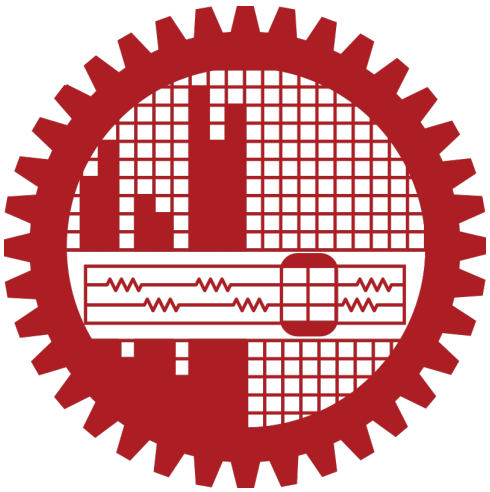

Strategic Stock Trading with Deep Reinforcement Learning Models

Submitted by:

1905095: Md Raihan Sobhan
1905115: Tahsin Wahid

Supervisor:
Dr. Atif Hasan Rahman
Associate Professor, Department of CSE, BUET

Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology (BUET)

December 18, 2024

Contents

1 Problem Definition

Traditional trading models rely heavily on historical data, making them rigid and slow to respond to real-time market fluctuations. This gap often leads to missed opportunities and increased risk for traders. Deep Reinforcement Learning (DRL) offers the potential to bridge this gap by enabling real-time adaptability and improving decision-making in volatile environments. By integrating real-time data and sentiment analysis, DRL provides robust, dynamic trading strategies that outperform traditional methods.

2 Literature Review

Recent advancements in Deep Reinforcement Learning (DRL) highlight its potential in financial trading. Key findings include:

- DRL models enable continuous learning and adaptability, making them ideal for volatile financial markets.
- Algorithms such as Proximal Policy Optimization (PPO) and Soft Actor-Critic (SAC) have shown superior performance in managing complex decision-making environments.
- Studies emphasize the integration of real-time data and sentiment analysis to enhance model robustness.
- The use of benchmarking techniques like Mean-Variance Optimization (MVO) ensures a balanced risk-reward strategy.

3 Reference Papers

- *FinRL: A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance*
- *Practical Deep Reinforcement Learning Approach for Stock Trading*

4 Datasets for Model Training

4.1 Datasets Used

- **Bangladesh — Dhaka Stock Exchange Dataset (DSEBD):** Daily stock price data from the Dhaka Stock Exchange, offering insights into a growing emerging market.
- **India — NIFTY-50 Stock Market Data (2000 - 2021):** Historical data covering 21 years of NIFTY-50.
- **US — USA 514 Stocks Prices NASDAQ NYSE:** Comprehensive pricing data for NASDAQ and NYSE.

4.2 Additional Datasets for Future Iterations

- **S&P 500 Stock Data:** Daily prices for S&P 500 stocks, offering a mature market perspective.
- **US Stocks and ETFs Price and Volume Data:** Comprehensive price and volume data across all US stocks and ETFs.
- **Brazilian Stock Market:** Daily updated data from Brazil.
- **Indian Stock Market Index Intraday Data (2008-2020):** Historical intraday data from India.

5 Proposed Solution

5.1 Solution Approach

The proposed solution involves developing a DRL-based trading model that dynamically adapts to real-time market conditions. By employing an optimal reward function, the model aims to:

- Maximize profits while carefully managing risks.
- Ensure a balanced strategy that prioritizes sustainable growth.

5.2 Key Techniques

The following DRL algorithms and optimization techniques were utilized for benchmarking:

- **A2C (Advantage Actor-Critic):** Ensures efficient exploration of action spaces.
- **DDPG (Deep Deterministic Policy Gradient):** Handles continuous action spaces effectively.
- **Proximal Policy Optimization (PPO):** Provides stability in decision-making.
- **TD3 (Twin Delayed DDPG):** Manages volatility and improves decision accuracy.
- **SAC (Soft Actor-Critic):** Enhances robustness and adaptability.
- **MVO (Mean-Variance Optimization):** Serves as a benchmarking tool.

6 Experimented Architectures

6.1 A2C

Algorithm 1 Advantage Actor-Critic (A2C)

//Assume global shared $\theta, \theta^-, and\ counter\ T = 0$.
Initialize thread step counter $t \leftarrow 0$, $\theta^- \leftarrow \theta$, $d\theta \leftarrow 0$.
Get initial state s.
repeat
    Take action a with ϵ-greedy policy base on Q$(s, a;\ \theta)$
    Receive new state s' and reward r
    $$y = \begin{cases} r & for\ terminal\ s' \\ r + \gamma max_{a'} Q(s', a'; \theta^-) & for\ non-terminal\ s' \end{cases}$$
    $s = s'$
    $T \leftarrow T + 1\ and\ t \leftarrow t + 1$
    If $T\ mod\ I_{target} == 0$ **then**
        Update the target network $\theta^- \leftarrow \theta$
    end if
    if t $mod\ I_{AsyncUpdate} == 0$ or s is terminal **then**
        Perform asynchronous update of θ using $d\theta$.
        Clear gradients $d\theta \leftarrow 0$.
  end if
until $T > T_{max}$

Figure 1: A2C Architecture Overview

6.2 DDPG

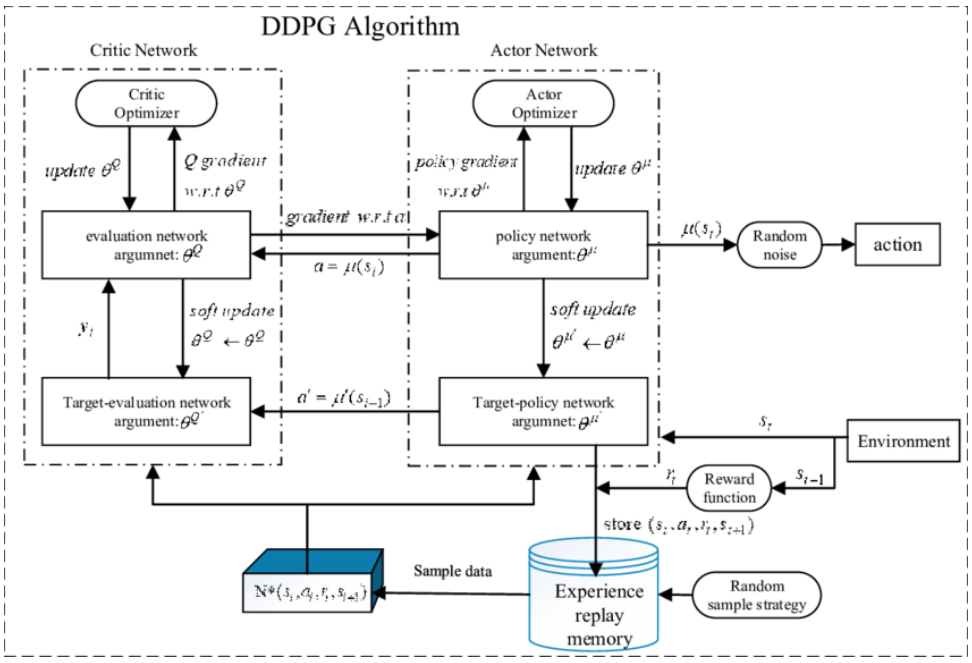

Figure 2: DDPG Architecture Overview

6.3 PPO

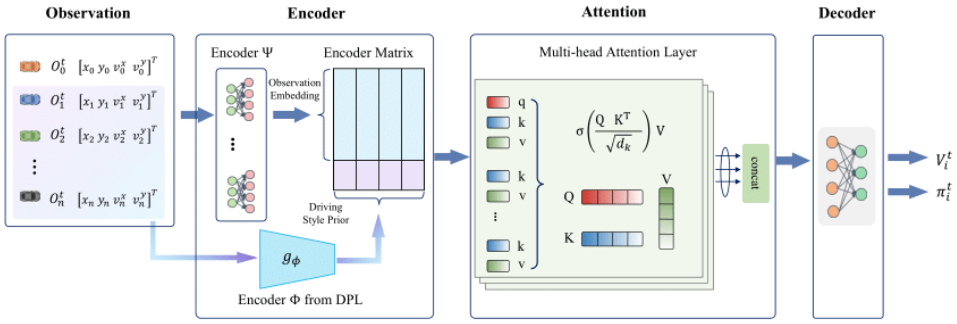

Figure 3: PPO Architecture Overview

6.4 TD3

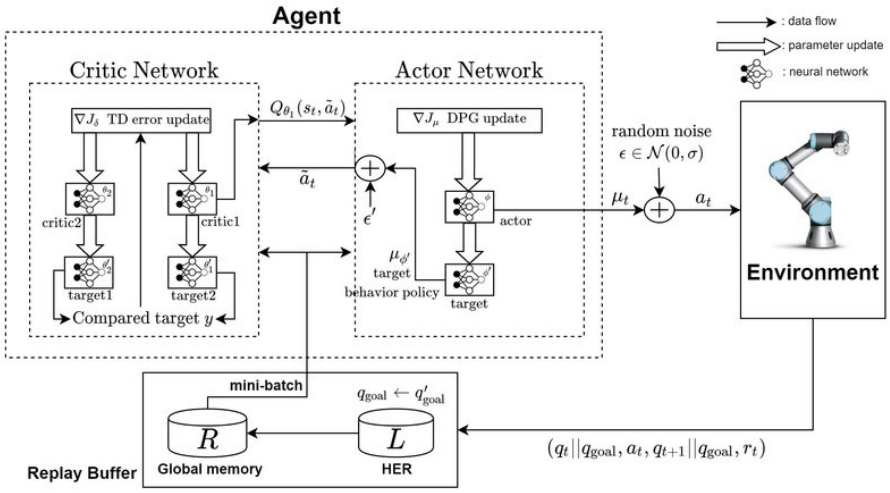

Figure 4: TD3 Architecture Overview

6.5 SAC

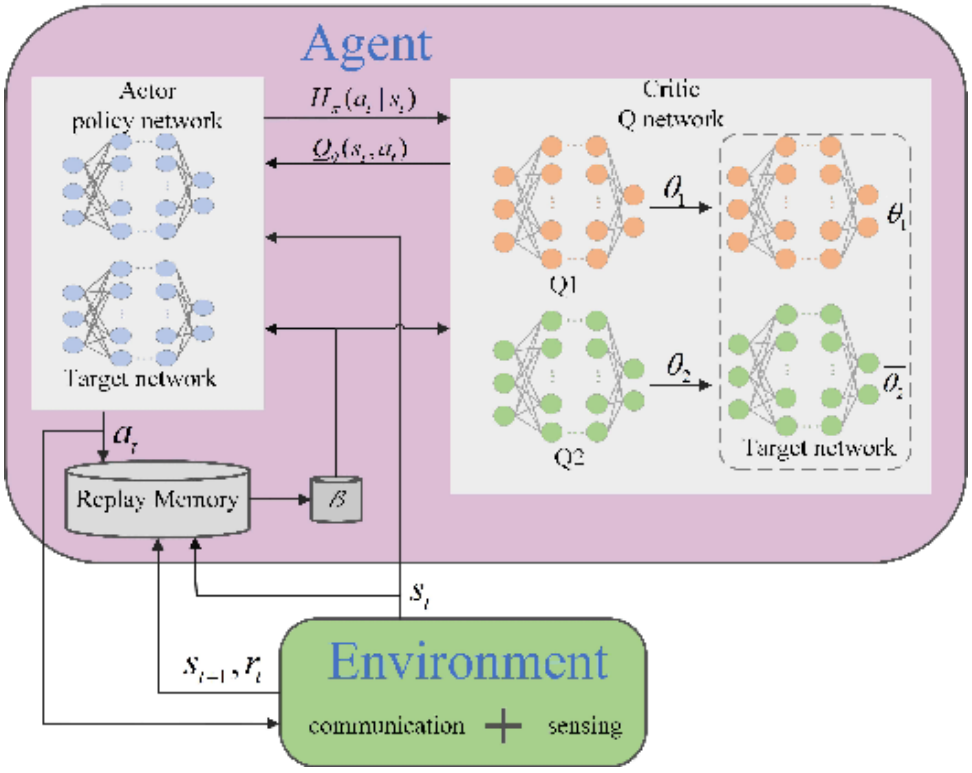

Figure 5: SAC Architecture Overview

7 Performance Metrics

To evaluate the performance of the DRL models, the following metrics were considered:

- Initial and Final Portfolio Value
- Annualized Return
- Annualized Standard Deviation
- Sharpe Ratio
- Max Drawdown
- Sortino Ratio
- Volatility
- Profitability
- Aggressiveness
- Recovery Time

8 Output Analysis

8.1 Key Factors

- **Risk (Max Drawdown)**
- **Profitability**
- **Consistency (Volatility of Returns)**
- **Aggressiveness**
- **Drawdown Recovery Time**
- **Final Value of Portfolio**

8.2 Risk (Max Drawdown)

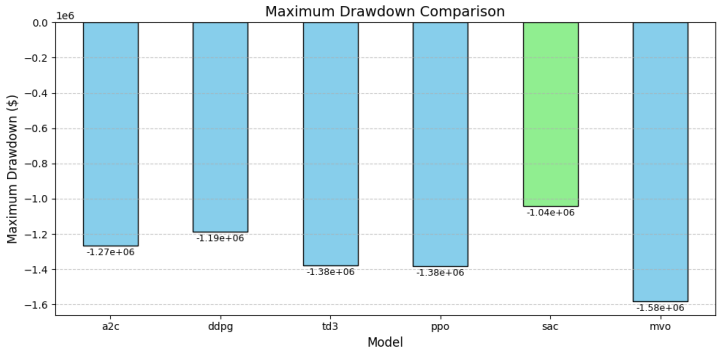

Figure 6: Max Drawdown Example

Max Drawdown (MDD) quantifies the largest peak-to-trough decline in stock or portfolio value, reflecting potential losses.

Mathematically:

$$\text{MDD} = \frac{\text{Peak} - \text{Trough}}{\text{Peak}}$$

MDD is vital for assessing downside risk. Lower values signify less risk and better stability.

8.3 Profitability

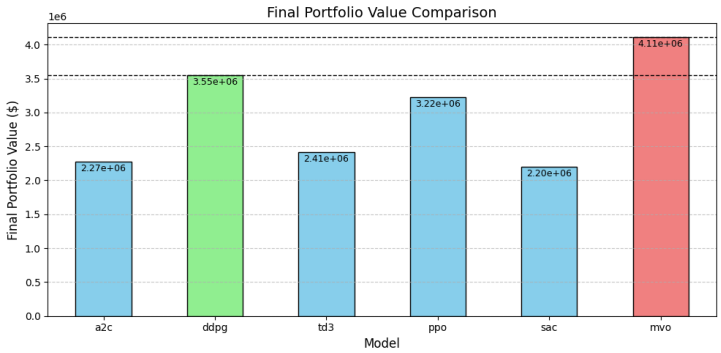

Figure 7: Profitability Example

Profitability measures the financial gain of a portfolio, stock, or investment, indicating the ability to generate returns.

Mathematically:

$$\text{Profit} = \text{Final Value} - \text{Initial Value}$$

This metric is crucial for assessing performance and growth potential. Higher profit indicates better outcomes.

8.4 Consistency (Volatility of Returns)

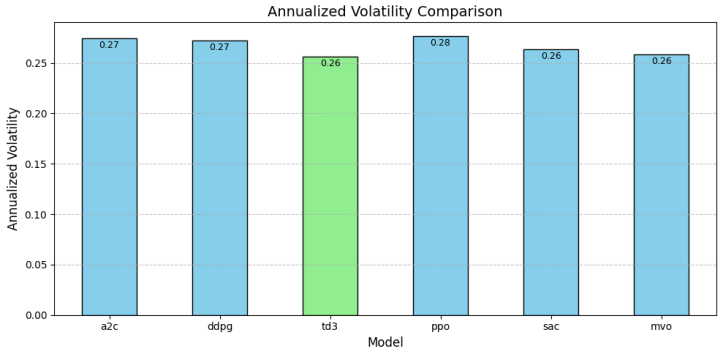

Figure 8: Consistency Example

Consistency refers to the stability of returns over time, often assessed using annualized volatility. It indicates how predictable an investment's performance is.

Mathematically:

$$\text{Volatility} = \text{Standard Deviation of Daily Returns} \times \sqrt{252}$$

Here, 252 represents the number of trading days in a year.

Lower volatility signifies more consistent returns, appealing to risk-averse investors.

8.5 Aggressiveness

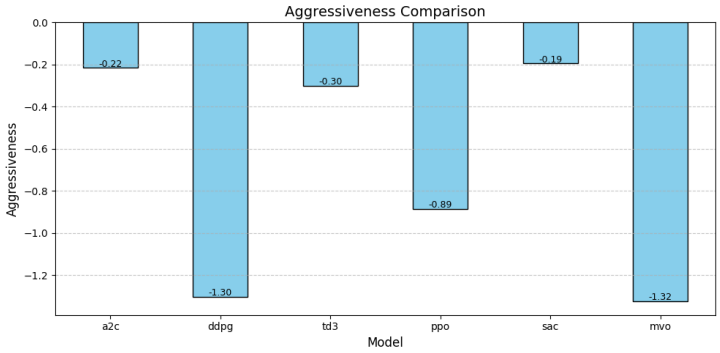

Figure 9: Aggressiveness Example

Aggressiveness measures the trade-off between profitability and risk (as represented by Max Drawdown). It indicates how much return is generated per unit of risk.

Mathematically:

$$\text{Aggressiveness} = \frac{\text{Profitability}}{\text{Max Drawdown}}$$

A higher aggressiveness value implies a better balance between achieving returns and managing risk, appealing to investors with a higher risk appetite.

8.6 Drawdown Recovery Time

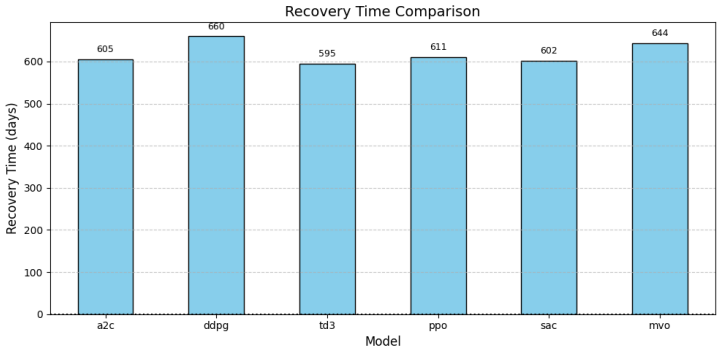

Figure 10: Drawdown Recovery Time Example

Drawdown Recovery Time quantifies the time taken to recover from a portfolio's lowest value back to its peak.

Mathematically:

$$\text{Recovery Time} = \text{Recovery Date} - \text{Lowest Point Date}$$

Where:

$$\text{Recovery Date} = \text{Date of Max Value}, \quad \text{Lowest Point Date} = \text{Date of Min Value}$$

Shorter recovery times are indicative of a more resilient portfolio, which is critical for long-term investment strategies.

8.7 Comparison of Factors

Factor	**High Value Suggests**	**Low Value Suggests**
Max Drawdown (MDD)	High risk and potential for significant losses.	Low risk and stable portfolio performance.
Profitability	Strong financial gains and performance.	Weak financial gains or potential losses.
Consistency (Volatility)	Unstable or unpredictable returns.	Stable and predictable returns.
Aggressiveness	High returns relative to risk.	Low returns relative to risk.
Drawdown Recovery Time	Slow recovery from losses.	Quick recovery, indicating resilience.

Table 1: Comparison of Factors: High vs. Low Values

9 Output for Each Dataset

9.1 DSE Dataset

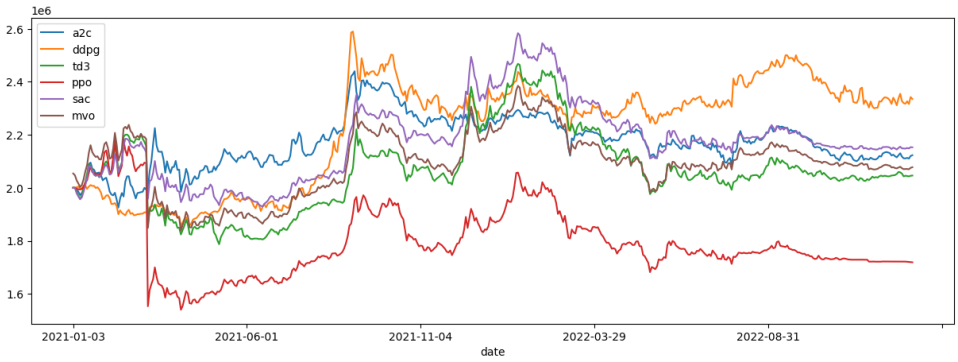

Figure 11: DSE Dataset Overview

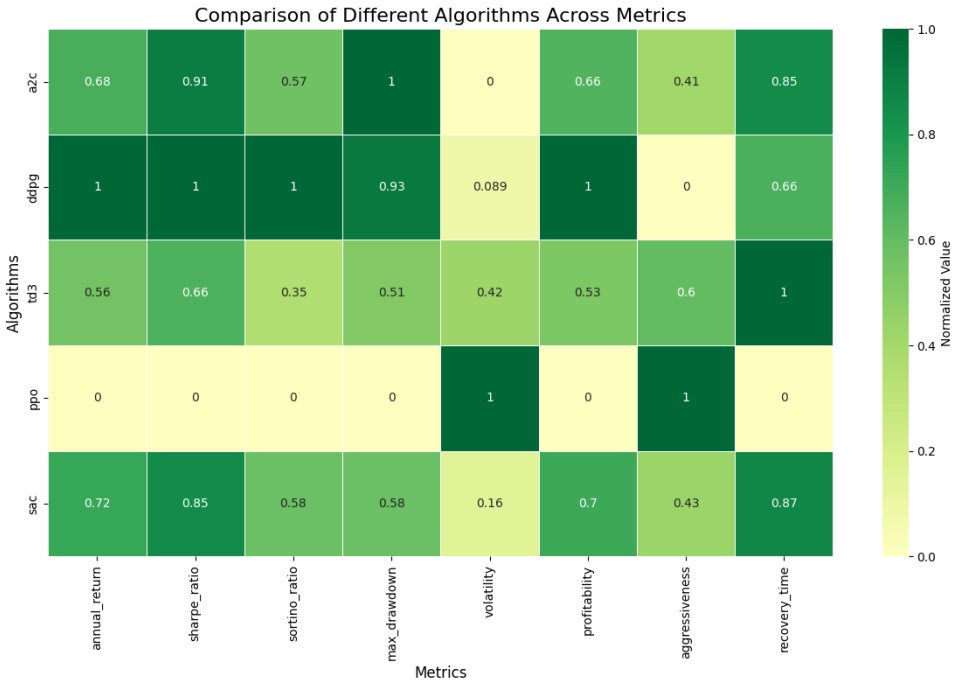

Figure 12: DSE Dataset Heatmap

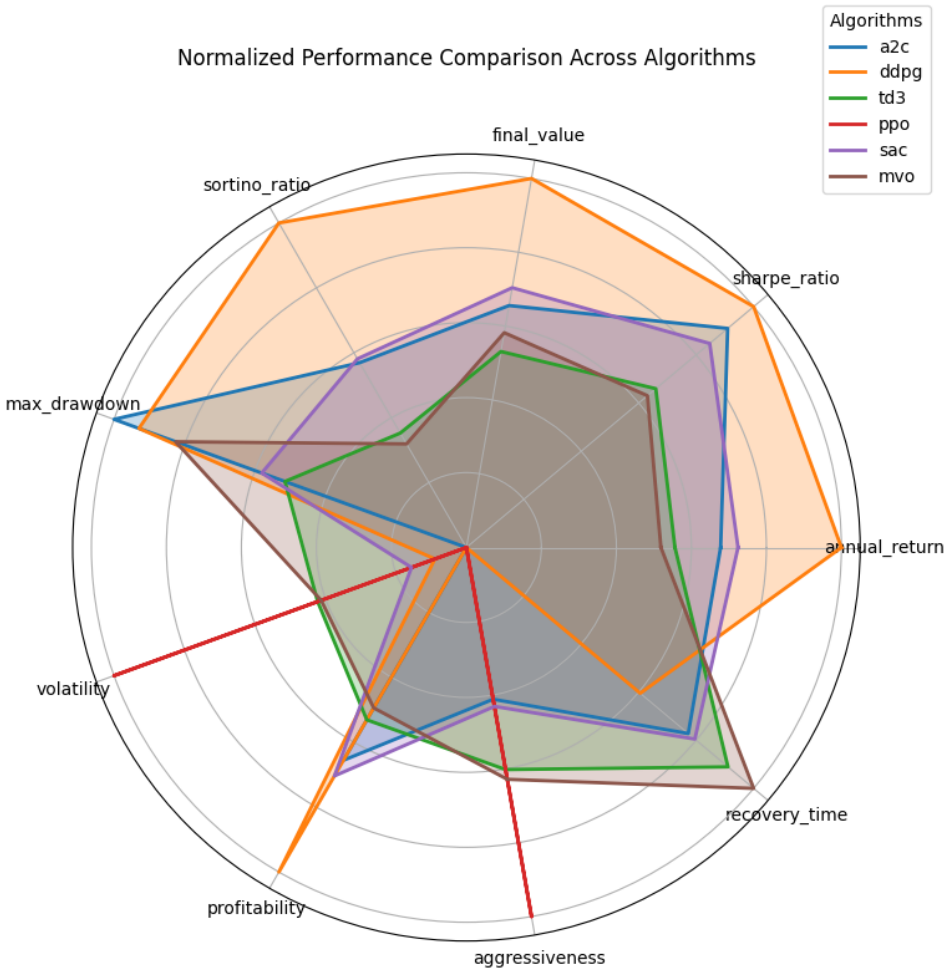

Figure 13: DSE Dataset Radar Plot

9.2 NIFTY Dataset

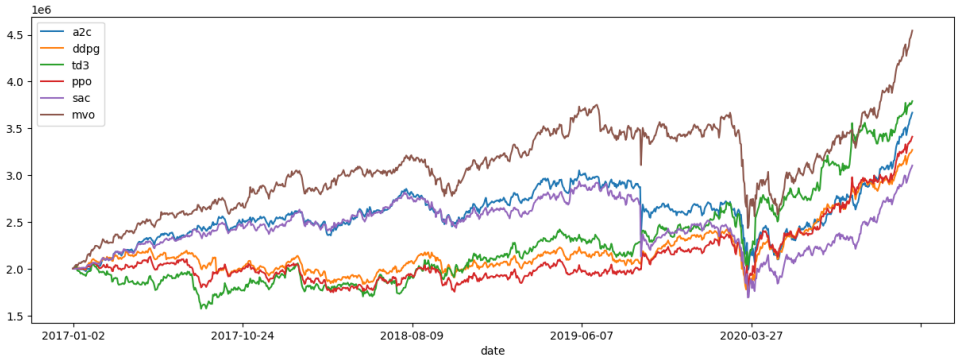

Figure 14: NIFTY Dataset Overview

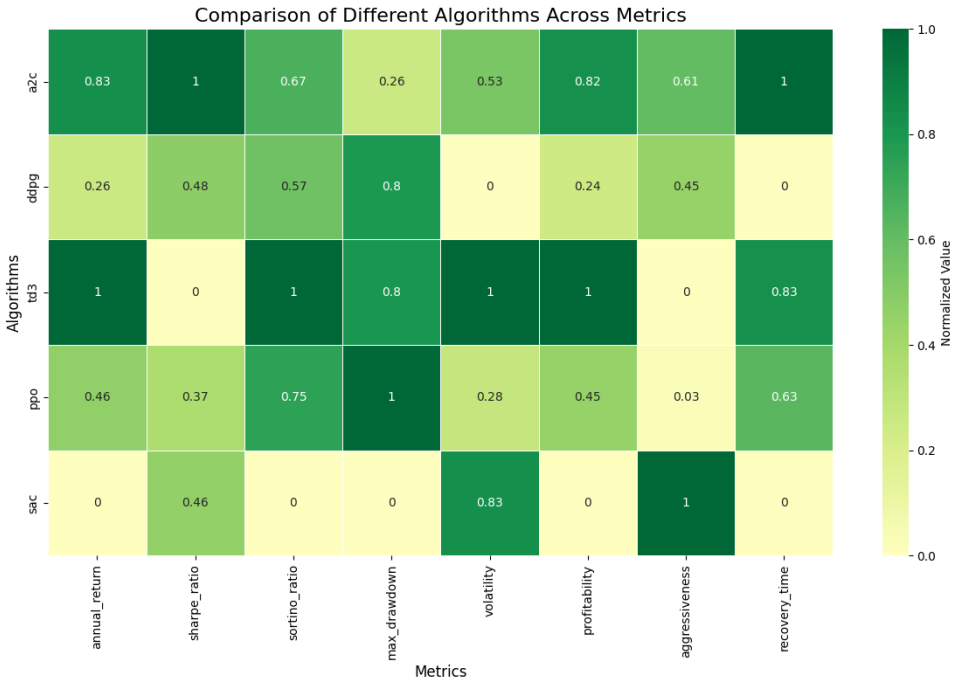

Figure 15: NIFTY Dataset Heatmap

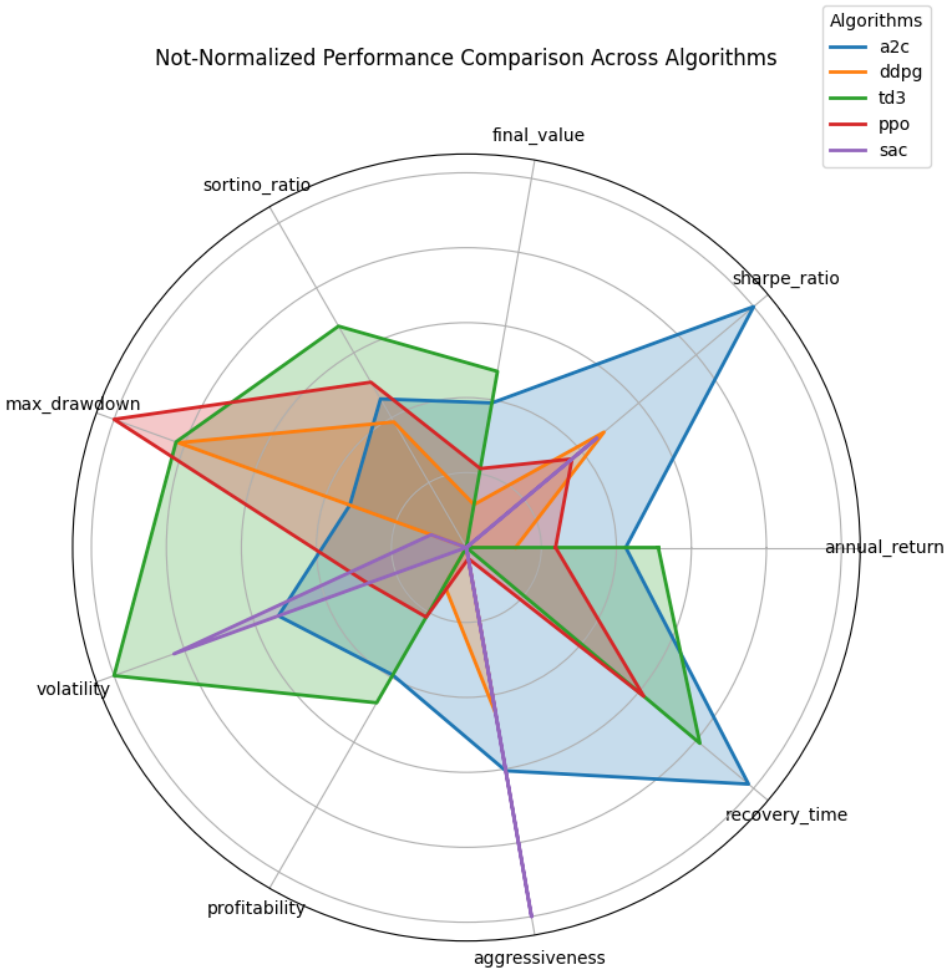

Figure 16: NIFTY Dataset Radar Plot

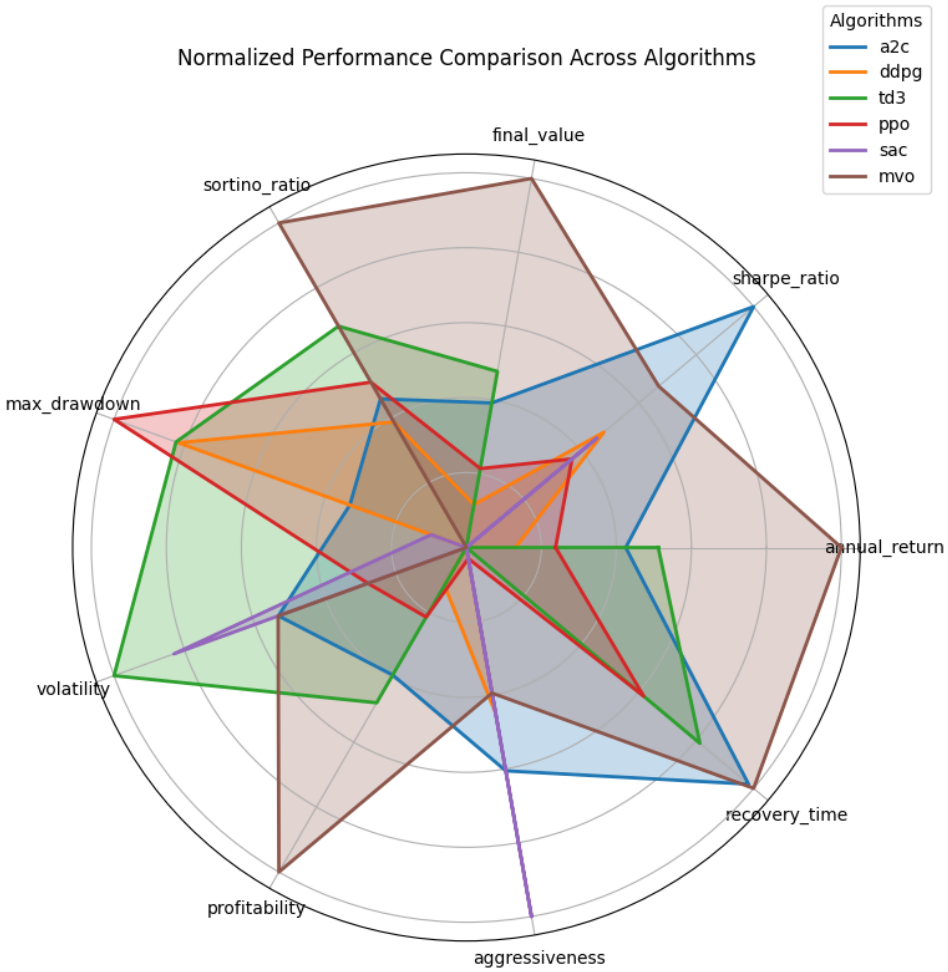

Figure 17: NIFTY Dataset Radar Plot

9.3 S&P 500 Dataset

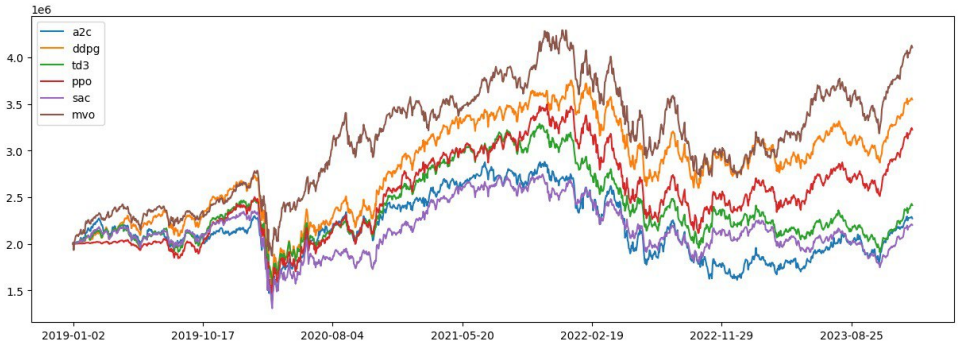

Figure 18: S&P 500 Dataset Overview

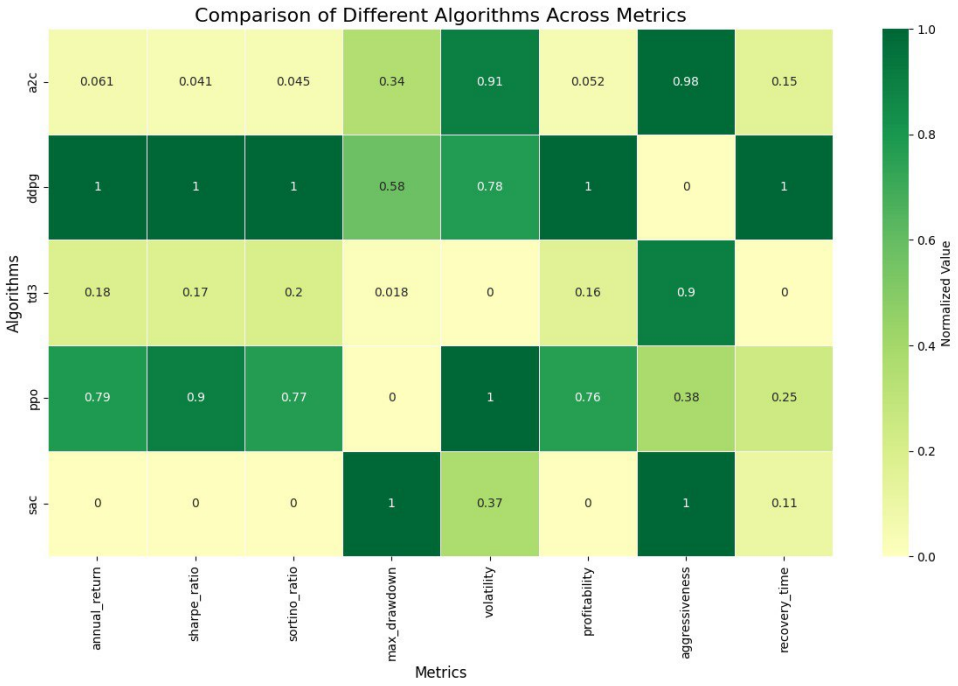

Figure 19: S&P 500 Dataset Heatmap

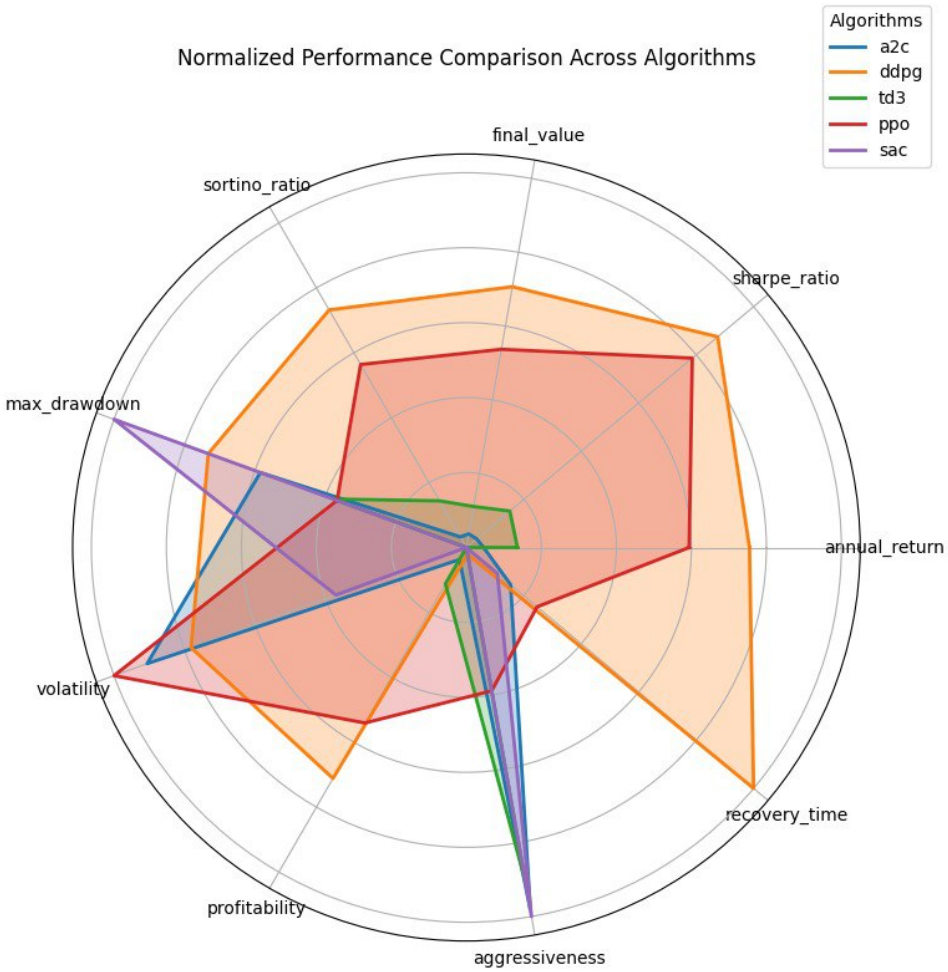

Figure 20: S&P 500 Dataset Radar Plot

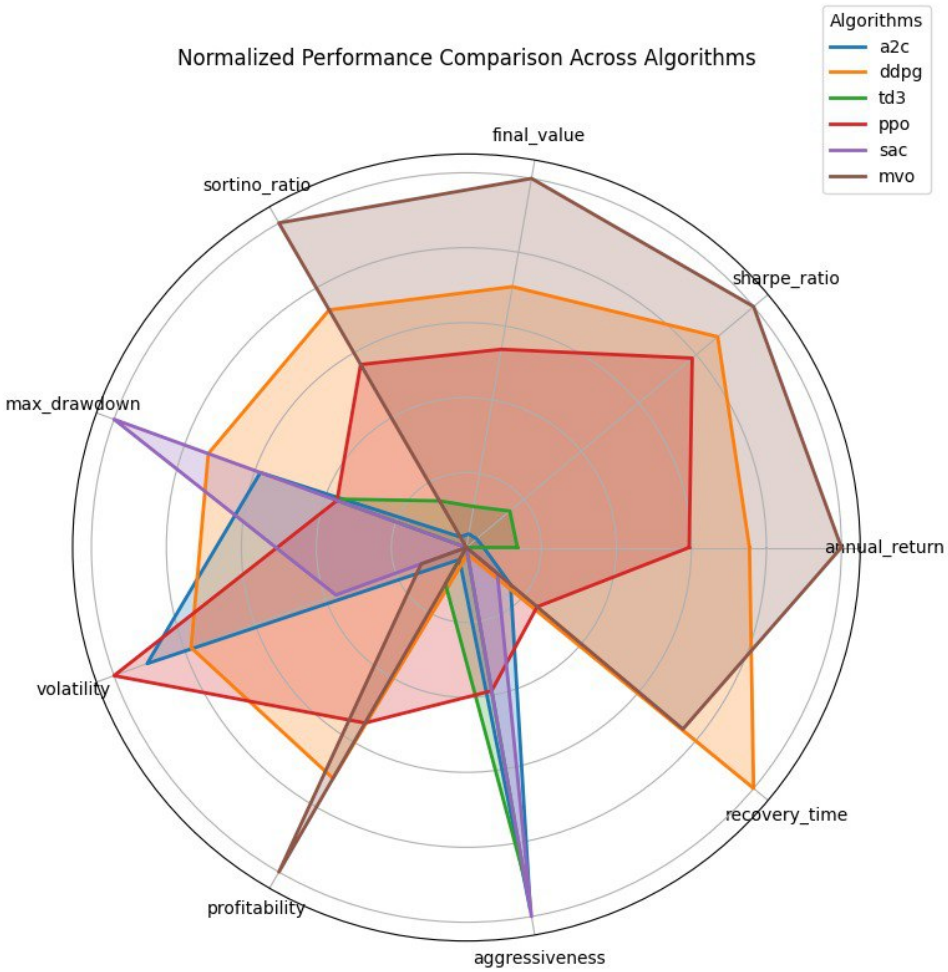

Figure 21: S&P 500 Dataset Radar Plot

10 GitHub Repository

To access the code and experiment on your own, visit the following GitHub repository:

`https://github.com/MdRaihanSobhan/`
`Strategic-Stock-Trading-with-Deep-Reinforcement-Learning-Models`

11 Conclusion

This project demonstrates the potential of Deep Reinforcement Learning (DRL) in addressing the challenges of traditional trading strategies. By leveraging real-time data, advanced DRL algorithms, and sentiment analysis, we developed adaptive trading strategies capable of outperforming conventional approaches.

The evaluation across diverse datasets from emerging and mature markets revealed key insights:

- Models like PPO and SAC excel in balancing profitability and risk, making them suitable for volatile markets.
- DDPG worked best for SP 500 companies due to its ability to handle high-dimensional state spaces and continuous action spaces. Its continuous nature aligns well with the complex dynamics of stock price fluctuations, maintaining stability and adaptability.
- Performance metrics such as Sharpe Ratio, Max Drawdown, and Consistency highlighted the robustness of DRL-based strategies.
- The inclusion of sentiment analysis and real-time adaptability significantly enhanced decision-making processes.

Overall, this study reinforces the viability of DRL as a transformative approach in financial trading, paving the way for further innovations. Future iterations could explore additional datasets, refine reward functions, and incorporate multi-agent systems to further enhance trading efficiency and scalability.