Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 10, NCREHAASTE - 2026

# PredictWise: An Intelligent Machine Learning Framework for Long-Term Stock Stability Prediction

Pradeep Kumar Gupta[1], Boby Saini[2], Harit Gaur[3], Amit Kumar[4], Vishal Jain[5]
*CCSIT, Teerthanker Mahaveer University, Moradabad*
*Corresponding Author Email: sainiboby08@gmail.com*

***Abstract*—This study presents PredictWise, a data-driven stock prediction platform designed to identify stable stocks for long-term investment over 10–20 years. The study analyzes 50 NIFTY50 companies using two decades of historical market data through data preprocessing, transformation, and exploratory analysis. Stock volatility, measured using standard deviation, is used as a key indicator of stability, with stocks showing consistently lower volatility classified as stable. Several machine learning models, including Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, and Logistic Regression, were evaluated. A hybrid Liquid Neural Network (LLN) and Random Forest model achieved the strongest results across accuracy, precision, recall, F1-score, and ROC-AUC. A Stream lit-based interface enables users to select stock symbols and obtain stability predictions using current market data. The findings highlight the potential of hybrid machine learning models for stock stability prediction and data-driven investment analysis.**

***Index Terms*—Standard Deviation (SD), Machine Learning (ML), Random Forest (RF, Support Vector Machine (SVM, Decision Tree (DT), Liquid Neural Network (LNN), Area under Curve (ROC-AUC).**

## I. INTRODUCTION

This paper describes the data, learning models, stability criterion, experimental workflow, and Streamlit implementation used by PredictWise.

### *A. Indian Stock Market and Long-Term Investment Trends*

India's equity market has become an important avenue for wealth creation, while NIFTY50 represents major firms from sectors such as IT, banking, pharmaceuticals, infrastructure, and FMCG. A 10–20 year horizon emphasizes consistency rather than short-term price movement. PredictWise therefore focuses on volatility-based stability using standard deviation and combines financial indicators with ML/DL models. The principal market variables are:

This paper presents a solution combining financial analytics, ML/DL techniques, and a user-friendly web interface. Evaluating stock stability requires understanding key market indicators:

- Open Price: First traded price of the day and an indicator of early market direction.
- Close Price: Final traded price of the day and a primary input for technical analysis.
- Volume: Number of shares traded; it reflects market activity and liquidity.

### *B. Machine Learning (ML)*

Machine learning learns patterns from historical observations and applies them to unseen data. In this study it is used to classify stocks from price, volume, and technical features.

### *C. Types of Machine Learning*

- Supervised Learning: Uses labeled examples for classification or regression.
- Unsupervised Learning: Uses unlabeled data to discover groups, patterns, or reduced representations, for example through clustering or PCA.

### *D. Types of Supervised ML Algorithms Used*

- Random Forest: An ensemble of decision trees that combines their outputs. Random feature and sample selection reduces variance and overfitting, while feature importance supports interpretation. It is useful for nonlinear financial patterns.

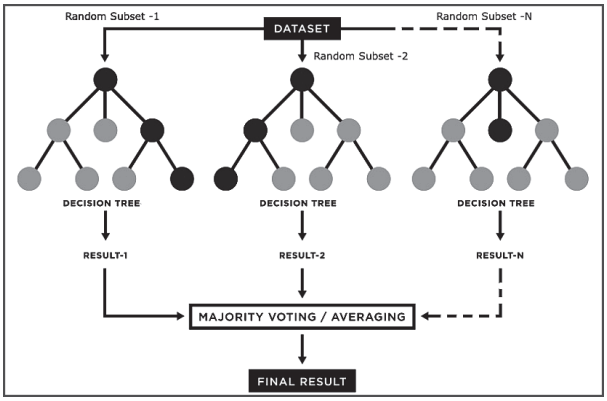



Fig. 1. Diagram of Random Forest

- Decision Tree: Splits observations through feature-based rules until a prediction is reached. It is easy to interpret but can over fit when trees become too deep.

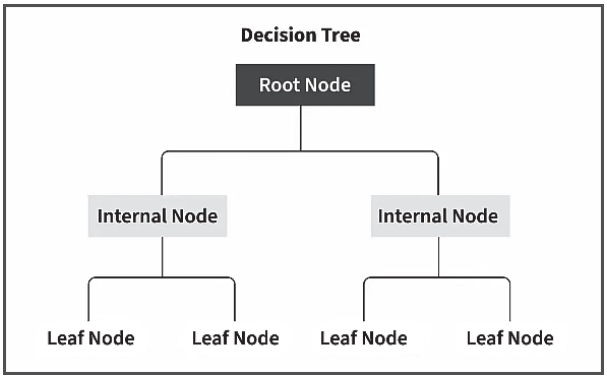



Fig. 2. Diagram of Decision Tree

- Logistic Regression: Estimates the probability of a binary class through the sigmoid function. It is computationally efficient and interpretable, although it has limited ability to represent complex nonlinear interactions.



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 10, NCREHAASTE - 2026

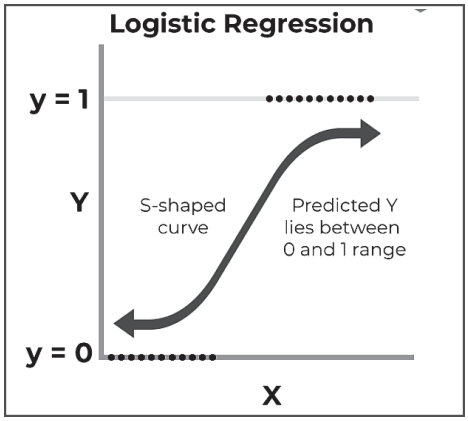



Fig. 3. Diagram of Logistic Regression

- Support Vector Machine (SVM): Finds a separating hyper-plane with maximum margin between classes. Kernel functions can model nonlinear boundaries; SVM is effective for structured, high-dimensional data.

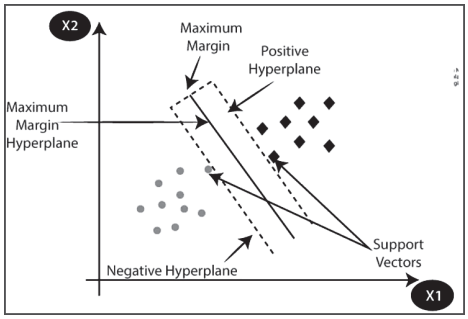



Fig. 4. Diagram of Support Vector Machine

- K-Nearest Neighbors (KNN): Classifies an observation from the labels of its nearest training points. It is simple and non-parametric but is sensitive to feature scaling, the value of k, and dataset size.

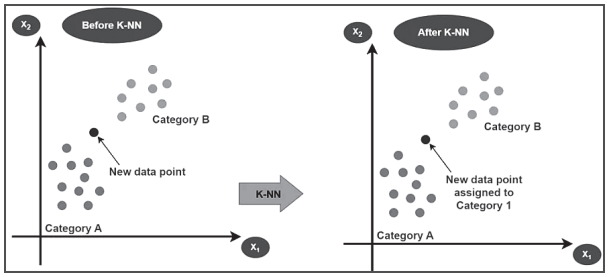



Fig. 5. Diagram of K-Nearest Neighbor

These models provide complementary baselines: trees offer interpretability, RF provides robustness, SVM handles high-dimensional boundaries, LR provides a probabilistic baseline, and KNN captures local similarity.

*E. Deep Learning (DL)*

Deep learning uses multilayer neural networks to learn complex nonlinear representations. It can model sequential financial patterns but generally requires more data and computational resources than conventional ML.

- Liquid Neural Networks (LNNs): LNNs use dynamic neuron states that evolve with incoming data, making them suitable for time-dependent problems. Here, LNN is paired with RF so temporal information and ensemble-based classification are combined.

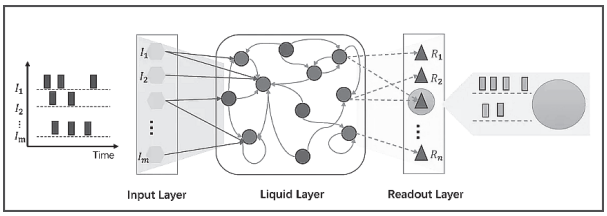



Fig. 6. Diagram of Liquid Neural Network

The LNN component was introduced to represent temporal changes that may be difficult for static classifiers to capture. Its output is combined with RF to improve the reported stability-classification performance and support practical deployment.

## II. LITERATURE REVIEW

Recent stock-forecasting research has moved from conventional statistical models toward recurrent, attention-based, transformer, and hybrid architectures. LSTM-based studies have reported advantages for sequential prediction and comparisons with ARIMA [1]. NARX models have also been used to represent nonlinear market dynamics [2]. These studies establish the value of temporal learning for financial series.

Research has also examined economic predictors and changes in their predictive behavior over time [3]. More recent hybrid architectures combine BiLSTM, Transformer, and TCN components and report improved forecasting measures on Chinese markets [4]. Market-style clustering with technical and sentiment information has likewise been used to improve prediction performance [5].

Hybrid learning remains a major direction. LSTM combined with Bayesian-optimized LightGBM has been investigated for stock-price changes [6], while graph-based LSTM methods model relationships among stocks for high-frequency settings [7]. Wavelet denoising with Extreme Learning Machine has been proposed to reduce short-term noise before trend prediction [8].

Other studies report improvements from GRU ensembles [9], optimized Random Forest models [10], frequency-decomposition GRU-Transformer networks [11], and multi-scale attention-based LSTM [12]. Attention-enhanced LSTM [13] and xLSTM [14] further target long-range dependencies.

Overall, the literature shows strong interest in temporal learning, ensemble methods, feature engineering, denoising, attention, and sentiment. PredictWise differs in emphasis by defining stability through long-term volatility and classifying NIFTY50 stocks rather than directly forecasting a future price.

Recent work continues to improve stock forecasting through frequency decomposition, attention mechanisms, and extended recurrent architectures [11]–[14]. These methods target noise reduction and long-range dependencies, showing the broader shift toward hybrid and adaptive sequence models.

Random Forest and related ensemble methods have also been used for medium- and long-term trend prediction [10], while recent studies combine neural models, sentiment features, and



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 10, NCREHAASTE - 2026

explainability techniques for financial decision support [15], [16].

The literature indicates that no single model is consistently sufficient for all market conditions. PredictWise therefore combines a temporal LNN with RF and focuses specifically on long-term stability classification rather than direct price forecasting[17-21].

## III. METHODOLOGY

PredictWise follows six linked stages: data acquisition, preprocessing, stability definition, feature construction, model training, and deployment. The workflow preserves chronological order so that the evaluation better reflects a forecasting setting.

### *A. Data Acquisition*

Historical data for 50 NIFTY50 companies covering 2004–2024 were collected from financial APIs and market-data providers. Daily open, high, low, close, volume, and derived VPA measures were organized on a common date-time basis.

### *B. Market Analysis*

Market analysis considered beta, drawdown, return variance, and sector behavior. Standard deviation of closing prices was selected as the main stability indicator because it represents long-term price dispersion.

### *C. Identifying Stability Parameters*

For each stock, 20-year closing-price volatility was calculated. The 15 stocks with the lowest deviations formed the reference group; their mean deviation was used as the threshold. Values at or below the threshold were labeled Stable (1), and the remaining stocks Unstable (0).

### *D. Data Preprocessing*

Preprocessing addressed missing observations, outliers, inconsistent formats, temporal alignment, scaling, and label generation.

- Missing values: isolated gaps were handled with forward/backward filling; longer gaps were interpolated where appropriate. Stocks with substantial irregular missing data were excluded.
- Outliers: Z-score and IQR checks were used. Genuine market events were retained, while data-entry anomalies were corrected or removed.
- Standardization: dates were converted to a common date time format and numeric fields to consistent types; invalid symbols and null placeholders were handled.
- Alignment: a common trading calendar was used so that stock series could be compared consistently across time.
- Scaling: Min-Max scaling and Z-score normalization were applied when required, with fitting performed on training data to reduce leakage.
- Label mapping: each stock received a binary stability label from the volatility threshold, which became the supervised-learning target.

### *E. Exploratory Data Analysis*

EDA examined price and volume series, VPA measures, correlations, and rolling volatility. The analysis was used to identify trends, co-movement, volatility changes, and patterns relevant to feature construction.

### *F. Feature Engineering*

Features included SMA, RSI, Bollinger Bands, rolling standard deviations over 30, 90, and 180 days, and VPA measures. These variables represent trend, momentum, price dispersion, and price–volume interaction.

### *G. Defining Stability*

Stability was defined from 20-year closing-price standard deviation. The mean deviation of the 15 lowest-volatility stocks formed the benchmark; stocks at or below it were Stable and the rest Unstable. This makes the target reproducible and directly tied to the study's volatility criterion.

### *H. Data Transformation*

Rolling windows converted time-series observations into feature-label records. Chronological train–test splitting reserved recent observations for testing, and no random shuffling was used. Scaling was applied according to model requirements.

### *I. Model Training*

RF, KNN, SVM, LR, and DT were trained on price, volume, and technical features. Grid search and cross-validation were used for model selection and generalization. LNN was trained on time-windowed inputs to capture temporal behavior.

The LNN output was combined with RF to form the proposed hybrid classifier. The reported results indicate that this ensemble provided the strongest balance of adaptability and classification performance among the evaluated models.

### *J. Model Evaluation*

Performance was assessed using accuracy, precision, recall, F1-score, and ROC-AUC. Time-aware validation was used to limit leakage. Precision and recall were considered alongside accuracy because the task contains two stability classes.

### *K. Model Deployment with Streamlit*

The Streamlit application lets users select a NIFTY50 stock and view market measures such as open, close, volume, and VPA. The trained LNN–RF model then returns a Stable/Unstable classification and a confidence value. SHAP integration is identified as a future enhancement for interpretability.


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 10, NCREHAASTE - 2026

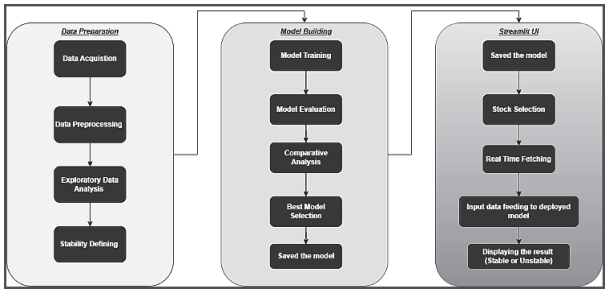



Fig. 7. Proposed Methodology of Predictwise

## IV. COMPARATIVE ANALYSIS

The models were compared using five measures: accuracy, precision, recall, F1-score, and ROC-AUC. Accuracy summarizes overall correctness; precision measures the proportion of predicted stable stocks that are actually stable; recall measures detection of actual stable stocks; F1 balances precision and recall; and ROC-AUC summarizes class discrimination across thresholds.

- **Accuracy**: Measures the overall correctness of the model by calculating the percentage of total predictions (both positive and negative) that were correct.
- **Precision**: Indicates how many of the stocks predicted as stable were actually stable. High precision reduces false positives.
- **Recall**: Represents the model's ability to identify all actual stable stocks. High recall reduces false negatives.
- **F1-Score**: The harmonic mean of precision and recall, balancing both metrics to give a single performance measure.
- **ROC-AUC Score**: Reflects the model's capability to distinguish between stable and unstable classes across various threshold levels. A higher AUC indicates better discriminatory power.

TABLE I
PERFORMANCE COMPARISON OF PROPOSED MODEL

| MODELS | ACCURACY | PRECISION (%) | RECALL (%) | F1-SCORE (%) | ROC-AUC SCORE |
|---|---|---|---|---|---|
| KNN | 81.45 | 79.12 | 76.80 | 77.94 | 0.83 |
| RN | 84.02 | 82.33 | 80.45 | 81.38 | 0.86 |
| DT | 83.20 | 81.25 | 78.60 | 79.96 | 0.85 |
| LR | 72.05 | 70.66 | 67.70 | 71.02 | 0.75 |
| SVM | 84.02 | 82.33 | 80.45 | 81.38 | 0.86 |
| **LNN+RN** | **95.88** | **94.42** | **91.35** | **90.33** | **0.95** |

## V. CONCLUSION

PredictWise presents a volatility-based framework for identifying relatively stable NIFTY50 stocks from 20 years of historical data. Standard deviation provides the main stability criterion, while preprocessing, technical features, chronological validation, and multiple ML/DL models form the analytical pipeline. Among the tested models, the LNN–RF hybrid achieves the strongest reported results across the evaluation measures.

The Streamlit implementation converts the research workflow into an accessible stability-analysis tool. Future work can extend the dataset, add macroeconomic and sentiment variables, improve explainability with SHAP, and investigate reinforcement-learning or other adaptive approaches. The system should be treated as an analytical aid rather than a guarantee of future investment performance.

## REFERENCES

[1] F. Qian and X. Chen, "Stock Prediction Based on LSTM under Different Stability," in 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Apr. 2019, pp. 483–486. doi: 10.1109/ICCCBDA.2019.8725709.

[2] "Stable Stock Market Prediction Using NARX Algorithm | Proceedings of the 2018 International Conference on Computing and Big Data." Accessed: Apr. 20, 2025. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3277104.3277120

[3] B. S. Paye and A. Timmermann, "How Stable are Financial Prediction Models? Evidence from US and International Stock Market Data," Jun. 2002, Accessed: Apr. 20, 2025. [Online]. Available: https://escholarship.org/uc/item/74v515fr

[4] S. Wang, "A Stock Price Prediction Method Based on BiLSTM and Improved Transformer," IEEE Access, vol. 11, pp. 104211–104223, 2023, doi: 10.1109/ACCESS.2023.3296308.

[5] X. Li and P. Wu, "Stock Price Prediction Incorporating Market Style Clustering," Cogn. Comput. vol. 14, no. 1, pp. 149–166, Jan. 2022, doi: 10.1007/s12559-021-09820-1.

[6] L. Tian, L. Feng, L. Yang, and Y. Guo, "Stock price prediction based on LSTM and LightGBM hybrid model," J. Supercomput., vol. 78, no. 9, pp. 11768–11793, Jun. 2022, doi: 10.1007/s11227-022-04326-5.

[7] T. Yin, C. Liu, F. Ding, Z. Feng, B. Yuan, and N. Zhang, "Graph-based stock correlation and prediction for high-frequency trading systems," Pattern Recognit., vol. 122, p. 108209, Feb. 2022, doi: 10.1016/j.patcog.2021.108209.

[8] D. Wu, X. Wang, and S. Wu, "A Hybrid Method Based on Extreme Learning Machine and Wavelet Transform Denoising for Stock Prediction," Entropy, vol. 23, no. 4, Art. no. 4, Apr. 2021, doi: 10.3390/e23040440.

[9] Z. He, H. Zhang, and L. Y. Por, "A Comparative Study on Deep Learning Models for Stock Price Prediction," in Image Processing, Electronics and Computers, IOS Press, 2024, pp. 540–551. doi: 10.3233/ATDE240502.

[10] L. Yin, B. Li, P. Li, and R. Zhang, "Research on stock trend prediction method based on optimized random forest," CAAI Trans. Intell. Technol., vol. 8, no. 1, pp. 274–284, 2023, doi: 10.1049/cit2.12067.

[11] C. Li and G. Qian, "Stock Price Prediction Using a Frequency Decomposition Based GRU Transformer Neural Network," Appl. Sci., vol. 13, no. 1, Art. no. 1, Jan. 2023, doi: 10.3390/app13010222.

[12] X. Teng, X. Zhang, and Z. Luo, "Multi-scale local cues and hierarchical attention-based LSTM for stock price trend prediction," Neurocomputing, vol. 505, pp. 92–100, Sep. 2022, doi: 10.1016/j.neucom.2022.07.016.

[13] S. Sang and L. Li, "A Novel Variant of LSTM Stock Prediction Method Incorporating Attention Mechanism," Mathematics, vol. 12, no. 7, Art. no. 7, Jan. 2024, doi: 10.3390/math12070945.

[14] X. Fan, C. Tao, and J. Zhao, "Advanced Stock Price Prediction with xLSTM-Based Models: Improving Long-Term Forecasting," in 2024 11th International Conference on Soft Computing & Machine Intelligence (ISCMI), Nov. 2024, pp. 117–123. doi: 10.1109/ISCMI63661.2024.10851608.

[15] J. Zheng, D. Xin, Q. Cheng, M. Tian, and L. Yang, "The Random Forest Model for Analyzing and Forecasting the US Stock Market in the Context of Smart Finance," Feb. 27, 2024, arXiv: arXiv:2402.17194. doi: 10.48550/arXiv.2402.17194.

[16] S. Deng, Y. Zhu, Y. Yu, and X. Huang, "An integrated approach of ensemble learning methods for stock index prediction using investor sentiments," Expert Syst. Appl., vol. 238, p. 121710, Mar. 2024, doi: 10.1016/j.eswa.2023.121710.

[17] Gupta, Mohan Vishal, Dwivedi, Rakesh Kumar & Kumar, Anil(2023) Enhancing remote sensing satellite image classification through parameter optimization in hybrid CNN-LSTM model, Journal of Discrete Mathematical Sciences and Cryptography, 26:7, 2061–2079, DOI: 10.47974/JDMSC-1855



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 10, NCREHAASTE - 2026

[18] Gupta, Mohan Vishal, Dwivedi, Rakesh Kumar & Kumar, Anil(2023) Impact analysis of convolutional neural network in classification of satellite imagery, Journal of Information and Optimization Sciences, 44:6, 1151–1166, DOI: 10.47974/JIOS-1453

[19] M. V. Gupta, R. K. Dwivedi and A. Kumar, "Comparative Analysis of Various Statistical Learning Algorithms in Classification of Satellite Imagery," 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2023, pp. 1256-1263, doi: 10.1109/ ICTACS59847. 2023.10390273

[20] M. V. Gupta, R. K. Dwivedi and A. Kumar, "Review of Various Learning Algorithms Applied to Satellite Image Classification," 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), MORADABAD, India, 2021, pp. 285-289, doi: 10.1109/SMART52563.2021.9676215

[21] Haripriya, Mohan Vishal Gupta, Nikita Nadkarni, Suraj Malik, Aditya Yadav, and Apoorva Joshi. 2024. Enhanced radial base deep learning algorithm for predicting multimedia security issues. J. Intell. Fuzzy Syst. 46, 2 (2024), 4829–4840. https://doi.org/10.3233/JIFS-237041



(This work is licensed under a Creative Commons Attribution 4.0 International License.)