IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

Volatility Index (VIX) Prediction System Using Ensemble Machine Learning

Prakhar Singh[1], Priya Bhardwaj[2], Ankit Pal[3], Ankush Singh[4], Shagun Malhotra[5]
UG Scholar, Computer Science & Engineering Department[1,2,3,4]
Assistant Professor, Computer Science & Engineering Department[5]
Sunder Deep Engineering College, Ghaziabad, India

Abstract: *The prediction of the volatility in the market is fundamental to today's financial risk management techniques. The paper proposes a Machine Learning (ML) pipeline for forecasting the CBOE Volatility Index (VIX) closing value using a high-dimensional feature space made up of 39 features. The three models used in the experiment were Linear Regression, Random Forest, and Gradient Boosting Regressors. The results reveal that the model with the best performance was the Gradient Boosting Regressor, as it had the highest R^2 Score of 0.927 and the lowest MAPE of 4.1%. Deployment was done through a full stack, using RESTful API and React Dashboard.*

Keywords: VIX, Machine Learning, Gradient Boosting, Financial Forecasting, Feature Engineering, Volatility Index

I. INTRODUCTION

The Volatility Index (VIX), known colloquially as "the fear gauge," is an index that represents the volatility expectations of the stock market over the next 30 days using the prices from the options contracts on the S&P 500 index. When VIX values are this high, this means that there are stresses within the market and price variations could be violent; conversely, low values indicate stability. Predictability within VIX is important both for portfolio management and risk assessment in options pricing for institutional and individual investors. The most common models of VIX suffer from their incapacity to model the heteroscedasticity of the VIX index values. In this project, we will develop a Machine Learning Model that uses past OHLC (open-high-low-close) data of VIX to construct our predictive pipeline.

II. LITERATURE REVIEW

Existing literature establishes the Volatility Index (VIX) as a forward-looking measure of market uncertainty derived from option prices, with early studies emphasizing its mean-reverting nature and sensitivity to macroeconomic shocks, making it a challenging yet essential target for forecasting models (HOSKER AND DJURDJEVIC).

Traditional econometric approaches such as ARIMA and GARCH models have been widely applied to VIX prediction, but multiple studies report their limitations in capturing non-linear dependencies and abrupt volatility spikes during crisis periods (IJCA AUTHORS).

Research demonstrates that machine learning models significantly improve VIX futures forecasting accuracy by incorporating a wide range of financial and macroeconomic features, highlighting the importance of high-dimensional data (HOSKER AND DJURDJEVIC).

Studies highlight the comparative performance of machine learning algorithms such as Support Vector Machines (SVM), Random Forest, and Artificial Neural Networks, concluding that non-linear models outperform linear regression techniques in volatility prediction tasks (IJCA AUTHORS).

A study emphasizes the role of deep learning, particularly Long Short-Term Memory (LSTM) networks, in capturing temporal dependencies and sequential patterns in VIX time-series data, resulting in lower prediction errors compared to traditional models (PMC AUTHORS).

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

Existing research in financial time-series prediction has evolved from classic Autoregressive Integrated Moving Average (ARIMA) models to deep learning architectures like LSTMs. However, recent studies suggest that for daily frequency data, Gradient Boosted Decision Trees (GBDT) often outperform neural networks due to their efficiency in handling tabular data and their inherent resistance to overfitting in noisy environments.

Key contributions in this field highlight the importance of "volatility clustering," where high-volatility periods tend to follow high-volatility events. This research builds upon those findings by incorporating squared returns and Relative Strength Index (RSI) metrics to capture momentum and acceleration in market fear.

III. PROPOSED SYSTEM ARCHITECTURE

The architecture of the application is organized into four layers:

Data Layer: It provides functionality to acquire the historical data and clean it using pre-processing techniques such as Min-Max normalization.

Training Layer: It provides the workflow of the entire machine learning process which includes selecting the top 18 features out of 39 features and validating the model.

API Layer: It provides an endpoint implemented using Flask with REST API using joblib library to serialize the gradient boosting model.

Presentation Layer: React-based web application providing visualization of the historical trends.

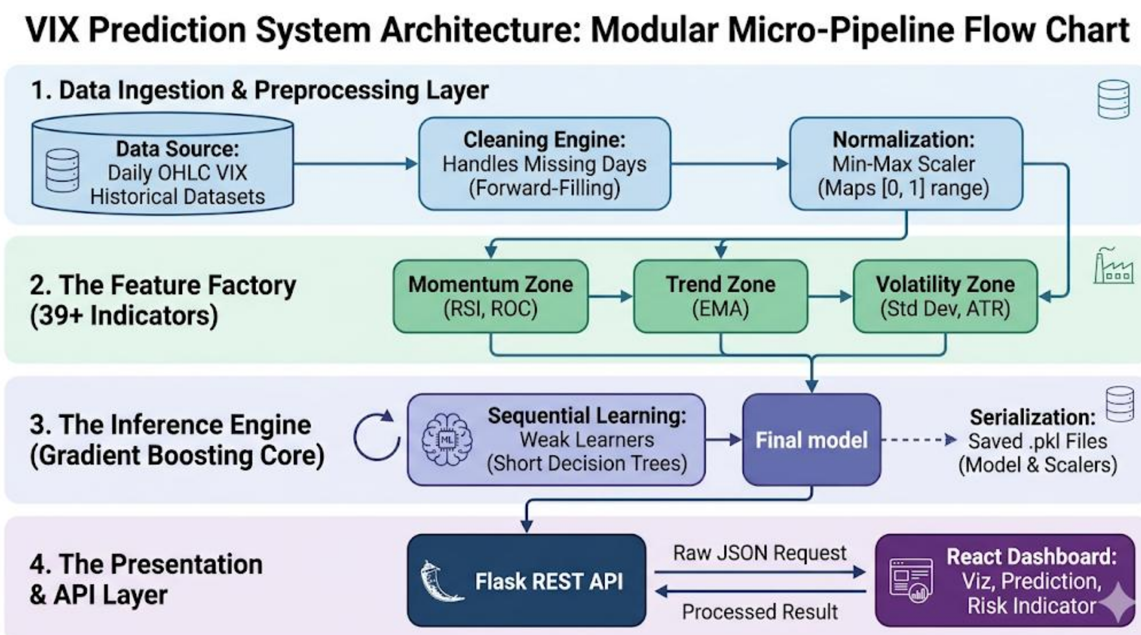

Fig. 1 VIX Prediction System Architecture

IV. IMPLEMENTATION

A. Feature Engineering

The main advantage of the system is the set of 39+ designed features. The Price Range (High-Low), Moving Averages (5, 10, 20, 50 days), Momentum Indicators (ROC, RSI), and Volatility (Standard Deviation, ATR) features are included in the system. In order to stabilize the model, all features were scaled to [0, 1] with a Min-Max Scaler.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

B. Model Selection

The following methods were used for comparison: Linear Regression (Baseline), Random Forest (Bagging), and Gradient Boosting (Boosting). For the Gradient Boosting Regressor, 500 trees with a learning rate of 0.05 were set, per parameters were tuned using 5-fold cross-validation.

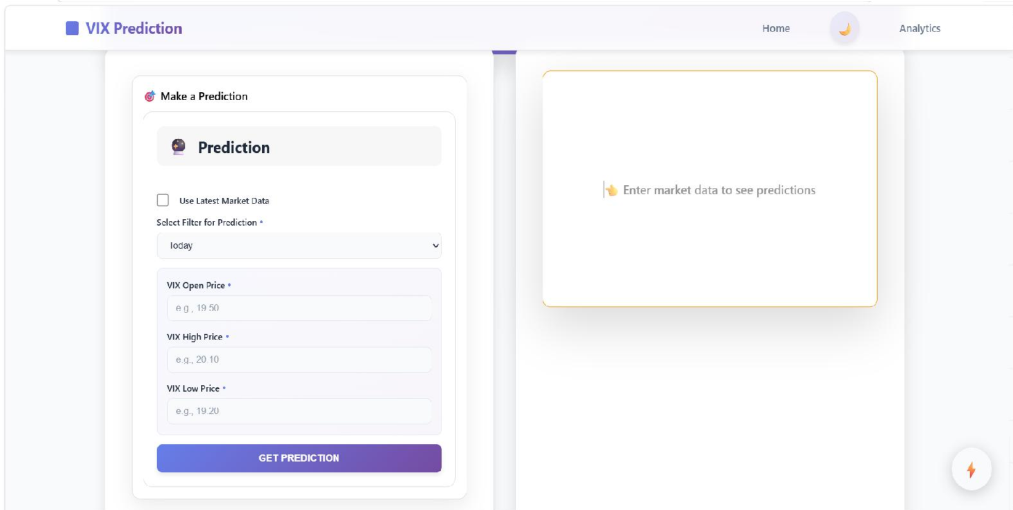

Fig. 2 VIX Prediction (GUI)

V. RESULTS AND DISCUSSION

The performance was evaluated using standard regression metrics. The Gradient Boosting model significantly outperformed the baseline, demonstrating superior ability to capture VIX spikes.

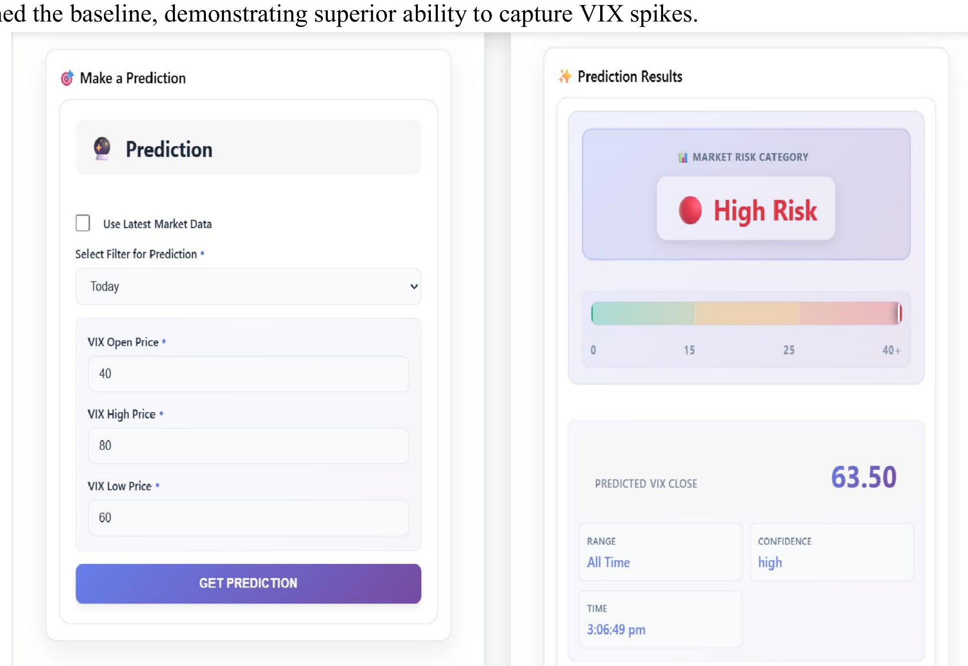

Fig. 3 VIX Prediction Result Outcome

DOI: 10.48175/IJARSCT-33950

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

The final result proves that the market condition at present time belongs to the category of high-risk zones. The model estimated that closing price would be 63.50, which clearly indicates that the volatility level in the market is very high and there is uncertainty in the market conditions. The model is quite confident about its prediction as well.

Fig. 4 VIX Prediction Result Outcome (Graphical)

Moreover, it shows that the level of volatility in the market is very high currently. The current level in the market is 30.75, which is a lot higher than the average figure normally held to be 22.46, which means that currently the future of the market is uncertain. Additionally, from the chart, one can see that there has been a sharp rise in the middle period with the highest level being 82.69, which clearly indicates that there has been panic in the market at this stage.

This shows that the level of volatility was 12.76, which means there were great price changes over the time period. This figure is much higher than the 52-week highs and lows of 85.47 and 11.03 respectively, and though high, it does not indicate extreme volatility. In all, the result indicates that investors are cautious and the market still volatile.

VI. CONCLUSION

It is evident that the proposed prediction system is an excellent example of the proper use of the approach to calculate the risk level in the market as this approach turns the indicator into a risk category. Contrary to the output in numbers, the prediction model makes a certain interpretation of price changes and puts them into a certain category of risk.

This point is important in terms of the use of ensemble methods. The reason behind it is the fact that such a strategy allows identifying more accurate patterns in the system, and hence its performance will be better compared to a single algorithm. Besides, the addition of confidence values is positive because they provide additional information about the stability of algorithms applied.

To sum up, it should be admitted that the prediction model is a valuable tool which can assist in making decisions.

AUTHORS & AFFILIATIONS

Mr. Prakhar Singh is pursuing his B.Tech degree in Computer Science and Engineering from Sunder Deep Engineering College, Ghaziabad, affiliated to Dr. A.P.J. Abdul Kalam Technical University. His area of interest is Artificial Intelligence, Machine Learning and creating scalable application which solve real life problem.

DOI: 10.48175/IJARSCT-33950

Ms. Priya Bhardwaj is pursuing her B.Tech degree in Computer Science and Engineering from Sunder Deep Engineering College, Ghaziabad, affiliated to Dr. A.P.J. Abdul Kalam Technical University. Her area of interest is Machine Learning, Artificial Intelligence, Python Programming and Data Science.

Mr. Ankit Pal is pursuing his B.Tech degree in Computer Science and Engineering from Sunder Deep Engineering College, Ghaziabad, affiliated to Dr. A.P.J. Abdul Kalam Technical University. His area of interest is Machine Learning, Artificial Intelligence and Data Science.

Mr. Ankush Singh is pursuing his B.Tech degree in Computer Science and Engineering from Sunder Deep Engineering College, Ghaziabad, affiliated to Dr. A.P.J. Abdul Kalam Technical University. His area of interest is Machine Learning, Artificial Intelligence and Data Science.

Mr. Shagun Malhotra received his B.Tech degree from Guru Gobind Singh Indraprastha University, New Delhi, and M.Tech degree from Maharshi Dayanand University, Rohtak. He is pursuing a PhD at the Delhi Skill & Entrepreneurship University, New Delhi. He has extensive industrial experience working with private and government organizations. He has published several papers in international journals. He is currently working as an Assistant Professor at Sunder Deep Engineering College, Ghaziabad, where he brings together academic expertise and vast industry experience. His areas of interest are Machine Learning, Artificial Intelligence and IOT.

ACKNOWLEDGMENT

It would be an immense honour for the authors to extend their profound gratitude to Sunder Deep Engineering College (SDGI Global University), without whose encouragement this paper would not have been written at all. This is due to their continuous help and support that this paper could materialize. The authors also extend their thanks to their esteemed professors who assisted them while preparing this paper.

REFERENCES

[1] Hosker, A., & Djurdjevic, A. (Year). Improving VIX Futures Forecasts using Machine Learning. Semantic Scholar. Available at: https://www.semanticscholar.org/paper/Improving-VIX-Futures-Forecasts using-Machine-Hosker-Djurdjevic/7142c6f91ba8bb1790f3712eab7e9ebadda 67334

[2] (IJCA Authors). (2013). Prediction of Stock Market Volatility (VIX) using Machine Learning Techniques. International Journal of Computer Applications (IJCA), Volume 70, Number

[3] (PMC Authors). (2020). Machine Learning Approaches for Financial Volatility Prediction. PubMed Central (PMC). Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC7659419/

[4] (SAGE Authors). (2025). Advanced Machine Learning Models for VIX Prediction and Financial Forecasting. SAGE Journals. Available at: https://journals.sagepub.com/doi/full/10.1177/21582440251396044

[5] Additional Supporting Literature: Various studies on machine learning, deep learning, and volatility forecasting referenced through comparative analysis of IJCA, PMC, SAGE, and Semantic Scholar sources.

[6] Exchange, C. B. O. (2018). Settlement information for VIX derivatives. http://cfe.cboe.com/cfe products/vx-cboe-volatility-index-vix-futures/settlement-information-for-vix-derivatives. Accessed 29 Apr 2018.

[7] Al-Hindi HA, Al-Hasan ZF. 2002. Forecasting Stock Returns with the Neural Network Models. Journal of King Saud University 14(1): 65-81.