Portfolio Optimization Model for Indian Market Using Modern Portfolio Theory

Ramish Jamal
IIIT Delhi
ramish22395@iiitd.ac.in

Mohd Masood
IIIT Delhi
masood22299@iiitd.ac.in

Akshat Karnwal
IIIT Delhi
akshat22052@iiitd.ac.in

Keshav Chhabra
IIIT Delhi
keshav22247@iiitd.ac.in

I. Motivation

The Indian financial market presents unique opportunities and challenges, necessitating sophisticated strategies for investors. The motivation for this project can be summarized in the following points:

- The Indian financial market offers distinct opportunities and challenges, requiring advanced investment strategies.
- The project aims to develop a portfolio optimization model combining Modern Portfolio Theory (MPT) with machine learning techniques specifically tailored for the Indian market.
- A critical aspect of this project is the creation of a predictive model to estimate stock price volatility over a 90 and 180 trading day span, providing valuable insights into future trends.
- This model will facilitate diversification across large-cap, mid-cap, and small-cap stocks, as well as various sectors of the Indian stock market.
- The ultimate goal is to empower investors with actionable insights and optimized strategies that enhance risk management while maximizing returns.
- Our primary objective is to predict stock price volatility (a measure of risk) to create an optimized portfolio.

***Index Terms*—Portfolio optimization, machine learning, stock price volatility, modern portfolio theory, financial market**

II. Introduction

The Indian financial market is rapidly evolving, presenting unique opportunities and challenges for investors. This project aims to develop a portfolio optimization model that leverages Modern Portfolio Theory (MPT) combined with machine learning techniques specifically designed for the Indian market. By predicting stock price volatility over different time frames (90 and 180 trading days), we seek to empower investors with actionable insights that optimize their portfolios while effectively balancing risk and returns.

III. Literature Survey

Previous studies have shown that machine learning significantly improves stock price predictions by identifying hidden patterns and non-linear relationships. Traditional methods, such as technical and qualitative analysis, often miss such patterns. Research on models like Artificial Neural Networks (ANN) and Random Forest (RF) has demonstrated better accuracy in predicting stock trends. These models have been applied to data from different sectors, making them an appropriate choice for this project.

IV. Data Description

The dataset used in this project comprises stock market data from 9 sectors of the Indian financial market, with each sector represented by 10 companies. The key steps and features of the data are as follows:

- **Sectors:** The 9 sectors include diverse industries such as IT, Banking, Pharmaceuticals, FMCG, Energy, Metals, Automobiles, Realty, and Consumer Durables.
- **Companies:** For each sector, we selected 10 representative companies based on their market activity.
- **Sectoral Closing Prices:** The closing prices for each sector were calculated as the average of the daily closing prices of the 10 companies within that sector. This provided a sector-level view of stock performance.
- **Time Span:** The dataset spans from the year 2000 to 2023, providing a comprehensive view of stock market trends over 23 years.
- **Features:**
  - **Date:** The trading date (YYYY-MM-DD).
  - **Sectoral Average Close:** The average closing price of the 10 companies within each sector for each trading day.

V. Exploratory Data Analysis

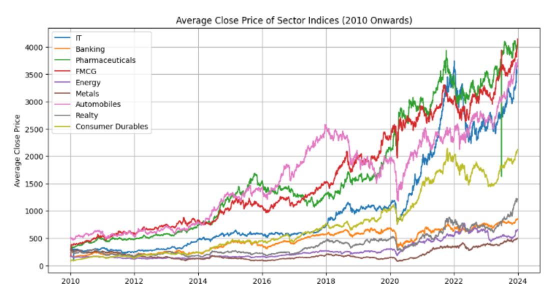

Fig. 1. Trends in Average Closing Prices

A. Trends in Closing Prices

The chart highlights sector-wise growth trends from 2010 to 2024. The IT sector exhibits a sharp upward trajectory post-2020, outperforming others, while Banking shows growth with fluctuations, particularly during the 2020 COVID-19 disruption. Metals and Energy are volatile, whereas FMCG and Pharmaceuticals show steady but slower growth. Realty has the lowest prices, reflecting challenges. Most sectors recovered post-2020, showcasing market resilience. Overall, the IT sector dominates in stock value, indicating robust performance in recent years.

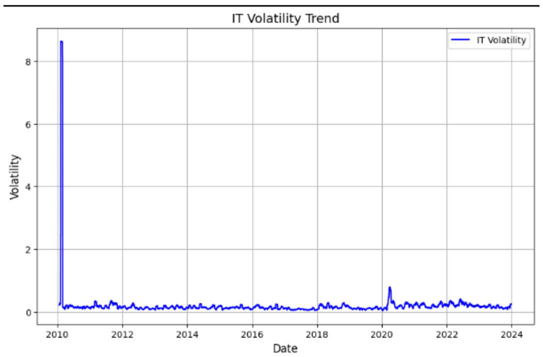

Fig. 2. Time-Series Trend of Stock Prices

B. Volatility in IT Sector

The IT sector's volatility has largely remained stable since 2010, with isolated spikes during major events such as the COVID-19 pandemic. The initial high volatility suggests potential structural or market-specific adjustments in 2010. This stability reflects investor confidence and the sector's maturity over time. ri

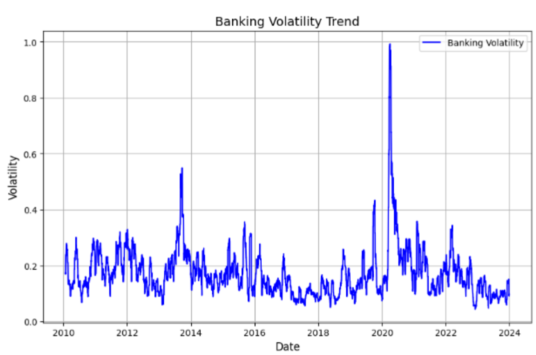

Fig. 3. Time-Series Trend of Stock Prices

C. Volatility in Banking Sector

The IT sector shows remarkable stability in volatility post-2010, with isolated disruptions during major global events like the pandemic. The large spike in 2010 could reflect early structural adjustments or market-specific factors that have since stabilized, reflecting maturity and consistency in the sector's performance.

VI. Data Preprocessing and Feature Engineering

This section outlines the preprocessing and feature engineering steps taken to prepare the dataset for analysis.

A. Preprocessing Steps

- **Combining Data:** All yearly data for each stock is consolidated into a single file for that stock.
- **Filling NaN Values:** Missing values are replaced by the averages of their respective columns.
- **Handling Discrete Dates:** Non-trading days are filled in to ensure the dataset is continuous.
- **Averaging Data in Sectors:** For each sector, the average of all companies' data is taken to generate a combined data point for that sector.
- **Distributing Companies into Sectors:** Companies are allotted to their respective sectors (e.g., FMCG, IT, etc.).
- **Creating Target Variables:** Target variables (i.e., stock price after 90, 180 days, or other specified frames) are created for the model.

B. Feature Engineering Steps

- **Returns:** Calculated percentage changes in closing prices compared to the previous day to measure daily price fluctuations and market performance.
- **Volatility:** Computed the annualized standard deviation of returns over the past 21 days to assess market risk over a short-term period.
- **Future Volatility:** Predicted the 21-day rolling volatility shifted 90 days forward to help forecast future market conditions.
- **Moving Average (21):** Averaged closing prices over the past 21 days to identify trends and smooth daily price movements.
- **RSI (14):** Computed the Relative Strength Index over a 14-day period to highlight overbought (¿70) or oversold (¡30) conditions, indicating potential market reversal opportunities.
- **MACD:** Calculated the difference between the 12-day and 26-day exponential moving averages to gauge market momentum.

C. Date-related Features

- **Day of the Week, Month, Quarter:** Temporal features capturing seasonal patterns and periodicity in the dataset.
- **Year Start/End Flags:** Created indicators for the start and end of the year to capture critical market milestones.

VII. Methodology

Our methodology for building an optimized portfolio involves two main steps: predicting volatility and constructing a portfolio using Modern Portfolio Theory (MPT). Additionally, we incorporated advanced preprocessing and feature engineering techniques to prepare the data for analysis.

A. Step 1: Volatility Prediction

- **Objective:** Predict sectoral volatility to quantify the risk associated with each sector.
- **Models Used:**

– Artificial Neural Networks (ANN): Utilized for its ability to capture complex, non-linear relationships in stock market data.
– Random Forest (RF): Provided a baseline for performance comparison.

- **Evaluation Metrics:**
  – Mean Squared Error (MSE): Measures prediction accuracy by penalizing large errors.
  – Mean Absolute Percentage Error (MAPE): Evaluates model accuracy as a percentage of the actual values.
  – Directional Accuracy: Assesses the model's ability to correctly predict the direction of volatility changes.

B. Step 2: Portfolio Optimization

- **Objective:** Allocate investments across the 9 sectors to minimize portfolio risk (volatility) while achieving optimal returns.
- **Modern Portfolio Theory (MPT):**
  – **Covariance Matrix:** Calculated the covariance matrix of sector-wise volatilities to quantify inter-sector dependencies.
  – **Optimization Problem:**

$$\text{Minimize } \sigma_p = \sqrt{w^T \Sigma w}$$

$$\text{Subject to: } \sum w_i = 1, \; 0 \leq w_i \leq 1$$

Here, σ_p is the portfolio volatility, w is the weight vector, and Σ is the covariance matrix.

- **Optimization Method:** Used Sequential Least Squares Quadratic Programming (SLSQP) for constrained optimization.
- **Portfolio Strategies Compared:**
  – Max Return: Focused all investments on the sector with the highest return.
  – Equal Allocation: Distributed investments equally across all sectors.
  – Optimized Portfolio: Balanced risk and return using MPT-based weights.

C. Validation and Comparison

The optimized portfolio achieved a 24% return with reduced risk, outperforming equal allocation (23%) and balancing risk better than the high-risk max return strategy (29%) during April-June 2023.

VIII. Results and Conclusion

A. Volatility Prediction Results

The predicted volatilities for various sectors are presented below . These results demonstrate the model's capability to estimate stock market risk, which is essential for portfolio optimization.

B. Portfolio Optimization Results

The optimized portfolio allocation outperformed the equal allocation strategy by providing a better return (24% vs. 23%) while reducing risk compared to the high-return strategy. Table I summarizes the outcomes.

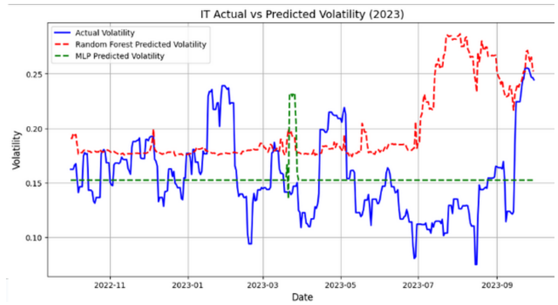

Fig. 4. Predicted vs Actual Volatility for the IT Sector.

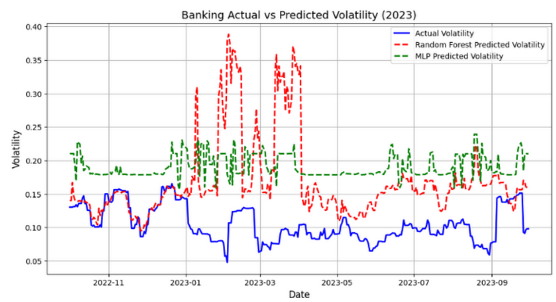

Fig. 5. Predicted vs Actual Volatility for the Banking Sector.

TABLE I
Portfolio Optimization Results Comparison(Apr-June 2023).

Strategy	Return (%)	Risk
Max Return Strategy	29	High
Equal Allocation Strategy	23	Medium
Optimized Portfolio	24	Low

C. Interpretation of Results

The results indicate that:

- The Artificial Neural Network (ANN) model consistently provided lower error metrics compared to Random Forest (RF), especially for the IT and Banking sectors.
- The optimized portfolio balances risk and return effectively, outperforming naive strategies.
- Incorporating sector-wise volatility predictions enhanced the allocation strategy's robustness.

D. Future Enhancements

Planned improvements include:

- Adding lag features and technical indicators such as RSI and EMA.
- Testing advanced machine learning models like XGBoost.
- Implementing ensemble methods and hyperparameter tuning.
- Conducting backtesting for more robust validation.

IX. Appendix

- **Code Snippets:** Key algorithms and implementations used in this project.
- **Data Visualization:** Additional visualizations that support our findings.

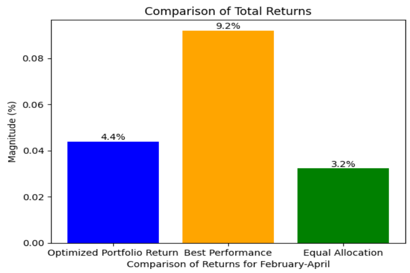

Fig. 6. Comparison of Portfolio Strategies(Feb-Apr 2023).

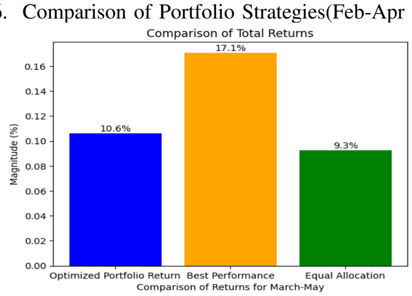

Fig. 7. Comparison of Portfolio Strategies(March-May 2023).

References

[1] Y. Huang, L. F. Capretz, and D. Ho, ”Machine Learning for Stock Prediction Based on Fundamental Analysis,” in *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1-10, doi: 10.1109/SSCI50451.2021.9660134.

[2] H. Ma, G. E. Ab, V. K. Menon, S. K. P. et al., ”NSE Stock Market Prediction Using Deep-Learning Models,” in *Proceedings of the 2018 International Conference on Computational Intelligence and Data Science (ICCIDS 2018)*, pp. 1–x.

[3] Y. Deshmukh, D. Saratkar, H. Hiratkar, S. Dhopte, S. Patankar, T. Jambhulkar, and Y. Tiwari, ”Stock Market Prediction Using Machine Learning,” in *International Journal of Computer Applications*, vol. 176, no. 4, pp. 34-39, 2021.

[4] H. Markowitz, ”Portfolio Selection,” *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952, doi: 10.1111/j.1540-6261.1952.tb01525.x.

[5] W. F. Sharpe, ”Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk,” *The Journal of Finance*, vol. 19, no. 3, pp. 425–442, 1964, doi: 10.1111/j.1540-6261.1964.tb02865.x.

[6] F. Black, M. Jensen, and M. Scholes, ”The Capital Asset Pricing Model: Some Empirical Tests,” in *Studies in the Theory of Capital Markets*, pp. 79-121, 1972.

[7] S. J. Brown and W. N. Goetzmann, ”Performance Persistence,” *The Journal of Finance*, vol. 49, no. 2, pp. 679–698, 1994, doi: 10.1111/j.1540-6261.1994.tb02437.x.

[8] J. Bender and J. D. Heisinger, ”Modern Portfolio Theory: A Guide to Portfolio Construction and Analysis,” *Wiley Finance*, 2009.

[9] P. Glasserman, ”Monte Carlo Methods in Financial Engineering,” *Springer*, 2003.

[10] X. Xu, W. Li, and X. Luo, ”Dynamic Portfolio Optimization Using Deep Reinforcement Learning,” in *Proceedings of the 2021 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pp. 1-7, 2021, doi: 10.1109/CIVEMSA54077.2021.9649020.

X. Contributions

- **Akshat Karnwal (2022052), Keshav Chhabra (2022247):** Conducted the literature review and contributed to model building.
- **Ramish Jamal (2022395):** Contributed to model building and evaluation.
- **Mohd Masood (2022299):** Handled data preprocessing and contributed to model building.