International Journal of Education, Business and Economics Research (IJEBER)

ISSN: 2583-3006

Vol. 5, Issue 3, May-June 2025, pp 422-436

To cite this article: Chibli Mayada (2025). Evaluating ML Models in Modern Portfolio Diversification – A theoretical Empirical Approach. International Journal of Education, Business and Economics Research (IJEBER) 5 (3): 422-436

EVALUATING ML MODELS IN MODERN PORTFOLIO DIVERSIFICATION – A THEORETICAL EMPIRICAL APPROACH

Chibli Mayada

PhD Candidate in The Bucharest University of Economic Studies, Bucharest, Romania
Faculty in the Lebanese International University, Tyre, Lebanon

https://doi.org/10.59822/IJEBER.2025.5322

ABSTRACT

The integration of machine learning (ML) into portfolio diversification has emerged as a transformative approach to address the limitations of classical financial frameworks. This study evaluates the efficacy of ML models, such as Gradient Boosted Decision Trees (GBDTs) and Long Short-Term Memory networks (LSTMs), against traditional methods like mean-variance optimization and risk parity. By adopting a theoretical-empirical approach, the research highlights ML's capability to capture non-linear dependencies in financial markets—such as regime-switching dynamics and tail risks—through advanced techniques like copula models and graphical dependency networks. The analysis reveals that ML models excel in high-dimensional data environments, offering superior risk-adjusted returns and diversification benefits, particularly in volatile regimes. However, challenges such as overfitting, interpretability trade-offs, and sensitivity to hyperparameter tuning persist. Classical models, while transparent and theoretically grounded, underperform in scenarios requiring adaptability to time-varying correlations and structural market breaks. The study further proposes enhancements to classical frameworks through ML-driven regularization and adaptive feature engineering, bridging the gap between computational innovation and financial theory. Empirical findings underscore ML's dominance in long-horizon forecasting and complex non-linear modeling, though hybrid approaches combining ML's predictive power with classical interpretability are advocated for balanced portfolio strategies. This work contributes to the evolving discourse on modern portfolio management by delineating scenarios where ML augments traditional practices and identifying critical areas for future research.

KEYWORDS: - Machine Learning, Portfolio Diversification, Risk Management, Non-linear Dependencies, Financial Markets.

1.0 INTRODUCTION

Machine Learning (ML) is a subfield of artificial intelligence (AI) dealing with the development of algorithms that automatically improve their performance or learn from previous experience. With input data of corresponding outputs, the supervised approach constructs a model by learning the mapping function from inputs to outputs by minimizing a defined loss function. Various methods can be distinguished based on the characteristics of the input data or the learning mechanism. The most common methods include support vector machines (SVM), artificial neural networks (ANN), Gaussian processes (GP), ensemble methods, and decision trees (DT). All ML approaches require feature selection or some engineering tasks to provide input data in an appropriate format. The framework of a traditional ML modeling process comprises problem definition, model design, data acquisition, data pre-processing, ML modeling, model fitting and validation, and analysis and use of the model to fulfill its purpose. A multitude of software solutions is available for realizing ML approaches, which are provided by libraries and frameworks such as Scikit-learn, Tensor Flow, Keras, or XGBoost. However, only a few of them are tailored for finance (Padhi et al., 2022).

The artificial neural network (ANN) framework is based on the work of McCulloch and Pitts and Rosenblatt et al. As a network of neurons, it undertakes more complex computations than a logistic regression model. It includes an input layer, hidden layers, and an output layer consisting of input-output transformations. Given the weights assigned to the interconnections between different layers, data from the input node pass to the other layers until they reach the output node and generate a prediction or decision on the problem. An ANN can approximate any nonlinear transfer function to an arbitrary accuracy by using a combination of input–output transformations with accurate multilinear functions.

ANN framework is based on the work of McCulloch and Pitts and Rosenblatt et al. As a network of neurons, it undertakes more complex computations than a logistic regression model. It includes an input layer, hidden layers, and an output layer consisting of input-output transformations. Given the weights assigned to the interconnections between different layers, data from the input node pass to the other layers until they reach the output node and generate a prediction or decision on the problem. An ANN can approximate any nonlinear transfer function to an arbitrary accuracy by using a combination of input–output transformations with accurate multilinear functions.

ML algorithms are frequently used to address financial problems. In particular, these algorithms have recently gained interest in various stock market prediction applications. One of the main areas of focus for global interdisciplinary engagement is the stock market, one of the high-stakes situations involving complex decisions. To maximize profits while minimizing risks, participants in the stock market require developing stock selection strategies. Analysts, traders, and investors make use of a wide range of information sources, from technical and statistical signals to macroeconomic and financial metrics (Padhi et al., 2022). Policies determining the acquisition and sales of stock are made in real-time—indeed, continuously—so they must be computed quickly, generally in minutes or seconds. This implies that any news or information, relevant or irrelevant, will be updated in the market Pricing gives the best perspective on the stock valuation.

ML techniques have been proposed to automate stock selection address challenges inherent to stock market prediction. Stock selection algorithms work by acquiring a set of signals to parameterize the stocks and selecting the stocks with resolving the best opportunity to profitability. Many ML approaches have been proposed for dealing with price movement forecasting via label susceptibility. Some notable methods are SVM, MLP, Back Propagation Networks, BPN, CNN, MLP and Random Forest.

2.0 METHODOLOGY

This study adopts a theoretical and empirical approach to evaluate the role of ML in modern portfolio diversification, with a focus on bridging gaps between classical financial theories and computational advancements. The methodology is structured around three core research objectives derived from the challenges and opportunities identified in existing literature.

The first objective is to systematically compare the risk-adjusted performance and diversification efficacy of ML models, such as Gradient Boosted Decision Trees (GBDTs) and Long Short-Term Memory networks (LSTMs), against classical frameworks like mean-variance optimization and risk parity. This involves analyzing how ML algorithms handle high-dimensional data and non-linear relationships, which are often oversimplified in classical models. The second objective investigates the impact of non-linear dependencies in financial markets, particularly during volatile regimes, and evaluates ML techniques like regime-switching models and graphical dependency networks in capturing these dynamics. The third objective addresses limitations inherent to classical diversification models, such as reliance on linear correlations and static assumptions, by exploring theoretical enhancements through ML-driven hyperparameter tuning, regularization strategies, and adaptive feature engineering.

The study integrates theoretical insights from ML literature, such as the ability of artificial neural networks (ANNs) to approximate non-linear functions and the use of copula-based models to capture tail dependencies. Non-linear relationships are further analyzed through regime-switching mechanisms, which theoretically adapt to structural breaks and market shocks by dynamically adjusting portfolio weights.

Comparative analysis between ML and classical models is grounded in theoretical principles of computational finance, such as the trade-off between model complexity and interpretability. Challenges such as data non-stationarity and overfitting are addressed through theoretical strategies like adversarial validation and incremental learning, which aim to enhance generalization without empirical implementation details.

3.0 LITERATURE REVIEW

3.1 Portfolio Diversification

Diversifying portfolios by holding several assets is one method to reduce risk for investors. An optimally diversified asset portfolio includes equities, bonds and commodities with selected weights. Important factors for optimal asset allocation arise from the interaction between the composites, mainly the cross-correlations and the volatilities. This is more difficult in the context of financial market time-varying volatility and correlations.

At present, financial market plummets occur, impacting market investors, large and small. Optimum asset allocation is essential to balance risk and return and remain in the game. Portfolio optimization is the science of producing the incidence matrix defined by weights, which transforms the vector of expected returns into an optimal risk-return trade-off. The method enables investors to achieve efficient portfolios using compositional assets characterized by weights. Current optimization techniques focus on standard quadratic programming while pay-off histories of assets are often presented in record time series. Such structured data provide a chance to use other state of the art paradigms for portfolio optimization, which result in an exceedingly easy implementation of complex optimization techniques.

Decision making in finance is a multi-attribute frontier problem associated with trade-offs between possibly conflicting magnitudes such as return and risk. Logistic regression and naive Bayes classifiers show a good learning behavior given the independence of attributes. Bayesian networks handle dependencies along with a massive parallel execution advantage over depth-first search algorithms. Feed forward neural networks surpass other algorithms providing a high learning performance achieved by error back propagation, but they are exposed to the so-called overfitting problem. These behaviors arise from the data distribution domain, choice of features and learning algorithms.

3.2 Classical Diversification Frameworks

The diversification of risky assets is an essential tool for long term investors to tame risk and achieve a desired level of return. Investors have a number of ways to diversify their portfolios, including investing in a larger set of (but similar) assets, in more risky assets with a smaller weight, or adding different types of assets such as stocks or bonds to the portfolio. Portfolios are constructed using alternative weights for each asset consisting of mean-variance weights, a range of heuristics and the extended CVaR approach, and evaluated using measures of the expected return, risk and expected shortfall. Diversification is found to depend on variables such as the country or asset class of investments and its risk level, this last one being generally a better measure than expected shortfall, standard deviation of returns or other more commonly used measures. Additionally it is found that heuristic combinations of assets, markets and risk levels diversify more than market capitalization weighted portfolios to a point where their added risk is small enough and the return close to that of the market portfolio. Bounds can be found on expected return for portfolios by constructing a set of portfolios from extreme weights and measuring the mean-variance frontiers of each one (Jacobs et al., 2009).

Optimal allocation of capital to risky assets involves a tradeoff between expected performance (returns) and risk. Returns are often a direct consequence of the investment policy and cannot be controlled directly. A widely adopted framework to measure and control risk is attributed to Harry Markowitz. The idea is straightforward; at a given expected return, the optimal portfolio minimizes risk, while at a given risk level, it maximizes expected return. Understanding the problems of portfolio construction in general brings useful insights for calibrating ML models and for understanding their different possible approaches regarding the asset class, investment horizon, and their investment policies.

Markowitz's seminal contribution outlined optimal portfolios in the 1952 paper "Portfolio Selection" and revolutionized how practitioners thought about risk and return (Boyd et al., 2024). The twin concepts of frontier optimal portfolios containing the maximum expected return for a given level of risk and comparable efficient allocation of portfolios formed the foundation for the discipline of risk modeling and investment decisions.

Considerable empirical evidence in the past few years has shown that one of the simplest models of asset return distribution, a bivariate normal model, can provide a misleading impression of certainty if applied without scrutiny. The broad conclusion is that a setting with constant volatilities and correlations is inherently unreliable. It is argued that such limitations are not unique to the specific case studied, and that there is no "get-rich-quick" scheme at the present time.

In recent years, several studies have analyzed risks of PTFs with regards to stocks, using daily data of countries indices traded at the NYSE. It was clearly shown that classical models significantly under-estimate risks in particular with regards to stocks. A systematic analysis of daily correlation and volatility coefficients, the relevant parameters for classical portfolio modeling, was performed and their instabilities over time were quantified. In particular, it was shown that using a bivariate normal model, the implied diversification effects can be very different, even for moderately different periods. It was shown that time scales of fluctuations could differ by an order of magnitude both in an inter-day stock data and in a ten-minute intraday data of government treasury bonds.

3.3 Non-linear Dependencies in Financial Markets

Adverse market shocks frequently provoke asset price drops and market volatility spikes, exacerbating the difficulty of portfolio diversification. Compared with recent past events, these shocks can lead to different tail behavior so that the use of fixed tail-dependence structure may not fit well to different time periods. In such a situation, adequately selecting the forecast horizon with the use of regime-switching (RS) modeling can improve the outcome of portfolio diversification and a better diversified portfolio may be achieved.

The modeling of time-varying tail-dependence structure, incorporating with the appropriate forecast horizon and fitting to the changing regime, provides a recommendation for more successful portfolio diversification. The implication of regime-switching models to diversify portfolios among extreme risks and forecast horizon selection is identified. It argues the advantage of time-varying approach over the fixed specification and the improved outcome of portfolio diversification by properly using forecast horizon and regime-switching modeling. It allows investigation of different changing features including interventions, market shocks or structural breaks, across assets or across different scales of data.

Possible applications of the approaches have been illustrated, pointing the need for proper modelling of filtering, regime-switching, and identification of forecast horizon. Non-linear effects have been identified through regime-switching mechanisms on the tail-dependence behavior and the tail-dependence structure of time-varying Gumbel-Gumbel copula indicates the need to monitor the forecasting horizon adjustment to diverse non-linear dependencies.

3.3.1. Understanding Non-linear Relationships

ML approaches thrive when dealing with non-linear relationships. For example, Support Vector Machines can describe any function if a proper kernel is chosen. The ability to identify such complex relationships is of particular interest when considering potential scalable alpha factors.

It is desirable to gain insight into the underlying relationship of the portfolio dynamics an alpha factor. In this context, an additional objective that fits ML well will be proposed: finding directional alpha factors. Such factors behave consistently with respect to the portfolio in question, flocking the weights to certain regimes or changing them smoothly. An interpretation concerning the dependability and usability of the alpha is provided. The analytical tasks will allow for insights on the individual alpha factors' mechanism of action, and thus robustness can be readily checked. This is particularly relevant as ML accuracy estimators tend to reflect and register bias in the training set, which is generally unfeasible in the financial domain. Together with the ability to describe the actions taken by the model, a better integration of such analytical factors into standard portfolio restoration workflows is rendered possible (Goldberg & Mouti, 2019).

While this thesis primarily focuses on high-level aspects of modern portfolio diversification, beyond its own scope, some efforts towards an interpretation of the alphas suggested in Section. In the scope of experimental findings, it becomes evident that the assumptions of constant regime behavior and smooth state switching do not hold in practice; rather, the behavior of financial series is multifractal. Although such effects cannot be reproduced natively, methods resembling can be used, resulting in a significant decrease in the volatility of the reconstruction compared to standard geometric Brownian motion.

3.3.2. Impact of Non-linearities on Portfolio Performance

Both academic finance and investments practitioners alike have expressed continued interest in ML due to advances in data availability and computing power. Alternative data becomes almost synonymous with ML in this context. Growing interest by regulators and growing understanding of market risks by practitioners create incentives for research into the potential consequences and limitations of ML methods for finance. Non-linearities in modern algorithms produce considerably better portfolios across multiple loss functions and tolerance to different parameterizations. Attempts at enforcing a certain distribution of weights, inspired by the use of regular over-fitting penalties, have although met with partial success, and it remains to be studied whether ML portfolios built with softer weight restrictions can nevertheless outperform the here-focused linear benchmark despite not emphasizing interpretability.

The extreme performance of sampled portfolios suggests a favorable non-linear fitting of the respective target loss function. Non-linearities are invited to be simply introduced to their portfolio allocation methods. Additionally, the result and methodology can be extended to a broader class of loss functions, such as maximum drawdown, reward-to-risk ratios, and non-quadratic preferences. ML and the wealth of available data continue to fundamentally change the way in which financial markets are analyzed. Structural modeling approaches have provided insights ranging from high-level systemic risk factors to the price impact of individual trades. The classic literature offers intuitive models capable of producing well-diversified and intuitive portfolios. Unfortunately, theoretical guarantees are lacking, though good results are often achieved in practice.

ML techniques have been employed to handle the resulting dimensionality in a more agnostic way. Despite largely improved diversification abilities of the ruled by numerical optimality criteria, it is troubling that storage and implementation of such models hinges on retaining exact numerical valuations and recommended allocation weights. Apart from improved performance, interpretable ML is invited to help understand the drivers behind performance and portfolio weights. The wide variety of methods is potentially complementary, and combining approaches in an ensemble context would be meaningful and desirable in recommending weights to the investor, management avenues, and prospective research.

3.4 ML Approaches to Diversification

There are various ways to mitigate risk through the portfolio return space. In Modern Portfolio Theories (MPT) risks of a portfolio is minimized by co-variances of the returns while by Asset pricing based approaches, asset pricing errors are minimized which equals risk on different levels.

In ML, risk is defined under various constraints in distributional, temporal and observational basis. There are various ways to evaluate the predictions of a ML Model in a stock portfolio/fund of securities context. For portfolios with a given constant number of assets there is the possibility to analyze the prediction's sufficiency/robustness under fire-drill exercises. Of course the number of stocks in the exploration universe would also imply a vast combinatorial space to rule out some unworthy combinations by simple checks.

ML's advantage in processing high dimensional data is clouded by the risk they uneasily disentangle into human interpretable dimensions; a ground of wariness alongside its unreliability in unseen regimes. This prudent approach could be enriched by still simple business performance measures capturing unfathomable aspects with systematic checks input space and hyper parameter's specifications.

In modern portfolio theory, investment strategies vary across investor profiles in terms of risk preferences. Here, considering an unconstrained mean-variance optimization problem first, the ending investment portfolio is a linear transformation of the predicted asset returns. Therefore, contemporary ML models designed solely for asset return prediction are directly applicable, predict the expected asset returns, and derive the strategy from posterior distributions. This strategy should prove advantageous over conventional models, considering different assumptions on investor preferences. Additionally, it is widely acknowledged that asset returns are in effect discrete observations of diffuse underlying state variables. All sorts of drivers that interest and affect the valuation of assets vehicles exist, and they come from a range of disciplines and sources, which may affect stock prices two-fold. Therefore, it is reasonable to speculate on how to incorporate multiple returns, features, and other data. A two-stage framework for long-term asset allocation is built for a range of models learned from past market behavior, and a theoretical analysis shows that coupling models in an ensemble master forecasting strategy could prove beneficial.

3.5 Addressing Limitations of Classical Models

A portfolio's diversification quality is commonly assessed by the Sharpe Ratio and the diversification effect. The latter can be accessed through the ratio between the portfolio's risk (σp) and the individual stocks' average risk (σ̄)/N. There exists a plethora of classical diversification portfolios, ranging from naive diversification to more sophisticated methods, such as the mean-variance approach, risk parity, and minimum risks (Ziggel et al ., 2011). For each portfolio , the parameters, σp, σ̄, are estimated through a rolling window of historical returns to test their diversification capabilities against their naive counterpart. These classical methods have proven to be outperformed by portfolios trained via decision trees and neural networks. However, they are constructed using established financial theories, so their parameters are transparent and may have intuitive appeal. Hence, it is worth investigating if an increase in predictive power through ML models diverges from classical portfolio construction strategies.

To provide fair and strict comparison with classical portfolios, GBDTs and LSTM models are embedded into a setup that closely resembles the rolling-window test with the classical methods. The resulting portfolios are score-carded on unseen test data from the same out-of-sample period as denoted in classical approaches. However, there are still several limitations of the presented ML models and portfolio construction framework. First, corresponding hyper parameters are picked on the test data of the whole out-of-sample period, while other classical models are trained individually per window on the training data only. To ensure a fair comparison, it is important to adopt a similar strategy in the future. A second avenue for further improvement concerns the ML models. Constraining the ensemble size of the GBDT models can mitigate some of the over fitting. Third, additional GBDT hyper parameters that dictate the tree structure, such as the maximum depth and the learning rate, can improve the overall performance of the decision trees. For LSTM, imposing constraints on the hidden size and fully connected layer can prevent over fitting and result in more robust portfolios. The breadth of stocks based on filtered data sets with the highest thin-trading ratios can be incrementally extended during training through data augmentation to further enhance portfolio stability.

3.5.1. Handling Non-linear Dependencies

The majority of classical diversifications are derived based on the linear dependencies of returns on the portfolio, which is insufficient since dependencies can be non-linear. Consequently, the market or assets that seem diversified can be non-diversified from other viewpoints (Hsing et al., 2004). Indeed, most importantly in the portfolio contexts, measuring the performance of different models, savings in transaction costs, explanatory capabilities, forecast intervals, and statistical robustness on auto-regressive time series is possible. However, an inappropriate model may lead to disaster with great differences in the properties, etc. Efforts made by trying different models individually are time-consuming. Furthermore, in the context of estimating the tail risks of portfolios, a pseudo-likelihood (alternative likelihood) based on a variational space partitioning was developed and validated.

Various approaches treat the dependencies among assets in the portfolios via graphs like correlation networks or graphical models (Zhan et al., 2021). Graphical models focus more on the "independence" from the graph view and algorithmically change the graph with time or among groups of stocks. There are variations of graphical models or Bayesian networks directly on the covariance matrix or precision matrix. Any ensemble estimation methods have been applied to stronger graphical models, measure the number of nonzero elements within the adjacency matrix, and decide a fixed “model size”. Additionally, moving-window approaches can also estimate independent graphs.

The time evolution of the dependencies or the graphical structures was learned through group invariance or temporal concentrations and used for portfolio selections or navigation decisions. These methods utilized prior knowledge of a few supervised edges or hyper-parameters. On the contrary, the Hidden Graphical Vector Autoregressive (HGVAR) model, inspired by the dynamic stochastic block-models, infers the time-evolving dependencies or graphs over historical observations with no extra supervision. Coupled with the dynamism and sparsity of natural dependence graphs, the HGVAR model can recover the underlying structures more initial coordination settings than state-of-the-art methods.

4.0 Comparative Analysis of ML Models and Classical Frameworks

In finance, diversification is regarded as one of the essential rules for investing. Numerous methods have been proposed to diversify proportions in the mean-variance framework. Although this framework is popular for its intuitive presentation and explicit diversification rule, calculating the weights of portfolios is not a trivial task, especially if there are many dimensions in the asset pool. Several methods and heuristics can calculate weights and handle multiple dimensions and complications. On the other hand, in recent years, deep learning networks have shown their worthiness in handling many tasks in natural language processing and image recognition. This has led to a few efforts that attempted to apply machine/deep learning-based methods to onshore equities of ETF complexes. So far, traditional econometric models have also worked reasonably well considering the volatility clustering and periods of calmness in the stock market.

It is found that, although ML models are capable of predicting more accurate returns and outperforming other traditional econometric frameworks, they generate excessively risky amounts of leverage on a few occasions despite very high Sharpe ratios. As such, portfolio remaining within risk tolerances with sound long-term performance is of utmost importance, given that risk management is crucial in high-frequency integral capital allocation. In addition, although both mean-variance-based approaches outperform static portfolios in the long-run with a more stable returns structure, while ML models tend to drive towards very few stocks on several occasions with excessive levels of leverage and very low diversify ratios, deep learning networks outperform traditional approaches in data-rich environments, especially for broader ETF complexes and sub-groups containing riskier assets.

4.1 Performance Metrics for Evaluation

The problem formulation involves the use of a regression model to calculate a portfolio of weight for the asset tickers listed. With the portfolio weights derived as prediction, they can be used to calculate the portfolio return. Prediction performance metrics based on the sequence of portfolio returns can then be applied to evaluate the performance of different models. Next, as the portfolio weight has to fulfill constraints, the second requirement is that the portfolio weights need to be valid for a portfolio selection. Different metrics are assessed in this section to evaluate how well different models fulfill these requirements.

When measuring the performance of a portfolio consisting of asset classes over more than one day, several metrics can be taken into account. One obvious choice is the Sharpe ratio, which uses the mean return divided by the standard deviation of the returns. This motivates the selection of the Sharpe ratio as a first performance metric. Furthermore, the information ratio is also used to evaluate the stock selection. As a second derived metric, the downside deviation of the Sharpe ratio from the risk-free rate is used. The downside deviation is calculated over the time period that losses occurred. The information ratio penalizes deviations smaller than 0, which represents streaks of bad performance.

In a subsequent study, (Jacobs et al., 2009) extend their previous research findings towards an exploration of how private investors should optimally construct a portfolio with an international stock market index while taking transaction costs and rebalancing effects into account. They explore the accuracy of equally weighted and market-capitalization weighted stock portfolios and compares them to mean-variance efficient portfolios. They show that, on average, the equally weighted portfolio outperforms the market-capitalization weighted portfolio over time, both without and after costs. In a further extension, they demonstrate that, by combining either of the above-mentioned simple strategies with a limited short-selling strategy, a minimum variance portfolio can be obtained that outperforms both the market-capitalization and the equally weighted portfolio in several stock market classes as well as in a combined portfolio comprising all indices. In a field test based on the announcement of the SHD index in Europe, the minimum variance strategy forecasts a market movement in the correct direction on 83 percent of the respective trading days, while an equally weighted strategy does not consistently obtain correct forecasts. In an independent paper, they analyze the additional forecasting error variance reduction that can be achieved by combining information from differing stock-selection strategies. They compare the 1987 Dow Jones Industrial Average (DJIA) with 14 other stock-selection strategies proposed in more than 50 papers over the past 70 years. By using a recursive out-of-sample estimation procedure, they provide an extensive empirical evaluation of portfolios composed of formalized stock-selection strategies. They show that a naive equal weight or market-capitalization weight combination obtains similar if not better results than various data mining corrections and also combinations based on forecasted risk optimization. In a further extension, they include a broad set of technical analysis strategies used in futures forecasting models. They find that, while forecasts based on recent return continuations barely outperform a naïve random walk forecast, forecasts based on either price or volume breakouts significantly outperform random walk forecasts. Combinations of this type of forecasts tend to perform even better. An equally weighted combination of the best performing price and volume breakouts comes close to the efficient portfolio in the mean-variance sense as illustrated in equity market comparisons.

5.0 Challenges and Considerations in ML Implementation

While the mentioned asset managers' machine and deep learning ideas all have merits, a few challenges and considerations need to be emphasized. First of all, most evaluation metrics for ML models require extensive back testing, which means fine-tuning the hyperparameters on historical data, making them prone to over-fitting. For instance, the parameter for the long short portfolio idea in (Kwak et al., 2022) should not be regarded as universal; during backtesting on a different period or in a different state, a different parameter should be employed. Simple models do not require this extensive testing, but excessive simplicity negates the utility of the models. Therefore, the proposed evaluation metric for these ML models should only be computed once at the end of backtesting (after hyper-parameter tuning).

Secondly, early noise filters in the raw input signals are crucial. Information should not be too fast so that the user may compute something hallucinatory, nor should it be too strong, in the sense of forming a regular relationship, so that the user can not compute anything new. It seems that the signals from non-traditional information, such as chain and sentiment data, are less predictable compared to a stock's historical price. It is quite possible that prior knowledge from a signal could become outdated due to interfering signals, regulatory restrictions, or improvement in other information sources. In some sense, learning from data (breaching other types of information sources) would be important, though too much weight should be put on it either.

5.1 Data Quality and Availability

Modern portfolio theories emphasizes that an investor, to optimize returns while avoiding risk, should allocate his or her wealth across different classes of securities and regularly rebalance their portfolio to hedge against market changes. The main parameter of the mean-variance model is means and variances quantified by returns and risks, which facilitates balancing expected return and reducing risk. A few studies have ensured that the asset selection and portfolio determination models work together. ML strategies, such as neural networks and hidden Markov models, are utilized for classification-related issues.

However, practical environments like the stock market do not present all information at once but provide continuous data. Thus, online learning is becoming increasingly important in dealing with never-ending data streams. Data everywhere is being generated: in sensors and smartphones, or as a result of the activity of users in social networking sites, among others. Information is captured using data streams, which are sequences of digital data that are generated in large quantities, with a time-varying distribution and speed, and that arrive very rapidly. A streaming application, capturing and possibly processing a data stream, can be generally characterized as a permanent data pipeline. Such an application is, thus, usually in continuous execution and only exceptionality stopped for maintenance or updates.

A data streaming model of computation incorporates input streams and output streams. Input streams are an unbounded sequence of elements, each of which is a character from a finite alphabet. Input elements are viewed as arriving online, meaning that the location of the next element is not known in advance but is revealed one at a time and incrementally. The first way to conduct portfolio prediction is a classification approach, which, however, only provides nominal investment decisions. On the other hand, a regression-based approach predicts the return of an asset, which is subsequently used to decide how to invest in that asset.

5.2 Model Interpretability and Transparency

The techniques and metrics designed to assess ML models can be thought of in two categories: specific model properties and general model properties (Mitros & Mac Namee, 2019). Specific model properties have a direct interpretation within a modeling method's architecture. Specific model-related methods assess the validity of properties that arise from the mechanism itself. On the other hand, general model properties apply to any mechanism without considering its definition. They are equated to tests of assumptions or tests of fit.

Model interpretability has become a highly sought-after quality of prediction systems. While the simplest models are inherently interpretable, the increasing use of black-box ML models necessitates post-hoc interpretability techniques as an alternative means of providing model comprehension. Interpretability also comes in various forms. Interpretable models can be black boxes but, despite a lack of transparency, it should still be possible to generate post-hoc interpretations. The latter fall into the category of post-hoc explanation or interpretability methods, while examples for the former include the use of sparse and compact models.

Ultimately, there exist many reasons for wanting smaller, simpler models, including to ensure the reliability and transparency of predictions, to discourage bias and discrimination, to facilitate human understanding, and to avoid cost or effort-intensive models. Transparency techniques should always ensure a data privacy constraint is respected. Conversely, model understanding may be impossible when there are too many components involved. This confirms the outlook that there are no single validity metric or sufficient validation properties. Instead, it seems promising to adopt a decision diagram for strategies to try and cover a large number of model ability properties. With regard to the toolbox of techniques devised in this paper, they remain agnostic as to whether a limited subset of the possible motifs or a wider set of motifs is needed for a modeling approach.

6.0 Future Directions in Portfolio Diversification

The performance of the ML methods is benchmarked against the combined performance of the traditional methods and the mean-variance model. The empirical results of examining the large basket of diversified equity indices in terms of better out-of-sample risk-return performance of ML models against traditional methods.

Prominent institutions, such as Citibank, JPMorgan, and Two Sigma, have utilized AI for a wide range of objectives, including developing strong anti-fraud systems, forecasting monetary policies, and identifying investment opportunities. In addition to well-known companies, academics have also made extensive efforts in many relevant fields. Considering the stock market as a complex adaptive system, neural networks are used to predict price trends in short-term trading. To avoid non-stationary time series issues, approaches based on the day's order book and trade data are explored as inputs to models. Investigate pertinent features that can represent stocks for prediction better. Their studies show that features derived from traditional financial theory, such as momentum and mean reversion, outperform features based on news events alone. However, most quantitative approaches are data-driven and lack scientific reasoning.

AI technology has numerous potential applications in the financial sector. Forecasting the movement of stocks and stock pools based on historical pricing or factor data is an important investment strategy in finance. Traditionally, statistical linear regression models, such as regression analysis based on the market model or CAPM, are employed to estimate the influence of various factors on the stock price. However, randomly down sampling many negatives makes it hard to explain the plausible relationship between factors and stock movement (e.g., Liu, 2024). Using support vector machines (SVM) for stock forecasting can better handle this instance imbalance issue, but the model still falls short of addressing the issue of explainability. To address the drawbacks of traditional models, a novel method that integrates the advantages of XGBoost and LIME is proposed. The two-step strategy of stacking models enables a multi-objective optimization model to be constructed for stock choice.

AI is developing rapidly in recent years and it has a profound influence and extensive applications in many fields including self-driving cars, visual recognition, intelligent voice and customer service. In the financial field, prominent institutions utilize AI to develop anti-fraud systems, forecast monetary policies, identify investment opportunities, etc. (Liu, 2024) summarizes AI applications in financial markets and institutions, covering forecast, strategy and execution, sentiment extraction, risk management, expert assistant systems, news/article/context understanding, and portfolio optimization.

Thanks to the significant progress and profound effects of AI development, it has become a focus point in the global competition of the new generation of science and technology. Many countries have taken AI as a strategic focus to be supported and guaranteed by the government. China has promulgated a series of AI policies from January 2020 to June 2023, among which 17 policies were released by several ministries of the Chinese central government targeting different industries. To be specific, "Guiding Opinions on Accelerating Scenario Innovation to Foster High-quality Economic Development through Advanced AI Applications" were issued in June 2022, emphasizing the importance of the application of advanced AI technology in key sectors including finance.

ML is one of the fundamental aspects of AI. The various scenarios that ML plays a crucial role in, which need thorough investigation, including stock selection, trading strategies, and stock portfolio diversification. During the last three decades, notable quantitative investment firms have invested substantial amounts of time and financial resources into using ML for stock selection and trading strategies. Many of these firms have gained significant profits from this effort. At the same time, much scholarly literature has investigated the portfolio investment.

7.0 CONCLUSION

AI has garnered attention in recent years, driven by the development of deep learning technology. This algorithm imitates the neural network structure of the human brain and achieves unparalleled success in fields such as image recognition, natural language processing, and historical weather and climate prediction (Jacobs et al., 2009). However, there is a lack of comparative literature on AI and traditional time series forecasting models in the application of financial forecasting. Therefore, the objective of this paper is to provide a comprehensive and fair comparison of the forecasting performance of ML and traditional forecasting models in the domain of financial forecasting.

To this end, classifiers, as GRNN, SMO, IBK, MLP, and Random Forest models, are selected, and the forecasting performance is compared with traditional time series models, such as regression, ARIMA, and Holt-Winters. The empirical study focuses on the forecasting performance instead of the construction of an ML forecasting model. Therefore, all classifiers are finely tuned with the available default parameters to establish the best performance.

Forecasting results show that traditional models outperform ML models for low prediction horizons, while ML models maintain a stable trend and are superior for high prediction horizons. When considering the overall forecasting performance of the two categories, the competitive elements are fundamental to promoting hybrid forecasting models in the future. Combined methods outperform individual models for both sides, indicating the effectiveness of a hybrid framework that extends the performance of all forecasting models.

The results reveal that the main statistical facts and forecasting mechanisms of time series are inherent properties that cannot be learned by a layer structure. ML models have great potential in addressing highly nonlinear forecasting problems. However, ML parameters greatly affect their performance, and most ML models are substantially complex with very slow fitting and forecasting speeds.

REFERENCES

[1] Boyd, S., Johansson, K., Kahn, R., Schiele, P., & Schmelzer, T. (2024). Markowitz portfolio construction at seventy. *arXiv preprint arXiv:2401.05080*.

[2] Goldberg, L. R., & Mouti, S. (2022). Sustainable investing and the cross-section of returns and maximum drawdown. *The Journal of Finance and Data Science*, *8*, 353-387.

[3] Hsing, T., Klüppelberg, C., & Kuhn, G. (2004). *Modelling, estimation and visualization of multivariate dependence for risk management* (No. 375). Discussion Paper.

[4] Jacobs, H., Müller, S., & Weber, M. (2009). How should private investors diversify?: An empirical evaluation of alternative asset allocation policies to construct a" world market portfolio". Available at: https://madoc.bib.uni-mannheim.de/2493/.

[5] Kumar Padhi, D., Padhy, N., Kumar Bhoi, A., Shafi, J., & Hassen Yesuf, S. (2022). An Intelligent Fusion Model with Portfolio Selection and ML for Stock Market Prediction. ncbi.nlm.nih.gov

[6] Kwak, J., Ahn, J., Lee, J., & Park, S. (2022). Shai-am: a machine learning platform for investment strategies. *arXiv preprint arXiv:2207.00436*.

[7] Liu, J. (2024). Predicting Chinese stock market using XGBoost multi-objective optimization with optimal weighting. *PeerJ Computer Science*, *10*, e1931.

[8] Mitros, J., & Mac Namee, B. (2019). A Categorisation of post-hoc explanations for predictive models. *arXiv preprint arXiv:1904.02495*.

[9] Padhi, D. K., Padhy, N., Bhoi, A. K., Shafi, J., & Yesuf, S. H. (2022). An intelligent fusion model with portfolio selection and machine learning for stock market prediction. *Computational Intelligence and Neuroscience*, *2022*(1), 7588303.

[10] Zhan, N., Sun, Y., Jakhar, A., & Liu, H. (2020, October). Graphical models for financial time series and portfolio selection. In *Proceedings of the first acm international conference on ai in finance* (pp. 1-6).

[11] Ziggel, D., Bissantz, N., & Bissantz, K. (2011). An empirical study of correlation and volatility changes of stock indices and their impact on risk figures. Acta Universitatis Danubius. Œconomica, 7(4).

Author Profile

Mayada Chibli is an instructor at the Lebanese International University in the Department of Banking and Finance. She holds a BBA and MBA in Finance, both with honors, and she is currently pursuing a PhD at the Bucharest University of Economic Studies, Romania. Her interests include FinTech, financial markets, and portfolio optimization. Her current research focuses on integrating deep learning with modern portfolio theory to enhance investment decision-making.

