Portfolio Optimization of Indian Stock Market Using Reinforcement Learning and Explainable AI

Rishabh Chorasiya[1] and Shakti Kinger[1]

[1]Department of Computer Engineering and Technology,

Dr. Vishwanath Karad MIT-World Peace University, Pune, Maharashtra, India.

Abstract:

This paper gives a hybrid AI system to optimize risk-adjusted portfolio for the Indian stock market with Deep Reinforcement Learning (DRL) methods and Explainable AI (XAI). The structure relies on an individual NSE dataset with 1900+ stock symbols and DRL models, such as Proximal Policy Optimization (PPO), Advantage Actor-Critic (A2C), Soft Actor-Critic (SAC) and Deep Deterministic Policy Gradient (DDPG). It has a dynamic reallocation policy and constraints to reduce interest exposure in high-performing stocks and over-concentration. Compared to the prior performance, risk-adjusted performance is stabilized using Sharpe Ratio, Sortino Ratio, Annualized Volatility and Maximum Drawdown in the model. Evaluation metrics indicate PPO was better compared to the other models with a Sharpe ratio of 1.3972 and Cumulative Returns of 34.04%. The framework demonstrates how DRL and XAI are effective where building transparent and robust investment portfolios, specific to emerging markets such as India, are concerned.

Keywords: Deep Reinforcement Learning, Portfolio Optimization, Explainable AI, Indian Stock Market, LIME, Proximal Policy Optimization.

1. Introduction:

Investments in the stock market may give huge returns and give losses when due risk procedures are not taken. Portfolio optimization forms one of the significant areas in finance that aims to optimize the returns over against the risks. The traditional ML methods like Markowitz mean-variance optimization are simple and even they do not tackle the new complexities of the markets that entailed volatility and trends. Innovative solutions to these problems need to be identified.

Another area in the field of machine learning is known as Deep Reinforcement Learning (DRL), which became acknowledged as the reliable method of dynamically active portfolio. As time passes by it will be able to react to the modern complexity of the market. However, the disadvantage of DRL is that it entails non-pragmatic and uninterpretable allocation strategies and model adaptation with high-dimensional and volatile market information. Such methods are prone to mistrust due to black-box character of applications. The DRL frameworks can be further improved through the incorporation of Explainable AI (XAI) elements, which will help to make them more recognizable and understandable to the investors, securing them a place.

XAI aids investors in knowing the decision-making process of DRL models, making them more confident in the portfolio. It gives an understanding of the inclusion of a stock as either one having better returns with attainable risks or low risks.

One research intends to use four Deep Reinforcement Learning agents to optimize portfolios on the Indian stock market, namely Proximal Policy Optimization, Advantage Actor-Critic, Soft Actor-Critic, and Deep Deterministic Policy Gradient. The model employs transparency and reasonability methods Local Interpretable Model-agnostic Explanations (LIME) thereby closing the gap on performance and transparency pertaining to statistics of using AI in and creating portfolio construction.

The remaining sections of the article are as follows: Section 2 contains a review of the literature. Section 3 illustrates the Proposed Methodology. Section 4 introduces Evaluation Metrics. Section 5 contains Results and Discussion and finally, Section 6 consists of Conclusion and Future Scope.

2. Literature review:

Escudero et al [1] demonstrated that in the Deep Reinforcement Learning (DRL), explainability mechanisms can be applied to make financial decisions more transparent. They exploit the financial information of Yahoo Finance and explainable ways of decision-making of agents such as SHAP or LIME to understand what motivates

decision-making of agents and what processes impact them. Hachaïchi et al[2], in their study, dwell upon the application of reinforcement learning strategies to optimize the portfolio management in finance. They contrast new reinforcement learning algorithms such as Advantage Actor-Critic and Proximal Policy Optimization with the older approach such as Mean-Variance Optimization. The research demonstrates that instead of the conventional approaches, reinforcement learning is superior when it comes to such measures as annual returns and risk management index.

Cortés et al [3] have designed an explainable reinforcement learning (RL) to manage financial portfolios, which deal with transparency and black boxes of classical machine learning models. The custom RL model exceeds weighted portfolios and points out the factors that ought to be considered in making investments. Their algorithm has an iterative process of learning and adjusting in a simulated trading environment. R. Sharma et al [4] comment on the use of the deep reinforcement learning (DRL) in market prediction and portfolio optimisation. They contrast DRL models with those of the baseline and conventional strategy models and conclude that the returns (definite) of investing based on the DRL models are more rewarding at overcoming adjusted returns and freestyle in the market than the conventional approach to investing.

M. Mortaji et al [5] presented a literature review that examines reinforcement learning (RL) as applied to the issue of portfolio optimization in terms of the performance of its main elements and the development of new trading rules. They mention the experiments in which the RL algorithms underpin the improved risk-adjusted returns and more desired portfolio choice, though various problems were tackled such as overfitting and potential transaction cost. N. Vodnala et al. [6] suggested Deep reinforcement learning techniques can be applied to a portfolio management approach that involves adapting to the market conditions. The model attains, in thirty-stock funds, a Sharpe Ratio of 1.45, 73% cumulative return, 13.3% annual return and 8.5% annual volatility through application of three models. R. Ozalp et al [7] in their article provide a look at recent contributions to, as well as approaches, and open challenges in DRL and IRL to robotic manipulation tasks. They divide the documents into applications and the need of credible AI in robotics that are directed to the improvement of efficiency and accountability.

K. Eikså et al [8] state that an Explainable AI methodology (including decision trees, ProtoDash, and SHAP) is applied to the decision-making process of Deep Reinforcement Learning (DRL) agents in the simulation of traffic situations. The study exposes the shortcomings of the policy and rationality of such agents which further makes it more commercial. Z. Wang et al [9] apply deep reinforcement learning (DRL) in portfolio optimization to trading in stocks, a combination of stock covariance and technical indicators. To trade they model the process as a Markov Decision Process and apply algorithms such as the Advantage Actor-Critic and Proximal Policy Optimization. Their solutions are more successful than benchmarks, which proves that DRL can be successfully applied in a volatile market. X. Jia et al [10] address that PG-PSO is an emerging portfolio optimization solution referred to as a particle swarm optimization optimizer with the policy gradient scheme. The process incorporates Particle Swarm Optimization in reinforcement learning and hence enhances accuracy and reduces manual tuning. Experiments concluded that PG-PSO outperforms conventional algorithms to improve Sharpe ratio of the portfolios of investments.

As per K. Zhang et al [11], Explainable AI (XAI) is very important in making Deep Reinforcement Learning (DRL) models of power system emergency control more readable. To facilitate the human operators in making better decisions and model transparency, they suggest the Deep-SHAP scheme, which is a Deep-SHAP version of the Shapley additive explanations framework. B. A. Luthfianti et al [12] aim to apply Deep Reinforcement Learning (DRL) to model stock portfolio allocation in the LQ45 index of the Indonesian Stock Exchange, putting the classic approaches, such as Mean-Variance to critical, and show superior risk-adjusted outcomes.

The research by P. S. Tan et al [13] applies LIME and SHAP algorithms in Explainable Artificial Intelligence (XAI), when applied to Deep Reinforcement Learning (DRL) controllers to make optimal decisions concerning the extraction of power in photovoltaic systems. They generate both unique agents and agents without duty cycle information with an eye to producing a greater sense of interpretability and to the application of DRL to renewable energy systems. G. Rjoub et al [14] constructed a trust-aware system integrating an approach of Federated Deep Reinforcement Learning and Explainable Artificial Intelligence to boost integrity and decision capabilities of the Autonomous Vehicles. The system safeguards the privacy of its users and gives greenhorn vehicles the capability to exercise intelligence in their decisions which makes it applicable in future transport systems.

A. Heuillet et al [15] use Shapley values to facilitate explainability of cooperative learning dynamics within multiagent reinforcement learning systems. They show that the agents can have different contributions, which imply that their contribution can very well be represented by Shapley values. This provides knowledge about behavior and dynamics within teams and emergent behavior. L. Wei et al [16] implemented an application of Deep Deterministic Policy Gradient algorithm on a dynamic reinforcement learning model of portfolio optimization in illiquid financial markets. They suggest policy networks that have LSTM and CNN layers,

disclosing that trading strategies are determined by learning rate and cash bias.

Y. -H. Miao et al [17] designed a new portfolio management approach developed on a deep reinforcement learning framework and adaptive sampling approach. This would do better than the usual random sampling yielding an increment of Sharpe ratio of 6-7%. The analysis study by K. Zhang et al [18] incorporates Shapley additive explanations (SHAP) into deep reinforcement learning models to make them more interpretable in power system emergency control operations particularly load shedding under-voltage conditions. This increases the clarity and interpretability of decision-making processes of managing power systems using AI.

L. Xucheng et al [19] explored the reinforcement learning (RL) on portfolio optimization and addressed exercises in dynamic financial problems that cannot be resolved using traditional approaches. They reveal the inferior characteristics of the current methods, such as risk aversion and market flexibility, and propose a company with the usage of the concept of RL, being able to optimize the investment performance. A. Tabrez et al [20] proposed Reward Augmentation and Repair via Explanation (RARE) framework to fix the cases of confusion over the tasks to be done and it also offers real-time feedback and clarity of the misunderstanding concerning the incentive elements through interpretations given to fix the human-robot cooperation.

3. Proposed Methodology:

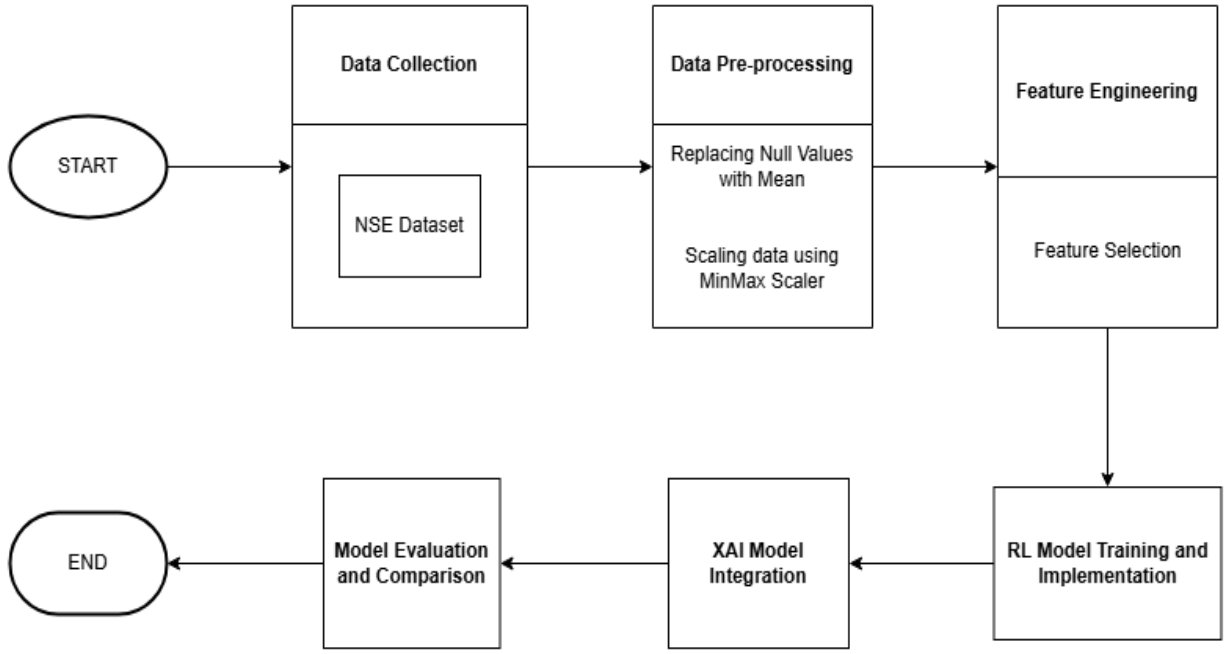

Fig. 1. Portfolio Optimization System architecture

The system design has the goal of building a portfolio that has a high risk-adjusted return. The system consists of the following parts: data collection, data pre-processing, feature engineering, model implementation, and model evaluation and comparison. The framework is specifically targeted to the Indian market and deals with multi-dimensionality and dynamics of the emerging markets. The data is pre-processed to remove noises and inconsistencies, and 14 major features are chosen to train and implement the model.

3.1. Data Preparation and Pre-processing

Data is the basis of research, and its validity, depth, coverage, and genuineness are essential in coming up with good, genuine and accurate results. Data integrity, privacy, and security data are guaranteed when preparing and pre-processing data.

3.1.1. Data Collection: The NSE data of the Indian stock market was developed with the help of the YFinance API, and the stock equity symbols csv file which contains more than 1900 stock symbols of the past year.

Dataset: The dataset generated has 38 features, which are as follows:

Date: Consists of the dates for the market data.

Symbol: Holds the name of the stock.

Open: The opening price of the stock for the day.

High: The highest price a stock reached on that day.

Low: The lowest price stock reached that day.

Close: The closing price of the stock on that day.

Volume: Volume of stock shares traded on that day.

The remaining features in the dataset are: Actual Price, Adj. Close, P/E Ratio, EPS, Market Cap, Dividend Yield, P/B Ratio, Beta, MA20, MA50, MA100, Volatility, RSI, EMA12, EMA26, MACD, Signal Line, OBV, VMA20, VMA50, CMF, VPT, Daily Return, Cumulative Return, Sharpe Ratio, Sortino Ratio, Drawdown, Maximum Drawdown, Annualized Volatility, Future Return and Signal.

3.1.2. Data Preprocessing: The process of data preprocessing is essential to the models so that they can remove noises as well as inconsistencies. SimpleImputer erases the null values, or the value will be filled with the mean, and MinMaxScaler normalizes data to the appropriate range.

3.1.3. Feature Engineering: The feature engineering step is to calculate the correlation between features and eliminate the highly correlated features and keep one of the sets to be executed.

After feature engineering, the number of features that remain are 14, viz. Date, Symbol, Close, Actual Price, Volume, MACD, Signal Line, Cumulative Return, Sharpe Ratio, Sortino Ratio, Maximum Drawdown, Annualized Volatility, Future Return and Signal.

3.2. Deep Learning Models:

1. Proximal Policy Optimization(PPO):

Introduction to PPO:

PPO is a policy gradient technique which uses a clipped surrogate objective to get a consistent and stability policy update. As compared to vanilla policy gradient algorithms, which may be unstable because of extremely large policy changes, PPO constrains the scale of policy change that may happen in one update.

How does PPO Model work:

In this project, PPO agent was used as a starting point of DRL to learn optimal asset allocation in the long run. Through interacting with a simulated stock trading system using NSE data, PPO learned to adjust its policy to yield long-term profits containing limited risk at the portfolio. PPO got a wonderful score upon testing stability and reward generalization, hence, viable in use in financial decision making.

2. Advantage Actor-Critic (A2C):

Introduction to A2C:

Synchronous and deterministic variant of the Asynchronous Advantage Actor-Critic (A3C) method is called A2C. It uses two concepts: an actor decides what to do and a critic analyses this decision. The benefit function minimizes variation allowing maximum learning.

How to work A2C Model:

At the portfolio optimization setting A2C was trained on value estimate (through critic) and policy-update (through actor) by evaluating the asset state and making a prediction of the following optimum allocation action. It exhibited a competitive performance especially in capturing the momentum of the reward in turbulent market circumstances.

3. Soft Actor-Critic (SAC)

Introduction to SAC:

SAC is an off-policy algorithm using actor critic architecture and making use of entropy maximization to provide policy improvement. It enables the model to investigate an increasing number of different strategies, and do not lead to premature convergence on less than purely optimal strategies.

How does SAC Model work:

Under the scenario of portfolio, SAC did use entropy-aware learning to perform a trade-off between exploration and exploitation by adjusting weights in a diverse stock universe. It also performed better with superior cumulative returns but with control over volatility and drawdown, which shows that it had the capability to take sound and confident decisions in an extreme situation.

4. Deep Deterministic Policy Gradient (DDPG)

Introduction to DDPG:

It is an off-policy model-free method of continuous action spaces that leans on the DDPG algorithm. It is run in an actor-critic architecture, when the actor suggests the actions and the critic verifies them along with target networks and experience replay to improve stability.

How does DDPG Model work:

In this implementation, DDPG was applied in seeing portfolio allocation as a continuous action problem and calculated fractionated investment weights on stocks. It did extremely well in initial learning, but was not more consistent than SAC and PPO, probably because it converges slowly in such high dimensional action spaces and is sensitive to reward signal noise.

The reward function is based on a weighted ensemble of Sharpe, Sortino, negative volatility, and drawdown and the penalties on weight concentration and the standard deviation.

3.3. Explainable AI (XAI) Model:

In the context of reinforcement learning (RL) based portfolio models, the aspects of model tractability can and usually do play a secondary role to model performance. But explainability is the requirement of trust-based and regulation-free financial decision-making. The strategies of explainable AI (XAI) provide the missing links between a high-performing model into a black-box and the explanations that can be communicated to people who use them to explain the logic behind the decision that was made by a given model. This would especially be of great essence in cases like trading in stocks whereby absurd or incomprehensible distributions can lead to suspicions among stakeholders.

1. Local Interpretable Model-agnostic Explanations:

Introduction to LIME:

The standard post-hoc XAI algorithm is LIME (Local Interpretable Model-agnostic Explanations) that will give a local fidelity explanation, which should approximate model behavior around a certain prediction. It directly distorts the input data and observes the reaction in the output and locally approximates the action of the complex design by an easy-to-understand design (say a linear regression) to give a cheap assessment of the local change in the output.

How does LIME work:

In this project we applied LIME for the final selection of stock in the PPO model to explain how some stocks were chosen, and some dropped. There are two imagined cases that have been examined:

Case 1: A ‘Selected’ Stock.

As LIME has mentioned, a high Sharpe Ratio, relatively low drawdown, and positive RSI were the reasons why one of the leading stocks had to be chosen. The model valued an asset that was less volatile, and this suited with its one-year horizon of risk-adjusted returns.

Case 2: A non-selected stock.

Conversely, in situations of a passing company by the PPO model, LIME revealed that unwanted volatility and negative Sortino Ratio were major reasons why the said company was rejected. The model will learn that it should not focus on return, although its attainment may experience short term spikes, due to unforeseeable downside risk.

4. Evaluation Metrics:

Our performance metrics, such as Cumulative Return, Sharpe Ratio, Sortino Ratio, Annualized Volatility, and Maximum Drawdown offers an insight about how portfolios perform in terms of risk and returns and helps us gauge the credibility of portfolio optimization systems.

1) **Cumulative Return (C.R.)**: Cumulative Return measures the percentage change in portfolio value over the full investing period.

   C.R = [(Portfolio Value at the end - Initial Portfolio Value) / Initial Portfolio Value] ×100%

2) **Sharpe Ratio:** In 1966, William F. Sharpe introduced the Sharpe ratio as a risk-adjusted return measure, evaluating an investment's excess return over the risk-free rate per unit of volatility. A higher ratio indicates better risk-adjusted performance, as it measures the portfolio's ability to reward investors for their risk level.

   Sharpe Ratio = (Return of Portfolio – Risk free rate) / Std. Deviation of Portfolio

3) **Sortino Ratio:** The Sortino ratio assesses risk-adjusted return by focusing just on downside risk. The downside deviation (DD) is the standard deviation of negative returns.

   Sortino Ratio = (Return of Portfolio – Risk Free Rate) / Downside Deviation

4) **Annualized Volatility (A.V.):** Annualized Volatility measures the standard deviation of the portfolio returns over the year.

   Annualized Volatility = √252 × Std. Deviation of Daily Returns

5) **Maximum Drawdown (MDD):** MDD identifies the worst-case scenario by measuring the biggest loss from the peak over the whole trading session.

   $M.D.D = max_{t\in[0,\ T]}[(Peak\ Portfolio\ Value – Portfolio\ Value\ at\ t) / Peak\ Portfolio\ Value]$

   where, T represents the total number of time periods.

5. Experimental results and analysis:

Four instances of the DRL algorithms were applied in the experiments of portfolio optimization that were trained and tested in a combined simulation environment based on final metrics and compared with each other.

5.1. Performance Comparison of DRL Models

To assess the effectiveness of the DRL algorithms for portfolio optimization, four models were trained - PPO, A2C, SAC and DDPG and were evaluated using the same dataset, reward function and portfolio rebalancing technique. Each model gave a final rebalanced portfolio, created from the learned stock preferences and capital allocation, through which the metrics were calculated.

Table 1. Comparison of Portfolio Performance Metrics across DRL Models

MODEL	Cumulative Return (%)	Sharpe Ratio	Sortino Ratio	Annualized Volatility (%)	Maximum Drawdown (%)
PPO	34.04	1.3972	1.9462	23.58	-46.93
A2C	27.05	1.1622	1.6553	23.61	-47.51
SAC	12.59	0.5735	0.8533	28.57	-35.14
DDPG	24.02	0.9995	1.4781	25.52	-47.72

Table-1 shows the Cumulative Return, Sharpe Ratio, Sortino Ratio, Annualized Volatility and Maximum Drawdown for all the four models. PPO and A2C attained the best risk-adjusted returns, with Sharpe Ratios of 1.39 and 1.16 respectively. On the contrary, SAC performed the worst amongst the four models, having weak metrics.

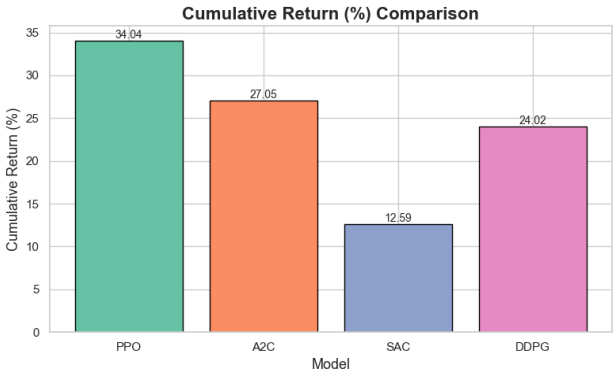

Fig 2. Cumulative Return Comparison across Models

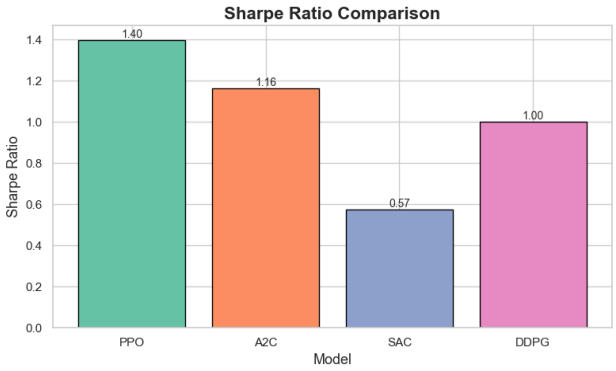

Fig. 3 Sharpe Ratio Comparison across Models

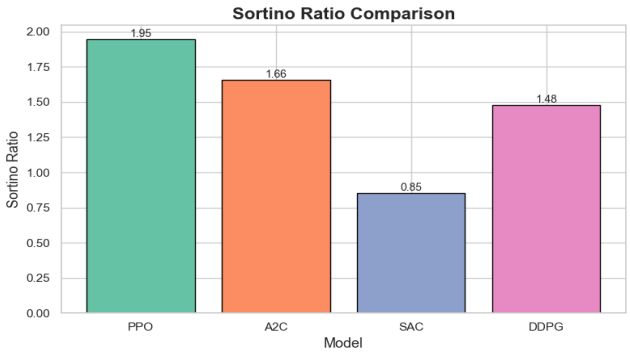

Fig. 4 Sortino Ratio Comparison across Models

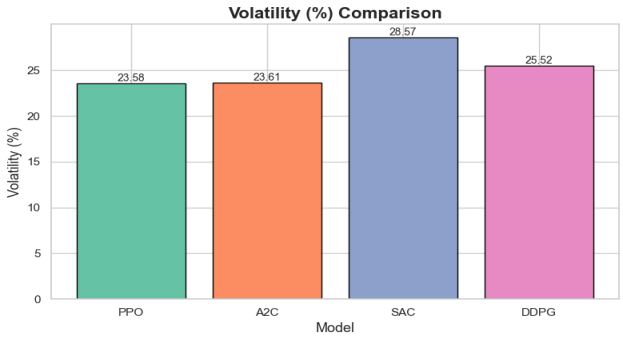

Fig. 5 Volatility Comparison across Models

Figures 2-5 show a comparison of the metrics of performance between the models, and it indicates that A2C has come up with moderate returns and competitive volatility whereas SAC is not as good. On policy and off policy methods are both useful however PPO provides the best performance and consistency.

5.2. Final Portfolio Composition (PPO)

The PPO model performance was better as a budget is spread on best performing stocks and budget stays at the best performing stocks after the score, filter, and rebalancing, hence maximizing return possibilities as well as diversification.

Table 2. Final PPO Portfolio: Stock Level Allocation

Stock Symbol	Share Price (in ₹)	Number of Shares	Capital Allocated (in ₹)
KAVVERITEL	47.53	131	6226.43
KERNEX	1296.0	4	5184.00
SHAKTIPUMP	995.65	6	5973.90
SHREERAMA	45.78	136	6226.08
NILASPACES	12.21	511	6239.31
SABTNL	591.85	10	5918.50
21STCENMGM	64.30	97	6237.10
TCIFINANCE	17.09	365	6237.85

HSCL	469.15	13	6098.95
TREJHARA	233.51	26	6071.26
GMRP&UI	108.74	57	6198.18
TIL	363.45	17	6178.65
RKDL	30.30	206	6241.80
MICEL	62.65	99	6202.35
SAMPANN	28.86	216	6233.76
ATLANTAA	36.14	172	6216.08
Total Invested: ₹97684.2 Remaining Balance: ₹2315.8 Final Total Asset: ₹134036.14			

Table-2 displays PPO's chosen stocks, as well as the number of shares owned, market prices, allocated values, Sharpe Ratios, Sortino Ratios, volatility, and drawdown data. The model excluded penny stocks (₹ < 10) and over-concentrated bets, creating a balanced portfolio structure.

5.3. Explainability with LIME (PPO only)

To improve interpretability, the Local Interpretable Model-Agnostic Explanations (LIME) framework was used for the PPO model output. LIME was used to investigate why specific stocks were included or removed from the final portfolio based on feature importance determined by the model's prediction inputs.

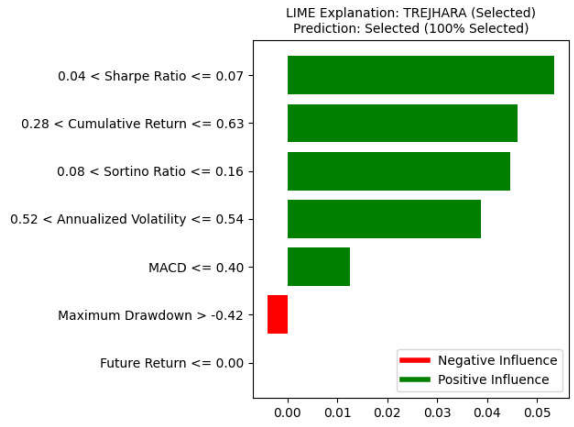

Fig.6. LIME Output – Feature Importance for a Selected Stock (PPO)

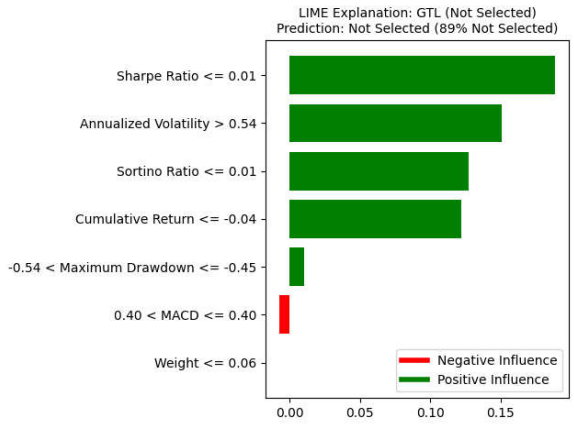

Fig.7. LIME Output – Feature Importance for a Rejected Stock (PPO)

Fig.6 is a LIME visualization of a PPO-sampled stock symbol, which strongly contributes to inclusion because of good Sharpe Ratio, Sortino Ratio, and Cumulative Return, and Fig.7 is a stock that has been rejected because of low measures, returns, and volatility.

These graphics show that Reinforcement learning with LIME would improve the accountability of the model and prove the financial feasibility of the policy and establish the trust of the investor.

6. Conclusion and Future Scope:

This paper presents the deep reinforcement learning (DRL) methods to optimize portfolio of stocks in the Indian stock market, and the best one is PPO. LIME has explainability and solves transparency, that is, when using AI-driven financial models.
Future work may take the models to a multi-agent context, consider multi-agent settings, state-of-the-art SHAP explanations to study explainability, and add macroeconomic signals to make them more consistent with the market outcomes. Overall, our study tried to execute a new and feasible form of Deep Learning implementation integrated with Explainable AI, in retail and institutional finance with a focus on the requirement of transparency and risk-adjusted assessment.

References:

1. Escudero, Alejandra de la Rica, Eduardo C. Garrido-Merchan, and Maria Coronado-Vaca. "Explainable Post hoc Portfolio Management Financial Policy of a Deep Reinforcement Learning agent." arXiv preprint arXiv:2407.14486 (2024).

2. Hachaïchi, Yassine & Lanwer, Amir. (2024). “Benchmarking Reinforcement Learning (RL) Algorithms for Portfolio Optimization.” 10.13140/RG.2.2.20482.70085.

3. Cortés, D. G., Onieva, E., Pastor, I., Trinchera, L., & Wu, J. (2024). “Portfolio construction using explainable reinforcement learning.”, Expert Systems, e13667. https://doi.org/10.1111/exsy.13667.

4. R. Sharma, A. Sharma, S. Hariharan and V. Jain, "Adaptive Investment Strategies: Deep Reinforcement Learning Approaches for Portfolio Optimization," 2024 4th International Conference on Intelligent Technologies (CONIT), Bangalore, India, 2024, pp. 1-5, doi: 10.1109/CONIT61985.2024.10627674.

5. M. Mortaji, A. Khiat and M. Benhouad, "Reinforcement learning application in portfolio optimization: a comprehensive literature review," 2024 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 2024, pp. 1-6, doi: 10.1109/ISCV60512.2024.10620129.

6. N. Vodnala, P. Yarlagadda, P. Vamsikrishna and L. Tejaswini, "Dynamic Portfolio Management Using Multi-Model Reinforcement Learning," 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2024, pp. 1-4, doi: 10.1109/ICITEICS61368.2024.10625331.

7. R. Ozalp, A. Ucar and C. Guzelis, "Advancements in Deep Reinforcement Learning and Inverse Reinforcement Learning for Robotic Manipulation: Toward Trustworthy, Interpretable, and Explainable Artificial Intelligence," in IEEE Access, vol. 12, pp. 51840-51858, 2024, doi: 10.1109/ACCESS.2024.3385426.

8. K. Eikså, J. E. Vatne and A. M. Lekkas, "Explaining Deep Reinforcement Learning Policies with SHAP, Decision Trees, and Prototypes," 2024 32nd Mediterranean Conference on Control and Automation (MED), Chania - Crete, Greece, 2024, pp. 700-705, doi: 10.1109/MED61351.2024.10566218.

9. Z. Wang, S. Jin and W. Li, "Research on Portfolio Optimization Based on Deep Reinforcement Learning," 2022 4th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Shanghai, China, 2022, pp. 391-395, doi: 10.1109/MLBDBI58171.2022.00081.

10. X. Jia and X. Cai, "A Policy Gradient Based Particle Swarm Optimizer for Portfolio Optimization Problem," 2022 41st Chinese Control Conference (CCC), Hefei, China, 2022, pp. 1991-1996, doi: 10.23919/CCC55666.2022.9901620.

11. K. Zhang, J. Zhang, P. -D. Xu, T. Gao and D. W. Gao, "Explainable AI in Deep Reinforcement Learning Models for Power System Emergency Control," in IEEE Transactions on Computational Social Systems, vol. 9, no. 2, pp. 419-427, April 2022, doi: 10.1109/TCSS.2021.3096824.

12. B. A. Luthfianti, D. Saepudin and A. F. Ihsan, "Portfolio Allocation of Stocks in Index LQ45 using Deep Reinforcement Learning," 2022 10th International Conference on Information and Communication Technology (ICoICT), Bandung, Indonesia, 2022, pp. 205-210, doi: 10.1109/ICoICT55009.2022.9914892.

13. P. S. Tan, T. B. Tang and E. T. W. Ho, "Explainable Artificial Intelligence Applied to Deep Reinforcement Learning Controllers for Photovoltaic Maximum Power Point Tracking," 2022 International Conference on Future Trends in Smart Communities (ICFTSC), Kuching, Sarawak, Malaysia, 2022, pp. 1-5, doi: 10.1109/ICFTSC57269.2022.10040061.

14. G. Rjoub, J. Bentahar and O. A. Wahab, "Explainable AI-based Federated Deep Reinforcement Learning for Trusted Autonomous Driving," 2022 International Wireless Communications and Mobile Computing (IWCMC), Dubrovnik, Croatia, 2022, pp. 318-323, doi: 10.1109/IWCMC55113.2022.9824617.

15. A. Heuillet, F. Couthouis and N. Díaz-Rodríguez, "Collective eXplainable AI: Explaining Cooperative Strategies and Agent Contribution in Multiagent Reinforcement Learning With Shapley Values," in IEEE Computational Intelligence Magazine, vol. 17, no. 1, pp. 59-71, Feb. 2022, doi: 10.1109/MCI.2021.3129959.

16. L. Wei and Z. Weiwei, "Research on Portfolio Optimization Models Using Deep Deterministic Policy Gradient," 2020 International Conference on Robots & Intelligent System (ICRIS), Sanya, China, 2020, pp. 698-701, doi: 10.1109/ICRIS52159.2020.00174.

17. Y. -H. Miao, Y. -T. Hsiao and S. -H. Huang, "Portfolio Management based on Deep Reinforcement Learning with Adaptive Sampling," 2020 International Conference on Pervasive Artificial Intelligence (ICPAI), Taipei, Taiwan, 2020, pp. 130-133, doi: 10.1109/ICPAI51961.2020.00031.

18. K. Zhang, P. Xu and J. Zhang, "Explainable AI in Deep Reinforcement Learning Models: A SHAP Method Applied in Power System Emergency Control," 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2), Wuhan, China, 2020, pp. 711-716, doi: 10.1109/EI250167.2020.9347147.

19. L. Xucheng and P. Zhihao, "Portfolio Optimization Under the Framework of Reinforcement Learning," 2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Qiqihar, China, 2019, pp. 799-802, doi: 10.1109/ICMTMA.2019.00180.

20. A. Tabrez, S. Agrawal and B. Hayes, "Explanation-Based Reward Coaching to Improve Human Performance via Reinforcement Learning," 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea (South), 2019, pp. 249-257, doi: 10.1109/HRI.2019.8673104.