Revolutionizing Active Investing with Machine Learning

Mukul Pal, Regina Fenesi, Oliviu David Cigan, Alexandru George Berciu, Ciprian Radu Tiric, Florina Pal, Dan Todor, Prof. Eva Dulf

Dec 25, 2023

Abstract

This paper introduces a novel machine learning method aimed at enhancing the capabilities of active asset managers in navigating the complexities of selection systems. These complexities encompass idiosyncratic factors, subpar prediction methods, and competition from entrenched systems like Market Capitalization (MCAP) weighted benchmarks, which have historically challenged the industry. Active managers often fail to outperform the market, leading to a decline in the active asset management sector's growth, contrasted with the rise of passive management, especially post the introduction of Index funds in 1976. The authors apply the [3N] non-linear systems approach, starting by redefining MCAP benchmarks as closed systems. They introduce "Discrete Decile Steps," a method that categorizes stock information into dynamic states that rise or fall, indicating performance trends. This approach is augmented with a Random Forest Regressor to predict these states, forming a prediction system independent of idiosyncratic elements, aiding managers in making more effective selections. The paper also presents a thought experiment using elevators as a metaphor for stock movements, illustrating the importance of predicting the direction of these 'elevators' to enhance selection and reduce underperformance risk.

The authors express their gratitude to the Adapted Research Group at the Technical University of Cluj-Napoca for their support.

I - Introduction and Literature Review

-The Underperformance

The problem of active managers underperforming against benchmarks is central to the stock market forecasting challenge. The underperformance problem has been well-documented in financial literature, notably Cowles' “Can Forecasters Forecast?” [1] and Fama's “Efficient Market Hypothesis,” [2] both highlight the limitations of predictive accuracy in financial markets. Granger's “Forecasting Challenge” [3] further explores these difficulties. Behavioral Finance theories from Kahneman [4], Thaler [5], Shiller [6] have articulated inefficiency but have failed to come up with work that can solve the underperformance problem. A few economists have even come with the theory of not using Physics [7] to solve the problem and rather relying on Adaptive Mechanisms [8], which also have proved inconclusive. Additionally, the S&P Dow Jones Indices' SPIVA reports regularly shed light on the percentage of active managers failing to outperform market benchmarks.

- Context and Non-Linear Mechanisms

Markowitz's "Portfolio Selection" from The Journal of Finance (1952) [9] provides a historical foundation, contrasting traditional Modern Portfolio Theory (MPT) with contemporary thoughts of context. Markowitz's seminal work on efficient frontier provided a risk-return context for investment strategies.

Benoit Mandelbrot's seminal paper, "The Variation of Certain Speculative Prices," [10] revolutionized the understanding of financial markets by highlighting their non-linear dynamics. He introduced concepts like fractal geometry and heavy-tailed distributions, demonstrating that market behaviors were far more complex and erratic than traditional linear models suggested. However, Mandelbrot’s work primarily focused on describing these irregular patterns, without developing a predictive framework. While it illuminated the intricate nature of market movements, it didn’t offer systematic methodologies for forecasting future trends or prices.

Though researchers like Kenneth Boulding [11] have talked about Informational context and how Information can flit between relevancy and irrelevancy, it was William Sharpe and Andre F Perold, who in their paper “Dynamic Strategies for Asset Allocation” [12], delve into the nuances of stock market investing through the lens of concave and convex payoff, a mechanism thinking suggesting that successful investing depended on the state of the market more than the prevalent news or information at a certain time.

-[3N] Systems

The [3N] Model, Method and Methodology [13] [14] [15] referred to as [3N] Systems presents a novel approach to understanding informational contexts, as market states driven by statistical factors that operate in multiple dimensions are designed to embrace failure as a probabilistic event, unlike conventional factors that fail or success in an absolute way. The non-linear mechanism framework blends concepts from quantum mechanics [16], statistics [17], physics [18], mathematics [19] and considers markets as a complex system, a combination of probabilistic states that dynamically interact with each other.

2 – Background and methodology

- Benchmarks as closed systems

MCAP weighted benchmarks [20] function as closed systems that self-regulate, maintain an internal order, and have a construction that is largely unaffected by extrinsic noise. This inherent structure creates a persistent challenge for active asset managers, whose success hinges on their ability to surpass these benchmarks. Acknowledging benchmarks as closed systems is crucial for active managers seeking to refine their selection systems. It shifts the focus from the unpredictable nature of individual stock performance to the patterns and rules that govern the collective behavior of stocks within a benchmark. This paradigm shift is central to constructing new selection systems that bypass the benchmarks' self-reinforcing nature, thereby enhancing the potential for outperformance.

-Probabilistic overhang for Active Managers

Active asset managers often face a probabilistic overhang in their selection process, akin to a gambler's dilemma. Each stock chosen carries inherent probabilities of success or failure, influenced by myriad market factors. This uncertainty is compounded by the benchmarks' propensity to reinforce existing winners, creating a cycle where today's winners are likely tomorrow's as well. To counteract this, active managers must not only select stocks but predict which will become the new winners before this becomes common knowledge and also underweight or exit the same winners before or at their peak outperformance. The challenge lies in the precision of these predictions, made more complex by the fluidity of market dynamics.

- Next generation predictive systems

To counter these challenges, a new age predictive system is essential. Such systems should harness the probabilistic nature of markets, discerning patterns within the apparent randomness. They should offer a strategic advantage by utilizing contextual information—data that provides insight into the market's behavior and state changes, rather than static, historical data alone. These systems should enable active managers to make informed decisions that correlate with the underlying market mechanisms, aligning more closely with benchmarks' systematic approaches. The goal is to develop a selection system that can anticipate and act on market movements, isolating potential for outperformance amidst the broader trends of the financial markets. This approach emphasizes the need for active managers to shift from a content-based analysis to one steeped in the deeper context of market behavior, utilizing algorithms and models that can learn and adapt to market changes, thus offering a more potent competitive edge.

- Discrete Decile States

Scoring systems are particularly effective in building predictive models for non-linear systems due to several key advantages. Firstly, they offer simplicity and interpretability, translating complex relationships into more straightforward formats, thereby enhancing user understanding of how various variables influence predictions [21]. Secondly, scoring systems are adept at handling non-linearities, a common characteristic in complex systems where variables interact in unpredictable ways. This capability allows them to accurately reflect the intricacies of such interactions [22]. Additionally, these systems tend to reduce the risk of overfitting. By simplifying relationships, they focus on underlying trends rather than noise in the data, enhancing the model's generalizability [23]. Ease of use is another significant advantage, as scoring systems are less resource-intensive and require less expertise to implement, making them ideal for settings where quick decision-making is essential [24]. Their flexibility is also notable, as they can accommodate diverse data types and variable categories, making them versatile across various applications [25]. Lastly, the transparency offered by scoring systems is a critical aspect, especially in fields where understanding the decision-making process is as important as the outcome itself. This transparency is often lacking in more complex models, such as deep neural networks [26].

The training data focused on relative returns, as beating the markets, or outperformance was about higher relative returns. There were 5 look back periods, which was a reduced range compared to the "[3N] Methodology" which considered 9 look back periods. The numeric relative ranking data was transformed into percentile rankings and further discretized into decile states.

Discretization reduced noise, reduced information, transformed continuous percentile (1 to 100%) into a non overlapping range 0-10, >10-20, >20-30, >30-40, >40-50, >50-60, >60-70, >70-80, >80-90, >90-100. As an example, 0-10 encompassed all data from the lowest ranking (0.2%) till 10%. The respective ranges were relabeled as 1 to 10, also called as the Discrete Decile States (DDS). An individual stock could look like the following sequence.

1,2,1,2,1,2,3,4,5,6,5,4,3,4,5,6,7,8,9,8,7,8,9,8,7,6,5,4,5,3,2,1…

Figure 1 explains the process of transforming relative ranking into percentile ranking and then into DDS.

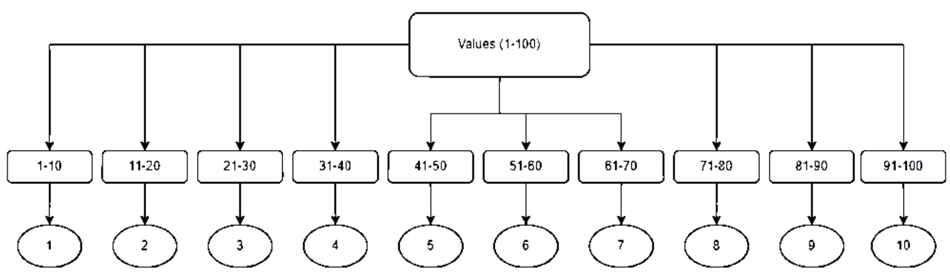


Figure 1: Training Data Preparation from Numeric Rankings to Discrete Decile States

Apart from stock states data, different portfolio states were built, each with different number of stock states. Sometime the stock states were randomly selected and sometime there was a constrained selection. An example was to select just from 1st decile all stock states to make a composite portfolio state. In other words, all components ranked at 1% and below were selected into the portfolio state. These portfolio states were arrays of list of stock states with their chronological sequence of decile states from a certain starting date to ending calendar date. The DDS were super selection systems where context mattered more than what the states contained. The composition of a state was a function of when the state was constructed and under what constraints. This made a state unique. A selection of selections.

Normalizing relative performance into DDS does not remove nonstationary elements from the data because it was rare for a component stock state to move in a complete cycle from bottom of the rankings to the top and back. Moreover, owing to multiple periodicities, the study involved different length and pattern of DDS sequences, which differed in starting point, ending point, a host of other constraints.

This research used S&P 500 Benchmark with the end of day closing prices of 500 stocks. The index measures the performance of 500 publicly traded companies in the United States. The data ranged from January 3, 2000, to March 10, 2023, a period of 23 years.

The S&P 500 is one of the most important [27] and representative stock market indices in the world. Many investors use S&P 500 as a reference benchmark to evaluate performance of investment funds, portfolios and their investment strategies. Therefore, it is an important reference point to monitor the general evolution of the stock market and to compare the performance of individual investments.

The database fields captured date, company names and 5 different relative rankings calculated from end of day price. No end of day price was stored in the database. The names were intentionally anonymized by giving a unique ID to each company. The primary dataset encompasses 296,893 DDS entries. These entries represent a wide range of stocks in various states and conditions.

The portfolio generation is divided into three distinct batches, each with its own unique characteristics and resulting dataset sizes. The first batch involves the creation of 500 portfolios, with each portfolio containing between 3 to 5 stocks. This batch expanded the dataset to approximately 1,137,614 DDS entries. Following this, the second batch consists of 100 portfolios, also comprising 3 to 5 stocks each. The data accumulation from this batch is substantial, amounting to around 229,216 stock DDS. Lastly, the third batch includes the generation of 200 portfolios, each again featuring 3 to 5 stocks. This resulted in a dataset increase to about 463,376 DDS entries.

The key features used to train were sequence of different periodicities (Yearly, Monthly, and Daily). The time spent (number of days) inside a decile state. The current and previous value of the DDS. The data was split into test data and train data. And for simplicity, this paper focused on predicting the immediate next step for the DDS for a certain unique id of stock or portfolio.

In the context of feature selection and data processing for stock market predictions, our approach with Discrete Decile Steps (DDS) method marks a significant departure from traditional methodologies. As outlined in Kumbure et al. (2022) [28], feature selection in stock market research has conventionally focused on filtering and reducing data to identify relevant variables for prediction tasks. These techniques, while effective in reducing overfitting and improving accuracy, often do not fully account for the dynamic and non-linear nature of financial markets. The DDS method, by contrast, not only refines the data but transforms it to capture the inherent dynamism and complexity of market movements. Utilizing machine learning, specifically the Random Forest Regressor, DDS adapts to the evolving patterns and probabilistic nature of market states, offering a more nuanced and effective tool for predicting stock market trends. This innovative approach aligns with the latest advancements in financial market analysis, representing a paradigm shift in how data is interpreted and utilized for investment strategies

-The Elevator and the Building

The prediction of DDS can be vividly illustrated using an enhanced elevator analogy. In this analogy, the stock market is depicted as a high-rise building, where each elevator represents a DDS, and the floors indicate the ranking. The top floor was the top performance and the bottom floor was the worst performance. The elevators, each carrying a set of stocks, move dynamically through these floors. And what mattered was getting in the fastest elevator to the top, given the constraints.

The [3N] systems excel in this environment because of their ability to adapt to non-stationary market data, where traditional models often fail. Unlike methods that transform data to fit stationary models, [3N] embraces the market's inherent non-linearity and evolving temporal patterns. It preserves and leverages these time-dependent characteristics, which is crucial for understanding the sequential nature of market states, similar to tracking an elevator's journey across different floors.

The prediction is akin to understanding how past movements of an elevator (past trend) can influence its future course. This recognition of long-term dependencies and patterns is essential for predicting future states.

The practical applications of solving this prediction challenge are manifold. First, a better prediction of an elevator (state) means the market outperformance of a certain selection i.e. alpha. Second; the predictive accuracy leads to a more informed and strategic investment decision. Third; the understanding of probable market state transitions aids in effective risk management, allowing for preparedness against potential market shifts. Fourth, investors can leverage these insights for strategic portfolio allocation, selecting the right 'elevator' or set of stocks that align with their investment objectives and risk tolerance. Fifth, furthermore, the adaptability to rapid market changes offers a significant advantage over traditional models, making it an indispensable tool in modern financial markets.

- Predicting DDS

Prediction has historically been seen as linked to prices of individual stocks and fundamental information related to them [29]. In the world of stock market forecasting, the industry has always focused on information that is specific to stock prices. What is happening fundamentally to the stock? How is it relatively valued to its local or global peers? What are its future prospects? However, these questions are not enough in a risk weighted alpha seeking world which focuses on Information ratio [30].

To look at prediction of states, which are dynamic, never with the same set of components, something that is contextual and not content specific, can move like elevators in a building, elevators that operates like a variable which can be changed for capacity or can be merged with another elevator, an abstract concept that only lends itself to machine interpretation, is a novel thinking, which challenges conventional thought.

Predicting all the 500 stocks of S&P 500 with precision both for entry and exit, underweight and overweight perspective, is not a monumental task, it’s an impossible task both for human brain and computational machines. Complexity by nature is unpredictable and throwing computation and all human capability to solve the problem is using sledgehammer to crack a nut.

Sometimes, all it requires is to relook at the puzzle. The challenge was never about seeing where a stock price or a specific index can go in the intermediate or longer-term. The challenge was always to understand how a stock is going to outperform the 499 other stocks, understanding the context was more important than interpreting the content. It was not enough to pick a great set of selections but to make sure those selections understood systematic risk i.e. make sure not to be thrown off the horse.

Hence the need for the thought experiment of elevators in the building. Selections that could not outperform the benchmark were not travelling in the right elevator. Any misalignment of selections with their state was destined for failure (underperformance vs. benchmark).

Selections were always secondary if they were not chosen by the states. Predicting states was the only way to beat the market, being fastest at the top of the building. This counter intuitive outcome of states superseding selections is why predictions had a sustained chance to deliver. Context over content, states over selections, objectivity over subjectivity, machine-human augmentation over human emotion, probability over heuristics, machine curated portfolio over stock selections, there are a ton of other reasons why forecasting states is easier than selections. Training data generated by a non-linear system for dynamic states that work as functions with parameters is bound to be richer than chronological data for stocks that is content biased.

The authors have demonstrated that machines can learn and predict states better than stocks and if this could be done, active investing could be revolutionized with Machine Learning, giving active managers a new way to think contextually about their preferences, differentiate themselves from their peers, and exploit the gold rush for sustained alpha.

-Random Forest Regressor

The authors of the paper tested AdaBoost Classifier [31], Random Forest Classifier and XGBoost Classifier, Random Forest Classifier and Random Forest Regressor [32]. The regressor had the best metrics as depicted in Table 1.

The Random Forest Regressor stands out as an exceptional tool for forecasting in the context of Discrete Decile States, primarily due to its sophisticated handling of non-linear data and its adaptability to the complex nature of financial markets.

Ensemble Learning Approach: The Random Forest Regressor employs an ensemble learning technique, combining multiple decision trees to form a more robust model. This methodology is particularly effective in capturing the non-linear relationships prevalent in financial data. By averaging the outcomes of various trees, it reduces the risk of overfitting, which is a common problem in models dealing with complex, non-linear mechanisms.

Flexibility with Data Types: It can process a diverse range of data types, a critical feature for financial markets where data comes in various formats. This versatility ensures that the model remains effective across different scenarios and data structures.

Dimensionality Reduction and Pattern Recognition: The model excels in reducing data complexity, making it easier to interpret and analyze. It also identifies and learns from patterns in historical data, which is crucial for forecasting future market states.

Probabilistic and Dynamic Nature of Markets: Financial markets are characterized by their probabilistic and dynamic nature. The Random Forest Regressor, with its ability to handle continuous and probabilistic data, aligns well with this aspect. It effectively predicts the likelihood of different market states, adapting to the evolving patterns seen in financial data.

Feature Importance Evaluation: In financial forecasting, understanding which features significantly influence the outcome is crucial. The Random Forest Regressor provides insights into feature importance, enabling better interpretation and understanding of the underlying factors driving market movements.

Handling Non-Linearity and Evolving Temporal Patterns: The inherent non-linearity and continuously evolving temporal patterns of financial markets require a model that can adapt and learn from these changes. The Random Forest Regressor is adept at managing these complexities, capturing nuanced behaviors and hidden patterns within the market data.

	MAPE	RMSE	R2 SCORE
AdaBoost Classifier	15.85%	1.31	57.51%
Random Forest Classifier	21.85%	1.48	59.00%
XGBoost Classifier	15.43%	1.059	75.20%
Random Forest Regressor	11.93%	1.02	78.00%

Table 1: Comparatives for various ML Processes

3 - Results and Observed Behaviors

The results obtained from the Machine Learning process were both fascinating and validating for the hypothesis set forth by the [3N] Systems. The analysis began with an examination of the training data frequency distributions. Here, a notable curvature as shown in Figure 2 was observed, particularly pronounced at the extremities of the scale – specifically at the values 10, 9, 2, and 1. This pattern indicated that the portfolios tended to linger longer in these boundary states than in the intermediate, or non-boundary, states. Such a distribution pattern is not typically observed in linear systems and hinted at a more complex interaction at play.

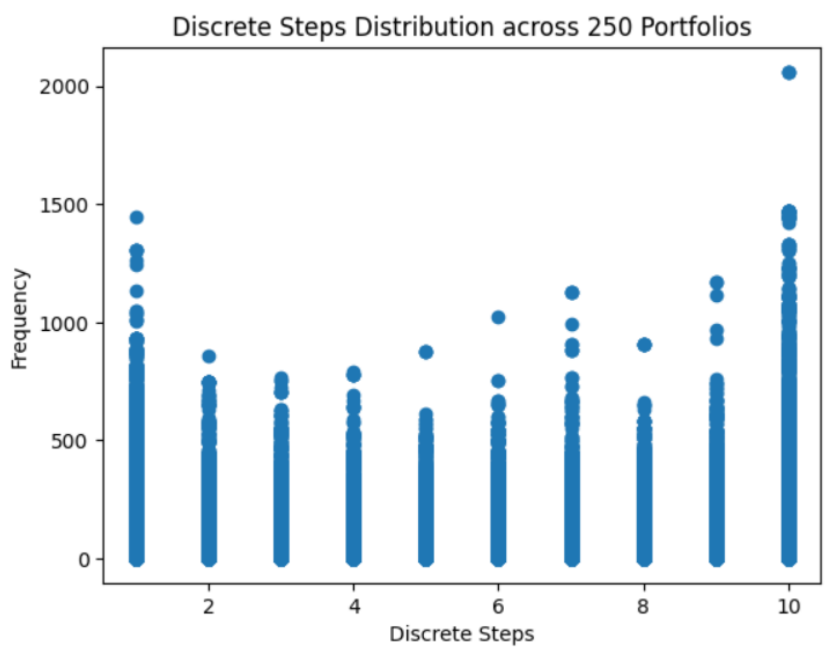


Figure 2 - Convexity in Training Data

This distinct behavior in the data highlighted the heterogeneity characteristic of a non-linear system. It was a clear indicator that the [3N] system was not static but rather dynamic in its nature. This dynamism was crucial, as it suggested the system's capability to identify and possibly isolate multiple equilibriums within the closed system of the S&P 500.

The ability to pinpoint these equilibriums is vital in understanding market behaviors and trends, especially in a system as complex and variable as the stock market. Further analysis revealed that the behavior of individual stock components within these portfolios was different from the behavior of the portfolios. However, a consistent pattern was observed across all data points in the form of curvatures, specifically convexity. These curvatures were not similar. They exhibited variability in shape which suggested that the [3N] non-linear mechanism employed could indeed be a composite of curvatures, encompassing both convexity and concavity.

Such a discovery is pivotal as it underscores the complexity and adaptability of the [3N] system, allowing it to accurately model and predict behaviors within the multifaceted environment of the S&P 500 with 500 stocks. The ML process was posed many predictive problems, like predicting positive boundary conditions established at the top quintile level.

Figure 3 illustrates the upper boundary condition across portfolios where the next step was predicted. If the next step was going to breach the upper boundary condition established at 8. And in case the state was already above 8, whether the next step will breach the respective boundary condition downward and head below 8.

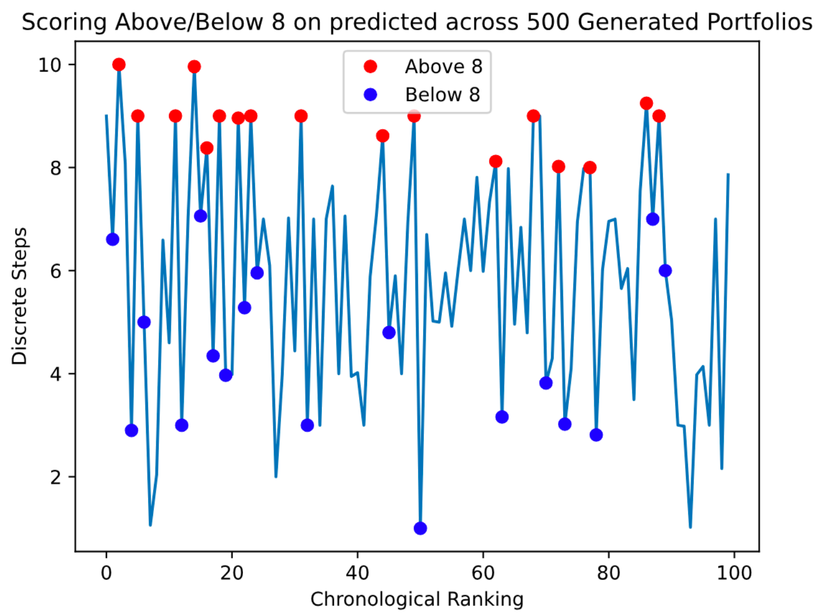


Figure 3 – Predictions on upper boundary conditions

Figure 4 illustrated the prediction of the next step at the lower boundary condition across portfolios. If the next step was going to breach the lower boundary condition established at 2. And in case the state was already below 8, whether the next step will breach boundary condition downward and head above 2.

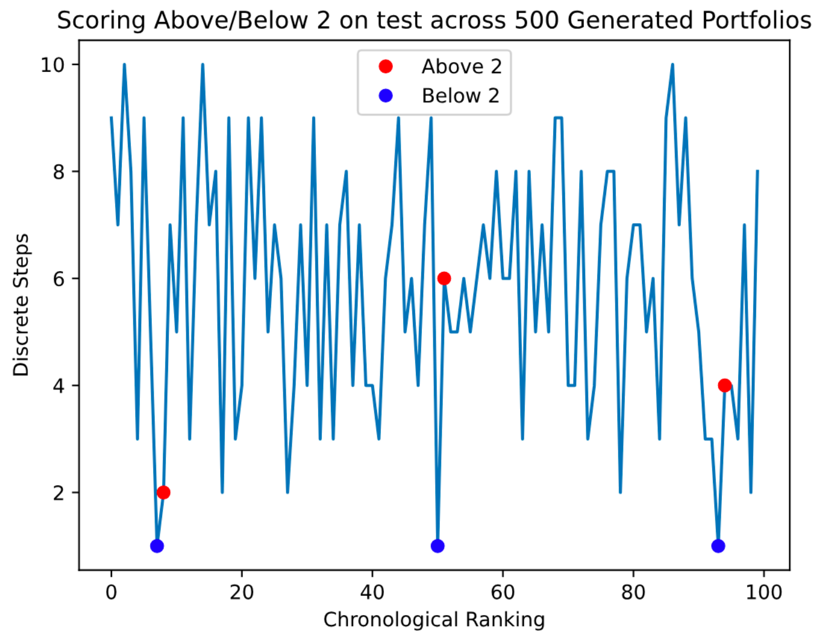


Figure 4 – Predictions on lower boundary conditions

Figure 5 illustrated the prediction of the next step across sample of chronological ranking including boundary conditions. With training the machine was able to predict the next step with a high degree of accuracy.

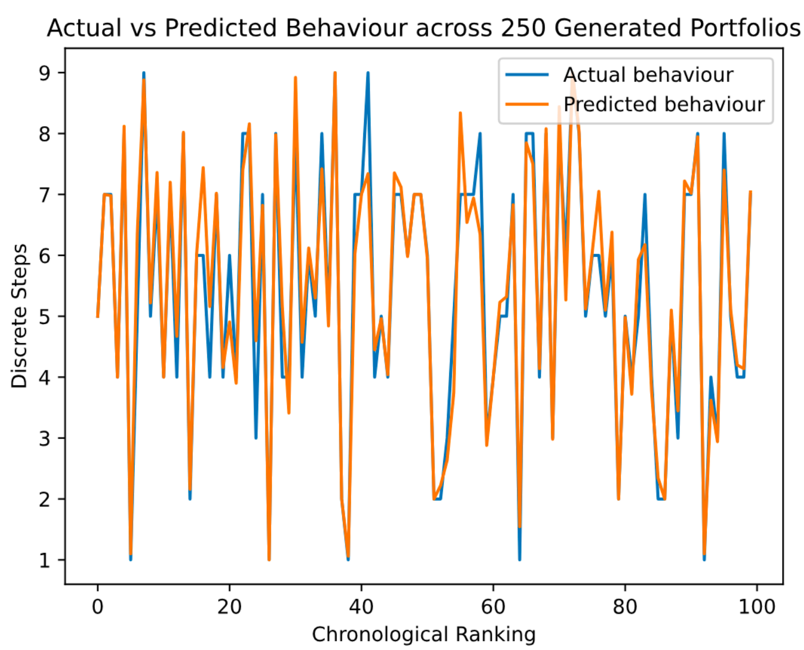


Figure 5 – Actual and Anticipated chronological ranking

The outcome was in line with initial expectations that noise reduction in portfolio states should generate better predictive outcome than stock states, even if intuitively the stock states seemed more meaningful than the portfolio states, which were a more diverse group of stock states.

The group over component debate in respect to the Discrete Decile States on one side confirmed that averages are more predictable than its components but on the other side indicated that even when we move away conventional informational analysis (idiosyncratic) to machine information (discrete decile states), even then a group prevailed over its components. Even machines appreciated noise reduction. The [3N] discrete steps method provided clear evidence that a group of states was more predictable than the individual components it consisted of.

Figure 6 overlays the prediction of stock states vs. portfolio states over many iterations. Unlike stock states, where the learning peaked early and did not see an increased learning, the portfolio states benefited from more iterations. Apart from the group vs. component reason, the training data for portfolio states was much more diverse than stock states. The portfolio training dataset captured a wider array of market dynamics and interactions, offering a richer context for the machine learning algorithms to analyze and learn from.

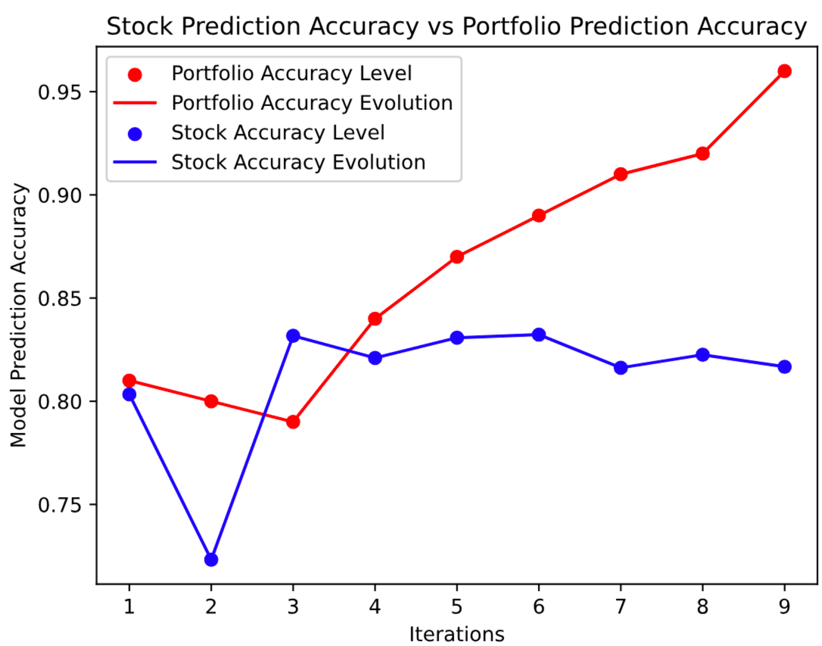


Figure 6 – Stock states vs. portfolio states prediction accuracy

The study's results became increasingly promising as the range and complexity of the training datasets were expanded. The training involved a wide variety of portfolio sizes, starting from small sets of under 10 portfolios to large sets comprising more than 200 portfolios. Additionally, these portfolios varied not just in size (referred to as basket size in the study) but also in the length of their chronological sequences of each individual component stock. This diversity in the training data was crucial, as it allowed the machine learning model to be exposed to and learn from a broad spectrum of market behaviors and trends. Different portfolio and stock states along with their average prediction accuracy are tabulated in Table 2.

Identifier	Accuracy	Iteration
PORT_A	0.81	1
ST_AV_A	0.8	1
PORT_B	0.8	2
ST_AV_B	0.72	2
PORT_C	0.79	3
ST_AV_C	0.83	3
PORT_D	0.84	4
ST_AV_D	0.82	4
PORT_E	0.87	5
ST_AV_E	0.83	5
PORT_F	0.89	6
ST_AV_F	0.83	6
PORT_G	0.91	7
ST_AV_G	0.81	7
PORT_H	0.92	8
ST_AV_H	0.82	8
PORT_I	0.96	9
ST_AV_I	0.81	9

Table 2 – Stock Prediction vs. Portfolio Prediction Accuracy Table

As the model was trained with this extensive and varied dataset, its predictive accuracy experienced a significant boost, surpassing the 90% mark for forecasts one step ahead. Predicting the immediate next step, or 'one step ahead', was identified as the relatively simpler task for the machine, largely due to the persistence observed in the data. This persistence implies that the immediate future of a portfolio often resembles its recent past, providing a somewhat stable basis for prediction.

However, achieving this level of accuracy was not without its challenges. The model had to effectively navigate through and overcome two major hurdles: noise and boundary conditions. Noise, or random fluctuations in the data, can obscure underlying patterns and mislead predictive models.

Boundary conditions, on the other hand, refer to the extreme values or states that portfolios can reach, which are often harder to predict due to their less frequent occurrence and greater volatility. Both these factors added layers of complexity to the predictability challenge.

Despite these challenges, the authors of the study found that the [3N] systems, particularly when augmented with discrete jumps, proved to be highly effective. These discrete jumps, or sudden changes in data patterns, helped in reducing informational dimensions – essentially simplifying the complex data into more manageable forms. This simplification, along with the system's ability to manage noise, enabled the machine learning model to learn and perform the forecasting task with remarkable efficiency.

The authors posit that as the forecasting task becomes more complex – moving from predicting a single step ahead to multiple steps, determining the duration of a state at a given step, or predicting when a state will reach a specific level – the accuracy of the model is likely to decrease.

However, the discrete step method has shown promising potential in stock market forecasting by focusing on contextual rather than informational content. This approach emphasizes machine interpretation over human interpretation, marking a significant shift in market analysis strategies.

5 – Proposed Application

This predictive system's unique focus on states rather than individual components, coupled with its ability to foresee several steps ahead, revolutionizes our understanding of information, benchmarks, and the broader market landscape.

Such systems are capable of discerning stock market biases and dynamically calibrating them to garner excess risk-weighted returns from a group of stocks without diverging from the overall systematic bias of the larger group.

This allows active asset managers to focus on more important tasks like mandate curation (e.g. selecting a preferred list of fundamental quality components from a larger universe), mandate maintenance, understanding market regimes and macro-economic aspects that are currently beyond the scope of DDS.

In the end, it's the active managers job to understand the capability of his machine and what differentiates one machine from the other, potentially leading to more efficient and profitable investment outcomes compared to passively indexed benchmarks.

6 - Conclusion

The paper articulates the underperformance challenge for active managers and why selection systems are unable to beat the market. The authors present a comprehensive approach to enhance active asset management industry through an innovative machine learning technique.

The authors utilize the [3N] systems to build a unique "Discrete Decile Step" method. This method uses a scoring system to dynamically groups stock information into states that rise or fall, indicating outperformance or underperformance, respectively. Then the authors utilize a Random Forest Regressor to predict these dynamic states, illustrating a predictive method that supersedes selections without ignoring systematic risk and hence potentially allowing active managers to make more informed selection decisions without the risk of underperforming traditional benchmarks.

The study reveals that portfolio states, rather than stock states, provide greater predictability due to noise reduction and the ability to capture a wider range of market dynamics. The paper highlights the relative importance of informational context over content, suggesting a paradigm shift from traditional approaches.

This shift is underpinned by the recognition that financial markets are probabilistic and dynamic in nature, thus necessitating predictive systems that can adapt to evolving market states and patterns. The authors emphasize the potential of such a predictive process to revolutionizing active investing by enabling selection systems to deliver sustained outperformance, i.e. alpha.

Bibliography

[1] Cowles, A. (1933). Can Stock Market Forecasters Forecast?. Econometrica, 1(3), 309-324.

[2] Fama, E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), 383-417.

[3] Granger, Clive W. J. (1992). "Forecasting stock market prices: Lessons for forecasters," International Journal of Forecasting, Elsevier, vol. 8(1), pages 3-13, June.

[4] Kahneman, Daniel, and Amos Tversky. "Prospect Theory: An Analysis of Decision under Risk." Econometrica 47, no. 2 (1979): 263-291.

[5] Thaler, R. H. (1999). The End of Behavioral Finance. Financial Analysts Journal, 55(6).

[6] Shiller, R. J. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? American Economic Review, 71(3), 421-436.

[7] Lo, A. W., & Mueller, M. T. (2010). WARNING: Physics Envy May Be Hazardous To Your Wealth! arXiv.org.

[8] Lo, A. W. (2004). The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective. Journal of Portfolio Management, 30, 15-29.

[9] Markowitz, Harry. "Portfolio Selection." The Journal of Finance 7, no. 1 (1952): 77-91.

[10] Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. The Journal of Business, 36(4), 394-419.

[11] Boulding, Kenneth E. (1966). "The Economics of Knowledge and the Knowledge of Economics." American Economic Review, Vol. 56, No. 1/2, March 1, 1966, pp. 1–13.

[12] Perold, André, and William F. Sharpe. "Dynamic Strategies for Asset Allocation." Financial Analysts Journal 44, no. 1 (January–February 1988): 16–27.

[13] Pal, Mukul, The [3N] Model of Life (April 19, 2021), SSRN.

[14] Pal, Mukul, The [3N] Method (February 17, 2023), SSRN.

[15] Pal, Mukul, The [3N] Methodology (Dec 25, 2023), SSRN.

[16] Schrödinger, E. (1944). What Is Life? The Physical Aspect of the Living Cell. Cambridge University Press.

[17] Galton, F. (1886). Regression Towards Mediocrity in Hereditary Stature. The Journal of the Anthropological Institute of Great Britain and Ireland, 15, 246-263.

[18] Simon, H. A. (1962). The Architecture of Complexity. Proceedings of the American Philosophical Society, 106(6), 467-482.

[19] Price, D. J. de Solla. (1976). A General Theory of Bibliometric and Other Cumulative Advantage Processes. Journal of the American Society for Information Science, 27(5), 292–306.

[20] Booth, D. G., & McQuown, J. A. (1982). Diversification Returns and Asset Contributions. Financial Analysts Journal, 38(3), 26-32.

[21] Molnar, C. (2020). Interpretable Machine Learning.

[22] Buzzi-Ferraris, G., & Manenti, F. (2013). Nonlinear Systems and Optimization for the Chemical Engineer: Solving Numerical Problems. Wiley-VCH.

[23] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics. Springer.

[24] Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.

[25] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

[26] Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K.-R. (Eds.). (2019). Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Springer.

[27] Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. The Journal of Portfolio Management, 11(3), 9-17.

[28] Kumbure, M. M., Luukka, P., & Collan, M. (2022). A review of machine learning applications in stock market prediction - State-of-the-art. Expert Systems with Applications, 197, 116659.

[29] Ball, R., & Brown, P. (1968). Empirical Evaluation Of Accounting Income Numbers. Journal of Accounting Research, 6(2), 159-178.

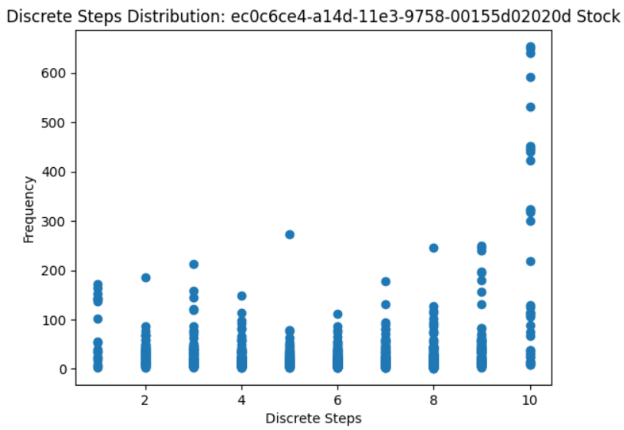
[30] Sharpe, W. F. (1994). The Sharpe Ratio. The Journal of Portfolio Management, 21(1), 49-58.

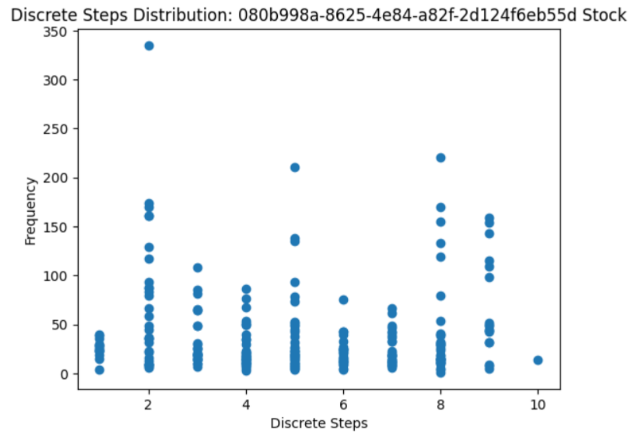
[31] Schapire, Robert E., and Yoav Freund. "Boosting: Foundations and Algorithms." MIT Press (2012).

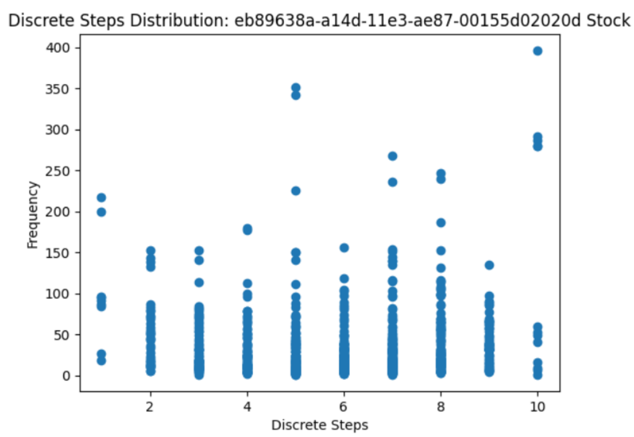
[32] Breiman, Leo. "Random Forests." Machine Learning 45, no. 1 (2001): 5-32.

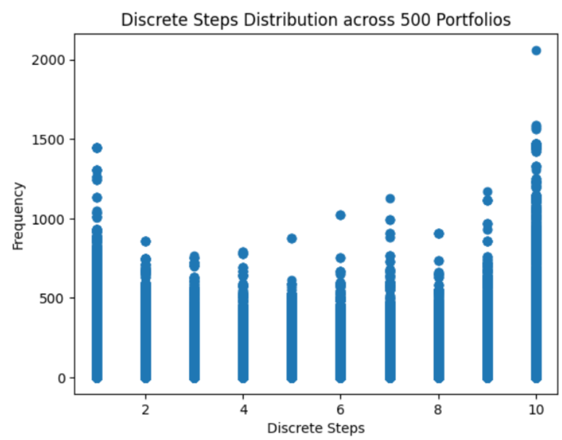
Annexure

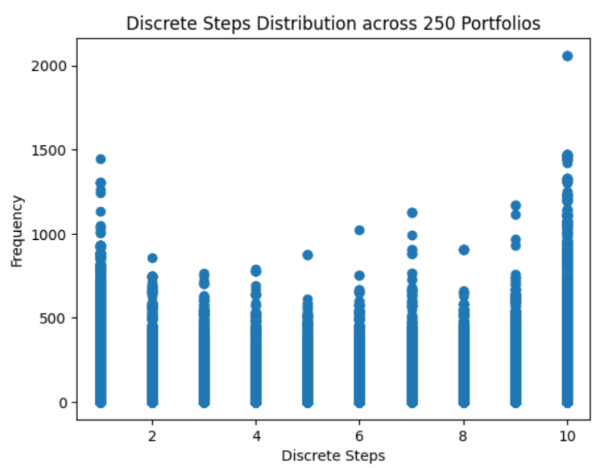
Curvatures in training data

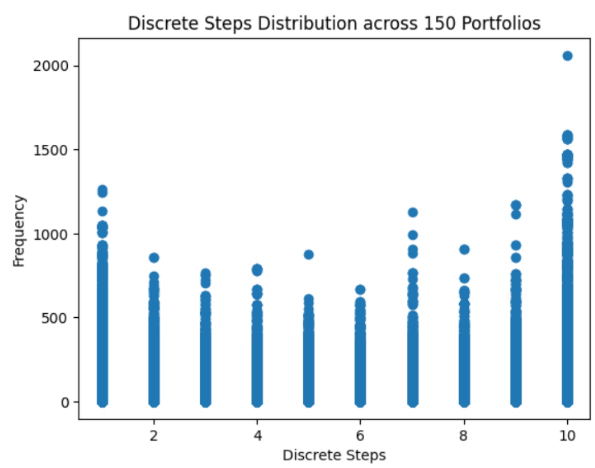




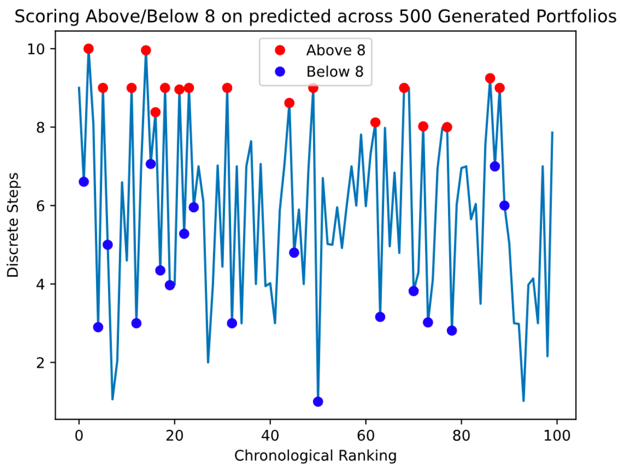


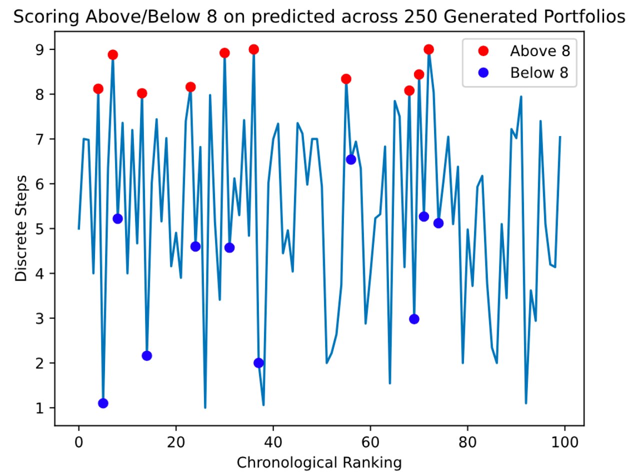


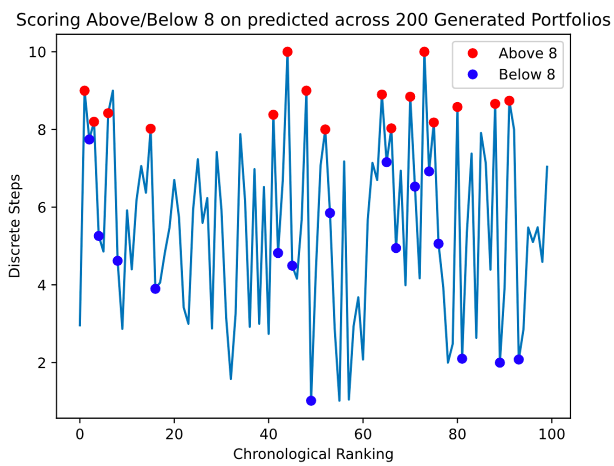


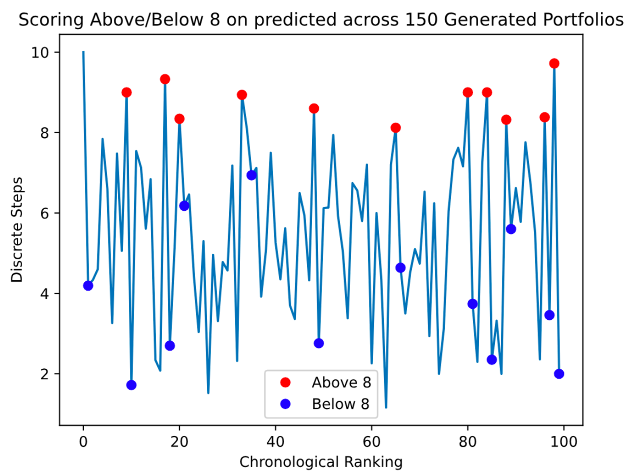


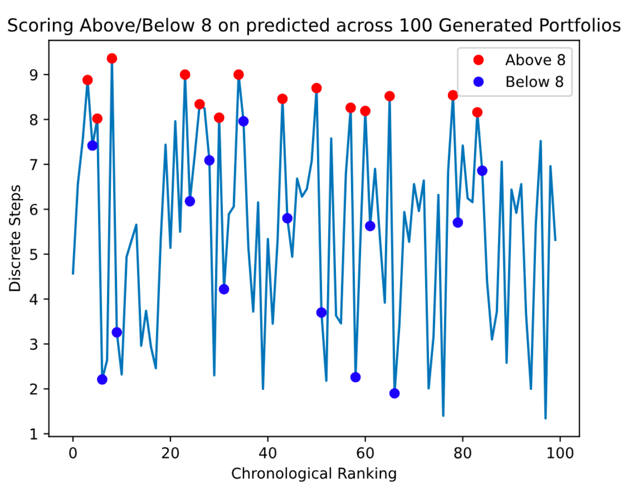
Prediction of Upper Boundary

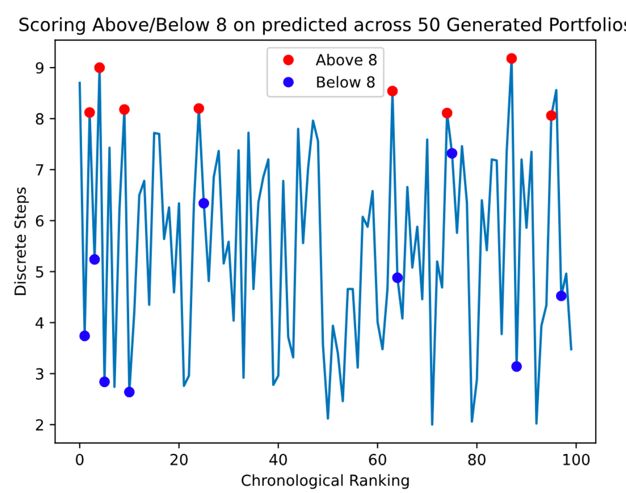




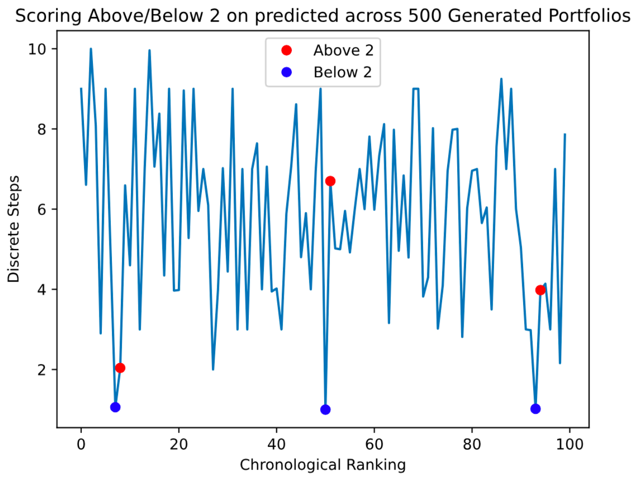


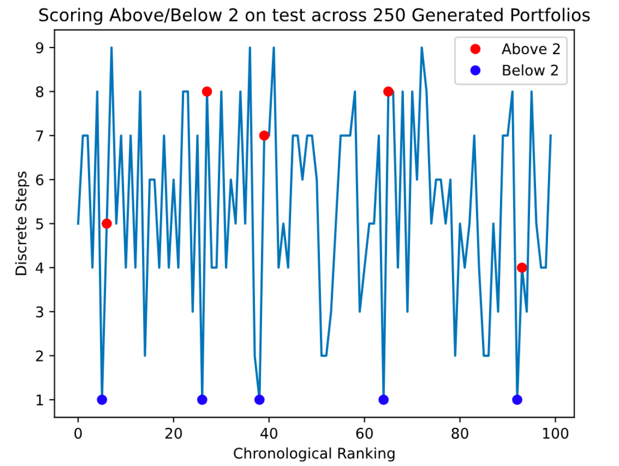


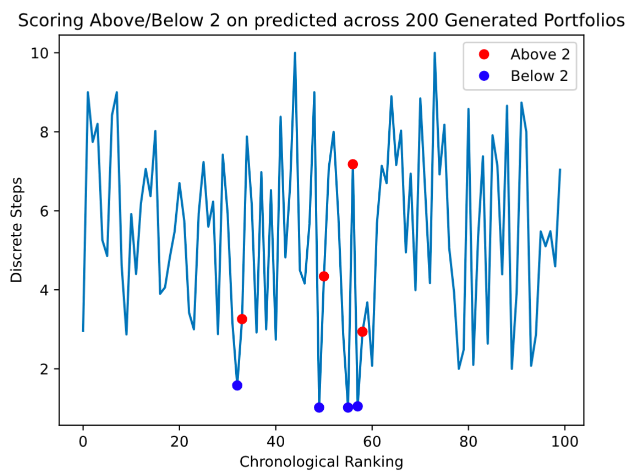


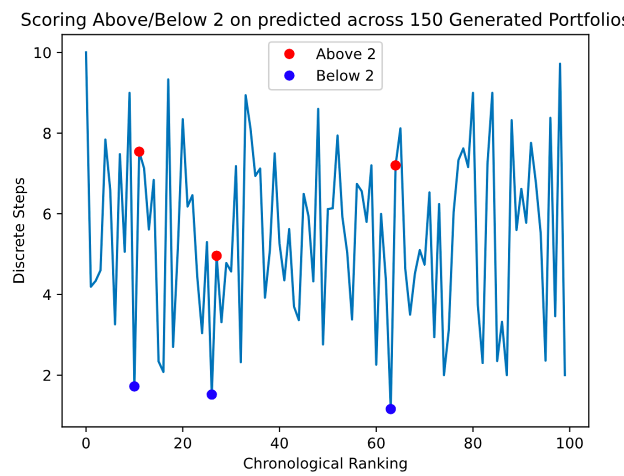


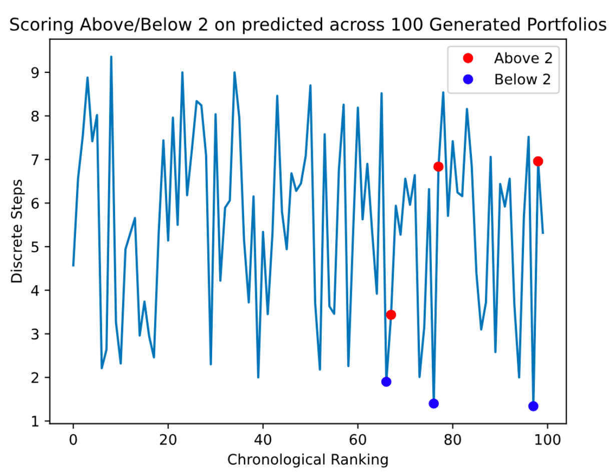
Prediction of lower Boundary

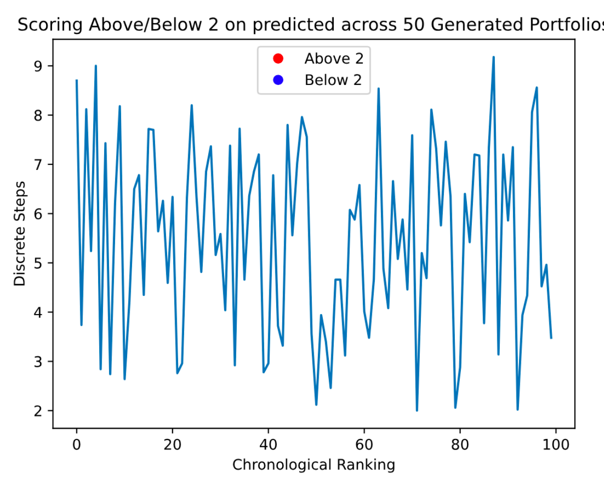




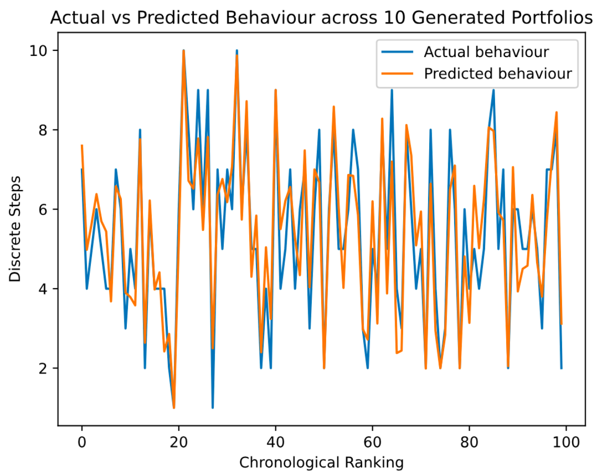


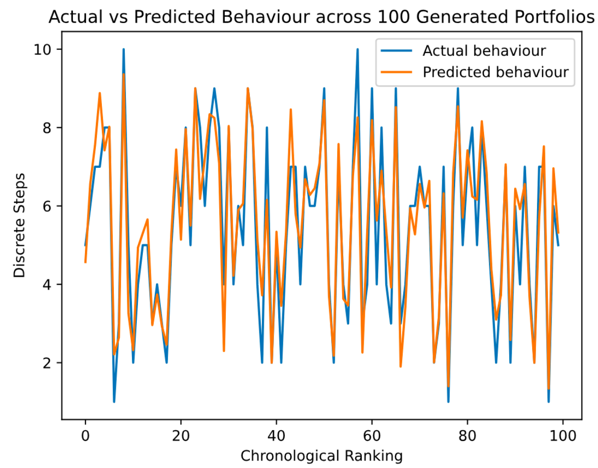


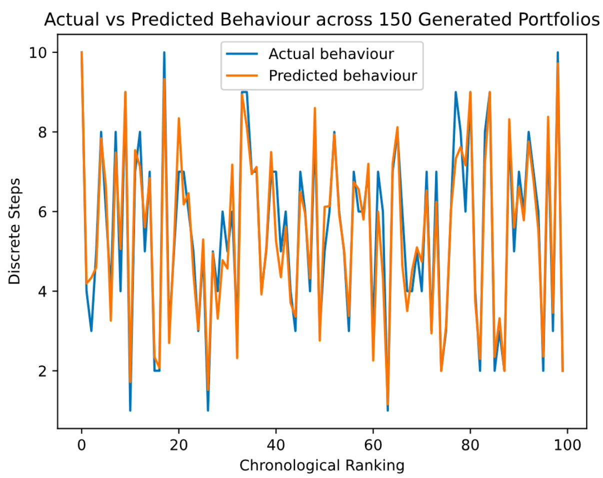


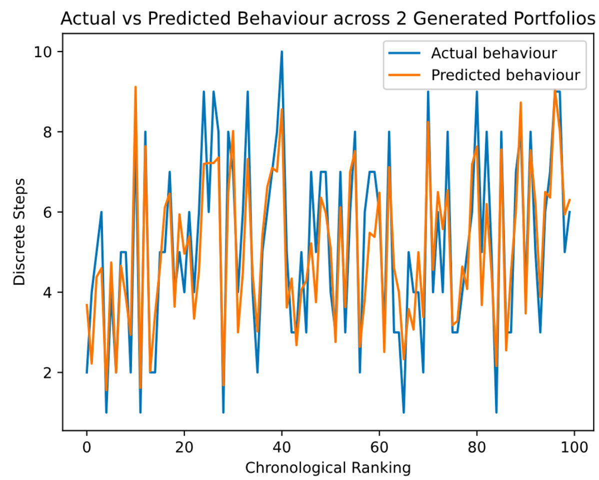


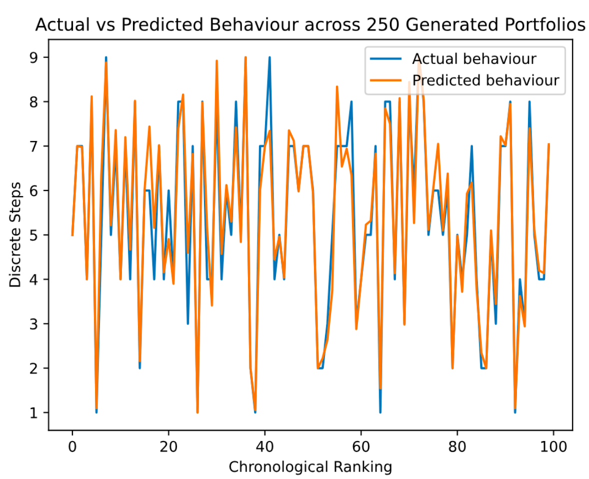
Actual vs. Predicted DDS Sequences

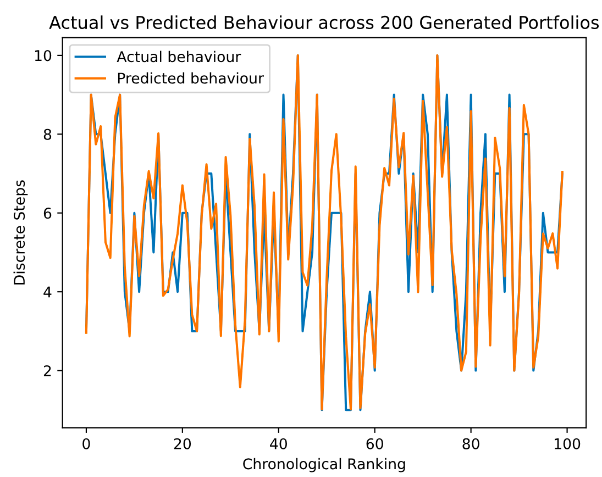




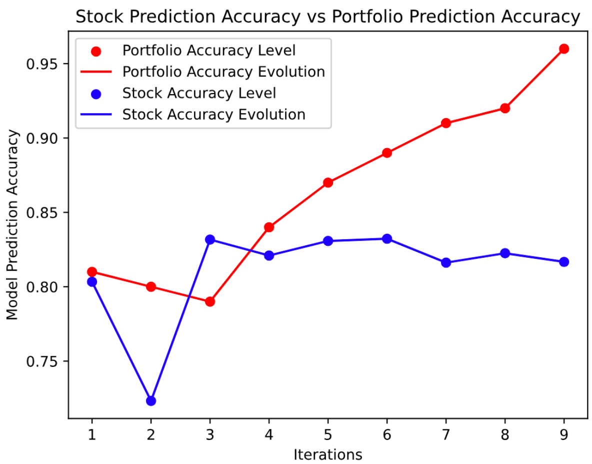


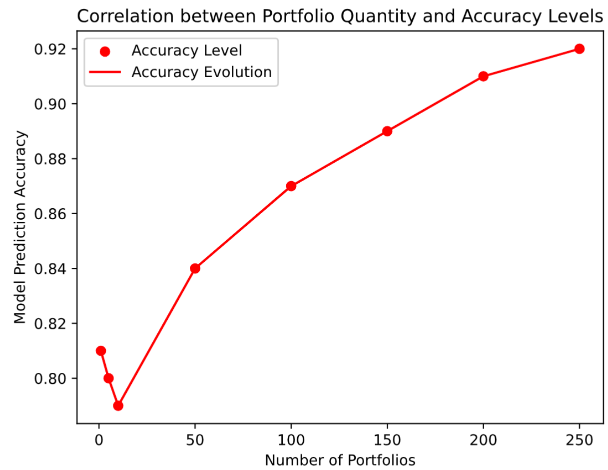


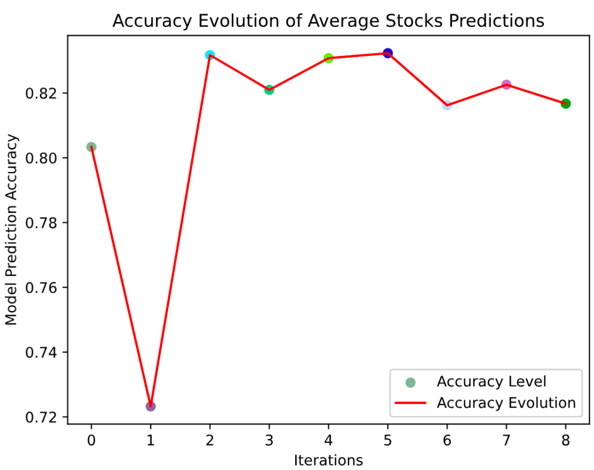


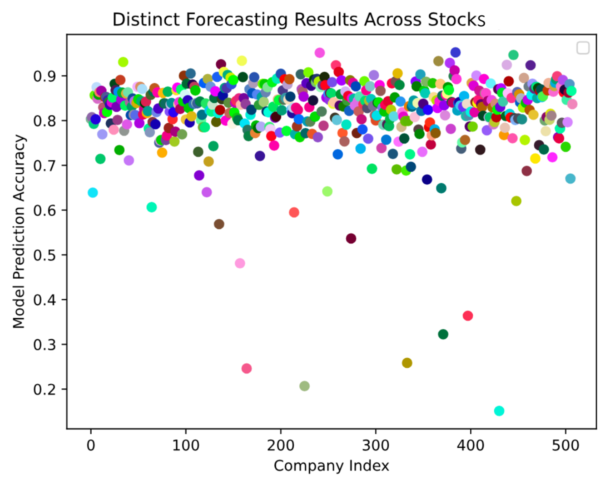


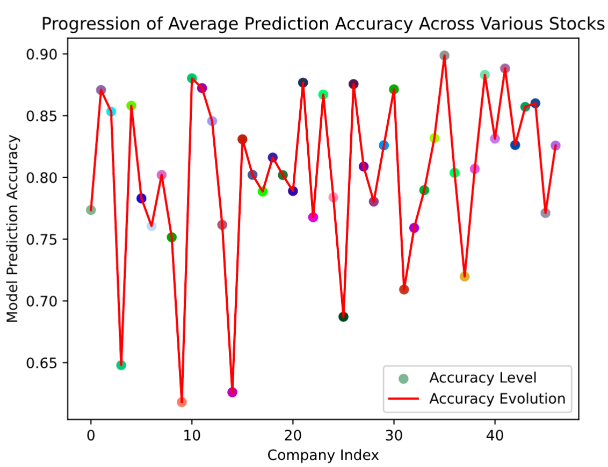
Learning Accuracy











Whenever the terms "stock", “stocks", "portfolio", and “portfolios” are used in Annexures, they specifically refer to the Discrete Decile Steps (DDS) of stock(s) and portfolio(s).