



VIT[®]
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Pareto Optimization for Portfolio Management Using NSGA-II

Saahil Ahmad (22BRS1023) , Md Rameez Haider (22BRS1372), Ramjan Khandelwal (22BRS1263)

School of Computer Science and Engineering, Vellore Institute of Technology, Chennai

BCSE306L: Artificial Intelligence

Dr. Suganya G.

April 17th 2025

Abstract

This paper presents a multi-objective approach to portfolio optimization using the Non-dominated Sorting Genetic Algorithm II (NSGA-II). We apply Pareto optimization principles to simultaneously maximize expected returns and minimize risk (measured by portfolio variance) in a financial portfolio comprising five stocks from India's Nifty 50 index:

Tata Steel, Titan, Axis Bank, HDFC Bank, and Bharti Airtel. Historical daily return data from April 17, 2024, to April 17, 2025, is utilized. The proposed methodology, implemented using Python and the DEAP library, demonstrates how evolutionary algorithms can effectively navigate the complex trade-offs in portfolio management. The results yield a diverse set of optimal portfolios along the Pareto front, including minimum risk, maximum return, and maximum Sharpe ratio portfolios, providing investors with actionable insights for asset allocation based on their risk tolerance.

1. Introduction

Portfolio optimization is a cornerstone of financial management, aiming to allocate capital across various assets to achieve desired investment objectives. The seminal work by Markowitz (1952) introduced Modern Portfolio Theory (MPT), providing a framework for balancing risk and return based on the concept of diversification. MPT typically identifies a single optimal portfolio on the efficient frontier based on an investor's specific risk aversion level. However, real-world investment decisions often involve multiple, often conflicting, objectives beyond just a single risk-return trade-off. Factors like maximizing returns while simultaneously minimizing volatility, or considering other metrics like liquidity or downside risk, necessitate a multi-objective optimization approach.

Pareto optimization provides a powerful framework for addressing such multi-objective problems. Instead of seeking a single "best" solution, it identifies a set of non-dominated solutions, known as the Pareto front. Each solution on this front represents an optimal trade-off, where improving one objective is impossible without degrading at least one other. This approach offers decision-makers a range of optimal choices, allowing them to select a solution that best aligns with their specific preferences and priorities.

Evolutionary algorithms, particularly genetic algorithms, have proven effective in solving complex multi-objective optimization problems. The Non-dominated Sorting Genetic Algorithm II (NSGA-II), proposed by Deb et al. (2002), is a widely recognized and efficient algorithm for finding Pareto optimal solutions. Its strengths lie in its fast non-dominated sorting mechanism, elitism (preserving the best solutions across generations), and explicit diversity preservation using crowding distance, making it well-suited for exploring the complex search space of portfolio optimization.

This study applies NSGA-II to the problem of optimizing a portfolio consisting of five prominent stocks from India's Nifty 50 index: Tata Steel (TATASTEEL), Titan Company (TITAN), Axis Bank (AXISBANK), HDFC Bank (HDFCBANK), and Bharti Airtel (BHARTIARTL). Using one year of historical daily data, we aim to generate the Pareto front representing the optimal trade-offs between maximizing expected portfolio return and minimizing portfolio risk (variance). The analysis identifies specific portfolio compositions corresponding to minimum risk, maximum return, and the highest Sharpe ratio, demonstrating the practical application of NSGA-II in providing diverse and optimal investment strategies.

2. Methodology

2.1 Problem Formulation

The core of the portfolio optimization problem is to determine the optimal weights (w_i) for each asset i in a portfolio of n assets. In this study, we consider two primary, conflicting objectives:

1. **Maximize the expected portfolio return ($E(R_p)$):** This objective seeks the highest possible average return based on the historical performance of the assets.
2. **Minimize the portfolio risk σ_p^2 :** This objective aims to reduce the volatility or uncertainty associated with the portfolio's return, typically measured by the portfolio variance.

Mathematically, these objectives are formulated as follows:

Maximize Expected Return: $E(R_p) = \sum (w_i * E(R_i))$, for $i = 1$ to n Where:

- $E(R_p)$ is the expected return of the portfolio.
- w_i is the proportion (weight) of the total capital invested in asset i .
- $E(R_i)$ is the expected return of asset i , calculated as the mean of its historical daily returns.
- n is the number of assets in the portfolio (in this case, $n=5$).

Minimize Portfolio Variance: $\sigma_p^2 = \sum \sum (w_i * w_j * \sigma_{ij})$, for $i, j = 1$ to n
Where:

- σ_p^2 is the variance of the portfolio's return.
- w_i and w_j are the weights of assets i and j , respectively.
- σ_{ij} is the covariance between the returns of assets i and j . If $i=j$, σ_{ii} represents the variance of asset i (σ_i^2).

These optimization objectives are subject to the following constraints:

1. **Full Investment Constraint:** The sum of all weights must equal 1, ensuring all capital is allocated.
$$\sum w_i = 1$$

Non-Negativity Constraint: The weights must be non-negative, prohibiting short selling.
$$w_i \geq 0, \forall i \in \{1, 2, \dots, n\}$$

2.2 Pareto Optimality

In multi-objective optimization, there is often no single solution that simultaneously optimizes all objectives. Instead, we seek a set of Pareto optimal solutions. A solution (x_1) is said to **dominate** another solution (x_2) if:

1. Solution (x_1) is no worse than (x_2) in all objectives.
2. Solution (x_1) is strictly better than (x_2) in at least one objective.

For our problem (maximizing return (f_1), minimizing risk (f_2)), portfolio $w^{(1)}$ dominates portfolio $w^{(2)}$ if:

1. $E(R_p(w^{(1)})) > E(R_p(w^{(2)}))$ (Return is not worse)
2. $\sigma_p^2(w^{(1)}) \leq \sigma_p^2(w^{(2)})$ (Risk is not worse)
3. At least one of these inequalities is strict (either return is strictly better OR risk is strictly better).

A solution is **Pareto optimal** (or non-dominated) if no other feasible solution dominates it. The set of all Pareto optimal solutions in the objective space (risk vs. return) constitutes the **Pareto front**. In portfolio optimization, this Pareto front is known as the **Efficient Frontier**, representing portfolios that offer the highest expected return for a given level of risk, or the lowest risk for a given level of expected return.

2.3 Non-dominated Sorting Genetic Algorithm II (NSGA-II)

NSGA-II is an evolutionary algorithm designed for multi-objective optimization problems. It employs concepts inspired by natural selection and genetics to iteratively improve a population of candidate solutions (portfolios) over several generations. Key components include:

1. **Initialization:** A population of candidate solutions (portfolios) is randomly generated. Each solution (individual) is represented as a list of weights w_1, w_2, \dots, w_n .
2. **Fitness Evaluation:** Each individual in the population is evaluated based on the defined objectives (expected return and variance) using the formulas in Section 2.1.
3. **Non-dominated Sorting:** The population is sorted into different non-domination levels (fronts). Front 1 contains all non-dominated individuals, Front 2 contains individuals dominated only by those in Front 1, and so on. This ranks solutions based on their Pareto optimality.

Crowding Distance: To maintain diversity within each front and prevent premature convergence to a small region of the Pareto front, a crowding distance metric is calculated for each individual. It estimates the density of solutions surrounding an individual in the objective space. The formula for the crowding distance

$$d_i = \sum ((f_{m(i+1)} - f_{m(i-1)}) / (f_{m(\max)} - f_{m(\min)})), \text{ over all objectives } m$$

Where:

- M : Number of objectives
 - $f_{m(i+1)}$ and $f_{m(i-1)}$: Objective values of neighboring individuals
 - $f_{m(\max)}$, $f_{m(\min)}$: Max and min values of objective m
- Solutions with larger crowding distances are preferred.

4. **Selection:** Individuals are selected for reproduction based on their non-domination rank (lower rank is better) and crowding distance (larger distance is better for the same rank). Tournament selection is commonly used, where two individuals are compared, and the one with better rank, or better crowding distance if ranks are equal, is chosen.
5. **Genetic Operators (Crossover and Mutation):** Selected parent individuals create offspring through:
 - o **Crossover:** Combines genetic material from two parents (e.g., Simulated Binary Crossover (SBX) or Blend Crossover (BLX- α)).
 - o **Mutation:** Introduces small random changes to an individual's genes (weights) (e.g., Polynomial Mutation or Gaussian Mutation).
6. **Elitism:** NSGA-II combines the parent population and the generated offspring population. The combined population is then sorted using non-dominated sorting and crowding distance, and the best individuals are

selected to form the next generation's population, ensuring that good solutions are preserved.

7. **Termination:** The process repeats (evaluation, sorting, selection, reproduction) for a predefined number of generations or until another stopping criterion is met. The final population provides an approximation of the true Pareto front.

3. Data Description

The dataset comprises historical daily closing prices for five stocks listed on India's National Stock Exchange (NSE) and part of the Nifty 50 index:

1. Tata Steel Ltd. (TATASTEEL)
2. Titan Company Ltd. (TITAN)
3. Axis Bank Ltd. (AXISBANK)
4. HDFC Bank Ltd. (HDFCBANK)
5. Bharti Airtel Ltd. (BHARTIARTL)

The data spans one year, from **April 17, 2024, to April 17, 2025**. The raw data was obtained in CSV format, potentially sourced from the NSE website or a financial data provider.

Data preprocessing involved:

1. Parsing the CSV files, which had a non-standard header/footer structure.
2. Extracting the 'Date' and 'Close Price' columns.
3. Converting prices to numerical format, handling commas within price strings.
4. Calculating daily returns using the percentage change formula:

$$R_{(i,t)} = (P_{(i,t)} - P_{(i,t-1)}) / P_{(i,t-1)}$$

where $P_{(i,t)}$ is the closing price of asset i at day t .

5. Aligning the return series for all stocks to ensure they cover the same time periods and have equal length, resulting in 248 data points for daily returns per stock.
6. Calculating the mean daily return ($E(R_i)$) for each stock and the 5x5 covariance matrix σ_{ij} from these daily returns.

4. Implementation Details

The NSGA-II algorithm for portfolio optimization was implemented using Python (version 3.10 or compatible) and the following libraries:

- **DEAP (Distributed Evolutionary Algorithms in Python):** For the core evolutionary algorithm framework, including NSGA-II selection, genetic operators, and population management.
- **NumPy:** For numerical operations, especially vector and matrix calculations for portfolio return and variance.
- **Pandas:** For data loading, manipulation, and calculation of returns and covariance matrix from the stock data.
- **Matplotlib:** For visualizing the results, specifically plotting the Pareto front and portfolio allocation pie charts.
- **re:** For parsing the non-standard CSV file format.

The NSGA-II algorithm was configured with the following parameters:

- **Population Size:** 100 individuals (portfolios)
- **Number of Generations:** 100
- **Crossover Probability:** 0.9 (using Blend Crossover with $\alpha=0.5$)
- **Mutation Probability:** 0.1 (using Gaussian Mutation with $\text{mean}=0$, $\text{std dev}=0.1$)
- **Selection Method:** tools.selNSGA2 (DEAP's implementation of NSGA-II selection)

Chromosome Representation: Each individual in the population represents a portfolio and is encoded as a list of floating-point numbers, where each number corresponds to the initial weight assigned to a stock.

Constraint Handling: The constraints ($w_i \geq 0$ and $\sum w_i = 1$) were enforced within the evaluation function and after genetic operations:

1. **Non-negativity:** After mutation, any negative weight was set to 0.
2. **Sum-to-One:** After ensuring non-negativity, the weights within each individual were normalized by dividing each weight by the sum of all weights in that individual. This ensures the weights always sum to 1 before fitness evaluation.

The implementation was executed within a dedicated Python virtual environment.

5. Results and Analysis

5.1 Statistical Analysis of Assets

Based on the historical daily returns from April 17, 2024, to April 17, 2025, the expected annual returns (annualized from mean daily returns assuming 252 trading days) and the covariance matrix were calculated:

Expected Annual Returns:

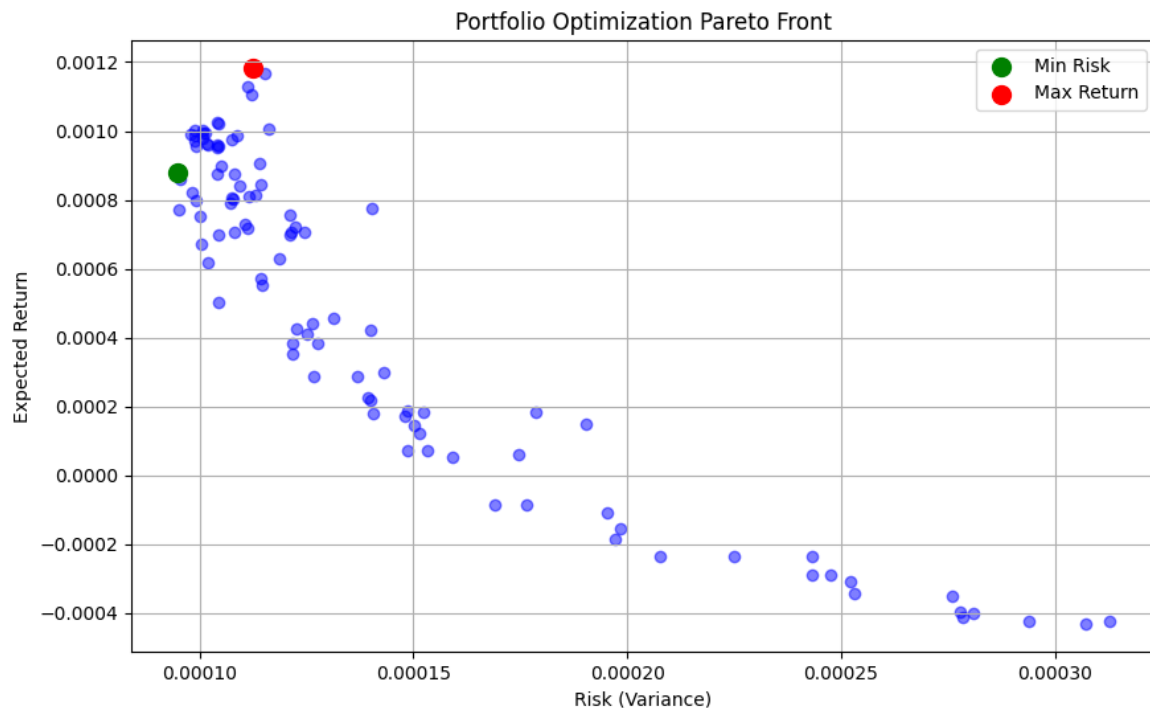
- TATASTEEL: -11.94%
- TITAN: -5.25%
- AXISBANK: 15.68%
- HDFCBANK: 25.05%
- BHARTIARTL: 39.87%

Covariance Matrix (Daily Returns):

	TATASTEEL	TITAN	AXISBANK	HDFCBANK	BHARTIARTL
TATASTEEL	0.000360	0.000082	0.000116	0.000094	0.000098
TITAN	0.000082	0.000224	0.000031	0.000027	0.000073
AXISBANK	0.000116	0.000031	0.000243	0.000104	0.000075
HDFCBANK	0.000094	0.000027	0.000104	0.000160	0.000047
BHARTIARTL	0.000098	0.000073	0.000075	0.000047	0.000216

Bharti Airtel and HDFC Bank exhibit the highest expected returns, while Tata Steel and Titan show negative expected returns over this specific period. Tata Steel displays the highest individual variance (risk), while HDFC Bank has the lowest. The covariance matrix reveals varying degrees of correlation between the stock returns.

5.2 Pareto Front Visualization



the expected trade-off relationship between risk (variance) and expected return. Each point on the front represents an optimal portfolio allocation.

5.3 Analysis of Optimal Portfolios

1. **Portfolio with Minimum Risk:** This portfolio aims to minimize volatility, regardless of the return.

- **Composition:**

- TATASTEEL: 1.64%
- TITAN: 18.37%
- AXISBANK: 9.25%
- HDFCBANK: 42.55%
- BHARTIARTL: 28.20%

- **Metrics:**

- Expected Daily Return: 0.0881%
- Expected Annual Return: 22.19%
- Risk (Variance): 0.000095

- **Analysis:** This portfolio achieves the lowest risk by diversifying significantly across HDFC Bank (highest weight due to low variance), Bharti Airtel, and Titan. It maintains minimal exposure to Tata Steel (highest variance) and a modest allocation to Axis Bank.

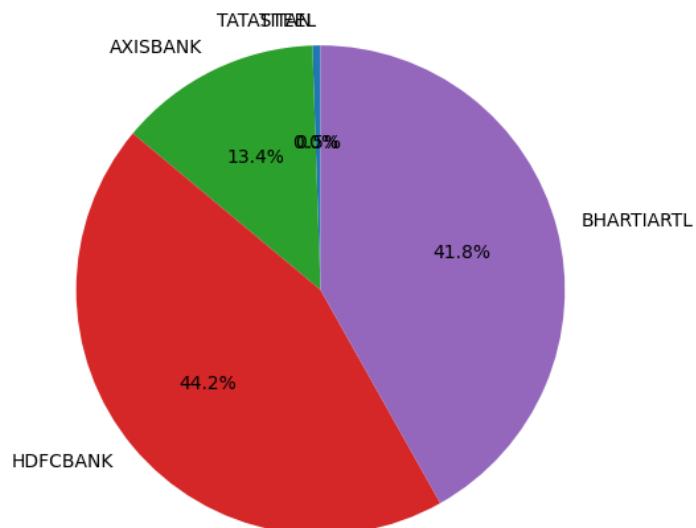
2. **Portfolio with Maximum Return:** This portfolio focuses solely on maximizing expected returns, accepting higher risk.

- **Composition:**

- TATASTEEL: 0.53%
- TITAN: 0.00%
- AXISBANK: 13.41%

- HDFCBANK: 44.21%
 - BHARTIARTL: 41.85%
 - **Metrics:**
 - Expected Daily Return: 0.1183%
 - Expected Annual Return: 29.80%
 - Risk (Variance): 0.000112
 - *Analysis:* This portfolio heavily favors the stocks with the highest historical returns (Bharti Airtel and HDFC Bank), along with a moderate allocation to Axis Bank. It completely avoids Titan and minimizes Tata Steel due to their negative expected returns in the analyzed period.
3. **Balanced Portfolio (Highest Sharpe Ratio):** This portfolio represents the best trade-off between risk and return, as measured by the Sharpe ratio (assuming a risk-free rate of 0). The Sharpe ratio is calculated as $(E(R_p) / (\sigma_p))$.
- **Composition:**
 - TATASTEEL: 0.53%
 - TITAN: 0.00%
 - AXISBANK: 13.41%
 - HDFCBANK: 44.21%
 - BHARTIARTL: 41.85%
 - **Metrics:**
 - Expected Daily Return: 0.1183%
 - Expected Annual Return: 29.80%
 - Risk (Variance): 0.000112
 - Sharpe Ratio (Daily Return / Daily Std Dev): 10.5148

Balanced Portfolio Allocation (Highest Sharpe Ratio)



6. Conclusion

This study successfully demonstrated the application of the Non-dominated Sorting Genetic Algorithm II (NSGA-II) for multi-objective portfolio optimization, specifically balancing expected return maximization and risk (variance)

minimization. Using one year of historical data for five Nifty 50 stocks, the NSGA-II implementation effectively generated a Pareto front of non-dominated portfolio solutions.

The analysis revealed distinct optimal allocation strategies based on risk preference. The minimum risk portfolio achieved diversification across several assets, notably HDFC Bank, Bharti Airtel, and Titan, while the maximum return and highest Sharpe ratio portfolios (which coincided in this instance) heavily favored HDFC Bank, Bharti Airtel, and Axis Bank, the highest-performing stocks in the observed period. The results underscore the ability of NSGA-II to provide a diverse set of optimal solutions, empowering investors to choose portfolios that align with their individual risk-return profiles, rather than relying on a single-point solution.

The findings highlight the importance of considering the specific time period and dataset, as the negative expected returns for Tata Steel and Titan significantly influenced their low or zero allocation in optimal portfolios. Future work could explore incorporating additional objectives (e.g., minimizing downside risk using metrics like CVaR, maximizing liquidity), incorporating transaction costs, applying different risk measures beyond variance, or testing the methodology on longer time horizons and different asset classes. Furthermore, integrating predictive models for returns and volatility, rather than relying solely on historical data, could enhance the practical applicability of the optimization framework.

7. References

1. Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182-197.
2. Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77-91.
3. Ponsich, A., Jaimes, A. L., & Coello, C. A. C. (2013). A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem and other finance and economics applications. *IEEE Transactions on Evolutionary Computation*, 17(3), 321-344.
4. Metaxiotis, K., & Liagkouras, K. (2012). Multiobjective evolutionary algorithms for portfolio management: A comprehensive literature review. *Expert Systems with Applications*, 39(14), 11685-11698.
5. Chang, T. J., Meade, N., Beasley, J. E., & Sharaiha, Y. M. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, 27(13), 1271-1302.
6. Fortin, F.-A., De Rainville, F.-M., Gardner, M.-A., Parizeau, M., & Gagné, C. (2012). DEAP: Evolutionary Algorithms Made Easy. *Journal of Machine Learning Research*, 13, 2171-2175.