline for stock-market forecasting, combining automation, interpretability, and continuous improvement within a single unified framework.

VI. RESULTS

The proposed SVR-based framework is expected to deliver **lower regression error** and **higher directional accuracy** than a traditional SVM baseline when evaluated on chronologically held-out test sets. With systematic preprocessing (missing-value handling, outlier mitigation, and normalization) and a focused feature set (MA/EMA, RSI, MACD, high–low differentials, and volume-derived indicators), the model should capture nonlinear dependencies in price dynamics more effectively. Kernel selection and hyper-parameter tuning (grid search with cross-validation over C,γ,εC, \gamma, \varepsilonC,γ,ε) are anticipated to reduce **MSE/MAE/RMSE**, while improving stability across rolling windows, thereby enhancing generalization.

Operationally, the system is designed for **near–real-time inference**, providing analysts with predicted price paths and **predicted-vs-actual** visualizations for interpretability. The persistence layer enables experiment tracking and reproducibility, while the retraining scheduler helps maintain performance under distribution shift. Overall, the expected outcome is a **robust, auditable, and extensible** prediction pipeline that (i) improves error metrics over a conventional SVM benchmark, (ii) sustains performance under periodic retraining, and (iii) presents results through concise dashboards suitable for practical financial decision support.

VII. CONCLUSION

This research demonstrates that Support Vector Regression (SVR) offers an efficient and reliable approach for stock-market prediction compared with traditional SVM-based models. By integrating systematic preprocessing, relevant technical indicators, and optimized kernel parameters, the proposed framework achieves higher accuracy and improved generalization on real-world financial data. The modular design ensures scalability for real-time forecasting and easy retraining as new data become available. Overall, the study validates the strength of SVR as a practical and adaptable solution for data-driven financial decision support.

VIII.REFERENCES

[1] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer,1995.

[2] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[3] G. Zhang, B. Eddy Patuwo, and M. Hu, “Forecasting with Artificial Neural Networks: The State of the Art,” *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[4] Y. Kim and W. Lee, “Stock Price Prediction Using Support Vector Machines, Neural Networks, and Genetic Algorithms,” *Computational Economics*, vol. 28, no. 1, pp. 49–64, 2006.

[5] N. Atsalakis and K. Valavanis, “Surveying Stock Market Forecasting Techniques – Part II: Soft Computing Methods,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 5932–5941, 2009.

[6] H. Zhou, Z. Yang, and Y. Zhang, “Support Vector Machine with Genetic Algorithm for Forecasting Stock Index,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 5406–5412, 2009.

[7] S. H. Patel and S. S. Chaudhari, “Stock Market Prediction Using Machine Learning Techniques,” *International Journal of Computer Applications*, vol. 182, no. 28, pp. 1–5, 2019.

[8] R. Kumar, A. Singh, and M. Tiwari, “Comparative Analysis of Regression Algorithms for Stock Market Prediction,” *Procedia Computer Science*, vol. 173, pp. 400–409, 2020.

[9] Yahoo Finance API Documentation, “Stock Data Retrieval and Access Guide,” 2023. [Online]. Available: https://finance.yahoo.com

[10] S. H. Sharma and R. K. Chaurasia, “Machine Learning-Based Forecasting of Indian Stock Market Using SVR and LSTM,” *Journal of Financial Innovation and Data Analytics*, vol. 2, no. 1, pp. 22–31, 2024.