Grenze International Journal of Engineering and Technology, January Issue

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

Rishabh Chorasiya[1] and Shakti Kinger[2]
[1-2]Department of Computer Engineering and Technology, Dr. Vishwanath Karad's MIT-World Peace University, Pune, India
Email: rishabh.chorasiya@mitwpu.edu.in, shakti.kinger@mitwpu.edu.in

***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 to high-performing stocks and over-concentration. Compared to the prior performance, risk-adjusted performance is evaluated, using Sharpe Ratio, Sortino Ratio, Annualized Volatility and Maximum Drawdown in the model. Evaluation metrics indicate PPO was superior in comparison to the other models, with a Sharpe Ratio of 1.3972 and Cumulative Return of 34.04%. The framework demonstrates how DRL and XAI are effective in building transparent and robust in-vestment portfolios, specific to emerging markets such as India are concerned.**

***Keywords*— Deep Reinforcement Learning, Explainable AI, Indian Stock Market, Portfolio Optimization, Proximal Policy Optimization.**

I. INTRODUCTION

Investments in the stock market give huge returns and give losses when due risk procedures are side-lined. Portfolio Optimization forms one of the significant areas in finance that aims to optimize the returns against the risks. The traditional ML methods like Markowitz Mean-Variance optimization are simple and they do not even tackle the new complexities of the markets that entail 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 im-proved through the incorporation of eXplainable AI (XAI) elements, which will help to make them more recogniza-ble 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 or low risks.

Grenze ID: 01.GIJET.12.1.558_1

Our research intends to use four DRL agents, viz. Proximal Policy Optimization (PPO), Advantage Actor-Critic (A2C), Soft Actor-Critic (SAC) and Deep Deterministic Policy Gradient (DDPG), to optimize portfolios on the Indian stock market. The model employs transparency and reasonability methods which is Local Interpretable Model-agnostic Explanations (LIME), thereby closing the gap on performance and transparency pertaining to statistics of using AI in creating portfolio construction.
The remaining sections of the article are as follows: Section II contains a review of the literature. Section III illus-trates the Proposed Methodology, while Section IV describes the Evaluation Metrics. Section V contains Results and Discussion and Section VI, the last section, consists of Conclusion and Future Scope.

II. 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 mak-ing of agents and what processes impact them. Hachaïchi et al.[2], in their study, dwell upon the application of rein-forcement 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, Reinforce-ment 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 DRL in market prediction and portfolio optimization. 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 invest-ing.
M.Mortaji et al.[5] presented a literature review that examines 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 experi-ments 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] suggest-ed DRL techniques can be applied to a portfolio management approach that involves adapting to the market condi-tions.
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] provides a look at recent contribu-tions 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 towards the improvement of efficiency and accountability.
K.Eikså et al.[8] state that an Explainable AI methodology (including Decision Trees, ProtoDash and SHAP) is ap-plied to the decision-making process of 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 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 A2C and PPO. 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 Optimiza-tion in RL and hence enhances accuracy and reduces manual tuning. Experiments concluded that PG-PSO outper-forms conventional algorithms to improve Sharpe Ratio of the portfolios of investments.
As per K.Zhang et al.[11], Explainable AI(XAI) is particularly important in making DRL models of power system emergency control more readable. To facilitate 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 frame-work. B.A.Luthfianti et al.[12] aim to apply DRL to model stock portfolio allocation in the LQ45 index of the Indo-nesian 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] explains LIME and SHAP algorithms in XAI, when applied to 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 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 DRL and XAI to boost integrity and decision-making capabilities of the Autonomous Vehi-cles. The system safeguards the privacy of its users and gives greenhorn vehicles the capability to exercise intelli-gence in their decisions which makes it applicable in future transport systems.

A.Heuillet et al.[15] use Shapley values to facilitate explainability of co-operative learning dynamics within multi-agent RL systems. They show that the agents can make different contributions, which implies that their contribution can very well be represented by Shapley values. This provides knowledge about their behavior and dynamics within teams and emergent behavior. L.Wei et al.[16] implemented an application of DDPG algorithm on a dynamic RL model of portfolio optimization in volatile 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 DRL 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 SHAP into DRL 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 RL on portfolio optimization and addressed exercises in dynamic financial problems 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 misunder-standing concerning the incentive elements through interpretations given to fix human-robot co-operation.

III. Proposed Methodology

The System Architecture of the proposed methodology, depicted in Figure 1, was developed with the aim of building a portfolio, which gives a high risk-adjusted return on investment. The proposed system was designed based on the proposed system developed by N. Vodnala et al.[6] in their research work, where they used the data of thirty selected stocks from the Dow Jones index, and fed that data to the three DRL models they selected for their research, i.e. PPO, A2C and DDPG. They decided to train their models for a 3-month period, so that they adapt to the market conditions, then validate them over the next 3-month period based on the Sharpe Ratio, and the model with the best result will be used for trading for the next 3 months. This process was repeated till the end of their data.

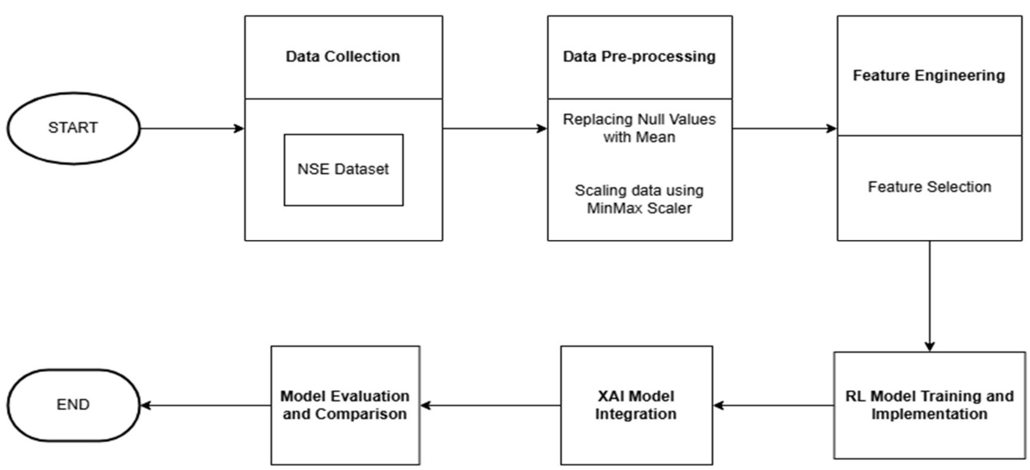

Figure 1. System Architecture for Portfolio Optimization

Taking inspiration from their model, the proposed system works on similar grounds. Initially, we generate the NSE dataset of about 1900 stocks with most of the required features for stocks over the last one year, using the YFinance API and the equity symbols. The dataset is bound to contain lot of noise and inconsistencies, therefore, to overcome these obstacles, the data pre-processing and normalization step becomes a vital part of architecture. The potential null/NaN/missing values are replaced by the mean, followed by scaling the features using MinMax Scaler to improve the performance of the models. Using correlation matrix, we observed that many features in the dataset were highly correlated, resulting in the conclusion that we can keep only one feature from the set of highly correlated ones, while dropping the other features which may not prove to be helpful in producing expected results.

After sorting and selecting the features which are to be fed to the DRL models, which gave us a set of 14 features for implementation, the architecture moves on to model training and implementation, where the four selected models, which are PPO, A2C, SAC and DDPG, will be trained on the complete dataset and learn the patterns and trends, which will be effective in producing the desired results: a set of top-n investible stocks which will give the reasona-ble returns with minimal risks.

The environment developed to train the four DRL models takes the data frame, the initial balance and number of stocks required in the portfolio as parameters. Constraints, like the total portfolio amount shouldn't exceed the initial balance, prevention of selecting penny stocks, over-concentration of capital to certain stocks and at least 80% of the required n stocks being selected in the portfolio, were diligently taken care of inside the environment for training the models.

This included the rebalancing of the capital among the stocks, which was a preventive measure for over-concentration of capital to only the top stocks while less or none of the capital was being allocated to other stocks in the portfolio. The environment returns the reward as the result or feedback to the models. All the four models were trained on 50000 timesteps, to maintain consistency across the models and evaluate their performance on similar grounds.

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

A. Proximal Policy Optimization (PPO)

Proximal Policy Optimization (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 constraints the scale of policy change that may happen in one up-date.

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 reasonable score upon testing stability and reward generalization, hence, viable in use in financial decision making.

B. Advantage Actor-Critic (A2C)

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

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

C. Soft Actor-Critic (SAC)

Soft Actor-Critic (SAC) is an off-policy algorithm using actor-critic architecture and making use of entropy maximi-zation to provide policy improvements. It enables the model to investigate an increasing number of different strate-gies and does not lead to premature convergence on less than purely optimal strategies.

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 re-turns but with control over volatility and drawdown, which shows that it had the capability to take sound and confi-dent decisions in an extreme situation.

D. Deep Deterministic Policy Gradient (DDPG)

Deep Deterministic Policy Gradient (DDPG) is an off-policy model-free method of continuous action spaces that lean on 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.

In this implementation, DDPG was applied in seeing portfolio allocation as a continuous action problem and calcu-lated 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 Ratio, Sortino Ratio, negative volatility and draw-down and the penalties on weight concentration and standard deviation.

E. Local Interpretable Model-agnostic Explanations (LIME)

The standard post-hoc XAI algorithm is LIME 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.

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 probable 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 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 ‘Not-Selected’ Stock.

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

IV. Evaluation Metrics

Our performance metrics, which are: Cumulative Return, Sharpe Ratio, Sortino Ratio, Annualized Volatility and Max-imum Drawdown, offer an insight into how portfolios perform in terms of risk and returns and help us gauge the credibility of portfolio optimization systems.

- Cumulative Return: Cumulative Return measures the percentage change in the portfolio value over the full in-vesting period.

$$Cumulative\ Return = \frac{Portfolio\ Value\ at\ the\ end - Initial\ Portfolio\ Value}{Initial\ Portfolio\ Value} \times 100\%$$

- 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 the investors for their risk level.

$$Sharpe\ Ratio = \frac{Return\ of\ Portfolio - Risk\ free\ rate}{Standard\ Deviation\ of\ Portfolio}$$

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

$$Sortino\ Ratio = \frac{Return\ of\ Portfolio - Risk\ free\ rate}{Downside\ Deviation}$$

- Annualized Volatility: Annualized Volatility measures the standard deviation of the portfolio returns over the year.

$$Annualized\ Volatility = \sqrt{252} \times Standard\ Deviation\ of\ Daily\ returns$$

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

$$Maximum\ Drawdown = max_{t\in[0,T]}\left(\frac{Peak\ Portfolio\ Value - Portfolio\ Value\ at\ t}{Peak\ Portfolio\ Value}\right)$$

where T represents the total number of time periods.

V. Results and Discussion

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

A. Performance Comparison of DRL Models

To assess the effectiveness of DRL algorithms for portfolio optimization, four models were trained – PPO, A2C, SAC and DDPG. They 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 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. Meanwhile, SAC performed the worst amongst the four models, having the weakest metrics.

TABLE I: Comparison of 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

Figure 2 to Figure 5 shows 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

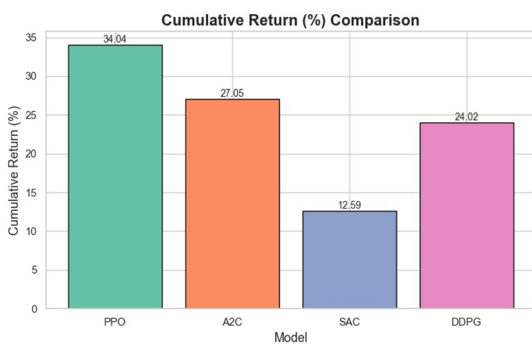

Figure 2. Cumulative Return Comparison across Models

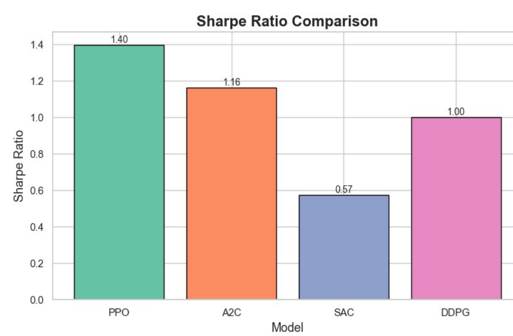

Figure 3. Sharpe Ratio Comparison across Models

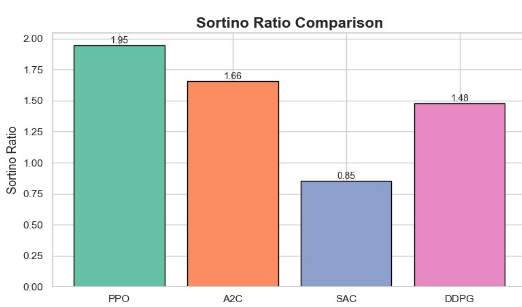

Figure 4. Sortino Ratio Comparison across Models

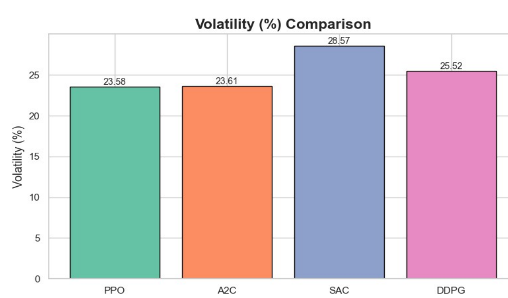

Figure 5. Volatility Comparison across Models

B. Final Portfolio Composition (PPO)

The PPO model performance was better as the budget is spread amongst the best performing stocks and budget stays at the best performing stocks after the score, filter and rebalancing, thereby, maximizing return possibilities as well as diversification. Table-2 displays PPO's chosen stocks, as well as the number of shares owned, market prices, allocat-ed values, Sharpe Ratio, Sortino Ratio, Volatility and Drawdown data. The model excluded penny stocks (share price < ₹10) and over-concentrated bets, creating a balanced portfolio structure.

TABLE II. 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			

C. Explainability with LIME (PPO only)

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

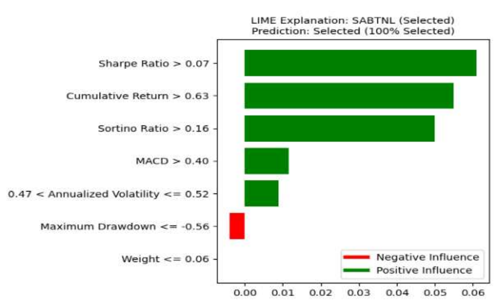

Figure 6. LIME Output - Feature Importance of a Selected Stock (PPO)

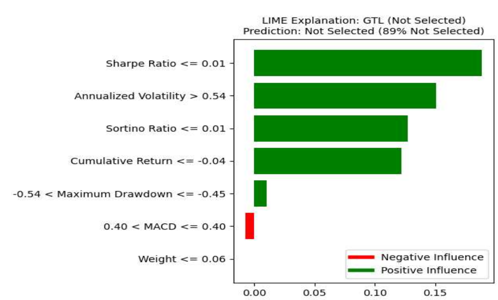

Figure 7. LIME Output -Feature Importance for Not-Selected Stock (PPO)

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

VI. CONCLUSION

This paper presents the DRL methods to optimize portfolio of stocks in the Indian stock market. PPO performed the best amongst the four DRL models selected. On the other hand, SAC had underwhelming performance metrics. We integrated XAI model to provide insights and transparency for the investors as to what features impacted the selec-tion of a certain stock in the portfolio and why a certain stock faced rejection. In our case, LIME helps in providing this explainability and solves transparency, that is, when using AI-driven financial models, by giving a visual representation of the features impacting the decisions taken by the DRL models, in this case the PPO model.

Future work may take the models to a multi-agent context, consider multi-agent settings, state-of-art SHAP explanations to study explainability and add macro-economic signals to make them more consistent with the market out-comes. Also, DRL models can be combined with other complex models like XGBoost, to enhance the selection process and extract even better returns and performance. Overall, our study tried to execute a new and feasible form of Deep Learning implementation integrated with XAI, 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 and 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., and 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 (CO-NIT), 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 G. Guzellis, "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), Chaina-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), Sanghai, 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), Hefel, 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 Rein-forcement 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 Communication and Mobile Computing (IWCMC), Dubrovnik, Croatia, 2022, pp. 318-323, doi:10.1109/IWCMC55113.2022.9824617.

[15] A. Heuillet, F. Couthouis and N. Diaz-Rodríguez, "Collective eXplainable AI: Explaining Cooperative Strategies and Agent Contribution in Multiagent Reinforcement Learning with Shapley Values," in IEEE Computational In-telligence 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 Ap-plied 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. Agarwal 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.