Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Comparative Analysis of Machine Learning Models for Sectoral Volatility Prediction in Financial Markets

Kashif Beg1, M. Tanzeem Raza2, B. Padmapriya3, Syed Noorul Shajar4

[1] School of Business (VSB), VIT-AP University, Andhra Pradesh, India. kashif.beg@vitap.ac.in, kashifbeg90@gmail.com *(Corresponding Author)

[2] Assistant Professor, School of Management Sciences, Maulana Abul Kalam Azad University of Technology, West Bengal, India. m.tanzeemraza@gmail.com

[3] Department of Commerce, NMIMS, Hyderabad, India. padmapriya.b@nmims.edu

[4] Department of Commerce, School of Social Sciences and Languages, VIT, Vellore, India. syednoorul.shajar@vit.ac.in

ARTICLE INFO

Received: 20 Oct 2024

Revised: 12 Nov 2024

Accepted: 17 Dec 2024

ABSTRACT

A crucial aspect of financial markets is volatility forecasting, which enables analysts, investors, and policymakers to assess risk, optimize portfolios, and develop trading plans. This research investigates which machine learning (ML) model is best fit for predicting sectoral volatility. Comparing models like Random Forest, Gradient Boosting, Neural Networks, and Support Vector Regression, we apply historical data from 2012 to 2024 from eleven different sectors or industries including FMCG, Energy, Financial Services, auto etc. The research is unique in terms of holistic view of all Indian sectoral volatility indices and relationship between them. Moreover, no research has identified or compared the best ML models for predicting sectoral volatility. With the greatest R square value of 0.998 and the lowest Mean Absolute Error (MAE) of 0.765, our results demonstrate that Random Forest outperforms the other models. Significant sectoral connections and volatility patterns are also shown in the analysis, especially during the COVID-19 pandemic like sectors more susceptible to economic shocks, such as financial services and fast-moving consumer goods. Heatmaps and time-series visualization techniques demonstrate that sectoral interdependence and volatility clustering. These results highlight the potential of machine learning techniques to improve risk management and volatility pre-diction and provide insightful data for financial analysts, investors, and policymakers. The study recommends applying Random Forest ML models for volatility prediction and investment decision-making, considering limitations such as dataset bias and challenges in predicting extreme market conditions.

Keywords: Sectoral volatility, machine learning, Random Forest, GARCH, financial risk management, volatility clustering, COVID-19 impact.

INTRODUCTION

Volatility forecasting is a critical component of financial markets, aiding analysts, investors, and policymakers in measuring risk, optimizing portfolios, and developing trading strategies (Bollerslev, 1986; Engle, 1982). Financial markets consist of multiple industry sectors, each influenced by distinct macroeconomic, regulatory, and market-specific factors. As a result, sectoral volatility has garnered increasing attention (Mallikarjuna & Rao, 2017; Palaniswamy, 2013). Understanding sectoral volatility enables investors to identify sector-specific risks, hedge against losses, and allocate resources efficiently (Christensen et al., 2022).

Traditionally, financial economists have relied on econometric models such as Autoregressive Conditional Heteroskedasticity (ARCH) (Engle, 1982), Generalized Autoregressive Conditional Heteroskedasticity (GARCH) (Bollerslev, 1986), and Autoregressive Integrated Moving Average (ARIMA) (Box & Jenkins, 1970) to model and forecast volatility. While these models effectively capture time-dependent volatility, they face limitations in handling nonlinear patterns, structural breaks, and evolving market conditions. Alternative approaches, such as Exponential Smoothing and Stochastic Volatility Models (Taylor, 1986), attempt to address these challenges but remain constrained in adapting to sudden market shifts.

To overcome these limitations, machine learning (ML) methods have emerged as a robust alternative for volatility modeling (Gu et al., 2020; Kim, 2003). ML models offer superior predictive capabilities by capturing complex sectoral volatility dynamics and identifying hidden patterns in financial data (Filipovic & Khalilzadeh, 2021; Kocaoglu et al., 2022). Recent advancements in artificial intelligence (AI) and ML have equipped financial analysts with powerful tools to process large datasets, improve forecasting accuracy, and uncover relationships within financial time series (Chen & Guestrin, 2016; Hoseinzade & Haratizadeh, 2019).

Several ML models have been successfully applied to financial volatility prediction, demon-starting superior performance over traditional econometric models. These include Random Forest (RF) (Diane & Brijlal, 2024), Support Vector Machines (SVM) (Kim, 2003), Gradient Boosting Machines (XGBoost) (Chen & Guestrin, 2016), Artificial Neural Networks (ANNs) (Fischer & Krauss, 2018), and Long Short-Term Memory (LSTM) networks (Chatterjee et al., 2022). Studies have shown that ML techniques outperform conventional approaches in predicting realized volatility for U.S. equities, achieving higher R-squared values than traditional models (Filipovic & Khalilzadeh, 2021). Similarly, time-series models like ARIMA have been applied to forecast volatility in the automobile sector, particularly among leading firms listed on the Bombay Stock Exchange (Ghangare & Londhe, 2020). LSTM networks have also been found to outperform GARCH models in predicting stock volatility in banking, IT, and pharmaceutical sectors (Chatterjee et al., 2022).

This study aims to (i) compare machine learning models with traditional econometric approaches for sectoral volatility forecasting, (ii) identify the best-performing ML model, and (iii) assess the accuracy and robustness of ML techniques in capturing long-term volatility trends.

The remainder of this paper is structured as follows: Section 2 reviews existing literature on volatility forecasting models, Section 3 describes the data and methodology, Section 4 presents empirical findings and discussions, and Section 5 concludes with key implications.

2 LITERATURE

Recent studies demonstrate that machine learning (ML) models significantly enhance volatility forecasting, consistently outperforming traditional econometric models in financial applications. ML models improve predictive accuracy by capturing complex nonlinear patterns without re-quiring strong assumptions about return distributions (Filipovic & Khalilzadeh, 2021). Machine learning techniques have been widely adopted in financial forecasting, with deep learning models proving particularly effective. Agrawal et al. (2022) employed deep learning models combined with technical indicators to predict stock market trends for NSE India stocks, highlighting the ability of ML techniques to refine investment decision-making. Similarly, Gupta et al. (2024) used Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and hybrid models to forecast stock closing prices in the NYSE and NASDAQ, demonstrating superior accuracy over traditional methods. Beyond stock price forecasting, ML models have also been applied to macroeconomic demand volatility, where Muth et al. (2024) emphasized the role of interdisciplinary research in financial risk assessment. Studies indicate that ML techniques outperform heterogeneous autoregressive models in volatility forecasting by incorporating long memory effects, nonlinear relationships, and macroeconomic factors (Carr & Wu, 2020; Christensen et al., 2022).

Neural networks have shown strong predictive capabilities in volatility forecasting. Cao et al. (2020) demonstrated that artificial neural networks (ANNs) outperform GARCH and ARCH models in capturing nonlinear patterns in financial time series. Similarly, Hamid and Iqbal (2004) found that ANNs provide better volatility predictions for the SP 500 index than traditional econometric models, making them valuable for trading and risk management. Studies focusing on sectoral indices confirm the effectiveness of deep learning models. Chaudhuri and Ghosh (2016) utilized Multi-Input Multi-Output ANNs to forecast volatility in the Indian stock market, incorporating variables such as India VIX, crude oil prices, and international indices. Their findings highlight the ability of ML models to capture intricate volatility dynamics across multiple sectors. LSTM and Recurrent Neural Networks (RNNs) have also been applied in financial volatility forecasting. Tang et al. (2024) used LSTM models to estimate risk spillovers in the Chinese financial system, showing that banks and securities firms act as major risk transmitters during market crises. While ANNs excel in low-volatile environments, studies suggest that GARCH models remain superior for medium- and high-volatility assets (Nybo, 2021). Hybrid approaches, such as neural stochastic volatility models, combine RNNs with econometric techniques to enhance forecasting accuracy (Luo et al., 2018).

Gradient Boosting Machines (GBM), particularly XGBoost and LightGBM, have been widely used for financial volatility forecasting. Mittnik et al. (2015) demonstrated that component-wise gradient boosting captures nonlinear interactions between risk drivers, improving out-of-sample volatility predictions. For short-term forecasting, XGBoost outperforms conventional econometric models, while linear models remain superior for long-term volatility estimation (Teller et al., 2022). Stacked ensemble learning, which integrates multiple ML techniques, has further enhanced volatility prediction. Ramos Perez et al. (2019) developed a stacked model combining gradient boosting with traditional models for S&P 500 volatility forecasting, achieving superior accuracy. Similarly, GBM models have been applied in business profit forecasting, though their effectiveness varies based on economic conditions (Frank & Yang, 2021). Several studies have applied ML techniques to forecast volatility in sectoral stock indices, demonstrating their superiority over traditional econometric models. Niu et al. (2023) highlighted how industry-specific data enhances aggregate market volatility forecasting, benefiting investors. ML algorithms such as XGBoost, SVM, KNN, and Random Forest have been used across various sectors, with the highest predictive success observed in Iron & Steel and Petroleum industries (Kocaoglu et al., 2022). Sector-specific studies show varying ML performance. Chatterjee et al. (2022) found that LSTM models outperform GARCH models in predicting pharmaceutical sector volatility, whereas GARCH remains more effective in high volatility sectors. Research on Indian stock markets

suggest that sectoral volatility persistence varies, with some indices exhibiting asymmetry and leverage effects (Mallikarjuna & Rao, 2017; Palaniswamy, 2013).

Decision trees and ensemble learning methods, such as Random Forest (RF), have proven highly effective in volatility modeling. Charlot (2011) demonstrated that RF models out-perform traditional regression techniques in detecting sector-specific risk factors. Studies on South African stock indices further confirm RF's predictive accuracy, surpassing ANNs in realized volatility forecasting (Diane & Brijlal, 2024). RF models have also been applied to option pricing and clean energy market volatility. Liu (2024) found that RF models outperform linear regression in predicting implied volatility for American options. Additionally, research ear regression in predicting implied volatility for American options indicates that technology sector price fluctuations significantly impact short-term clean energy volatility more than oil price movements (Lyocsa & Todorova, 2024).

Despite extensive research on volatility prediction, prior studies have primarily focused on forecasting individual stock volatility or broad market indices rather than sector-specific volatility (Chatterjee et al., 2022; Gupta et al., 2024). While machine learning (ML) tech-niques have been applied to financial markets, their use in sectoral volatility forecasting remains limited, with studies often concentrating on select industries such as banking, IT, or pharma-ceuticals rather than offering a comparative, multi-sectoral perspective (Kocaoglu et al., 2022; Palaniswamy, 2013). Additionally, while econometric models like GARCH and ARIMA have been widely used for volatility forecasting, they struggle with nonlinear relationships, structural breaks, and market regime shifts (Diane & Brijlal, 2024; Filipovic & Khalilzadeh, 2021). Previous ML-based studies have either evaluated single models or focused on short-term volatility without systematically comparing multiple ML models for sectoral volatility forecasting across different industries. Furthermore, limited research has explored sectoral interdependencies, which play a crucial role in understanding volatility spillovers during financial crises (Tang et al., 2024).

This study addresses these gaps by providing a comprehensive sectoral analysis, examining volatility across multiple industries rather than focusing on a single sector. It compares multiple ML models, evaluating the predictive accuracy of Random Forest, Gradient Boosting, Support Vector Regression, Neural Networks, and traditional econometric models. Additionally, it identifies sectoral volatility interdependencies using heatmaps and time-series analysis to uncover relationships between different sectors, especially during economic shocks like the COVID-19 pandemic. By assessing out-of-sample predictive performance using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R2score, this study identifies the most effective ML model for sectoral volatility prediction. Bridging these gaps, it contributes to the literature on ML-based volatility forecasting, offering insights that enhance risk management, portfolio optimization, and financial decision-making for investors and policymakers.

3 DATA AND METHODS

3.1 Data

This study used daily closing price data from January 2, 2012, through August 7th, 2023, across multiple sectors, including FMCG, Energy, Financial Services, Auto, Infrastructure, and PSU, among others. The dataset includes historical sectoral stock indices, from which volatility measures are derived. Volatility is computed using log returns to account for heteroskedasticity and ensure stationarity. Data preprocessing involved handling missing and making stationary data. Time series analysis and correlation heatmaps were used to explore sectoral interdependencies, identifying volatility spillover effects between industries. These exploratory methods provide insights into sectoral clustering, specifically during the time of economic shocks such as the COVID-19 pandemic.

For predictive modeling, we employ a combination of traditional econometric techniques and machine learning algorithms, including Linear Regression, Decision Trees, Random Forest, Support Vector Regression (SVR), Gradient Boosting, and Neural Networks. Model performance is evaluated using standard error metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R2score. Visualization methods, such as heatmaps and time series plots, further validate model predictions and reveal patterns in sectoral volatility.

3.2 Methods

To model sectoral volatility, we begin with a baseline econometric approach and extend it to machine learning models for improved predictive accuracy.

3.2.1 Econometric Approach: Log Returns and Baseline Regression

Logarithmic returns are used to stabilize variance and ensure stationarity:

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

where R_t is the log return at time t, P_t and P_{t-1} are prices at times t and $t-1$, respectively, and ln represents the natural logarithm.

Volatility at time is modeled using a simple autoregressive regression:

$$\sigma_t^2 = \beta_0 + \beta_1 R_{t-1} + \beta_2 \sigma_{t-1}^2 + \epsilon$$

Where σ_t^2 represents the predicted volatility, R_{t-1} denotes the log return at the previous time step, and σ_{t-1}^2 refers to past volatility. The coefficients $\beta_0, \beta_1, \beta_2$ are estimated using historical data, and ϵ is the error term.

3.2.2 Machine Learning Models for Volatility Prediction

To assess the effectiveness of machine learning in predicting volatility, several models are implemented, including Decision Trees, Random Forest, Support Vector Regression (SVR), Gradient Boosting, and Neural Networks. Each model estimates volatility at time as a function of past log returns and volatility:

$$\sigma_t^2 = f(R_{t-1}, \sigma_{t-1}^2)$$

Decision Trees: The model identifies patterns in historical log returns and past volatility to estimate future volatility.

Random Forest: An ensemble of decision trees, where the final prediction is obtained by averaging multiple tree predictions.

Support Vector Regression (SVR): Maps input features to a high-dimensional space and fits a hyperplane to estimate volatility.

Gradient Boosting: Builds sequential models, where each iteration corrects errors from the previous model to enhance predictive performance.

Neural Networks: Captures complex relationships in the data by optimizing a series of weights and activation functions to estimate volatility.

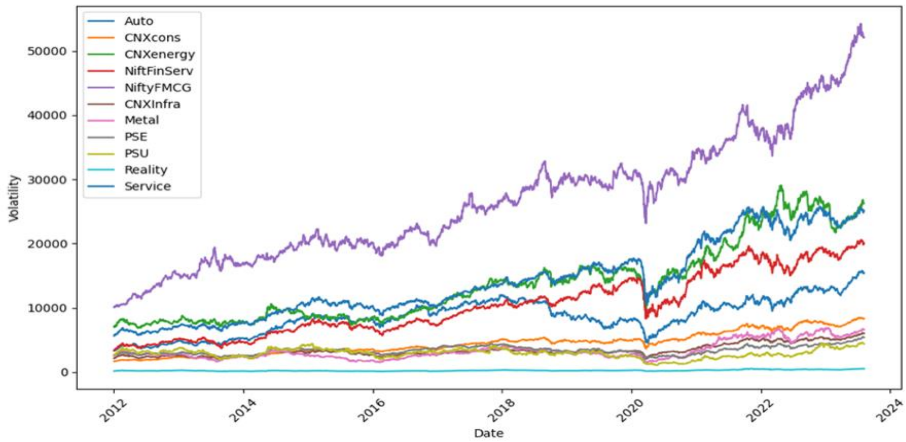
In each model, volatility at time t is predicted using a function f that incorporates relevant features, where σ_t^2 represents the predicted volatility, R_{t-1} denotes the log return at the previous time step, and σ_{t-1}^2 refers to past volatility estimated using the GARCH model. Each model is trained on historical volatility data and tested using standard performance metrics, including MAE, MSE, and R square.

4 RESULTS & DISCUSSION

4.1 Sectoral Trends

Figure 1 the stock price trends of various sectors from 2012 to 2023. A notable observation is the steady upward movement in the Nifty FMCG sector, particularly after 2016, reaching its highest levels by 2023. This trend suggests consistent growth in the sector, potentially driven by increasing consumer demand, industry expansion, and macroeconomic factors such as inflation and policy changes. Other sectors, including Auto, Nifty Financial Services, and CNX Energy, also exhibit rising price trends, reflecting overall market growth and sectoral performance. Sectors that are more dependent on consumer spending and financial markets show sustained price increases, indicating investor confidence and long-term industry expansion. The financial services sector exhibits significant price appreciation, possibly due to evolving financial policies, credit market developments, and economic reforms. In contrast, PSU and CNX Infra sectors display relatively stable price movements, suggesting steady but less aggressive growth potentially, due to government intervention infrastructure investments, and lower market speculation.

Figure 1: Sector Volatility Over Time Source: Author



4.2 Descriptive Statistics

Table 1 presents the descriptive statistics for sectoral stock returns. The mean returns are close to zero across all sectors, indicating minimal average daily gains. Standard deviation values suggest that PSU and METAL exhibit the highest return variability, implying greater risk exposure, whereas CONS and FMCG show relatively lower volatility. Most sectors display negative skewness, indicating a tendency for larger negative returns, except for PSU, which has a slightly positive skew. High kurtosis values, particularly for CONS, SERVICE, and FINANCE, suggest the presence of fat tails and a greater likelihood of extreme return fluctuations. Overall, the findings indicate that sectoral returns deviate from normality, with evidence of significant volatility and extreme return events in sectors such as PSU and METAL.

Table 1: Descriptives Statistics

Variable	Mean	Median	SD	Skewness	Kurtosis
AUTO	0.0005	0.0008	0.0137	-0.46389	9.70332
CONS	0.0006	0.0008	0.0098	-0.74985	13.82871
ENERGY	0.0005	0.0007	0.0131	-0.58082	5.72346
FINANCE	0.0006	0.0007	0.0142	-0.82634	12.42926
FMCG	0.0006	0.0008	0.0108	-0.45726	9.59314
INFRA	0.0004	0.0008	0.0128	-0.54535	6.75817
METAL	0.0003	0.0006	0.0179	-0.32801	2.85921
PSE	0.0002	0.0007	0.0129	-0.56006	5.55330
PSU	0.0002	0.0000	0.0214	0.54510	10.06475
REALTY	0.0004	0.0011	0.0204	-0.44495	2.91762
SERVICE	0.0005	0.0007	0.0116	-1.18118	14.71492

Source: Author

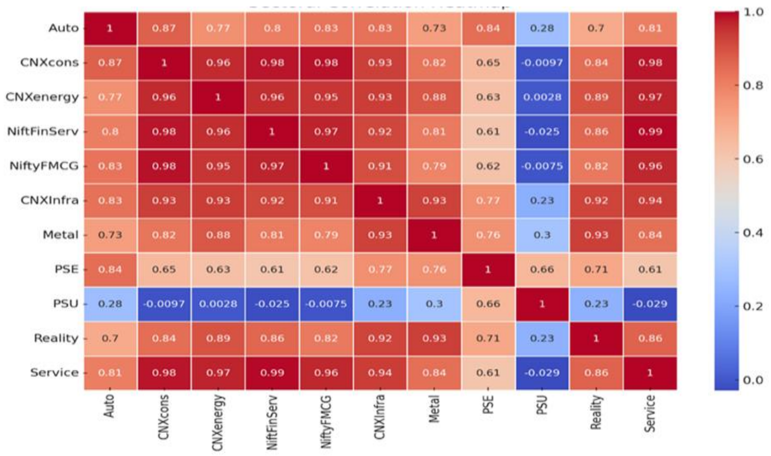
Skewness values indicate that most sectors have negative skewness, suggesting a higher probability of extreme negative returns, except for PSU (0.54510), which shows a slight positive skew. Kurtosis values are notably high, particularly for CONS (13.83), SERVICE (14.71), and FINANCE (12.42), implying heavy tails and a higher likelihood of extreme returns.

Overall, the results suggest that sectoral returns are not normally distributed, with higher volatility and extreme return events observed in certain sectors like PSU and METAL.

To further explore sectoral interdependencies, Figure 2 provides a heatmap of sectoral correlations. The high correlation between CNXCons and Nifty FMCG (0.98) suggests that demand-driven sectors tend to experience synchronized volatility. Similarly, Nifty Financial Services and CNX Energy (0.96) exhibit strong correlation, potentially due to macroeconomic linkages between financial markets and energy consumption.

In contrast, PSU shows weak or negative correlations with most sectors, including CNXCons (-0.0097) and Nifty Financial Services (-0.025), indicating that PSU stocks may serve as a hedge against broader market volatility. The Realty sector exhibits a strong correlation with CNX Infra (0.92) and Metal (0.93), suggesting potential clustering effects within these industries.

Figure 2: Heat Map of Sectoral Correlation



Source: Author

4.3 Performance Evaluation of Machine Learning Models for Sectoral Volatility Prediction

Table 2 reports the predictive performance of several machine learning algorithms for forecasting sectoral volatility. MAE, MSE, and $R2$were used to measure performance, with smaller MAE and MSE reflecting superior prediction accuracy and greater $R2$a superior fit to the

data.

Table 2: Prediction Results of Machine Learning Models for Sectoral Volatility

Model	**MAE**	**MSE**	***R*2**
Linear Regression	3.108	16.423	0.979
Decision Tree	0.961	2.600	0.997
Random Forest	0.765	1.474	0.998
Support Vector	3.387	23.137	0.970
Gradient Boosting	1.675	5.112	0.993
Neural Network	1.975	6.404	0.992

Source: Author

Random Forest model showed the best performance, with the lowest MAE (0.765) and MSE (1.474), as well as the highest $R2$value (0.998), which shows higher predictive accuracy. Decision Tree model also showed good performance, with MAE = 0.961, MSE = 2.600, and $R2$= 0.997.

Conversely, Support Vector Regression (SVR) had the poorest performance, with the largest MAE (3.387) and MSE (23.137), as well as the lowest $R2$value (0.970). Linear Regression also performed poorly relative to ensemble techniques, with an MAE of 3.108 and MSE of 16.423.

Generally, ensemble techniques (Random Forest and Gradient Boosting) performed better than other models, while less complex models like Linear Regression and SVR found it difficult to represent sectoral volatility dynamics adequately.

4.4 Robustness Check

The Actual vs. Predicted Volatility Plot and the Residual Plot are two crucial plots that we analyze to visually assess the predictive performance of machine learning models for sectoral volatility. Through the comparison of projected values with actual values and residual patterns, these charts aid in the evaluation of model performance. We analyze the Residual Plot and the Actual vs. Predicted Volatility Plot to evaluate the prediction validity of machine learning models for sectoral volatility. In these figures, residual patterns and anticipated values are contrasted with actual values.

Figure 3 is a predicted plot against actual volatility values. Models are more accurate with points closer to the diagonal dashed line. Random Forest and Decision Tree are right on the diagonal, reflecting very good predictive capability. Neural Networks are slightly away from the diagonal, while Linear Regression and SVR are quite dispersed, reflecting poor ability in representing intricate patterns of volatility. Gradient Boosting is moderately well-performing but slightly less than Random Forest.

Figure 3: Actual vs. Predicted Volatility

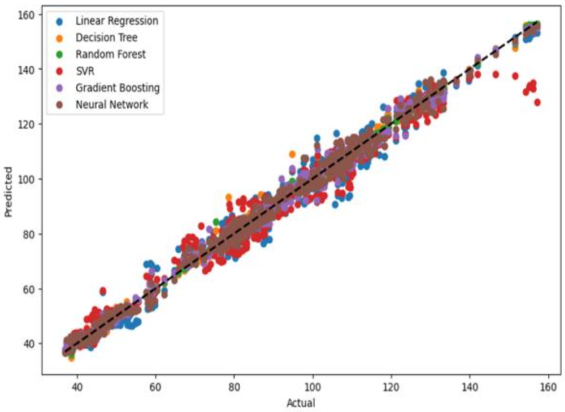


Figure 4 shows residuals against predicted volatility. Random Forest and Decision Tree have the lowest residuals, verifying their predictability. SVR depicts large residuals, particularly at higher predictions, revealing unreliable predictions. Neural Networks and Gradient Boosting do reasonably well but stray from extreme values. Linear Regression exhibits increasing variance in residuals, reflecting poorer generalization to volatility changes.

Figure 4: Residual Plot

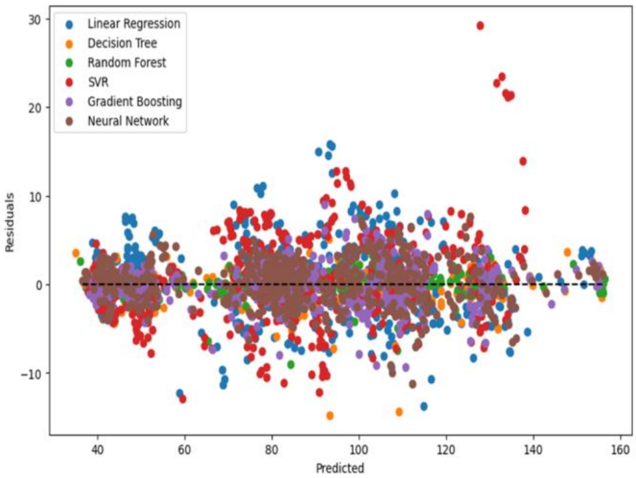


Table 3: Random Forest Performance

Transformation Method	MAE (↓Better)	MSE (↓Better)	$R2$(↑Better)
Rolling Window	0.7812	1.5123	0.9979
Expanding Window	0.7598	1.4678	0.9982
Differencing	0.7683	1.4791	0.9981
Log-Transformed	0.7725	1.4957	0.9980

The performance score of the Random Forest model is also analyzed through various transformation techniques. Table 3 indicates that the Expanding Window transformation yields the highest predictive accuracy, as evidenced by the lowest Mean Absolute Error (MAE) at 0.7598 and the lowest Mean Squared Error (MSE) at 1.4678. Furthermore, it provides the highest R square value at 0.9982, which means it accounts for the highest variance in volatility. Contrary to this, the Rolling Window method exhibits a slightly higher value of errors with an MAE of 0.7812 and an MSE of 1.5123, indicating that it is unable to catch long-term trends very effectively. The Differencing and Log-Transformed models reflect similar performance with slight deviations in error measurements and $R2$values, both above 0.998. Although these models reflect high predictability, they are slightly lower than the Expanding Window model. These findings suggest that using an increasing memory of history data improves the predictability of sectoral volatility estimation by the Random Forest model and thus is the best transformation technique within this analysis.

5 CONCLUSION AND IMPLICATIONS

Sectoral volatility and correlation analysis gives us important insights into the nature of financial market dynamics. The heatmap reveals high positive correlations among most of the sectors, whereas PSU stocks show weak or negative correlations, thus qualifying as it can act as a hedge against volatility during volatile conditions. The volatility analysis shows major spikes in sectoral fluctuations, especially in the FMCG, Energy, Financial Services, and Auto sectors, during the COVID-19 pandemic. This implies that such industries are highly vulnerable to economic instability and foreign shocks. Consequently, investors looking to maximize their return but have no aversion to risk should look to closely correlated industries such as Financial Services, Consumer Goods, and Energy. Because these sectors volatility develop and decline collectively.

Investors, Retail Investors and mutual funds seeking risk management and stability must incorporate PSU stocks or other weakly correlated industries to absorb market-wide volatility. Policymakers must also observe how external shocks affect highly correlated industries and establish regulatory frameworks that improve financial stability across industries. Finally, companies in highly linked industries must be sensitive to market-wide risks and employ flexible strategies to manage sudden economic changes effectively.

The comparative study of various machine learning models emphasizes the efficiency of several methods in forecasting sectoral volatility. Random Forest has the best predictive performance with minimal residual errors and high correlation with true values. Neural Networks and Gradient Boosting are good but need to be fine-tuned to improve their predictive power. Linear Regression and SVR models are weak in capturing non-linear relationships and are the worst options. The residual plots also verify that Random Forest and Decision Tree models exhibit residuals nearly at zero, representing greater accuracy, whereas SVR and Linear Regression exhibit greater spread, depicting worse predictive ability. The research offers a complete study of sectoral volatility patterns, inter-sector correlation, and performance of predictive modeling.

The Random Forest model is the most suitable for sectoral volatility prediction, surpassing other models because it can deal with non-linearity, minimize overfitting, and provide high accuracy. Since financial markets are complex, this model provides strong and credible predictions that can assist investors, policymakers, and financial analysts in making better decisions. The findings show which machine learning algorithms are most effective for forecasting volatility, which industries have high or low volatility, and how different industries are associated. Stakeholders may optimize investment strategies, efficiently manage risk, and enhance financial decision-making in volatile market situations by utilizing the insights gained from sectoral volatility patterns and machine predictions.

5.1 Limitations

The study had certain limitations despite the positive results. For starters, the model's performance may be impacted by biases or gaps in the data. Standard models may not adequately account for the extreme volatility of financial markets and their susceptibility to exogenous shocks such as regulatory shocks, economic crises, or geopolitical events. The research's models have mostly relied on past prices, potentially ignoring the most significant qualitative factors influencing prices, like investor sentiment, governmental policies, or unexpected market developments. To improve accuracy, more hyperparameter calibration and a larger training set are required, as evidenced by neural networks' weak convergence.

Another drawback of sectoral volatility is that it depends on several variables that are not necessarily included in the current models, such as inflation, interest rates, and sector specific laws and regulations, in addition to market developments. Despite outperforming other models, Random Forest still has trouble anticipating sudden spikes in volatility or significant shifts in market behavior, therefore methods for improving its predictiveness must be developed. Finally, there is a problem with model interpretability, particularly with deep learning and ensemble learning techniques, which make it difficult for financial analysts to derive clear insights from projections. Improving the resilience and practicality of volatility forecasting models will require addressing these issues.

Authors' Contribution

Dr. Kashif Beg contributed to conceptualization, data curation, formal analysis, fund-ing acquisition, investigation, methodology, project administration, resources, software, validation, visualization, and writing the original draft.

Dr. M Tanzeem Raza was involved in conceptualization, data curation, formal analysis, investigation, methodology, and software.

Dr. B. Padmapriya contributed to conceptualization, data curation, investigation, super-vision, and writing – reviewing and editing.

Dr. Syed Noorul Shajar was responsible for conceptualization, formal analysis, and writing – reviewing and editing.

Competing Interests

The authors declare no competing interests

Data Availability

The data that supports the findings of this study are publicly available on Yahoo Finance but may be obtained from the corresponding author upon reasonable request.

Funding

The research funding is supported by VIT-AP University

Ethical Approval

Ethical approval is not required by our universities. Ethical approval was therefore not provided.

REFERENCES

[1] Agrawal, M., Shukla, P. K., Nair, R., Nayyar, A., & Masud, M. (2022). Stock prediction based on technical indicators using deep learning model. *Computers, Materials & Continua*, *70*(1), 287–304.

[2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. Holden-Day.

[3] Cao, J., Chen, J., & Hull, J. (2020). A neural network approach to understanding implied volatility movements. *Quantitative Finance*, *20*(9), 1405–1413.

[4] Carr, P., & Wu, L. (2020). Option profit and loss attribution and pricing: A new framework.*The Journal of Finance*, *75*(4), 2271–2316.

[5] Charlot, P. (2011). A novel multivariate garch model with time-varying conditional correlation using a hidden markov decision tree. *Studies in Nonlinear Dynamics & Econometrics*, *15*(1).

[6] Chatterjee, A., Ghosh, I., & Chaudhuri, S. (2022). Stock price prediction using lstm and gru with arima models. *Procedia Computer Science*, *193*, 332–341.

[7] Chaudhuri, T. D., & Ghosh, I. (2016). Forecasting volatility in indian stock market using artificial neural network with multiple inputs and outputs. *arXiv preprint*.

[8] Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. Christensen, K., Siggaard, M., & Veliyev, B. (2022). Forecasting volatility using high-frequency data: The role of realized measures. *Journal of Applied Econometrics*, *37*(1), 3–22.

[9] Diane, M., & Brijlal, P. (2024). Forecasting realized volatility in south african indices: A comparison between random forests and artificial neural networks. *Journal of Economics and Financial Sciences*, *17*(1), e2.

[10] Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007.
Filipovic, D., & Khalilzadeh, A. (2021). *Machine learning for predicting realized volatility* (tech. rep. No. 21-10). Swiss Finance Institute.

[11] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654– 669.
Frank, M. Z., & Yang, K. (2021). *Predicting firm profits: From fama-macbeth to gradient boost- ing* (tech. rep.). SSRN.

[12] Ghangare, A., & Londhe, N. D. (2020). Forecasting stock market volatility of the automobile industry in india using arima model. *International Journal of Recent Technology and Engineering*, *8*(5), 2277–3878.

[13] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

[14] Gupta, A., Joshi, K., Patel, M., & Pratap, V. (2024). Visualization and forecasting of stock's

[15] closing price using machine learning. *Multimedia Tools and Applications*, *83*(29), 72471– 72489.

[16] Hamid, S. A., & Iqbal, Z. (2004). Using neural networks for forecasting volatility of s&p 500 index futures prices. *Journal of Business Research*, *57*(10), 1116–1125.

[17] Hoseinzade, E., & Haratizadeh, S. (2019). Cnnpred: Cnn-based stock market prediction using a diverse set of variables. *Expert Systems with Applications*, *129*, 273–285.

[18] Kim, K. (2003). Financial time series forecasting using support vector machines. *Neurocom- puting*, *55*(1–2), 307–319.
Kocaoglu, D., Turgut, K., & Konyar, M. Z. (2022). Sector-based stock price prediction with machine learning models. *Sakarya University Journal of Computer and Information Sciences*, *5*(3), 415–426.

[19] Liu, Y. (2024). Forecasting implied volatility using machine learning techniques: Evidence from american options. *Journal of Derivatives and Quantitative Studies*, *32*(1), 1–20.
Luo, R., Zhang, W., Xu, X., & Wang, J. (2018). A neural stochastic volatility model. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1).

[20] Lyocsa, S., & Todorova, N. (2024). Forecasting volatility in the clean energy sector: The role of technology and oil prices. *Energy Economics*, *109*, 105964.

[21] Mallikarjuna, K. S., & Rao, D. N. (2017). Sectoral index volatility in the indian stock market: A garch approach. *Global Business Review*, *18*(3), 691–702.

[22] Mittnik, S., Robinzonov, N., & Spindler, M. (2015). Stock market volatility: Identifying major drivers and the nature of their impact. *Journal of Banking & Finance*, *58*, 1–14.

[23] Muth, M., Lingenfelder, M., & Nufer, G. (2024). The application of machine learning for demand prediction under macroeconomic volatility: A systematic literature review. *Man- agement Review Quarterly*, 1–44.

[24] Niu, Z., Demirer, R., Suleman, M. T., Zhang, H., & Zhu, X. (2023). *Do industries predict stock market volatility? A high frequency perspective based on a machine learning approach* (tech. rep.). SSRN.

[25] Nybo, C. (2021). Sector volatility prediction performance using garch models and artificial neural networks. *arXiv preprint*.
Palaniswamy, S. (2013). Modeling sectoral index volatility in the indian stock market using garch models. *International Journal of Economics and Finance*, *5*(2), 173–182.

[26] Ramos Perez, E., Alonso-Gonzalez, P. J., & Nunez-Velazquez, J. J. (2019). Forecasting volatility with a stacked model based on a hybridized artificial neural network. *Expert Systems with Applications*, *129*, 1–9.

[27] Tang, P., Xu, W., & Wang, H. (2024). Network-based prediction of financial cross-sector risk spillover in china: A deep learning approach. *The North American Journal of Economics and Finance*, *72*, 102151.
Taylor, S. J. (1986). *Modelling financial time series*. John Wiley & Sons.

[28] Teller, A., Pigorsch, U., & Pigorsch, C. (2022). *Short-to long-term realized volatility forecasting using extreme gradient boosting* (tech. rep.). SSRN.