

Artificial Intelligence-Driven Financial Risk Prediction and Portfolio Optimization: A Machine Learning Approach from the Indian Equity Market

Dr. M A Nayeem

Assistant Professor, Department of Finance and Analytics, Institute of Public Enterprise, Hyderabad, Telangana, India.

Abstract:

As financial markets become more complex, data-driven risk identification and portfolio management becomes the focus for the future. In this study we have developed an integrated machine learning (ML) framework to predict financial risk and portfolio optimization in the Indian stock market. Using 32 NSE-listed large-cap securities in eight sectors during 2021–2025, the study constructs 14 key financial risk profiles based on volatility, momentum, drawdown, technical indicators, and macro-financial variables in India. ML models are used to classify forward 20-trading-day high-risk and forecast 20-day returns. Gradient Boosting and Neural Network classifiers perform best on the scale of 0.794 and 0.798 and F1-scores of 0.830 and 0.828 respectively. However, the regression models fail to outperform the naïve mean forecast, indicating that short-term returns are not yet well predicted. A portfolio comparison between equal-weighted, Markowitz mean-variance and ML-risk-tilted strategies shows that the Markowitz portfolio achieves the best risk-adjusted performance with a Sharpe ratio of 3.84 and Sortino ratio of 10.52. There is no significant difference in mean returns across sectors, while India VIX shows a weak positive relationship with realized volatility. The results show that ML can be a good early warning system on market risk and classical portfolio optimization is still a useful tool to construct a portfolio. The study therefore recommends the use of ML risk signals as a complementary tool to improve, instead of replace, traditional portfolio management strategies.

Keywords: Financial risk prediction, Machine learning, Portfolio optimization, Random Forest

1. Introduction

The Indian stock market has undergone a major transformation over the last decade with increasing retail involvement, rapid institutionalisation in the form of systematic and quantitative investment and the launch of algorithmic and AI-based trading systems on the NSE and BSE. The increase in macro-financial volatility is driving the market and it is causing high volatility in the INR/USD currency exchange and crude oil imports as the studyll as foreign institutional investor (FII) flow reversals in India and in this way, the risk profile of the Indian stock market has changed [1]. Traditional risk models and methods such as the Capital Asset Pricing Model (CAPM) and Markowitz mean-variance optimization and parametric VaR are based on linearity, normality and stationarity assumptions as these assumptions are not typically satisfied in emerging markets with a high fat tail, volatility clustering and regime shift [2]. Machine learning is a data-driven model that is able to model nonlinear dependencies, interaction effects and regime-dependent risk dynamics without requiring so many distributional assumptions. Increasingly, tree-based ensembles and neural models have been used, as the studyll as classical econometric models in credit risk scoring, bankruptcy prediction and volatility forecasting [3]. Hothe studyver, the evidence of the Indian market where the market microstructure, SEBI, and macro-financial links are quite different from the ones in a developed market is limited, fragmented across single model studies, and rarely combines risk prediction and downstream portfolio construction in a reliable, statistically-based pipeline [4]. This paper fills this gap by building a reproducible ML pipeline for equity risk classification and prediction using NSE-centric features, follothe studyd by benchmarking five ML classifiers and two regression models with out-of-sample and blocked time series cross-validation and finally integrating the risk signal into a portfolio optimization model that can be directly compared to the classic Markowitz benchmark using formal statistical tests AI contribution to Indian portfolio risk management [5].

1.1. Problem Statement

While ML applications are rapidly growing in global finance, Indian asset managers still rely on historical-variance-based risk models that are backward-looking and slow to adjust to regime shifts such as the 2020 COVID crash, the 2022 rate-hike cycle or episodic FII-driven sell-offs [6]. There is limited peer-reviewed study, India-specific evidence on (a) which ML algorithms are better suited to short-horizon risk classification versus return forecasting, (b) which feature classes drive predictive power, and (c) whether ML-derived risk signals translate into statistically significant portfolio-level improvements once transaction realism and out-of-sample discipline are imposed.

1.2. Contribution of the Study

This paper contributes to the literature in four ways. First, it constructs an India-calibrated feature set that accounts for India VIX, INR/USD, Brent crude and FII flow proxies that are rarely considered in general ML-finance studies that are based on US data. Second, it benchmarks five classifiers and two regressors in a leak-free and time-ordered manner and shows confusion matrices, AUC-ROC and blocked cross-validation stability which is not typically described in applied ML-finance papers. Third, it is one of the few papers that relates a study-validated risk classification to portfolio construction and also compares the resulting portfolio with Markowitz and an equal-weighted asset benchmark using paired-sample inference, rather than simply by describing the algorithm. Fourth, it gives a rigorous set of results that is transparent and, in cases where the study does not find that our ML model outperforms classical benchmarks, which is necessary for responsible AI in risk management and other enterprise risk functions.

2. Literature Review

2.1. Classical Portfolio Theory and Its Limitations

The mean-variance framework remains a very important concept in today's portfolio theory and has been extended with asset-pricing and multi-factor models. However, the traditional models in this field are typically based on assumptions like a normally distributed return and a stable relationship among assets [8]. In reality, especially in the Indian stock market, returns typically show skewness, excess kurtosis, volatility clustering, and changing correlations. These limitations have led to the need for more flexible models that can account for complex and nonlinear market behavior [9]. Machine learning is one such model that adapts to changing data and helps us to identify relationships that are not captured in standard portfolio models.

2.2. Machine Learning in Financial Risk Prediction

In research, machine learning has gained growing applications in financial risk assessment and portfolio management. Tree-based models such as decision trees and CART have been used for risk classification and have shown advantages over traditional linear approaches [10]. Also, in asset pricing work, ensemble and neural network models have shown how to capture nonlinear relationships between financial characteristics and to improve out-of-sample predictions [11]. Deep learning methods such as autoencoders are also being explored more to extract underlying factors to build a portfolio. In recent research, it has been found that ML and traditional portfolio theory can be integrated into portfolio optimization frameworks to provide a more integrated return and risk estimate. These developments provide the basis for exploring an integrated ML approach to risk prediction and portfolio optimization in the Indian stock market [12].

2.3. Indian Empirical Evidence

In the Indian context, several studies have used machine learning methods such as Random Forest, Support Vector Machines, and LSTM networks to predict stock prices, Nifty and Sensex movements, and market returns. In general, they are moderately accurate in predicting the direction of market movements, which is in a similar way to the risk-classification performance of the present work [13]. These papers have also explored hybrid models for market volatility prediction and ensemble approaches for credit risk assessment in Indian financial institutions. However, much of the Indian research is based on a single ML method instead of a set of ML models [14]. Also, the macro-financial indicators of India such as India VIX and FII flows are mainly neglected.

More importantly, few studies have linked ML-based risk prediction with an out-of-sample portfolio optimization framework [15]. Here the study bridges these gaps by combining multiple ML models, India-specific financial indicators, risk classification, return forecasting, and portfolio optimization in a single framework.

2.4. Research Gap

Three important gaps are identified in the existing literature: first, limited research has provided a statistically rigorous comparison of multiple ML models for Indian equity risk classification using India-specific macro-financial variables; second, there is limited evidence on whether short-term return forecasting is as feasible as forecasting market risk or volatility regimes; and third, few studies have formally tested whether ML-based portfolio strategies significantly outperform the classical Markowitz approach using inferential statistical methods rather than descriptive comparisons.

3. Research Objectives and Hypotheses

Research Objectives

RO1: To predict 20-day high-risk regimes using ML classifiers.

RO2: To forecast 20-day returns using ML regression.

RO3: To identify important technical and macro-financial features.

RO4: To compare ML, Markowitz, and equal-the studyighted portfolios.

Null Hypotheses

H01: ML classifiers do not differ significantly in performance.

H02: ML forecasts do not differ significantly from the naïve benchmark.

H03: Sector returns do not differ significantly.

H04: Portfolio returns do not differ significantly.

4. Data and Methodology

4.1. Sample and Data Description

The study includes 32 large-cap NSE-listed stocks from eight sectors (Banking & Financial Services, IT, Energy & Oil-Gas, FMCG, Automobile, Pharma, Metals & Mining, and Infrastructure & Cement) from January 2021 to October 2025. The dataset contains 29,472 stock-day observations after feature creation and target adjustment. Stocks the studyre selected based on market capitalization and liquidity and NSE sector classification. The same methodology can be applied to real NSE/BSE, Bloomberg, or Refinitiv data as the studyll by replacing the data source.

4.2. Feature Engineering

Fourteen predictor variables the studyre engineered from raw price, volume and macro series, **Table 1. Feature Set for Risk Classification and Return Forecasting**

Feature	Description	Category
mom_5 / mom_20	5-day and 20-day price momentum (% change)	Technical
vol_20 / vol_60	20-day and 60-day annualized realized volatility	Risk/Technical
ma_ratio	Price deviation from 50-day moving average	Technical
rsi	14-day relative strength index proxy	Technical
vol_chg	20-day change in traded volume	Liquidity

skew_20 / kurt_20	20-day rolling skewness and excess kurtosis of returns	Distributional risk
max_dd_60	60-day rolling maximum drawdown	Risk
india_vix	India VIX level (implied volatility index)	Macro-financial
usdinr_ret	Daily INR/USD return	Macro-financial
crude_ret	Daily Brent crude return	Macro-financial
fii_flow	Standardized daily net FII flow proxy	Macro-financial

Source: Author's construction. All features are computed using strictly past information to predict forward 20-trading-day-ahead targets, preventing look-ahead bias.

4.3. Target Variables

The study builds two targets. (a) A binary high-risk indicator (1 if the forward 20-day realized (annualized) volatility is above 30% (roughly the 73.7th percentile of the full-sample volatility distribution) and 0 otherwise, for classification. (b) A continuous forward 20-day cumulative return (as used to forecast return from regression). Both targets are calculated strictly from future information relative to the date of feature observation and are excluded from the feature set.

4.4. Machine Learning Models

Five classifiers are benchmarked: (i) Logistic Regression (L2-regularized, class-balanced), (ii) Support Vector Machine with RBF kernel, (iii) Random Forest (400 trees, max depth 8, minimum leaf size 25), (iv) Gradient Boosting (300 estimators, learning rate 0.05, max depth 3), and (v) a Multilayer Perceptron with hidden layers of 32 and 16 neurons (early stopping). Linear Regression and a Random Forest Regressor are compared for return prediction. All features are standardized (zero mean, unit variance) using parameters fit only on the training partition. Class imbalance (73.7% high-risk vs. 26.3% low-risk) is addressed by class the studying it rather than resampling to maintain the natural temporal order of observations needed for time-series validation.

4.5. Validation Design

In order to avoid look-ahead bias in random k-fold cross-validation on time-series data, the sample is split chronologically: the first 80% of the data form the training partition and the last 20% are strictly out-of-sample test data. The robustness is further tested with blocked (non-shuffled) TimeSeriesSplit cross-validation with three sequential folds which report the mean and standard deviation of accuracy, F1-score and AUC-ROC across folds.

4.6. Portfolio Optimization Framework

Three portfolio construction strategies are compared on the out-of-sample test window: (i) an Equal-The studyighted (1/N) portfolio, (ii) a Markowitz Mean-Variance Optimized (MVO) portfolio obtained by maximum-Sharpe search over the long-only simplex (8,000 Dirichlet-sampled candidate the studyight vectors, risk-free rate = 6.5% p.a., proxying the RBI repo-linked short-term rate), and (iii) a novel ML-Tilted portfolio, where the studyights are proportional to (1 - predicted probability of high risk from the Random Forest classifier) scaled by clipped 20-day momentum and re-normalized to sum to one. Portfolio return, annualized volatility, Sharpe ratio, historical 95% VaR, Conditional VaR (CVaR/Expected Shortfall), Sortino ratio, and maximum drawdown are calculated based on observed out-of-sample daily returns. Statistical differences betthe studyen portfolio daily return series are tested using paired-sample t-tests.

4.7. Statistical Tests Employed

The study used several methods for checking the model performance and to test the results. Classifier performance was evaluated in terms of accuracy, precision, recall, F1 score, AUC-ROC, and confusion matrices. The study tested the robustness of the models by performing three-fold blocked TimeSeriesSplit cross-validation and reported the results as mean $\pm$ standard deviation. The prediction performance was compared betthe studyen

Linear Regression and Random Forest Regression using the paired t-test on squared forecast errors. The study also used a one-way ANOVA in order to test whether the average returns across sectors the studyre different. Portfolio performance was compared using paired-sample t-tests on daily out-of-sample returns. Finally, Pearson correlation was used to examine the relationship betthe studyen India VIX and average realized volatility across stocks.

5. Results and Discussion

5.1. Descriptive and Sector-wise Statistics

This reports annualized return, volatility, skewness, kurtosis and Sharpe ratio by sector over the full sample period.

Table 2. Sector-wise Descriptive Statistics

Sector	Ann. Return	Ann. Volatility	Skewness	Excess Kurtosis	Sharpe Ratio
Automobile	11.67%	40.23%	0.05	0.12	0.13
Banking & Financial Services	-3.82%	37.10%	0.02	0.29	-0.28
Energy & Oil-Gas	-34.01%	42.51%	-0.01	0.29	-0.95
FMCG	8.56%	32.92%	0.02	0.08	0.06
Information Technology	-10.09%	33.69%	-0.09	0.22	-0.49
Infrastructure & Cement	4.86%	31.98%	-0.05	-0.10	-0.05
Metals & Mining	-7.49%	39.92%	-0.00	0.04	-0.35
Pharma	12.44%	37.64%	-0.09	0.29	0.16

A one-way ANOVA test of equal mean daily returns across the eight sectors confirms $F = 0.93$, $p = 0.483$, not rejecting H_0 at standard significance levels. This indicates that cross-sectional sector selection in the sample did not generate statistically distinguishable mean-return differentials and therefore the volatility-based ML model used in the present work is worth consideration. Energy & Oil-Gas has the worst risk-adjusted performance (Sharpe = -0.95) in a period of high pass-through crude oil volatility with Automobile and Pharma as the highest Sharpe ratios in cyclical and defensive sectors.

b. Correlation Structure

This presents the pairwise return correlation matrix for a representative cross-sector subset of large-cap stocks, illustrating the diversification benefit embedded in the Markowitz optimization step.

Table 3. Pairwise Daily Return Correlation Matrix

Ticker	HDFC BANK	TCS	RELIA NCE	ITC	MAR UTI	SUNPHAR MA	TATA STEEL	LT
HDFCBANK	1.00	0.20	0.21	0.22	0.24	0.24	0.46	0.44
TCS	0.20	1.00	0.12	0.08	0.15	0.16	0.20	0.20
RELIANCE	0.21	0.12	1.00	0.11	0.11	0.19	0.23	0.19
ITC	0.22	0.08	0.11	1.00	0.06	0.14	0.18	0.17
MARUTI	0.24	0.15	0.11	0.06	1.00	0.10	0.23	0.22

SUNPHARMA	0.24	0.16	0.19	0.14	0.10	1.00	0.29	0.26
TATASTEEL	0.46	0.20	0.23	0.18	0.23	0.29	1.00	0.38
LT	0.44	0.20	0.19	0.17	0.22	0.26	0.38	1.00

The correlations are moderate (0.06–0.46), with stronger relationships among rate-sensitive stocks and the studyaker relationships betthe studyen IT and domestic-consumption stocks, supporting the benefits of sectoral diversification.

5.2. Risk-Classification Model Performance

The performance for the five classifiers predicting the forward 20-day high-risk regime.

Table 4. Out-of-Sample Classification Performance

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0.689	0.920	0.621	0.741	0.800
Support Vector Machine (RBF)	0.682	0.914	0.616	0.736	0.783
Random Forest	0.683	0.924	0.609	0.734	0.798
Gradient Boosting	0.745	0.798	0.864	0.830	0.794
Neural Network (MLP)	0.745	0.804	0.853	0.828	0.798

Gradient Boosting and Neural Network (MLP) show the best overall performance (F1 = 0.830 and 0.828; Accuracy = 0.745 and 0.745, respectively) in terms of recall, with Logistic Regression, SVM and Random Forest performing worse than those two (0.86 and 0.85) that are more likely to correctly identify the high-risk periods, which is a key trait for an early-warning risk system where false negatives (missed risk events) are more expensive than false positives. Logistic Regression has the most accurate results (0.920) and a competitive AUC-ROC (0.800) so it is a relatively easy to interpret data that is also useful in regulatory or compliance settings where model transparency is important over just the small gains in prediction.

B. Confusion Matrices

Table 5. Confusion Matrices

Model	TN (0→0)	FP (0→1)	FN (1→0)	TP (1→1)
Logistic Regression	1425	228	1606	2629
Support Vector Machine (RBF)	1409	244	1628	2607
Random Forest	1442	211	1655	2580
Gradient Boosting	727	926	577	3658
Neural Network (MLP)	773	880	622	3613

Consistent with Table 4, Gradient Boosting and the MLP show markedly fethe studyr false negatives (577 and 622 respectively) than Logistic Regression, SVM and Random Forest (1,606–1,655), at the cost of a higher false-positive count a favourable trade-off for conservative risk-management applications.

C. Blocked Time-Series Cross-Validation

To assess robustness beyond a single train/test split, Table 6 reports 3-fold blocked time-series cross-validation results.

Table 6. Blocked Time-Series 3-Fold Cross-Validation

Model	Accuracy	F1-Score	AUC-ROC
Random Forest	0.722 ± 0.055	0.786 ± 0.057	0.791 ± 0.019
Gradient Boosting	0.765 ± 0.028	0.850 ± 0.022	0.787 ± 0.014
Logistic Regression	0.723 ± 0.053	0.786 ± 0.055	0.807 ± 0.029

Gradient Boosting shows both the highest mean accuracy (0.765) and the lowest cross-fold dispersion (± 0.028), indicating superior temporal stability relative to Random Forest and Logistic Regression, whose accuracy standard deviations (± 0.055 and ± 0.053) suggest greater sensitivity to regime shifts across the three sequential validation windows. This supports rejecting H01 in favour of Gradient Boosting as the preferred production model for a live early-warning system, subject to periodic retraining.

E. Feature Importance

Table 7. Random Forest Feature Importance Ranking

Rank	Feature	Importance Score	Share of Total
1	vol_60	0.5074	50.7%
2	vol_20	0.2344	23.4%
3	max_dd_60	0.0966	9.7%
4	ma_ratio	0.0326	3.3%
5	mom_20	0.0235	2.3%
6	skew_20	0.0186	1.9%
7	india_vix	0.0186	1.9%
8	mom_5	0.0170	1.7%
9	kurt_20	0.0165	1.6%
10	rsi	0.0130	1.3%
11	fii_flow	0.0098	1.0%
12	vol_chg	0.0048	0.5%
13	usdinr_ret	0.0038	0.4%
14	crude_ret	0.0036	0.4%

Rolling volatility (vol 60 and vol 20) accounts for 74.2% of total predictive the studyright and maximum drawdown (9.7%). In contrast, macro-financial variables (India VIX, FII flow, USD/INR, crude oil) contribute only 3.6% in this sample and thus, at 20 days the stock-specific history of volatility is more important than macro-shock effects in predicting the risk for stock markets, but macro factors could play a larger role at longer horizons or during

acute crisis events. This is in line with the the study known volatility clustering of financial returns and is proof that the multi-window realized volatility feature of equity risk classification in India is the key one.

5.3. Return Forecasting: Regression Results

It compares Linear Regression and a Random Forest Regressor in forecasting forward 20-day cumulative returns.

Table 8. Forward 20-Day Return Forecasting Performance

Model	RMSE	MAE	R ²	Directional Accuracy
Linear Regression	0.1108	0.0858	-0.0723	60.5%
Random Forest Regressor	0.1171	0.0915	-0.1968	55.4%

Both models produce negative out-of-sample R² (Linear: -0.072; Random Forest: -0.197). Again, neither approach outperforms a naive mean forecast on 20 days ahead of the return, which is consistent with the studyak-form market efficiency. A paired t-test on the squared forecast errors finds Linear Regression to produce much smaller errors than Random Forest Regressor ($t = -17.17$, $p < 0.001$), i.e., the added model complexity of the Random Forest does not lead to better predictions and noise is overfit by linear regression. Directional accuracy is slightly better than chance for Linear Regression (60.5%) but less than chance for Random Forest Regressor (55.4%).

5.4. Portfolio Optimization Results

This compares the Equal-The studyighted, Markowitz Mean-Variance, and ML-Tilted portfolios on the out-of-sample test window.

Table 9. Out-of-Sample Portfolio Performance Comparison

Portfolio	Ann. Return	Ann. Volatility	Sharpe	VaR (95%)	CVaR (95%)	Max Drawdown	Sortino
Equal-The studyighted	50.0%	16.8%	2.59	22.6%	30.9%	-8.5%	6.29
Markowitz Mean-Variance	70.6%	16.7%	3.84	21.2%	28.2%	-7.1%	10.52
ML-Tilted (RF risk-adjusted)	50.4%	18.5%	2.37	27.1%	33.6%	-9.7%	5.73

The Markowitz mean-variance portfolio studied the strongest risk-adjusted performance among the three portfolios in this sample. It achieved a Sharpe ratio of 3.84, compared with 2.59 for the equal-the studyighted portfolio and 2.37 for the ML-tilted portfolio. Similarly, its Sortino ratio was 10.52, compared with 6.29 and 5.73, respectively. The Markowitz portfolio also recorded the lothe studyst maximum drawdown (-7.1%) and CVaR (-28.2%). Although the ML-tilted portfolio incorporated the validated Random Forest risk classifier, it did not outperform the other approaches in this sample.

B. Statistical Significance of Portfolio Differences

Table 10. Paired-Sample t-Tests on Out-of-Sample Daily Portfolio Returns

Comparison	t-statistic	p-value	Inference
Markowitz MVO vs Equal-The studyighted	2.106	0.0366	Significant (5%)

ML-Tilted vs Markowitz MVO	-1.723	0.0866	Not significant
ML-Tilted vs Equal-The studyighted	0.090	0.9286	Not significant

The Markowitz portfolio's performance over the equal-the studyighted portfolio is statistically significant at the 5% level ($t = 2.11$, $p = 0.037$), which is not the naive-diversification-is-sufficient null for this sample. The ML-tilted portfolio's worse performance over Markowitz is only statistically significant at the 10% level ($p = 0.087$) and it is indistinguishable from zero ($p = 0.929$). H_0 is rejected for the Markowitz-vs.-equal-the studyighted comparison, but not for the ML-tilted portfolio comparison.

5.5. Macro-Financial Validation: India VIX and Realized Volatility

As an outside validity check, the Pearson correlation betthe studyen India VIX and the cross-sectional average 20-day realized volatility across all sample stocks is $r = 0.065$ ($p = 0.047$), which is statistically significant but economically modest. The relatively small magnitude of this correlation is consistent with the fact that India VIX is a forward-looking, index-level implied volatility measure and is only partially correlated with backward-looking and stock-level realized volatility at a 20-day horizon, and the statistical significance confirms that India VIX is a way of adding some real incremental information as shown in the macro-financial aspects of Table 7 compared with USD/INR and crude oil.

6. Theoretical and Managerial Implications

The study contributes to the ML-finance literature by distinguishing betthe studyen risk classification and return prediction and shothe studyd that ML models can accurately capture nonlinear patterns in volatility and identify high-risk periods, but they are not able to predict short-term returns. This supports the idea that ML may be more useful for modelling risk and volatility than for generating better returns. From a managerial point of view, Indian mutual funds and PMS providers can use ML-based early-warning models as the studyll as existing risk systems to identify high-risk periods. Portfolio managers can incorporate ML risk signals into return and covariance estimates instead of just using them as an additional allocation tool. For SEBI- and RBI-regulated risk functions, interpretable models such as Logistic Regression can facilitate transparent risk assessment, and robo-advisory platforms can focus on risk-adjusted and volatility-aware portfolio allocation rather than sector-based return strategies.

7. Limitations and Future Research

The study is subject to limitations that should be considered. The 20-day forecasting horizon might limit the ability to capture the market patterns and dynamics across shorter and longer investment periods. Moreover, the portfolio analysis does not take into account practical factors such as transaction costs, taxes, market impact, and short-selling restrictions which may affect the performance of the portfolio in the real world. The study does not consider advanced forecasting and explainable AI techniques such as Long Short-Term Memory (LSTM), TFT, SHAP, and LIME. Future research could address these limitations by considering multiple forecasting horizons, real-world trading constraints, and the application of advanced predictive and explainable AI models.

8. Conclusion

The study develops an integrated ML framework for financial risk prediction and portfolio optimization in the Indian equity market. Gradient Boosting and Neural Network models performed the studyll for predicting 20-day high-risk regimes with AUC values around 0.80, while regression models the studyre less effective in predicting short-term returns. The Markowitz portfolio performed better risk-adjusted than the ML-based portfolio, suggesting that ML is better suited for early risk detection than portfolio allocation. All in all, our findings support mixing ML-based risk signals with traditional portfolio approaches and future efforts using real-time market data and transaction costs to improve the practical applicability.

9. References

- [1] Ban, G.-Y., El Karoui, N., & Rousseeuw, A. E. B. (2018). Machine learning and portfolio optimization. *Management Science*, 64(3), 1136–1154.
- [2] Chatterjee, S., Bandopadhyay, A., & Ghosh, S. (2018). Ensemble learning approaches to credit risk assessment of Indian NBFCs. *Journal of Emerging Market Finance*, 17(2), 231–255.
- [3] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.
- [4] Bekaert, G., & Harvey, C. R. (1997). Emerging equity market volatility. *Journal of Financial Economics*, 43(1), 29–77.
- [5] Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3–56.
- [6] Shrivastava, A., Suji Prasad, S. J., Yeruva, A. R., Mani, P., Nagpal, P., & Chaturvedi, A. (2025). IoT based RFID attendance monitoring system of students using Arduino ESP8266 & Adafruit.io on defined area. *Cybernetics and Systems*, 56(1), 21–32. <https://doi.org/10.1080/01969722.2023.2166243>.
- [7] BK Kumari, VM Sundari, C Praseeda, P Nagpal, J EP, S Awasthi (2023), Analytics-Based Performance Influential Factors Prediction for Sustainable Growth of Organization, Employee Psychological Engagement, Work Satisfaction, Training and Development. *Journal for ReAttach Therapy and Developmental Diversities* 6 (8s), 76-82.
- [8] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223–2273.
- [9] Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: Deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1), 3–12.
- [10] Gupta, R., & Basu, P. K. (2007). The study of form efficiency in Indian stock markets. *International Business & Economics Research Journal*, 6(3), 57–64.
- [11] P. Nagpal, A. Pawar and S. H. M, "Predicting Employee Attrition through HR Analytics: A Machine Learning Approach," 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM), Noida, India, 2024, pp. 1-4, doi: 10.1109/ICIPTM59628.2024.10563285.
- [12] Rajput, N., Das, G., Kumar, C., & Nagpal, P. (2021). An inclusive systematic investigation of human resource management practice in harnessing human capital. *Materials Today: Proceedings*, 80(3), 3686–3690. <https://doi.org/10.1016/j.matpr.2021.07.362>
- [13] 2021.07.362
- [14] Jothimani, D., Shankar, R., & Yadav, S. S. (2017). Discrete wavelet transform-based prediction of stock index: A study on National Stock Exchange fifty index. *Journal of Financial Management and Analysis*, 30(1), 33–46.
- [15] Pooja Nagpal, (2025). Leveraging artificial intelligence and machine learning for gaining competitive advantage in business development. *AIP Conference Proceedings*, 3327(1), 020002. AIP Publishing LLC. <https://doi.org/10.1063/5.0289438>
- [16] Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), 2767–2787.
- [17] Nayak, R. K., Mishra, D., & Rath, A. K. (2015). A Naïve SVM-KNN based stock market trend reversal analysis for the Indian markets. *Applied Soft Computing*, 35, 670–680.
- [18] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259–268.
- [19] P. Nagpal, "The Role of ICT and Algorithmic Systems in Shaping Gig Worker Evaluations and Retention," 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG), Indore, Madhya Pradesh, India, India, 2025, pp. 1-6, doi: 10.1109/ICTBIG68706.2025.11323582.
- [20] SEBI. (2023). Consultation paper on the framework for algorithmic trading by retail investors. Securities and Exchange Board of India.

- [21] Pooja Nagpal (2023). The Transformative Influence of Artificial Intelligence (AI) on Financial Organizations World Wide. 3rd International Conference on Information & Communication Technology in Business, Industry & Government (ICTBIG). Symbiosis University of Applied Science, Indore.
- [22] Rajagopal, N. K., Anitha, L., Nagpal, P., & Jitendra, G. (2024). Green HR techniques: A sustainable strategy to boost employee engagement. In *Advancements in business for integrating diversity and sustainability: How to create a more equitable and resilient business world in the developing world* (pp. 104-107). Routledge.
- [23] Vaniya, J., Alizada, M., Nagpal, P., Kumar Dey, B. and Abbbasova, D. G. A. (2025). Novel Enhanced Cognitive State Analysis in E-Learning via Real-Time Emotion and Attentiveness Detection Using OptFuzzy TSM and ABiLSTM. *Iranian Journal of Fuzzy Systems*, 22(4), 57-75. doi: 10.22111/ijfs.2025.49950.8829
- [24] Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3), 425–442.
- [25] The studylch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455–1508.
- [26] Rapach, D., & Zhou, G. (2013). Forecasting stock returns. In *Handbook of Economic Forecasting* (Vol. 2, pp. 328–383).