Journal of Computer Science

Review

Economic Theory and Machine Learning Integration in Asset Pricing and Portfolio Optimization: A Bibliometric Analysis and Conceptual Framework

Patrick Kevin Aritonang, Sudarso Kaderi Wiryono and Taufik Faturohman

School of Business and Management, Institut Teknologi Bandung, Bandung, Indonesia

Article history
Received: 11-05-2024
Revised: 25-05-2024
Accepted: 06-06-2024

Corresponding Author:
Patrick Kevin Aritonang
School of Business and Management, Institut Teknologi Bandung, Bandung, Indonesia
Email: patrick_aritonang@sbm-itb.ac.id

Abstract: The integration of Machine Learning (ML) with economic theory has transformed financial market analysis, particularly in asset pricing and portfolio optimization. This study synthesizes existing research, identifies gaps and elucidates ML's impact on enhancing economic models and strategies through a Systematic Literature Review (SLR) of 401 relevant documents using bibliometric methodology and VOS viewer. Results show an increasing trend in publications, with the United States and China leading and institutions like the Shanghai University of Finance and Economics and Massachusetts Institute of Technology playing crucial roles. Cluster analysis reveals five main themes: Asset pricing and predictive analysis, algorithmic trading, data-driven portfolio management and optimization, reinforcement learning and adaptive strategies in financial markets and cryptocurrency market predictions. The proposed conceptual framework encompasses data acquisition and management, preprocessing and feature engineering, model selection and training, constraint formulation, theory-driven validation and real-time adaptation and monitoring, highlighting the potential synergy between ML and economic theory. This study provides insights into ML and economic theory integration, offering a structured pathway for enhancing asset pricing models and portfolio optimization techniques. Future research should refine this integration, focusing on practical application and adaptability to new financial instruments and market conditions.

Keywords: Machine Learning, Economic Theory, Asset Pricing, Portfolio Optimization

Introduction

Machine Learning (ML) has emerged as a transformative force in the finance sector, enabling the processing of large volumes of diverse data and capturing complex non-linear relationships. The field of asset pricing and portfolio optimization is undergoing a significant transformation, driven by the integration of ML technologies alongside traditional economic theory. This paradigm shift is a response to the limitations of conventional methods, which are increasingly unable to handle the vast amounts of data and complex relationships present in today's financial markets (Hannsgen, 2012). Studies by Rundo *et al.* (2019); Carta *et al.* (2021) have highlighted the superiority of machine learning techniques over traditional strategies, demonstrating enhanced performance in statistical arbitrage trading strategies through better return and risk management. The blend of economic theory with cutting-edge machine learning algorithms, such as Long Short-Term Memory (LSTM), has opened new avenues for predicting stock returns and optimizing portfolios (Choudhary and Arora, 2024). This approach leverages modified asset selection models, hyperparameter tuning kernels and advanced portfolio optimization frameworks like the Markowitz mean-variance model, both in its classic and modified forms, to achieve superior market performance, even in volatile conditions (Ferguson-Cradler, 2023).

There is currently a very limited number of papers that effectively combine machine learning with appropriate economic theory. This gap in the literature underscores the necessity of this study, which aims to bridge this gap by integrating ML with traditional economic theories to enhance financial modeling and analysis. Specifically, this study seeks to evaluate the research landscape, identify influential works, authors and institutions and uncover trends in applying ML to economic and financial issues through bibliometric data analysis. This analysis includes citation counts, co-citation networks, keyword frequencies and the proposal of a conceptual framework to integrate machine learning and economic theory. This study addresses the following research questions:

- RQ1: How can bibliometric analysis reveal trends and patterns in integrating machine learning with traditional economic theories in asset pricing and portfolio optimization
- RQ2: How can the proposed conceptual framework for integrating machine learning with economic theory enhance financial modeling and analysis

The analysis highlights the most impactful contributions, reveals key methodologies, datasets and evaluation metrics employed, sheds light on collaborative networks among researchers and institutions and identifies limitations and challenges in this research area, such as data scarcity, computational complexity and the interpretability of ML models in economic contexts. The conceptual framework proposed in this study is central to understanding how ML can be effectively integrated with economic theories.

Methods

This study uses Scopus as the main database due to its comprehensive coverage and wide selection of multidisciplinary articles, making it a reliable source for in-depth information across various fields globally. Scopus offers superior indexing quality compared to other databases, providing a robust dataset for analysis. It ensures high-quality metadata, which is crucial for accurate bibliometric studies (Mongeon and Paul-Hus, 2016). The literature search in Scopus is analyzed based on document types such as articles, proceedings, book chapters, books and other relevant document types and topics that align with the research focus. The selection of Scopus over other databases, such as Web of Science or Google Scholar, is justified by its superior indexing quality and the breadth of its multidisciplinary coverage. The flow of research literature selection can be seen in Fig. 1.

The bibliometric methodology was chosen for its ability to systematically quantify and analyze the landscape of academic literature. Bibliometric analysis is particularly effective for examining the intersection of fields such as machine learning and finance. It helps identify trends, key publications and influential authors in the domain (Zapata and Mukhopadhyay, 2022). This study approach uncovered 401 relevant documents, illustrating the significant interest and contributions in this niche area. The study period was confined to 2015-2024 to capture recent developments and trends.

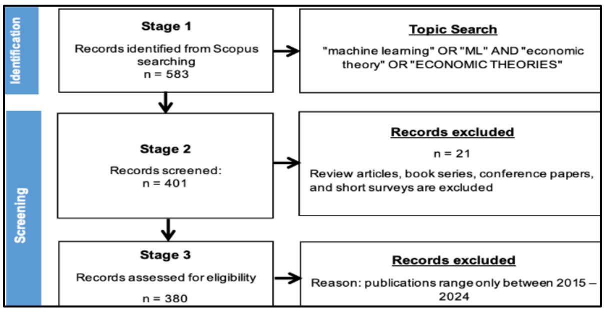


Fig. 1: Flow chart of the literature selection

By leveraging bibliometric indicators such as temporal trends, the prolificacy of countries and institutions and the impact of highly cited documents, this research utilizes VOS viewer as an advanced tool for bibliometric analysis. Bibliometric studies using VOS viewer help identify key contributors and influential works in specific research areas, such as the application of machine learning in finance (Sánchez García and Cruz Rambaud, 2022). The size of a node reflects the significance of the item it represents, while the thickness of a link between two nodes illustrates the intensity of their relationship, enabling a nuanced examination of collaboration patterns and influence within this field.

Despite the robust methodology, this study acknowledges several limitations that may impact the validity and generalizability of the findings. The reliance on Scopus as the sole database might introduce selection bias by excluding relevant studies indexed in other databases like Web of Science or Google Scholar and potentially omitting literature from lesser-known sources or non-English publications. A comparative analysis showed that significant variations exist in journal coverage between Scopus and Web of Science, impacting the comprehensiveness of bibliometric evaluations depending on the database used (Singh *et al.*, 2021). Bibliometrics analysis depends on citation data, which can be influenced by self-citations, the time lag between publication and citation and varying citation practices across disciplines, potentially introducing biases in perceived impact and influence.

Results

General Trends

The use of Machine Learning (ML) in asset valuation and portfolio optimization has received attention from researchers with the number of publications increasing from 2015 to 2023 (Fig. 2). To keep up with the development of related writings, this article categorizes the discussion of results in five main periods; specifically "the foundational a long time" (2015-2016), "extension and application" (2017-2018), "profound learning and enormous information analytics" (2019-2020), “integration and interdisciplinary approaches” (2021-2022), “towards autonomous finance” (2023-2024).

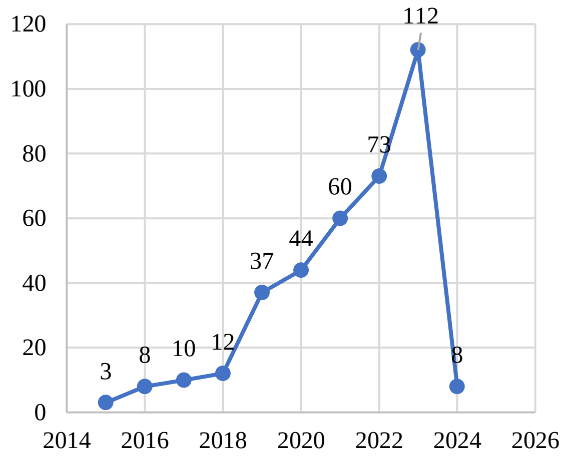


Fig. 2: Number of publications

2015-2016: The Foundational Years

The integration of Machine Learning (ML) into the financial sector during the initial phase of 2015-2016 marked a pivotal moment in the evolution of financial technologies, driven by the pioneering efforts of researchers like Bertsimas and Takeda (2015); Ghoddusi *et al.* (2019); Bonissone (2015). These early contributions laid a robust foundation for the application of ML in enhancing traditional financial models and optimizing portfolio management. Bertsimas and Takeda (2015) demonstrated the transformative potential of robust mixed-integer optimization approaches in financial decision-making, illustrating how ML could tackle complex datasets to optimize asset allocation and risk management strategies. Concurrently, Ghoddusi *et al*. (2019) explored the augmentation of the Black-Litterman model with alternative data sources, showcasing ML's ability to incorporate a wider array of information, from market data to social media insights, thereby refining investment strategies with enhanced market predictions.

Building on this foundation, the work of Deplano *et al*. (2016) further underscored the practical applications of ML in financial markets, particularly in the realm of portfolio optimization under budget constraints. Their research highlighted ML's capacity to navigate financial challenges with unprecedented precision, offering solutions that maximize returns while adhering to fiscal limitations. Additionally, Bonissone (2015) emphasized the importance of ensemble methods and model fusion in creating more accurate and resilient financial analysis tools, capable of adapting to market volatilities. Together, these early explorations into the integration of ML within finance not only validated the theoretical benefits but also paved the way for future advancements, establishing a precedent for the utilization of data-driven and computational techniques in revolutionizing asset pricing, risk assessment and portfolio management strategies.

Implications: These foundational efforts set the stage for future advancements by validating the theoretical benefits of ML and demonstrating its practical applications in finance. They highlight the importance of integrating diverse data sources and advanced optimization techniques to improve financial decision-making.

2017-2018: *Expansion and Application*

During the years 2017-2018, the finance sector witnessed an expansion in the application of Machine Learning (ML) techniques, heralding a new era of analytical capabilities in market analysis, risk management and asset allocation. This period was characterized by a significant shift towards leveraging alternative data sources to enhance asset pricing models, illustrating the growing recognition of ML's versatility and effectiveness in capturing complex market signals beyond traditional financial indicators. Studies like those by Houlihan and Creamer (2017) emphasized the potential of sentiment analysis from social media and options volume as predictive tools for future asset returns, showcasing the innovative ways in which ML could integrate diverse data types to improve investment strategies and model performance.

Furthermore, the exploration of intraday online investor sentiment by Renault (2017) and the development of portfolio optimization models by Ha *et al*. (2017) exemplified the practical applications of ML in real-time market conditions and portfolio management. These advancements underscored the ability of ML to not only analyze vast datasets but also to derive actionable insights that could anticipate market movements and optimize investment portfolios with a degree of precision and efficiency previously unattainable. The integration of quantitative data analysis techniques, as discussed by Shi *et al*. (2017), along with the innovative application of ML algorithms for multi-objective loan portfolio optimization by Srinivas and Rajendran (2017), further highlighted the transformative impact of ML on financial decision-making processes. This period marked a crucial phase in the evolution of finance, setting the stage for more sophisticated and data-driven approaches to investment management and financial analysis.

Implications: The expansion during this period underscores the importance of utilizing alternative data sources and advanced ML techniques to gain deeper insights into market dynamics. It suggests that integrating sentiment analysis and other non-traditional data can significantly enhance the predictive power of financial models.

2019-2020: Deep Learning and Big Data Analytics

The years 2019 and 2020 marked a transformative period in the financial sector, distinguished by groundbreaking strides in the application of deep learning and big data analytics. This era ushered in novel methodologies that profoundly reshaped traditional asset pricing and portfolio management frameworks. Aldridge's (2019) seminal work shed light on the untapped potential of big data in enhancing portfolio allocation strategies. By challenging the conventional wisdom of mean-variance portfolio theory, Aldridge demonstrated that optimizing the inverse of the correlation matrix rather than the matrix itself could significantly amplify the efficacy of portfolio selection strategies. This innovative approach was shown to deliver substantial outperformance over traditional strategies, including a remarkable 400% improvement over equally weighted allocations in a 20-year S&P 500 portfolio analysis. Concurrently, the research by Cong *et al*. (2021) delved into the realm of deep sequence modeling for asset pricing, showcasing the capability of deep learning to decipher complex datasets and extract pivotal insights for financial modeling.

These advancements signified a pivotal shift toward data-centric finance, where machine learning algorithms became instrumental in refining asset pricing models and optimizing investment strategies. The exploration into residential asset pricing prediction by Luo (2019) and the neural network-based framework for financial model calibration by Liu *et al*. (2019) exemplified the burgeoning role of machine learning in enhancing the precision of financial predictions. These methodologies leveraged diverse and intricate data sets, ranging from micro-level property features to broad market indicators, demonstrating an unparalleled ability to forecast asset prices with high accuracy. Furthermore, the integration of machine learning into portfolio optimization processes not only improved the strategic allocation of investments but also fostered a more dynamic and responsive investment management approach. This period of innovation laid the groundwork for a new era in finance, characterized by an emphasis on data-driven decision-making and the adoption of sophisticated analytical models. As these technologies continue to evolve, their potential to revolutionize financial analysis and investment strategies is boundless, promising ever-more refined and efficient solutions in the complex landscape of modern finance.

Implications: The advancements in deep learning and big data analytics highlight the transformative potential of these technologies in finance. They suggest that deep learning models can significantly improve the accuracy and efficiency of financial predictions and portfolio optimization, paving the way for more sophisticated data-driven approaches.

2021-2022: Integration and Interdisciplinary Approach

The years 2021-2022 further solidified the integration of Machine Learning (ML) into the financial sector, focusing on developing holistic and interdisciplinary frameworks that blend economic theory with advanced computational techniques. This period underscored the significance of leveraging the strengths of both fields to enhance the accuracy and efficiency of financial models. Notably, sustainability and ethical considerations began to play a crucial role in portfolio management, with ML models incorporating these factors to create more responsible investment strategies. For instance, William *et al*. (2021) highlighted the advantages of deep sequence modeling in asset pricing, demonstrating how ML can capture complex historical dependencies that traditional time-series models might overlook, thereby offering a more nuanced approach to predicting asset returns and measuring risk premiums.

In 2022, the field witnessed innovative explorations into the realms of smart beta and factor investing through the lens of Hsu *et al*. (2022) critiqued the performance of smart beta products, advocating for a diversified approach across a broader set of global factors. Their research underscored the effectiveness of ML models, such as linear ridge and gradient boosting, in generating significant excess returns by fitting expected returns to a comprehensive set of factors. This period also saw the application of ML in identifying companies with lasting competitive advantages (Jiménez-Preciado *et al*., 2022), demonstrating ML's potential to enhance stock portfolio optimization by identifying key financial ratios indicative of a company's moat. These advancements exemplify the increasing sophistication of ML applications in finance, highlighting a trend towards more dynamic, informed and ethically conscious investment strategies that leverage the predictive power and analytical depth of ML algorithms.

Implications: This integration period emphasizes the importance of interdisciplinary approaches that combine economic theories with ML techniques. It highlights the growing trend towards incorporating sustainability and ethical considerations into financial models, reflecting a broader shift towards responsible and informed investment strategies.

2023-2024: Towards Autonomous Finance

In the years 2023-2024, the financial industry has embarked on a transformative journey towards autonomous finance, a paradigm where Artificial Intelligence (AI) and Machine Learning (ML) technologies take the helm of financial decision-making and strategy formulation. This shift is largely fueled by significant advancements in reinforcement learning and AI-driven portfolio management systems. Such technologies promise to revolutionize the landscape of asset pricing and portfolio optimization, offering unprecedented levels of efficiency and precision. The introduction of fully automated trading strategies and decision-making processes not only enhances the speed and adaptability of financial operations but also brings forth a new level of personalization tailored to individual investor preferences and market dynamics. This move towards autonomous finance signifies a departure from traditional, manual intervention in financial markets, steering towards a future where algorithms and AI systems can dynamically adjust to real-time market changes, optimize investment returns and mitigate risks more effectively than ever before.

The evolution towards autonomy in finance during 2023-2024 marks a pivotal moment in the industry's history, reflecting a broader trend of digital transformation across global markets. As these AI-driven systems continue to mature, they are expected to handle increasingly complex financial tasks, from asset allocation and risk management to predictive analytics for asset pricing. This development not only heralds a new era of financial technology but also poses new challenges and opportunities for investors, asset managers and regulatory bodies. The implications of autonomous finance extend beyond mere operational efficiency, hinting at a future where financial markets operate with greater transparency, reduced human error and enhanced resilience against market volatility. As we move forward, the integration of AI and ML in finance promises to redefine the principles of asset management and investment strategy, making autonomous finance a cornerstone of the next-generation financial services industry.

Implications: Autonomous finance represents a significant leap in the evolution of financial technology. It suggests that AI-driven systems can dynamically adjust to real-time market changes, optimize investment returns and mitigate risks more effectively than traditional methods. This development poses new challenges and opportunities for investors, asset managers and regulatory bodies, heralding a future where financial markets operate with greater transparency and reduced human error.

Furthermore, the integration of economic theories with machine learning in asset pricing and portfolio optimization showcases a significant global trend towards leveraging computational approaches for financial market analysis. Leading the charge in this innovative field is the United States with a notable engagement value of 98, indicating a profound commitment to advancing financial technology research and application (Fig. 3). China follows closely with a value of 81, reflecting its substantial investments in machine learning applications for financial markets. India, with a value of 35 and the United Kingdom, at 28, also demonstrate significant contributions, underlining the widespread interest in this area across both advanced and emerging economies.

The trend continues with countries across various economic backgrounds engaging in this research domain. For instance, Germany (21), Italy (14) and France (13) contribute notably, highlighting the interest in merging traditional economic principles with modern computational techniques in Europe. Similarly, South Korea's engagement, valued at 10, aligns with its known penchant for technological innovation, particularly in applying machine learning to financial markets. Other countries, including Brazil and Canada (each with a value of 10), also participate actively, showing the universal appeal and applicability of these advanced techniques in understanding and navigating the complexities of global financial markets.

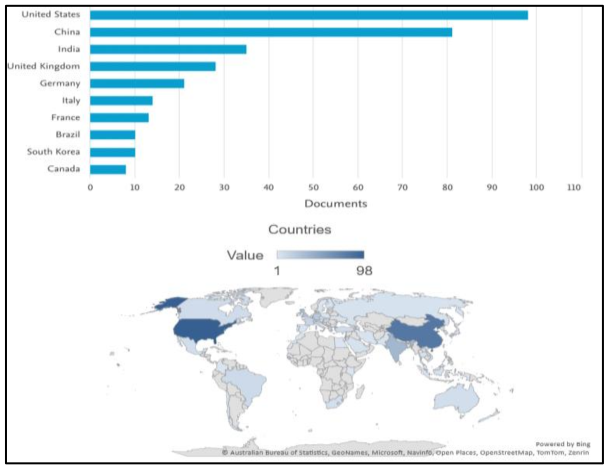


Fig. 3: Publications by countries

This broad spectrum of engagement from the United States and China to countries like South Korea and beyond emphasizes a global shift towards data-driven financial decision-making. It reflects a collective recognition of the potential that machine learning and artificial intelligence hold for enhancing the accuracy of asset pricing models and the efficiency of portfolio optimization strategies. This paradigm shift is not confined to the financial sectors of highly developed economies but is also evident in emerging markets, suggesting a future where technology and economics are increasingly intertwined to address the challenges and opportunities presented by global financial systems.

In Europe, countries like Germany, Italy and France are merging traditional economic principles with modern computational techniques, further highlighted by Germany's engagement in integrating solvency capital requirements with machine learning for portfolio optimization. Similarly, South Korea's investment in machine learning for the financial sector reflects its penchant for technological innovation, particularly in financial markets. This collective movement towards leveraging artificial intelligence across the financial sector suggests a future where technology and economics are increasingly interwoven to navigate global market complexities. In the United States, deep learning innovations and the identification of novel factors influencing asset prices are forefront, as evidenced by research from Chen *et al*. (2024), which showcases the use of neural networks for dynamic asset pricing models. This is complemented by Maasoumi *et al*. (2024), who apply debiased machine learning for high-dimensional risk factor identification and (Peng and Linetsky, 2022), proposing new portfolio optimization frameworks that incorporate economic theories with machine learning. In contrast, Chinese research delves into market dynamics and investor behavior, employing data fusion techniques to enhance asset pricing and portfolio strategies, as seen in the works by Jin and Sui (2022); Henrique *et al*. (2019)

who explore the predictive capabilities of machine learning models under varying market conditions. Similarly, in Korea, innovative approaches like those of Kim *et al*. (2023) leverage conditional autoencoders to improve the explanatory power of asset pricing models, indicative of a broader trend in the Korean financial research community towards adopting advanced computational techniques for financial modeling. This global engagement in machine learning applications for financial analysis underscores a collective recognition of its potential to revolutionize asset pricing accuracy and portfolio optimization efficiency, marking a significant paradigm shift in the financial sectors of both developed and emerging markets worldwide.

The analysis of institutions (Fig. 4) appearing in the list provides an overview of the academic entities playing a crucial role in integrating economic theory and machine learning in the context of asset pricing and portfolio optimization. Shanghai University of Finance and Economics, with the highest frequency of 7, stands out as one of the leaders in this field. This institution likely maintains a specific focus on financial economics research and the application of machine learning in asset and portfolio analysis. Additionally, tecnológico de Monterrey, Massachusetts Institute of technology, Stevens Institute of technology, the University of Chicago and The University of Chicago booth school of business, each with a frequency of 6, also play significant roles in developing methods that combine economic theory with cutting-edge technology in asset and portfolio management. Furthermore, institutions like the University of Johannesburg, Southwestern University of Finance and economics and the University of California, Berkeley, each with a frequency of 5, have a substantial presence in the thinking and research related to the topic. These institutions provide an ideal environment for interdisciplinary collaboration between economics and computer science, enabling the development of innovative solutions in asset analysis and portfolio management. Through this frequency distribution, it is evident that these institutions play crucial roles in advancing the integration of economic theory and machine learning for the purpose of asset valuation and portfolio optimization.

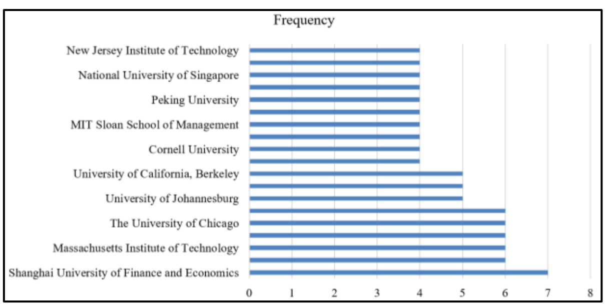


Fig. 4: Number of publications by institutions

Table 1: Number of publications by journal

Journal	Number of citations
Lecture notes in computer science	16
Expert systems with applications	14
Quantitative Finance	8
Lecture notes in Networks and Systems	7
Journal of Financial Data Science	7
Annals of Operations Research	5
ACM international conference proceeding series	5
Mathematics	4
Journal of risk and financial management	4
Journal of finance and data science	4

The academic landscape in various fields is often characterized by the prominence of journals that serve as important platforms for disseminating research findings (Table 1). In the realm of computer science and related disciplines, the "lecture notes in computer science" journal stands out with a notable citation count of 16, indicating its significant impact on the field. Similarly, "expert systems with applications" garners attention with 14 citations, underscoring its influence in the realm of applied artificial intelligence. In the realm of finance, "quantitative finance" is recognized with 8 citations, reflecting its importance in advancing quantitative methodologies for financial analysis and decision-making. Additionally, "lecture notes in networks and systems" and the "journal of financial data science" both receive 7 citations, highlighting their contributions to the understanding of network systems and financial data analysis, respectively. Furthermore, journals such as the "annals of operations research" and the "ACM international conference proceeding series" each receive 5 citations, signifying their roles in disseminating research on operational and computational aspects of various disciplines. Lastly, "mathematics," "journal of risk and financial management," and "journal of finance and data science" all receive 4 citations, demonstrating their relevance in publishing research at the intersection of mathematics, risk management and financial analytics. Collectively, these journals play vital roles in shaping scholarly discourse and advancing knowledge within their respective domains.

Furthermore, Table 2 summarizes ten articles that explore the integration of machine learning techniques in various aspects of finance. The first article by Gu *et al*. (2020) demonstrates substantial economic gains for investors through machine learning forecasts, identifying tree and neural network methods as particularly effective. Heaton *et al.* (2017) investigate the application of deep learning hierarchical models for financial prediction, showing improvements over traditional methods. Renault (2017) derives investor sentiment from social media and its impact on intraday stock returns, outperforming conventional sentiment analysis approaches. Zhang *et al*. (2019) propose a Generative Adversarial Network (GAN) for stock market prediction, yielding promising results in predicting closing prices. Beck and Jentzen (2019) introduce a method for solving high-dimensional fully nonlinear PDEs, demonstrating its efficiency in financial modeling. Elmachtoub and Grigas (2022) present a framework called Smart "Predict, then Optimize" (SPO), which leverages optimization problem structures to design better prediction models. Adapt machine learning methods for portfolio optimization, showing superiority over traditional benchmarks. Lamperti *et al*. (2018) combine machine learning and iterative sampling for efficient calibration of agent-based models, resulting in accurate model approximations. Buehler *et al*. (2019) introduce deep hedging, a framework for portfolio hedging using reinforcement learning methods, which shows promising results across different scenarios.

Last *et al*. (2020) propose a decision-making model for day trading investments using machine learning and portfolio selection, achieving significant performance improvements. These articles collectively contribute to advancing the use of machine learning in finance, offering insights and methodologies for various financial applications.

Based on Fig. 5, each cluster seems to represent a thematic focus area within the broader context of finance, economics and machine learning. Here are the proposed themes for each cluster.

Table 2: Most-cited articles related to the economic aspects of MSW management systems

Reference	Journal	Number of citations	Main result
Gu *et al*. (2020)	Review of financial studies	551	Demonstrated economic gains to investors using machine learning forecasts, identifying best-performing methods (trees and neural networks)
Heaton *et al*. (2017)	Applied Stochastic models in business	280	Explored deep learning hierarchical models for financial prediction, showing improved results over standard methods in finance and industry
Renault (2017)	Journal of Banking and finance	194	Derived investor sentiment from social media and its relation to intraday stock returns, outperforming standard methods in sentiment analysis
Zhang *et al*. (2019)	Procedia Computer science	163	Proposed a Generative Adversarial Network (GAN) for stock market prediction, showing promising performance in closing price prediction
Beck and Jentzen (2019)	Journal of Nonlinear science	155	Proposed a method for solving high-dimensional fully nonlinear PDEs, demonstrating their efficiency in financial models
Elmachtoub and Grigas (2022)	Management Science	147	Introduced a framework, Smart "Predict, then Optimize" (SPO), for designing better prediction models by leveraging optimization problem structure
Ban *et al*. (2018)	Management Science	142	Adapted machine learning methods, regularization and cross-validation, for portfolio optimization, showing dominance over benchmark methods
Lamperti *et al*. (2018)	Journal of Economic dynamics and control	140	Combined machine-learning and iterative sampling for efficient calibration of agent-based models, showing accurate model approximations
Buehler *et al*. (2019)	Quantitative Finance	138	Presented a framework, deep hedging, for portfolio hedging using reinforcement learning methods, showing promising results in various scenarios
Paiva *et al*. (2019)	Expert systems with applications	128	Proposed a decision-making model for day trading investments using machine learning and portfolio selection, showing significant results

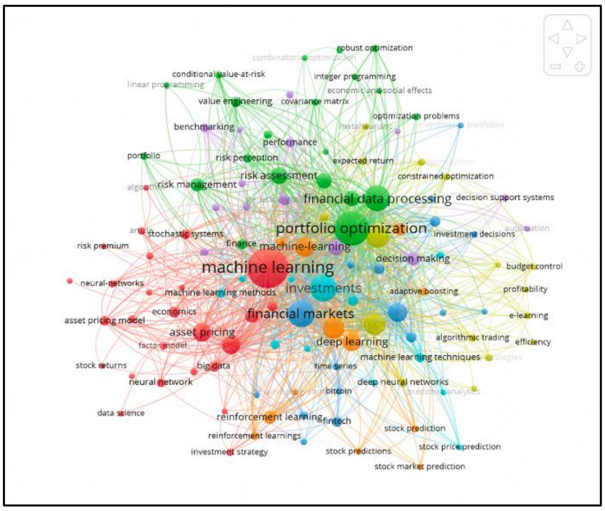


Fig. 5: Segmentation clustering

Cluster 1: Asset Pricing and Predictive Analysis

This cluster offers a comprehensive insight into the integration of ML techniques in asset pricing. It delves into the effectiveness of various ML models in predicting asset prices, compares these models with traditional asset pricing methods and critically evaluates empirical findings in the field. It explores the effectiveness of various ML models in forecasting asset prices, juxtaposing these models against traditional asset pricing methods while offering a critical examination of empirical evidence within the domain. Notably, ML predictions have been shown to significantly enhance performance beyond traditional regression-based strategies, with models like trees and neural networks standing out for their efficacy Gu *et al*. (2020). Furthermore, Fang and Taylor (2021) spotlight the superiority of ML approaches-including regularized linear, support vector machines, neural networks and tree-based-over basic ordinary least squares linear regression in the context of asset pricing. The LSTM model, when integrated with realized skewness, markedly boosts asset pricing performance, surpassing classical and other ML methods across various metrics (Choudhary and Arora, 2024). Additionally, autoencoder neural networks have demonstrated the ability to produce out-of-sample pricing errors that are significantly lower and generally negligible when compared to leading factor models (Gu *et al*., 2021).

The selection of the ML model is crucial and varies depending on the asset pricing task at hand, suggesting a preference for different ML methodologies in forecasting stock prices, bond yields, or real estate values. Machine learning algorithms have consistently outperformed traditional benchmark models like ARIMA and random walk-in predicting prices of Bitcoin futures (Akyildirim *et al*., 2021) and neural network classifiers have been remarkably accurate in foreseeing significant stock price changes (Kamalov, 2020). ML techniques also enhance linear asset pricing models by facilitating multiple hypothesis testing, thereby reducing false positives and improving hedge fund evaluation performance (Giglio *et al*., 2018). These findings affirm ML models as transformative in asset pricing, offering unparalleled predictive accuracy and performance across different asset classes and market scenarios relative to traditional models.

In practical settings, employing ML in asset pricing necessitates thoughtful consideration of data sources, preprocessing techniques and model selection. ML effectively integrates economic indicators, media content and momentum spillovers from related firms to improve asset pricing performance (Huang *et al*., 2022). However, the quality of financial data plays a pivotal role in the success of ML models, addressing challenges like missing data, noise and non-stationarity common in financial datasets. Data preprocessing methods, including normalization, feature engineering and dimensionality reduction, are essential for preparing data for effective ML analysis. Furthermore, sorting techniques are a cornerstone methodology in asset pricing, as they ascertain the contribution of specific attributes to the understanding of the variability in returns across different securities. Typically constructed around specific corporate attributes like market size or value-to-market ratio, these techniques can also incorporate predictive assessments to gauge forecast precision (Coqueret, 2020).

Blitz *et al*. (2023) note that traditional ML studies in asset pricing commonly aim to predict 1-month ahead stock returns, favoring variables with strong short-term predictive power. This approach, while not inherently unsuitable for long-term return predictions, faces challenges due to the need for extensive training data to achieve a comparable number of independent observations. For instance, a decade's training data yields 120 independent monthly observations but only 10 annual ones, making long-term return prediction more challenging due to limited data history. The transition from short-term to long-term return predictions underscores the significance of characteristics in ML predictions, with (Coqueret, 2020) documenting the importance of memory in ML-based models that utilize firm characteristics for asset pricing, finding that predictive algorithms perform optimally when trained on extensive samples with long-term returns as dependent variables.

Table 3 provides a structured overview of various machine learning models and their corresponding authors who have contributed to the field of asset pricing. Models such as Neural Networks, Random Forests and Support Vector Machines, among others, are listed to represent the diverse algorithmic approaches employed in financial analysis.

Table 3: Machine learning for asset pricing

Best-performing model	Authors
Neural network	Ceffer *et al*. (2019); Drobetz and Otto (2021); Fang *et al.* (2023); Gu *et al*. 2020; Kamalov (2020)
Random Forest	Alanis (2022); Escribano and Wang (2021); Götze *et al*. (2020); Yu *et al*. (2010)
Support Vector Machine	Aggarwal *et al*. (2020); Başoğlu Kabran and Ünlü (2021); Sedighi *et al*. (2019)
Recurrent Neural Network (RNN)	Koudjonou and Rout (2020); Lamothe-Fernández *et al*. (2020); Liu (2019)
Convolutional Neural Network (CNN)	Korade and Zuber (2023); Lee and Ko (2019); Liu and Wu (2023); Li *et al*. (2022)
Long Short-Term Memory (LSTM)	Fischer and Krauss (2018); Gao *et al*. (2021); Liu and Zhang (2023);
Bayesian Neural Network	Cocco *et al*. (2021); Jang and Lee (2019)
K-Nearest Neighbors	Gu *et al*. (2022); Tang *et al*. (2020a-b)
XGBoost	Akyildirim *et al*. (2023a-b); Ben Jabeur *et al.* (2021); Jiang *et al*. (2020)

Source: own work (2023)

Theoretical Foundation and Framework Integration

The theoretical foundation of this first cluster includes the Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT), which are foundational models in asset pricing that describe the relationship between systematic risk and expected return for assets, particularly stocks. This theoretical framework aids in understanding the transformative impact of ML on asset pricing models, data preprocessing techniques and the overall financial decision-making process.

CAPM is a foundational model in finance that describes the relationship between systematic risk and expected return for assets, particularly stocks (Sharpe, 1964). The limitation of CAPM is the assumption of a linear relationship and it often fails to account for complex market dynamics and nonlinearities.

CAPM Formula:

$$E(R_i) = \mathrm{R_f} + \beta_\mathrm{i}(\mathrm{ER_m} - \mathrm{R_f})$$

Arbitrage Pricing Theories (APT) is an extension of CAPM that considers multiple factors in determining asset prices (Ross, 1976), allowing for a more nuanced understanding of risk and return. APT has a limitation in identifying the correct factors and their influence on asset prices and APT also still assumes linear relationships.

APT Formula:

$$E(r_p) = r_f + \beta_1\lambda_1 + \beta_2\lambda_2 + \cdots + \beta_k\lambda_k$$

APT extends CAPM by incorporating multiple factors, providing a more nuanced understanding of asset prices. However, identifying the correct factors and their influence can be challenging and APT still assumes linear relationships.

Machine Learning (ML) techniques offer a transformative approach to asset pricing by addressing the limitations of CAPM and APT. Unlike traditional models, ML algorithms can capture complex, nonlinear relationships and interactions within financial data. For instance, neural networks, Support Vector Machines (SVM) and random forests can process large volumes of data and identify patterns that traditional models might miss. Empirical studies have demonstrated the superior performance of ML models over traditional methods in predicting asset prices, highlighting their ability to enhance predictive accuracy and uncover insights from diverse data sources.

The connection of this cluster with the technical aspects of machine learning can be seen in Chapter 4.

Cluster 2: Algorithmic Trading

Algorithmic trading has been redefined through the innovative contributions of Bogousslavsky *et al*. (2024), who introduced a pioneering approach for quantifying informed trading. By developing an Informed Trading Intensity (ITI) metric and employing machine learning techniques on a data set rich with informed trades, their work marks a significant advance in the field. This measure becomes particularly insightful around significant market events such as earnings announcements, mergers and acquisitions and news releases, highlighting its potential in predicting return reversals and implications for asset pricing. The strength of ITI lies in its ability to capture complex nonlinear relationships and interactions between informed trading, trade volume and market volatility, offering a nuanced understanding of the mechanics of informed trading. Such an approach not only sheds light on the dynamics of informed trading but also explores facets like impatient trading behaviors, commonality in informed trading across different assets and theoretical models of informed trading, making it a significant tool for analyzing market efficiency.

Complementing this, Tan *et al*. (2024) have unveiled a Context-Aware Hierarchical Attention Mechanism (CHARM), designed to encode unstructured textual media information. This innovative tool adeptly traces the tangible impact of news on stock movements influenced by media. By integrating encoded media content with other structured market data through tensor-based learning, CHARM facilitates the exploration and visualization of their combined effects on stock market fluctuations. Furthermore, a method for pre-estimating turning point locations for trading clues is employed, enhancing the efficiency of each investment opportunity. It not only augments the clarity of investment strategies but also significantly boosts predictive accuracy and investment returns when contrasted with existing models like AZFinText, TeSIA, eLSTM, CMT, MAC and SA-DLSTM.

Further enriching the discourse, Zhao *et al.* (2016) focus on forecasting return reversals through a two-tiered approach. Initially, a Dynamical Bayesian Factor Graph (DBFG) is utilized to distill key factors closely associated with return reversals from a broad spectrum of economic factors. Subsequently, these crucial factors are fed into various predictive models- Artificial Neural Network (ANN), Support Vector Machine (SVM) and Hidden Markov Model (HMM)-to forecast return reversals. Their analysis across the U.S. stock market reveals that while the pivotal factors for return reversals vary annually, elements tied to the liquidity effect theory consistently emerge as significant. Notably, the DBFG-ANN model outperforms its counterparts, achieving prediction accuracies above 60.

These studies have substantially advanced the field of algorithmic trading. Through innovative metrics, sophisticated data processing tools and advanced predictive models, these studies offer new avenues for understanding market movements, enhancing investment strategy effectiveness and improving the predictive accuracy of market analysis.

Theoretical Background and Framework Integration

The theoretical foundation of this cluster is rooted in the Efficient Market Hypothesis (EMH). The Efficient Market Hypothesis (EMH) defines that asset prices fully reflect all available information, meaning that it is impossible to consistently achieve returns higher than the overall market average (Malkiel and Fama, 1970). The formal expression of EMH can be summarized as follows.

EMH formula:

$$P^{e}_{t+1}=P^{e}_{t+1} \rightarrow R^{e} = R^{of} \rightarrow R^{of} = R^{*}$$

A limitation of EMH is that it assumes all market participants have access to all available information and interpret it in the same way, which often does not hold true in reality. Markets can be influenced by irrational behavior, information asymmetry and other anomalies that EMH does not account for.

By combining the concepts from EMH with Machine Learning (ML) techniques, algorithmic trading can enhance market efficiency and liquidity while reducing transaction costs. The use of ML models can help quantify informed trading and predict market movements more accurately, thereby aligning trading strategies with the theoretical underpinnings of EMH.

The connection of this cluster with the technical aspects of machine learning can be seen in chapter 4.

Cluster 3: Data-Driven Portfolio Management and Optimization

Machine learning techniques have demonstrated significant effectiveness in portfolio optimization for financial assets. These techniques range from clustering analysis, genetic algorithms and neural networks, to adaptive Bayesian optimization, each offering unique advantages in portfolio management. For instance, the application of Long Short-Term Memory (LSTM) networks for forecasting financial time series, in tandem with mean-variance models, significantly optimizes portfolio formation and asset preselection (Gu *et al.*, 2022). Integrating unsupervised machine learning methods, like clustering analysis, into mean-variance portfolio optimization models enhances the selection of assets in a portfolio (Tolun Tayalı, 2020). Novel frameworks like the DSRG Network have demonstrated superior performance in terms of returns compared to traditional strategies (Qin *et al.*, 2022). Additionally, the application of clustering analysis and in-band optimization in dynamic portfolio optimization presents a competitive advantage over classical approaches in financial mathematics (Mahdavi-Damghani *et al.*, 2021). Genetic algorithms in machine learning can optimally combine funds to create diversified portfolios that outperform market benchmarks. This method is particularly effective in combining various funds to achieve optimal portfolio diversification (Onek, 2019). Neural networks can enhance portfolio performance by incorporating macroeconomic conditions. They have shown superior results in terms of annualized return, volatility, Sharpe ratio and 99% Conditional Value at Risk (CVaR), compared to alternative methods (Chang and Yu, 2014; Yuan *et al.*, 2020). Adaptive Bayesian optimization in machine learning can manage risks effectively while achieving positive investment outcomes. This is achieved by adaptively tuning parameters to optimize the portfolio (Nyikosa *et al.*, 2019). Furthermore, portfolios constructed using k-means clustering have yielded returns exceeding those of leading indices and mutual funds, affirming its strength in portfolio construction (Kedia *et al.*, 2018). The integration of machine learning in portfolio theory offers advanced analytical tools for optimizing investment strategies and enhancing traditional models with data-driven insights and predictive capabilities.

DOI: 10.3844/jcssp.2024.1291.1309

Furthermore, Heaton *et al*. (2017) showcased the application of deep learning in constructing portfolios through a four-step algorithm that emphasizes model development and validation. This process consisting of auto-encoding, calibration, validation and verification presents a novel, model-independent approach to predictive analytics. This methodology aims to outline a comprehensive process for achieving specific investment goals, such as outperforming a benchmark, with success heavily reliant on the market framework defined by historical data. Here, portfolio optimization and inefficiency identification emerge as predominantly data-driven tasks, offering a fresh perspective divergent from classical portfolio theories.

The integration of ML is transforming decision-making processes through data-driven insights and optimization. Research in data-driven portfolio management, such as the works of Bogousslavsky *et al*. (2024); Cui *et al*. (2024), highlight ML's capability to not only dissect the micro-level intricacies of market dynamics like informed trading intensity but also to navigate the macro-level complexities of multi-period portfolio optimization through Deep Reinforcement Learning (DRL). These studies illustrate ML's broad applicability, from predicting market movements based on informed trades to optimizing investment strategies over time with significant efficiency gains.

In practical settings, ML techniques have evolved beyond theoretical constructs to become essential elements of real-world financial strategy formulation. These approaches harness data-driven insights and predictive analytics to redefine investment strategy landscapes. A key feature of these methodologies is their adaptability, offering the capacity for real-time updates and adjustments in response to evolving investor risk profiles.

Table 4 shows a range of machine learning applications in portfolio optimization. These studies employ various techniques that contribute to the evolving landscape of asset management. The listed methodologies reflect a trend toward leveraging complex algorithms for enhanced financial decision-making, signifying a paradigm shift in how data analytics can drive investment strategies.

Theoretical Foundation and Framework Integration

The most prominent theoretical basis of this cluster is Modern Portfolio Theory (MPT), Key concepts of MPT include diversification, the trade-off between risk and return, the efficient frontier, expected return and portfolio variance, all aimed at creating an optimal investment strategy. This foundational theory in finance was introduced by Harry Markowitz in his seminal paper "portfolio selection" (Markowitz, 1952).

MPT portfolio return:

$$R_p = \sum_{t=1}^{n} x_1 R_1$$

MPT portfolio risk:

$$\sigma_p = \sqrt{\mathrm{X}_1^2\sigma_1^2 + \mathrm{X}_2^2\sigma_2^2 + 2\mathrm{X}_1\mathrm{X}_1(r_{12}\sigma_1\sigma_2)}$$

The limitation of MPT is that asset returns are normally distributed, this assumption often does not hold true in real markets, where returns can exhibit skewness and kurtosis. This misalignment can lead to inaccurate risk assessments (Leahy, 2001). Furthermore, MPT involves a static optimization process, assuming that the statistical properties of returns, such as mean and variance, remain constant over time. This static nature fails to accommodate the dynamic and evolving conditions of financial markets. MPT also assumes homogeneous expectations among investors, which is unrealistic as investors often have varying information, risk preferences and investment horizons.

Table 4: Machine learning for portfolio optimization

No.	Author	Year	Machine learning algorithm
1	Pun and Wang (2021)	2019	Range-based risk using SVR
2	Naveed *et al*. (2023)	2019	Artificial Neural Network (ANN)
3	Kim *et al*. (2019)	2019	GA-rough set theory
4	Guo *et al*. (2020)	2020	Deep matching algorithm and deep stock profiling method
5	Kim and Kim (2020)	2020	Deep latent representation learning
6	Chen *et al*. (2020)	2020	Sparse-group lasso regularization
7	Ta *et al*. (2020)	2020	Long-short term memory network
8	Zhang *et al*. (2022)	2021	Listwise learn-to-rank algorithm
9	Zhang and Chen (2021)	2021	Double-screening Socially Responsible Investment (DSSRI)
10	Liu *et al*. (2021a-b)	2021	MGC algorithm and MGC-EM
11	Padhi *et al*. (2022)	2022	Intelligent fusion model
12	Chaweewanchon and Chaysiri (2022)	2022	Hybrid machine learning model

Source: own work (2023)

Machine Learning (ML) techniques, particularly clustering analysis, offer significant enhancements to MPT by addressing these limitations. Clustering algorithms, such as k-means clustering, can group assets based on their return characteristics and risk profiles, improving the asset selection process and enhancing portfolio diversification. This method goes beyond the simplistic linear relationships assumed by MPT, capturing more complex interactions and dependencies among assets.

The connection of this cluster with the technical aspects of machine learning can be seen in chapter 4.

Cluster 4: *Reinforcement Learning and Adaptive Strategies in Financial Markets*

The burgeoning field of financial technology has witnessed significant advancements through the application of Reinforcement Learning (RL), offering dynamic and adaptable solutions for portfolio optimization and strategic decision-making in the volatile realms of financial markets. Sarkar *et al*. (2024) pioneered the use of RL to enhance portfolio allocation strategies, showcasing its superiority over traditional methods by adeptly adapting to market fluctuations to maximize returns and minimize risks. This approach, validated through rigorous testing against historical data, marks a departure from static investment strategies, demonstrating RL's potential to achieve superior risk-adjusted returns in turbulent markets. Complementing this, Cui *et al*. (2024) introduced a Deep Reinforcement Learning (DRL) hyper-heuristic framework aimed at multi-period portfolio optimization, addressing scalability challenges and leveraging domain knowledge to outperform existing models. Furthermore, Khemlichi *et al*. (2020) expanded the horizon by integrating Multi-Agent Reinforcement Learning (MARL) with Proximal Policy Optimization (PPO) to foster a more realistic simulation of market conditions, thereby enhancing the robustness of investment strategies. Collectively, these studies underscore the transformative impact of RL and DRL in navigating the complexities of financial markets, offering innovative, data-driven approaches that promise to redefine the landscape of investment strategies and portfolio management with unprecedented adaptability and strategic depth.

Theoretical Background and Framework Integration

This cluster, like the previous ones, is anchored in the same foundational financial theories such as Modern Portfolio Theory (MPT). However, the primary distinction lies in the application and enhancement of these theories through Reinforcement Learning (RL) and its advanced variants.

MPT portfolio return:

$$R_p = \sum_{t=1}^{n} x_1 R_1$$

MPT portfolio risk*:*

$$\sigma_p = \sqrt{X_1^2\sigma_1^2 + X_2^2\sigma_2^2 + 2X_1X_1(r_{12}\sigma_1\sigma_2)}$$

Reinforcement Learning (RL) builds on the concepts of MPT by not only considering the trade-off between risk and return but also incorporating the ability to adapt and optimize portfolios continuously. MPT provides a framework for constructing portfolios that maximize expected return for a given level of risk through static optimization, typically assuming that the statistical properties of returns remain constant over time.

The connection of this cluster with the technical aspects of machine learning can be seen in chapter 4.

Cluster 5: *Cryptocurrency Markets Predictions*

Cryptocurrency markets are at the forefront of the application of machine learning predictions, given their volatile and data-rich nature. Lorenzo and Arroyo (2023) present an innovative risk-based portfolio allocation method that utilizes a prototype-based clustering algorithm specifically tailored for the cryptocurrency domain. This approach addresses the sensitivity of mean-variance portfolio optimization models to the uncertainties in risk-return estimates that often result in subpar performance. By focusing on a selected cluster of crypto assets that align with an investor's risk aversion and applying machine learning models such as Random Forest, the research offers a strategic advantage in managing investments in the unpredictable cryptocurrency market.

Concurrently, Erfanian *et al*. (2022) delve into predicting Bitcoin prices within an economic theory framework, employing machine learning methods like support vector regression to discern the influence of macroeconomic, microeconomic, technical and blockchain indicators. Notably, their research identifies certain technical indicators as critical for short-term Bitcoin price forecasts, thereby validating the relevance of technical analysis. They found that macroeconomic and blockchain indicators emerge as significant predictors for the long term, suggesting that theories of supply, demand and cost-based pricing fundamentally underpin Bitcoin price predictions. Their findings underscore the efficacy of machine learning over traditional statistical analysis in forecasting prices, contributing significantly to asset pricing and investment decision-making.

Saad *et al*. (2020) delve into Bitcoin and Ethereum, examining network features that correlate with price movements. Through data analysis on user and network activity and its significant impact on cryptocurrency prices, they draw connections to economic theories. The study identifies critical network features by analyzing correlations with variables such as hash rate, user count, transaction rate

DOI: 10.3844/jcssp.2024.1291.1309

and total bitcoins, shedding light on the demand and supply dynamics within the cryptocurrency market. Employing ML techniques, including regression, LSTM networks and the conjugate gradient algorithm, they develop models capable of predicting cryptocurrency prices with remarkable accuracy. Validation across two extensive datasets confirms the efficacy of their models, demonstrating up to 99% accuracy in forecasting Bitcoin and Ethereum prices.

These studies collectively illuminate the potential of machine learning in deciphering the complex dynamics of cryptocurrency markets, offering insightful methodologies for portfolio management, price prediction and investment strategy enhancement. Through the use of ML and economic theory, these contributions pave the way for more sophisticated and effective approaches to navigating the intricacies of cryptocurrency investment.

Theoretical Background and Framework Integration

This cluster, like the previous ones, is grounded in the Efficient Market Hypothesis (EMH). The EMH posits that asset prices fully reflect all available information, making it impossible to consistently achieve higher returns than the market average through trading strategies based on available information.

EMH formula:

$$P_{t+1}^{e} = P_{t+1}^{e} \rightarrow R^{e} = R^{of} \rightarrow R^{of} = R^{*}$$

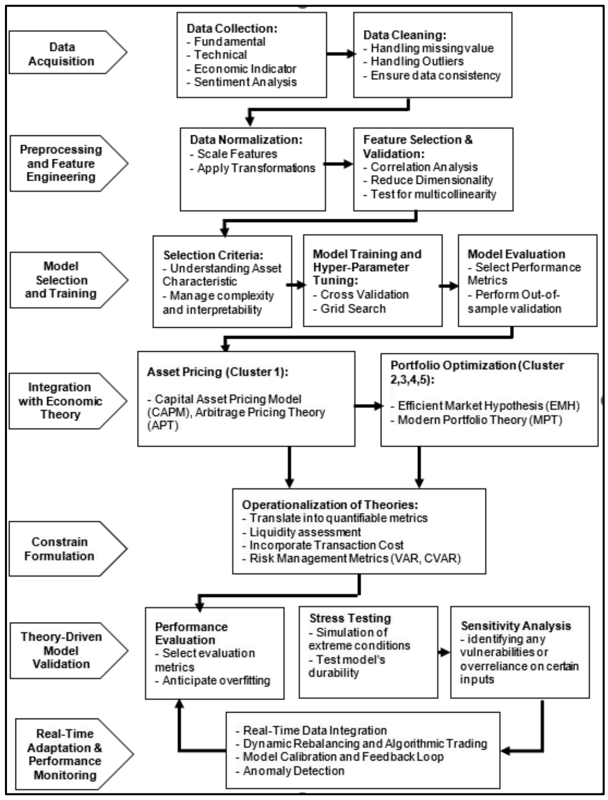
The application of EMH to cryptocurrency markets presents unique challenges. Cryptocurrency markets are newer, less regulated and more susceptible to price manipulation, information asymmetry and extreme volatility. These factors can lead to deviations from the ideal market efficiency postulated by EMH.

Machine Learning (ML) techniques significantly enhance the ability to navigate the challenges of cryptocurrency markets, as evidenced by the studies in this cluster. By employing methods such as prototype-based clustering, random forest, Support Vector Regression (SVR) and Long Short-Term Memory (LSTM) networks, ML models can effectively manage the high volatility and unpredictability of these markets.

The connection of this cluster with the technical aspects of machine learning can be seen in chapter 4.

Conceptual Framework

The comprehensive conceptual framework presented in the image aligns with the key findings and themes discussed in the provided document content. The conceptual framework in Fig. 6 presents a comprehensive approach for integrating Machine Learning (ML) with economic theory in the domain of asset pricing and portfolio optimization.



Source: own work (2024)

Fig. 6: Conceptual framework for machine learning and economic theory integration

The framework begins with data acquisition, which involves collecting fundamental, technical, economic and sentiment analysis data. The collected data then undergoes preprocessing and feature engineering, including data cleaning, handling missing values, dealing with outliers and ensuring data consistency. The preprocessed data is normalized and relevant features are selected and validated using techniques such as correlation analysis, reduced dimensionality and testing for multicollinearity.

The next stage is model selection and training, where appropriate ML models are chosen based on the asset characteristic, complexity and interpretability. The selected models are trained using cross-validation, grid search and hyper-parameter tuning techniques and then evaluated using performance metrics such as out-of-sample validation and perform validation. This stage involves choosing appropriate ML models based on asset characteristics, complexity and interpretability, resonating with the document's findings on the importance of model selection, with various ML models showing superior performance in asset pricing and portfolio optimization tasks.

The integration with the economic theory stage incorporates well-established financial theories such as the Capital Asset Pricing Model (CAPM), Arbitrage Pricing Theory (APT), the Efficient Market Hypothesis (EMH) and Modern Portfolio Theory (MPT), anchoring the ML models in sound economic principles. The CAPM, introduced by Sharpe (1964), describes the relationship between systematic risk and expected return for assets, particularly stocks. The APT, proposed by Ross (1976), extends CAPM by considering multiple factors in determining asset prices, allowing for a more nuanced understanding of risk and return. The EMH, developed by Malkiel and Fama (1970), posits that asset prices fully reflect all available information, making it impossible to consistently achieve higher returns than the overall market average. MPT, introduced by Markowitz (1952), suggests that investors can construct an optimal portfolio that maximizes expected return for a given level of risk through diversification. This stage emphasizes the synergy between ML and economic theory, consistently highlighted in the document.

Theory-driven model validation ensures the models' outputs are consistent with economic theory through stress testing under extreme conditions, testing the model's durability and conducting sensitivity analysis to identify vulnerabilities or overreliance on certain inputs. This aligns with the document's emphasis on ML models adhering to the economic rationale and the significance of model validation techniques. Real-time adaptation and performance monitoring ensures the models remain up-to-date and relevant by integrating real-time data, employing dynamic rebalancing and algorithmic trading, calibrating models through a feedback loop and detecting anomalies, guaranteeing the models adapt to changing market conditions and investor preferences while continuously enhancing performance. This resonates with the document's discussion on the evolution towards autonomous finance and the importance of adaptability to changing market conditions and investor preferences.

The 5 main periods discussed in the document showcase the progressive integration of ML and economic theory in asset pricing and portfolio optimization, well-represented in the framework's interconnected stages. The document's cluster analysis reveals 5 main themes: Asset pricing and predictive analysis, algorithmic trading, data-driven portfolio management and optimization, reinforcement learning and adaptive strategies in financial markets and cryptocurrency market predictions, which are captured in the framework's components. In conclusion, the conceptual framework comprehensively encapsulates the key findings, trends and themes discussed in the provided document content, offering a structured approach for the integration of ML and economic theory in asset pricing and portfolio optimization.

Conclusion

This study demonstrates the transformative impact of integrating Machine Learning (ML) with economic theory on financial market analysis, particularly in asset pricing and portfolio optimization. The proposed conceptual framework, supported by comprehensive bibliometric analysis, encapsulates the synergistic potential of ML and economic theory, providing a structured pathway for enhancing traditional models and techniques. The framework encompasses data acquisition and management, preprocessing and feature engineering, model selection and training, constraint formulation, theory-driven validation and real-time adaptation and monitoring, ensuring the resulting models are robust, adaptive and attuned to the complexities of modern financial markets.

In comparison to existing frameworks such as the traditional Capital Asset Pricing Model (CAPM), Arbitrage Pricing Theory (APT), Efficient Market Hypothesis (EMH) and Modern Portfolio Theory (MPT), the proposed framework integrates advanced ML techniques that allow for more dynamic and data-driven approaches to asset pricing and portfolio management. Traditional models often rely on static assumptions and linear relationships, which can be limiting in capturing the complexities of financial markets. In contrast, our framework leverages the predictive power of ML to adapt to changing market conditions, enhancing the accuracy and robustness of financial models.

For practitioners, this integration offers more sophisticated, data-driven financial strategies, enhancing decision-making and risk management and potentially leading to higher returns. Policymakers can leverage these insights to craft regulations that ensure market stability and protect investors, addressing ethical considerations, data privacy issues and potential biases in algorithmic trading.

The study reveals the growing interest and contributions in this interdisciplinary field, with the United States and China emerging as leaders and identifies five main themes: Asset pricing and predictive analysis, algorithmic trading, data-driven portfolio management and optimization, reinforcement learning and adaptive strategies in financial markets and cryptocurrency markets predictions. As the financial landscape becomes increasingly complex, this study paves the way for advancements that blend ML agility with the depth of economic insight, leading to more informed, effective and efficient asset pricing and portfolio management strategies.

Future research should focus on refining this integration and adapting the framework to emerging financial instruments and market conditions. This includes developing more sophisticated models that can handle the complexities of new financial products, such as cryptocurrencies and addressing the ethical and regulatory challenges that arise with the increased use of ML in finance. By continuing to explore and enhance the synergy between ML and economic theory, researchers can contribute to a more efficient and adaptable approach to financial modeling and analysis.

Acknowledgment

We would like to express our gratitude to the reviewers, Publisher and all those who support this research.

Funding Information

This research was funded by the authors' own resources.

Author's Contributions

Patrick Kevin Aritonang: Data acquisition, data analysis and paper written.

Sudarso Kaderi Wiryono and Taufik Faturohman: Supervision, conceptualization and review.

Ethics

This article is original and contains unpublished material.

References

Aggarwal, D., Chandrasekaran, S., & Annamalai, B. (2020). A complete empirical ensemble mode decomposition and support vector machine-based approach to predict Bitcoin prices. *Journal of Behavioral and Experimental Finance*, *27*. https://doi.org/10.1016/j.jbef.2020.100335

Akyildirim, E., Goncu, A., & Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research*, *297*(1–2). https://doi.org/10.1007/s10479-020-03575-y

Alanis, E. (2022). Forecasting betas with random forests. *Applied Economics Letters*, *29*(12). https://doi.org/10.1080/13504851.2021.1912278

Ban, G.-Y., El Karoui, N., & Lim, A. E. B. (2018). Machine Learning and Portfolio Optimization. *Management Science*, *64*(3), 1136–1154. https://doi.org/10.1287/mnsc.2016.2644

Başoğlu Kabran, F., & Ünlü, K. D. (2021). A two-step machine learning approach to predict S&P 500 bubbles. *Journal of Applied Statistics*, *48*(13–15). https://doi.org/10.1080/02664763.2020.1823947

Beck, C., Weinan, E., & Jentzen, A. (2019). Machine Learning Approximation Algorithms for High-Dimensional Fully Nonlinear Partial Differential Equations and Second-order Backward Stochastic Differential Equations. *Journal of Nonlinear Science*, *29*(4). https://doi.org/10.1007/s00332-018-9525-3

Ben Jabeur, S., Khalfaoui, R., & Ben Arfi, W. (2021). The effect of green energy, global environmental indexes, and stock markets in predicting oil price crashes: Evidence from explainable machine learning. *Journal of Environmental Management*, *298*. https://doi.org/10.1016/j.jenvman.2021.113511

Bertsimas, D., & Takeda, A. (2015). Optimizing over coherent risk measures and non-convexities: a robust mixed integer optimization approach. *Computational Optimization and Applications*, *62*(3), 613–639. https://doi.org/10.1007/s10589-015-9755-3

Blitz, D., Hoogteijling, T., Lohre, H., & Messow, P. (2023). How Can Machine Learning Advance Quantitative Asset Management? *The Journal of Portfolio Management*, *49*(9), 78–95. https://doi.org/10.3905/jpm.2023.1.460

Bogousslavsky, V., Fos, V., & Muravyev, D. (2024). Informed Trading Intensity. *The Journal of Finance*, *79*(2), 903–948. https://doi.org/10.1111/jofi.13320

Bonissone, P. P. (2015). Machine Learning Applications. In *Springer Handbook of Computational Intelligence* (pp. 783–821). Springer. https://doi.org/10.1007/978-3-662-43505-2_41

Buehler, H., Gonon, L., Teichmann, J., & Wood, B. (2019). Deep hedging. *Quantitative Finance*, *19*(8), 1271–1291. https://doi.org/10.1080/14697688.2019.1571683

Carta, S. M., Consoli, S., Piras, L., Podda, A. S., & Recupero, D. R. (2021). Explainable Machine Learning Exploiting News and Domain-Specific Lexicon for Stock Market Forecasting. *IEEE Access*, *9*, 30193–30205. https://doi.org/10.1109/access.2021.3059960

Ceffer, A., Levendovszky, J., & Fogarasi, N. (2019). Applying Independent Component Analysis and Predictive Systems for Algorithmic Trading. *Computational Economics*, *54*(1), 281–303. https://doi.org/10.1007/s10614-017-9719-z

Chang, Y.-P., & Yu, C.-T. (2014). Bayesian confidence intervals for probability of default and asset correlation of portfolio credit risk. *Computational Statistics*, *29*(1–2), 331–361. https://doi.org/10.1007/s00180-013-0453-2

Chaweewanchon, A., & Chaysiri, R. (2022). Markowitz Mean-Variance Portfolio Optimization with Predictive Stock Selection Using Machine Learning. *International Journal of Financial Studies*, *10*(3), 64. https://doi.org/10.3390/ijfs10030064

Chen, J., Dai, G., & Zhang, N. (2020). An application of sparse-group lasso regularization to equity portfolio optimization and sector selection. *Annals of Operations Research*, *284*(1), 243–262. https://doi.org/10.1007/s10479-019-03189-z

Chen, L., Pelger, M., & Zhu, J. (2024). Deep Learning in Asset Pricing. *Management Science*, *70*(2), 714–750. https://doi.org/10.1287/mnsc.2023.4695

Choudhary, A., & Arora, A. (2024). Assessment of bidirectional transformer encoder model and attention-based bidirectional LSTM language models for fake news detection. *Journal of Retailing and Consumer Services*, *76*, 103545. https://doi.org/10.1016/j.jretconser.2023.103545

Cocco, L., Tonelli, R., & Marchesi, M. (2021). Predictions of bitcoin prices through machine learning based frameworks. *PeerJ Computer Science*, *7*, e413. https://doi.org/10.7717/peerj-cs.413

Cong, L. W., Tang, K., Wang, J., & Zhang, Y. (2021). Deep Sequence Modeling: Development and Applications in Asset Pricing. *The Journal of Financial Data Science*, *3*(1), 28–42. https://doi.org/10.3905/jfds.2020.1.053

Coqueret, G. (2020). Stock-specific sentiment and return predictability. *Quantitative Finance*, *20*(9), 1531–1551. https://doi.org/10.1080/14697688.2020.1736314

Cui, T., Du, N., Yang, X., & Ding, S. (2024). Multi-period portfolio optimization using a deep reinforcement learning hyper-heuristic approach. *Technological Forecasting and Social Change*, *198*, 122944. https://doi.org/10.1016/j.techfore.2023.122944

Deplano, I., Squillero, G., & Tonda, A. (2016). Portfolio Optimization, a Decision-Support Methodology for Small Budgets. In *Applications of Evolutionary Computation* (Vol. *9597*, pp. 58–72). Springer International Publishing. https://doi.org/10.1007/978-3-319-31204-0_5

Drobetz, W., & Otto, T. (2021). Empirical asset pricing via machine learning: Evidence from the European stock market. *Journal of Asset Management*, *22*(7), 507–538. https://doi.org/10.1057/s41260-021-00237-x

Elmachtoub, A. N., & Grigas, P. (2022). Smart "Predict, then Optimize." *Management Science*, *68*(1), 9–26. https://doi.org/10.1287/mnsc.2020.3922

Erfanian, S., Zhou, Y., Razzaq, A., Abbas, A., Safeer, A. A., & Li, T. (2022). Predicting Bitcoin (BTC) Price in the Context of Economic Theories: A Machine Learning Approach. *Entropy*, *24*(10), 1487. https://doi.org/10.3390/e24101487

Escribano, Á., & Wang, D. (2021). Mixed random forest, cointegration and forecasting gasoline prices. *International Journal of Forecasting*, *37*(4), 1442–1462. https://doi.org/10.1016/j.ijforecast.2020.12.008

Fang, M., & Taylor, S. (2021). A machine learning based asset pricing factor model comparison on anomaly portfolios. *Economics Letters*, *204*, 109919. https://doi.org/10.1016/j.econlet.2021.109919

Fang, Y., Chen, Y., & Ren, H. (2023). A factor pricing model based on machine learning algorithm. *International Review of Economics & Finance*, *88*, 280–297. https://doi.org/10.1016/j.iref.2023.06.012

Ferguson-Cradler, G. (2023). Narrative and computational text analysis in business and economic history. *Scandinavian Economic History Review*, *71*(2), 103–127. https://doi.org/10.1080/03585522.2021.1984299

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

Gao, Z., He, Y., & Kuruoglu, E. E. (2021). A Hybrid Model Integrating LSTM and Garch for Bitcoin Price Prediction. *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*, 1–6. https://doi.org/10.1109/mlsp52302.2021.9596429

Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics*, *81*, 709–727. https://doi.org/10.1016/j.eneco.2019.05.006

Giglio, S., Liao, Y., & Xiu, D. (2018). Thousands of Alpha Tests. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3259268

Götze, T., Gürtler, M., & Witowski, E. (2020). Improving CAT bond pricing models via machine learning. *Journal of Asset Management*, *21*(5), 428–446. https://doi.org/10.1057/s41260-020-00167-0

Gu, A., Wang, Z., Yin, Z., Zhou, M., Li, S., Xun, Q., & Dong, J. (2022). Empirical Research for Investment Model Based on VMD-LSTM. *Mathematical Problems in Engineering*, *2022*, 1–16. https://doi.org/10.1155/2022/4185974

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Gu, S., Kelly, B., & Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics*, *222*(1), 429–450. https://doi.org/10.1016/j.jeconom.2020.07.009

Guo, G., Rao, Y., Zhu, F., & Xu, F. (2020). Innovative deep matching algorithm for stock portfolio selection using deep stock profiles. *PLOS ONE*, *15*(11), e0241573. https://doi.org/10.1371/journal.pone.0241573

Ha, M., Yang, Y., & Wang, C. (2017). A portfolio optimization model for minimizing soft margin-based generalization bound. *Journal of Intelligent Manufacturing*, *28*(3), 759–766. https://doi.org/10.1007/s10845-014-1011-7

Hannsgen, G. (2012). Infinite-variance, alpha-stable shocks in monetary SVAR. *International Review of Applied Economics*, *26*(6), 755–786. https://doi.org/10.1080/02692171.2012.686484

Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry*, *33*(1), 3–12. https://doi.org/10.1002/asmb.2209

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, *124*, 226–251. https://doi.org/10.1016/j.eswa.2019.01.012

Houlihan, P., & Creamer, G. G. (2017). Can Sentiment Analysis and Options Volume Anticipate Future Returns? *Computational Economics*, *50*(4), 669–685. https://doi.org/10.1007/s10614-017-9694-4

Hsu, J., Liu, X., Viswanathan, V., & Xia, Y. (2022). When Smart Beta Meets Machine Learning and Portfolio Optimization. *The Journal of Beta Investment Strategies*, *13*(4), 123–146. https://doi.org/10.3905/jbis.2022.1.015

Huang, J., Xing, R., & Li, Q. (2022). Asset pricing via deep graph learning to incorporate heterogeneous predictors. *International Journal of Intelligent Systems*, *37*(11), 8462–8489. https://doi.org/10.1002/int.22950

Jang, H., & Lee, J. (2019). Generative Bayesian neural network model for risk-neutral pricing of American index options. *Quantitative Finance*, *19*(4), 587–603. https://doi.org/10.1080/14697688.2018.1490807

Jiang, Z., Ji, R., & Chang, K.-C. (2020). A Machine Learning Integrated Portfolio Rebalance Framework with Risk-Aversion Adjustment. *Journal of Risk and Financial Management*, *13*(7), 155. https://doi.org/10.3390/jrfm13070155

Jiménez-Preciado, A. L., Venegas-Martínez, F., & Ramírez-García, A. (2022). Stock Portfolio Optimization with Competitive Advantages (MOAT): A Machine Learning Approach. *Mathematics*, *10*(23), 4449. https://doi.org/10.3390/math10234449

Jin, L. J., & Sui, P. (2022). Asset pricing with return extrapolation. *Journal of Financial Economics*, *145*(2), 273–295. https://doi.org/10.1016/j.jfineco.2021.10.009

Kamalov, F. (2020). Forecasting significant stock price changes using neural networks. *Neural Computing and Applications*, *32*(23), 17655–17667. https://doi.org/10.1007/s00521-020-04942-3

Kedia, V., Khalid, Z., Goswami, S., Sharma, N., & Suryawanshi, K. (2018). Portfolio Generation for Indian Stock Markets Using Unsupervised Machine Learning. *2018 4th International Conference on Computing Communication Control and Automation (ICCUBEA)*, 1–5. https://doi.org/10.1109/iccubea.2018.8697771

Khemlichi, F., Chougrad, H., Khamlichi, Y. I., el Boushaki, A., & Ben Ali, S. E. (2020). Deep Deterministic Policy Gradient for Portfolio Management. *2020 6th IEEE Congress on Information Science and Technology (CiSt)*, 424–429. https://doi.org/10.1109/cist49399.2021.9357266

Kim, E., Cho, T., Koo, B., & Kang, H.-G. (2023). Conditional autoencoder asset pricing models for the Korean stock market. *PLOS ONE*, *18*(7), e0281783. https://doi.org/10.1371/journal.pone.0281783

Kim, J., Shin, S., Lee, H. S., & Oh, K. J. (2019). A Machine Learning Portfolio Allocation System for IPOs in Korean Markets Using GA-Rough Set Theory. *Sustainability*, *11*(23), 6803. https://doi.org/10.3390/su11236803

Kim, S., & Kim, S. (2020). Index tracking through deep latent representation learning. *Quantitative Finance*, *20*(4), 639–652. https://doi.org/10.1080/14697688.2019.1683599

Korade, N. B., & Zuber, Mohd. (2023). Stock Forecasting Using Multichannel CNN and Firefly Algorithm. In *Proceedings of the 2nd International Conference on Cognitive and Intelligent Computing* (pp. 447–458). Springer. https://doi.org/10.1007/978-981-99-2742-5_46

Koudjonou, K. M., & Rout, M. (2020). A stateless deep learning framework to predict net asset value. *Neural Computing and Applications*, *32*(14), 1–19. https://doi.org/10.1007/s00521-019-04525-x

Lamothe-Fernández, P., Alaminos, D., Lamothe-López, P., & Fernández-Gámez, M. A. (2020). Deep Learning Methods for Modeling Bitcoin Price. *Mathematics*, *8*(8), 1245. https://doi.org/10.3390/math8081245

Lamperti, F., Roventini, A., & Sani, A. (2018). Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control*, *90*, 366–389. https://doi.org/10.1016/j.jedc.2018.03.011

Last, C. (2020). Global Commons in the Global Brain. In *Global Brain Singularity* (pp. 107–147). Springer. https://doi.org/10.1007/978-3-030-46966-5_7

Leahy, R. (2001). Depressive Decision Making: Validation of the Portfolio Theory Model. *Journal of Cognitive Psychotherapy*. *15*. 341-362.
https://doi.org/10.1891/0889-8391.15.4.341

Lee, H.-C., & Ko, B. (2019). Fund Price Analysis Using Convolutional Neural Networks for Multiple Variables. *IEEE Access*, *7*, 183626–183633. https://doi.org/10.1109/access.2019.2959022

Li, Z., Zhou, B., & Hensher, D. A. (2022). Forecasting automobile gasoline demand in Australia using machine learning-based regression. *Energy*, *239*, 122312. https://doi.org/10.1016/j.energy.2021.122312

Liu, D., & Wu, Y. (2023). Option pricing using deep convolutional neural networks enhanced by technical indicators. *2023 IEEE 9th International Conference on Cloud Computing and Intelligent Systems (CCIS)*. 2023 IEEE 9th International Conference on Cloud Computing and Intelligent Systems (CCIS), Dali, China. https://doi.org/10.1109/ccis59572.2023.10262865

Liu, F., Li, Y., Li, B., Li, J., & Xie, H. (2021a). Bitcoin transaction strategy construction based on deep reinforcement learning. *Applied Soft Computing*, *113*, 107952. https://doi.org/10.1016/j.asoc.2021.107952

Liu, S., Borovykh, A., Grzelak, L. A., & Oosterlee, C. W. (2019). A neural network-based framework for financial model calibration. *Journal of Mathematics in Industry*, *9*(1), 9. https://doi.org/10.1186/s13362-019-0066-7

Liu, S., Wu, S., Xie, F., & Yuan, G. (2021b). Computer simulation of investment efficiency function model based on GMM method and artificial intelligence. *Journal of Ambient Intelligence and Humanized Computing*, 1–11. https://doi.org/10.1007/s12652-021-03112-0

Liu, Y. (2019). Novel volatility forecasting using deep learning–Long Short-Term Memory Recurrent Neural Networks. *Expert Systems with Applications*, *132*, 99–109. https://doi.org/10.1016/j.eswa.2019.04.038

Liu, Y., & Zhang, X. (2023). Option Pricing Using LSTM: A Perspective of Realized Skewness. *Mathematics*, *11*(2), 314. https://doi.org/10.3390/math11020314

Lorenzo, L., & Arroyo, J. (2023). Online risk-based portfolio allocation on subsets of crypto assets applying a prototype-based clustering algorithm. *Financial Innovation*, *9*(1), 25. https://doi.org/10.1186/s40854-022-00438-2

Luo, Y. (2019). Residential Asset Pricing Prediction using Machine Learning. *2019 International Conference on Economic Management and Model Engineering (ICEMME)*, 193–198. https://doi.org/10.1109/icemme49371.2019.00046

Maasoumi, E., Wang, J., Wang, Z., & Wu, K. (2024). Identifying factors via automatic debiased machine learning. *Journal of Applied Econometrics*, *39*(3), 438–461. https://doi.org/10.1002/jae.3031

Mahdavi-Damghani, B., Mustafayeva, K., Buescu, C., & Roberts, S. (2021). Portfolio optimization for cointelated pairs: SDEs vs Machine learning. *Algorithmic Finance*, *8*(3–4), 101–125. https://doi.org/10.3233/af-200311

Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, *7*(1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

Naveed, H. M., HongXing, Y., Memon, B. A., Ali, S., Alhussam, M. I., & Sohu, J. M. (2023). Artificial Neural Network (ANN)-based estimation of the influence of COVID-19 pandemic on dynamic and emerging financial markets. *Technological Forecasting and Social Change*, *190*, 122470. https://doi.org/10.1016/j.techfore.2023.122470

Nyikosa, F. M., Osborne, M. A., & Roberts, S. J. (2019). Adaptive Configuration Oracle for Online Portfolio Selection Methods. *ArXiv*, arXiv:1908.08258.

Onek, T. (2019). A Genetic Algorithm Model for Financial Asset Diversification. *2019 SoutheastCon*, 1–6. https://doi.org/10.1109/southeastcon42311.2019.9020556

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*, 1–18. https://doi.org/10.1155/2022/7588303

Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2019). Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Systems with Applications*, *115*, 635–655. https://doi.org/10.1016/j.eswa.2018.08.003

Peng, Y., & Linetsky, V. (2022). Portfolio Selection: A Statistical Learning Approach. *Proceedings of the Third ACM International Conference on AI in Finance*, 257–263. https://doi.org/10.1145/3533271.3561707

Pun, C. S., & Wang, L. (2021). A cost-effective approach to portfolio construction with range-based risk measures. *Quantitative Finance*, *21*(3), 431–447. https://doi.org/10.1080/14697688.2020.1781237

Qin, Y., Gu, F., & Su, J. (2022). A Novel Deep Reinforcement Learning Strategy for Portfolio Management. *2022 7th International Conference on Big Data Analytics (ICBDA)*, 366–372. https://doi.org/10.1109/icbda55095.2022.9760348

Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance*, *84*, 25–40. https://doi.org/10.1016/j.jbankfin.2017.07.002

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, *13*(3), 341–360. https://doi.org/10.1016/0022-0531(76)90046-6

Rundo, F., Trenta, F., di Stallo, A. L., & Battiato, S. (2019). Machine Learning for Quantitative Finance Applications: A Survey. *Applied Sciences*, *9*(24), 5574. https://doi.org/10.3390/app9245574

Saad, M., Spaulding, J., Njilla, L., Kamhoua, C., Shetty, S., Nyang, D., & Mohaisen, D. (2020). Exploring the Attack Surface of Blockchain: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials*, *22*(3), 1977–2008. https://doi.org/10.1109/comst.2020.2975999

Sánchez García, J., & Cruz Rambaud, S. (2022). Machine Learning Regularization Methods in High-Dimensional Monetary and Financial VARs. *Mathematics*, *10*(6), 877. https://doi.org/10.3390/math10060877

Sarkar, P., Dwibedi, P., Deore, S. S., Gawali, T., Kulkarni, M. D., & Saha, A. (2024). Portfolio Optimization in Dynamic Markets: Reinforcement Learning for Investment. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(13s), 386–395. https://ijisae.org/index.php/IJISAE/article/view/4605

Sedighi, M., Jahangirnia, H., Gharakhani, M., & Farahani Fard, S. (2019). A Novel Hybrid Model for Stock Price Forecasting Based on Metaheuristics and Support Vector Machine. *Data*, *4*(2), 75. https://doi.org/10.3390/data4020075

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Shi, X., Zhang, P., & Khan, S. U. (2017). Quantitative Data Analysis in Finance. In *Handbook of Big Data Technologies* (pp. 719–753). Springer International Publishing. https://doi.org/10.1007/978-3-319-49340-4_21

Singh, V. K., Singh, P., Karmakar, M., & *et al*. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*(6), 5113-5142. https://doi.org/10.1007/s11192-021-03948-5

Srinivas, S., & Rajendran, S. (2017). A Data-Driven Approach for Multiobjective Loan Portfolio Optimization Using Machine-Learning Algorithms and Mathematical Programming. In *Big Data Analytics Using Multiple Criteria Decision-Making Models* (1st Ed., pp. 175–210). CRC Press. https://doi.org/10.1201/9781315152653-8

Ta, V.-D., Liu, C.-M., & Tadesse, D. A. (2020). Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading. *Applied Sciences*, *10*(2), 437. https://doi.org/10.3390/app10020437

Tan, J., Deveci, M., Li, J., & Zhong, K. (2024). Asset pricing via fused deep learning with visual clues. *Information Fusion*, *102*, 102049. https://doi.org/10.1016/j.inffus.2023.102049

Tang, L., Pan, H., & Yao, Y. (2020a). Computational Intelligence Prediction Model Integrating Empirical Mode Decomposition, Principal Component Analysis and Weighted k -Nearest Neighbor. *Journal of Electronic Science and Technology*, *18*, 341–349. https://doi.org/10.11989/JEST.1674-862X.80124016

Tang, H., Xu, Y., Lin, A., Heidari, A. A., Wang, M., Chen, H., Luo, Y., & Li, C. (2020b). Predicting Green Consumption Behaviors of Students Using Efficient Firefly Grey Wolf-Assisted K-Nearest Neighbor Classifiers. *IEEE Access*, *8*, 35546–35562. https://doi.org/10.1109/access.2020.2973763

Tolun Tayalı, S. (2020). A novel backtesting methodology for clustering in mean–variance portfolio optimization. *Knowledge-Based Systems*, *209*, 106454. https://doi.org/10.1016/j.knosys.2020.106454

William, C., Ke Tang, J. Wang Yang Zhang, L., Capponi, A., Han, B., Karolyi, A., Kozak, S., Neuhierl, A., Grennan, J., Nagel, S., Pelger, M., Prado, M., Shams, A., Sheng, J., Tompaidis, S., Shi, Z., Weber, M., Xiu, D., Zhang, L., Zhou, G., & Zhou, L. (2021). AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI. *Social Science Research Network*, 3554486. http://dx.doi.org/10.2139/ssrn.3554486

Yu, P. L. H., Li, W. K., & Jin, S. (2010). On some models for value-at-risk. *Econometric Reviews*, *29*(5–6), 622–641. https://doi.org/10.1080/07474938.2010.481972

Yuan, X., Yuan, J.f, Jiang, T., & Ain, Q. U. (2020). Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market. *IEEE Access*, *8*, 22672–22685. https://doi.org/10.1109/access.2020.2969293

Zapata, H. O., & Mukhopadhyay, S. (2022). A Bibliometric Analysis of Machine Learning Econometrics in Asset Pricing. *Journal of Risk and Financial Management*, *15*(11), 535. https://doi.org/10.3390/jrfm15110535

Zhang, J., & Chen, X. (2021). Socially Responsible Investment Portfolio Construction with a Double-Screening Mechanism considering Machine Learning Prediction. *Discrete Dynamics in Nature and Society*, *2021*, 1–15. https://doi.org/10.1155/2021/7390887

Zhang, K., Zhong, G., Dong, J., Wang, S., & Wang, Y. (2019). Stock Market Prediction Based on Generative Adversarial Network. *Procedia Computer Science*, *147*, 400–406. https://doi.org/10.1016/j.procs.2019.01.256

Zhang, X., Wu, L., & Chen, Z. (2022). Constructing long-short stock portfolio with a new listwise learn-to-rank algorithm. *Quantitative Finance*, *22*(2), 321–331. https://doi.org/10.1080/14697688.2021.1939117

Zhao, S., Tong, Y., Meng, X., Yang, X., & Tan, S. (2016). Predicting return reversal through a two-stage method. *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 341–344. https://doi.org/10.1109/icsess.2016.7883081