5-6-2024

A Hierarchical Decision Model for Evaluating the Strategy Readiness of Quantitative Machine Learning/Data Science-Driven Investment Strategies

Mohammadsaleh Saadatmand
Portland State University

A Hierarchical Decision Model for Evaluating the Strategy Readiness of Quantitative Machine Learning/Data Science-Driven Investment Strategies

by

Mohammadsaleh Saadatmand

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Technology Management

Dissertation Committee:
Tugrul U. Daim, Chair
Judith Estep
Carlos Mena
Haydar Yalcin

Portland State University
2024

Abstract

Big data and computational technologies are increasingly important worldwide in asset and investment management. Many investment management firms are adopting these data science methods and technologies to improve performance across all investment processes. Researchers actively use these methods to develop more effective systematic investment strategies and produce more reliable outcomes less vulnerable to human decision-making biases. However, the success of such a strategy depends heavily on the scientific rigor applied throughout the process. "Best practices involve understanding how to make better decisions in the research design process. A good question is whether we can make better decisions in developing quantitative strategies. Therefore, the decisions made in the research process are crucial to developing successful quantitative strategies." Additionally, as this field is inherently multidisciplinary, it requires a system thinking approach to consider multiple perspectives to provide a clearer understanding of the strategies often referred to as "black boxes."

Therefore, the main objective of this research is to develop a multi-criteria assessment framework and scoring decision support system to evaluate quantitative investment strategies that apply machine learning and data science techniques in their research and development. Subject matter experts will assess all framework perspectives from a systematic literature review to approve their reliability. The perspectives consist of economic and financial foundations, data perspective, features perspective, modeling perspective, and performance perspective. The research methodology applied is the Hierarchical Decision Model (aka HDM) to provide a 360-degree view of the quantitative

investment strategy and improve and generalize the concept to other asset classes and regions. Finally, this research helps investment researchers and professionals to focus on research process decisions in generating more hypotheses and developing financial theories to be tested empirically rather than cherry-picking investment strategies based on historical simulations.

DEDICATION

“I would rather have questions that cannot be answered than answers which cannot be questioned.”

-Richard Feynman-

Dedicated to the pillars of my life – my beloved family, whose boundless love and unwavering support have fueled my aspirations. To my esteemed advisor and mentors, your guidance and wisdom have illuminated my path. And to the relentless innovators and champions of progress in technology management, may this work serve as a beacon of inspiration and a testament to our collective pursuit of excellence.

ACKNOWLEDGEMENTS

This research endeavor has been a collective journey molded by the unwavering support and inspiration of numerous individuals for whom I am deeply grateful. Attempting to adequately convey my appreciation to all those who have played a pivotal role in this academic expedition is a humbling task.

First and foremost, I express my heartfelt thanks to my family, whose unwavering support, understanding, and encouragement have served as the bedrock of my academic pursuit. To my beloved wife, Oria, your enduring patience and unwavering support have guided me through moments of uncertainty. Your steadfast belief in my abilities has strengthened and motivated me. I also sincerely appreciate my extended family in Iran – my parents, siblings, and relatives – whose unwavering confidence in my academic aspirations has been a constant source of inspiration.

I am profoundly grateful to my esteemed committee members for your invaluable guidance, encouragement, and insightful discussions that have shaped my research trajectory. Your expertise and resolute advocacy have played an indispensable role in refining my ideas and ensuring their scholarly rigor.

Dr. Daim, your willingness to serve as my chairperson, mentor, and guide on this academic journey has been instrumental in navigating the complexities and challenges we encountered along the way. Your wise counsel, unwavering encouragement, and steadfast support have been a beacon of light in the scholarly pursuit. I am deeply appreciative of your mentorship and guidance.

I sincerely thank Dr. Mena, Dr. Yalcin, and Dr. Estep for your invaluable expertise, guidance, and insightful contributions to this research endeavor. Your scholarly insights have significantly enriched the depth and quality of my work.

I also sincerely thank the esteemed technology management and financial data science luminaries who generously shared their insights and expertise through collaborative endeavors and enlightening panel discussions. Your scholarly contributions have infused vitality and relevance into my research, enhancing its academic impact.

Lastly, to my fellow students turned friends, your camaraderie, support, and shared zeal for learning have been a constant source of inspiration on this academic journey. The bonds forged and the shared experiences have enriched our collective intellectual pursuit.

In conclusion, I am profoundly grateful to each and every individual who has contributed to this scholarly odyssey in their unique way. Your support, guidance, and encouragement have been instrumental in shaping this intellectual journey and bringing it to fruition.

TABLE OF CONTENTS

Abstract i
DEDICATION iii
ACKNOWLEDGEMENTS iv
LIST OF TABLES ix
LIST OF FIGURES xi
LIST OF EQUATIONS xiii
CHAPTER 1. INTRODUCTION 1
1.1 Background 1
1.2 Quantitative Investing Primer 6
1.3 Research Motivation 11
1.4 Problem Statement 16
CHAPTER 2. LITERATURE REVIEW 20
2.1 AI and Machine Learning 20
2.1.1 Supervised 21
2.1.2 Unsupervised 22
2.1.3 Reinforcement 22
2.1.4 Deep Learning 23
2.2 AI in Finance and Asset Management 27
2.2.1 Current Trends 29
2.2.2 Machine Learning Techniques in Asset Management 31
2.2.3 Financial Data Science and Machine Learning 32
2.3 Primer of Empirical Asset Pricing and Factor Investing 44
2.3.1 Machine Learning in Empirical Asset Pricing and Factor Investing 48
CHAPTER 3. RESEARCH GAPS, GOALS, AND OUTPUTS 53
3.1 Research Gap 53
3.1.1 Overview 53
3.1.2 Gap Analysis 58
3.2 Research Goals 63
3.3 Research Outputs 64

CHAPTER 4. RESEARCH FRAMEWORK 66
4.1 Background and Literature Review.. 67
4.2 Research Model Development 68
4.3 Panel Formation, Model Validation and Quantification 69
4.4 Model Application and Results Analysis 70
4.5 Discussion and Conclusion 71
CHAPTER 5. RESEARCH METHODOLOGY 72
5.1 HDM Model 73
5.1.1 Hierarchical Decision Model (HDM) Overview 73
5.1.2 Desirability Curves 77
5.1.3 Inconsistency, Disagreement, and Sensitivity Analysis 78
5.1.4 HDM Benefits and Limitations 80
5.1.5 Justification of The Method.. 82
5.1.6 Model Generalizability 84
5.2 Experts Judgement 87
5.2.1 Expert's Judgement Overview.. 87
5.2.2 Expert Panel Formation 95
5.2.3 Experts Inconsistencies.. 101
5.2.4 Experts Disagreement... 106
5.2.5 Sensitivity Analysis 110
5.3 Methodology Comparison with Other Methods 114
CHAPTER 6. RESEARCH MODEL DEVELOPMENT AND RESULTS.................................. 119
6.1 The Initial HDM Model 119
6.1.1 Model Perspectives... 121
6.1.2 Model Criteria... 122
6.2 MODEL VALIDATION AND QUANTIFICATION 133
6.2.1 HDM Model Validation... 133
6.2.2 HDM Model Quantification 145
6.3 Results Analysis 155
6.3.1 Inconsistency and Disagreement Analysis 155
6.3.2 Final Model Weights 156

6.4 Desirability Curves Validation and Quantification 160
CHAPTER 7. RESEARCH MODEL APPLICATION 178
Case Study 1: The US Hedge Fund Investing in European Equity Market 178
Case Study 2: The Mexican Quant Firm Investing in the US Equity Market............. 181
CHAPTER 8. CASE STUDIES AND SENSITIVITY ANALYSIS 183
8.1 Readiness Assessment Scores 183
8.2 Strengths, Weaknesses, and Improvement Simulation 187
8.3 Sensitivity Analysis 191
8.4 Recommended Improvements.. 200
CHAPTER 9. RESEARCH VALIDATION.. 204
CHAPTER 10. DISCUSSION AND CONCLUSION 205
10.1 Discussion and Recent Research 205
10.2 Conclusion.. 209
10.3 Recommendations 212
10.4 Expected Contributions 214
10.4.1 Academic Contributions 214
10.4.2 Professional Contributions 216
10.5 Limitations 221
10.6 Future Research.. 223
REFERENCES 225
APPENDICES 247
Appendix A: Letter of Invitation to Experts 247
Appendix B: Letter of Model Validation to Experts.. 248
Appendix C: Letter of Model and Desirability Curve Quantifications 249
Appendix D. Supplemental File Information... 250

LIST OF TABLES

Table 1: Machine Learning Characteristics 26
Table 2: AI in Finance Classification 29
Table 3: Investment Questions and Associated Data Analysis Techniques 30
Table 4: Benefits of Machine Learning and Data Science in Investment Research 38
Table 5: Challenges and Considerations of Machine Learning and Data Science in Investment Research 40
Table 6: Related Work to The Financial Data Science Research Process 42
Table 7: The Main Risk Factors 49
Table 8: Ranking of Papers by The Number of Citations 58
Table 9: Measure Description 79
Table 10: Expert Panels 100
Table 11: Expert Panels by Background 101
Table 12: Comparison of Methods 120
Table 13: HDM Model Perspectives 121
Table 14: HDM Model Criteria 122
Table 15: The Expert Panels' Roles in The Validation Phase 133
Table 16: Perspective's Validation Summary 135
Table 17: Detailed Perspectives Validation 135
Table 18: Economic Foundations and Research Criteria Summary 137
Table 19: Detailed Summary of Economic Foundations and Research Criteria 137
Table 20: Data Criteria Summary 138
Table 21: Detailed Summary of Data Criteria 138
Table 22: Feature Criteria Summary 140
Table 23: Detailed Summary of Features Criteria 140
Table 24: Modeling Criteria Summary 141
Table 25: Detailed Summary of Modeling Criteria 141
Table 26: Performance Criteria Summary 143
Table 27: Detailed Summary of Performance Criteria 143
Table 28: The Final HDM Model Validation 144
Table 29: The Expert Panels' Roles in The Quantification Phase 145
Table 30: Experts' Distribution Across The Quantification Panels 146
Table 31: Perspectives Detailed Summary 147
Table 32: Economic Foundations and Research Factors Detailed Summary 149
Table 33: Data factors detailed summary 150
Table 34: Features Factors Detailed Summary 151
Table 35: Modeling Factors Detailed Summary 153
Table 36: Performance Factors Detailed Summary 154
Table 37: HDM Model Final Weights 157
Table 38: Desirability Curves for The Model Criteria 160
Table 39: Strategy (1) Assessment Score 184

Table 40: Strategy (2) Assessment Score 185
Table 41: Strategy Application Overall Assessment Score 186
Table 42: Strengths and Weaknesses of Case 1 187
Table 43: Strengths and Weaknesses of Case 2 188
Table 44: Implemented Scenarios 191
Table 45: Scenario 1 Outcomes for Case 1 and Case 2 193
Table 46: Scenario 2 Outcomes for Case 1 and Case 2 194
Table 47: Scenario 3 Outcomes for Case 1 and Case 2 195
Table 48: Scenario 4 Outcomes for Case 1 and Case 2 197
Table 49: Scenario 5 Outcomes for Case 1 and Case 2 198
Table 50: Summary of Scenario Analysis 199
Table 51: Improvement Simulation – Case 1 202
Table 52: Improvement Simulation – Case 2 203
Table 53: Summary of The Research Gaps and The Research Contributions 215
Table 54: Summary of the research outputs and the research contributions 216

LIST OF FIGURES

Figure 1: Global Asset Under Management Growth 1
Figure 2: Global AI Adoption by Asset Management 2
Figure 3: Global AUM Projection for 2020 to 2025... 4
Figure 4: Evolution of Quantitative Investing 7
Figure 5: The Complementary Relationship of Econometrics and Financial Data Science 13
Figure 6: The Theoretical Model 15
Figure 7: Neural Network Architectures.. 25
Figure 8: Number of Papers Published Using AI Technology Over Time, 1996 – 2018. 31
Figure 9: Primary ML Techniques Commonly Used in Asset Management.................... 33
Figure 10: Left: Statistics vs. Machine Learning Terminology, Proper: Hypothesis-Based vs. Data-Driven Analysis 37
Figure 11: Asset Class Breakdown into Factors 50
Figure 12: The Growing Trend of Publications 55
Figure 13: Top 10 Journals by The Number of Publications 56
Figure 14: Co-occurrence of Keywords and Growing Trend of Deep Learning Research in Finance 57
Figure 15: Bibliographic Coupling Network of Papers 59
Figure 16: Gap Analysis Logic 60
Figure 17: Research Gaps, Goals, and Outputs 65
Figure 18: Research Design 66
Figure 19: Example of an HDM Hierarchy 74
Figure 20: HDM Framework 76
Figure 21: Sample Desirability Curve 79
Figure 22: Social Network Analysis (SNA) Steps... 93
Figure 23: Network Visualization of Scientific Papers and Their Citation Importance... 94
Figure 24: Citation Network Visualization of Papers on Web of Science 94
Figure 25: The Initial HDM 121
Figure 26: The Pre-Validation HDM Model 134
Figure 27: Perspectives Validation 135
Figure 28: Economic Foundations and Research Validation... 136
Figure 29: DataValidation.. 138
Figure 30: Feature Validation 139
Figure 31: Modeling Validation... 141
Figure 32: Performance Validation.. 142
Figure 33: The Post-Validation HDM Model 145
Figure 34: Relative Weight of Perspectives in Descending Order 147
Figure 35: Relative Weight of Factors in Descending Order 149
Figure 36: Relative Weight of Factors in Descending Order 150
Figure 37: Relative Weight of Factors in Descending Order 151

Figure 38: Relative Weight of Factors in Descending Order 152
Figure 39: Relative Weight of Factors in Descending Order 154
Figure 40: Factors With Their Impact Weight on ML-DS-based Investment Strategies 158
Figure 41: Local Criteria Weights 159
Figure 42: Global Criteria Weights... 159

LIST OF EQUATIONS

Equation 2: Inconsistency Formula 105
Equation 3: Disagreement Formula 109
Equation 4: Global and Local Criterion Contribution in a Sensitivity Analysis 112

CHAPTER 1. INTRODUCTION

1.1 Background

Change in the asset and investment management industry is now accelerating exponentially. Technology advances in artificial intelligence, big data, and machine learning will drive fast transformation across business models and investment processes, including valuation, portfolio management, risk management, and investment execution. How well firms adopt new artificial intelligence technologies will help distinguish leaders from laggards in this competitive industry. [1], [2], [3] As *PWC* illustrated in a report on asset and wealth management, by 2025, assets under management (hereafter AUM) will almost double to reach a new record high of $145 Trillion. 60% of global AUM sees active management and 25% passive management. This trend demonstrates new opportunities for firms that act now and are open to new technologies.

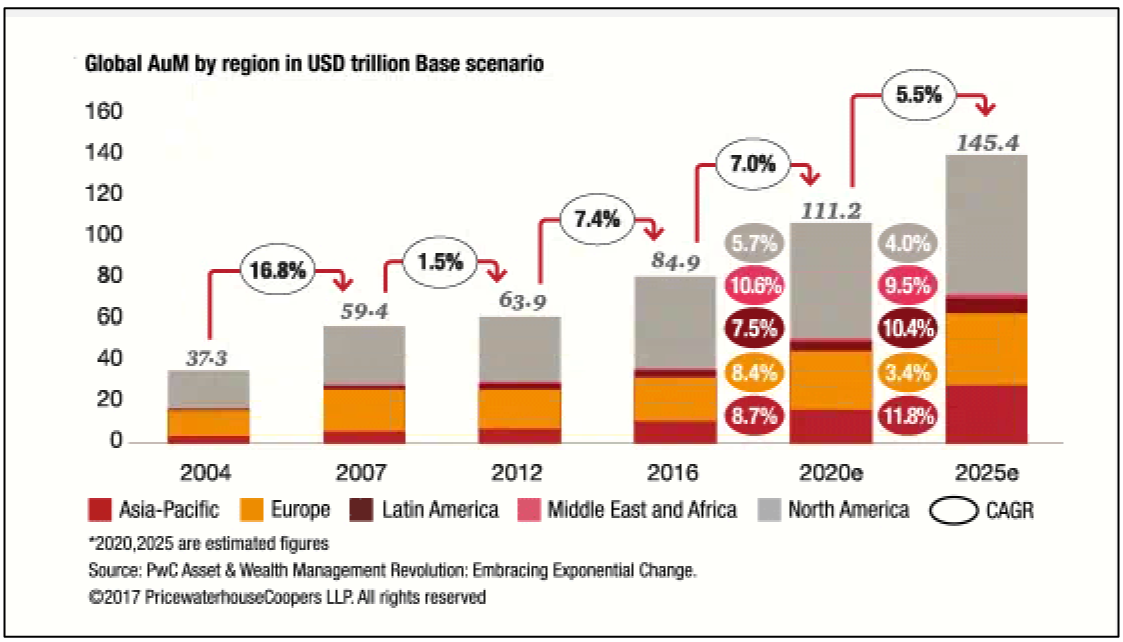


Figure 1: Global Asset Under Management Growth

In dealing with such changes, "investment in technology and data management will need to be maintained or increased … and technology will become mission-critical to driving" prosperity. [4, p. 20] Successful investment firms will proactively and strategically incorporate data and machine learning technologies into their investment processes. They will be able to exploit opportunities by applying new technologies. To that end, reports also indicate the increasing market size growth for the global AI adoption by asset management that anticipates reaching $8.3 billion by 2026, rising at the market growth of 41.1% CAGR. [5]

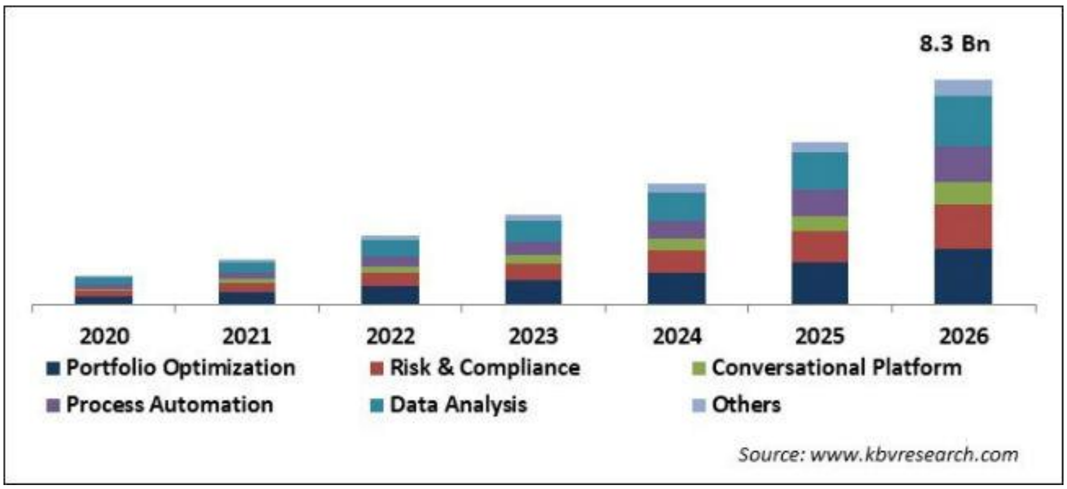


Figure 2: Global AI Adoption by Asset Management

Despite great potential value in data and AI technologies, few investment professionals use AI/big data techniques in their investment research and processes. [2] One of the main reasons is that a big part of the investment community is still unsure how to evaluate the relevance of artificial intelligence and machine learning, and many are still just watching the industry's transformation via big data and high-performance computing. However, due to the fast development of machine learning systems and the emergence of abundant alternative data sources in the investment landscape, people in different roles, including investment analysts, traders, portfolio managers (PM), and chief investment officers (CIO), will eventually need to get familiar with machine learning and data technologies. Machine learning can improve investment activities such as portfolio construction, idea generation, alpha-factor design, asset allocation, bet sizing, and portfolio optimization. In demonstrating this rapid development of such technologies in the investment management industry, J.P. Morgan's global quantitative and derivatives strategy team has reported three trends that have enabled the start of the data and machine learning revolution: "1. Exponential increase in the amount of data available 2. Increase computing power and data storage capacity at reduced cost 3. Advancement in Machine Learning methods to analyze complex datasets." [6] Combining machine learning systems with big data and advanced statistical and computational modeling will likely form the future frontiers of investment management. Some scholars believe that applying data science and machine learning techniques in finance is not just a temporary trend but a discipline per se, and they have called it "financial data science." [7] they hold this perspective for three reasons: "First, finance brings a unique set of problems and puzzles that distinguish it from standard applications of data science, especially those in the natural sciences. Practitioners' challenges in devising trading strategies, asset allocation, and financial risk management, for example, require specific solutions. Second, financial time series pose unique characteristics that reflect their origins in human action and intentionality. The defining properties of financial time series, such as volatility clustering, momentum, and mean reversion, are prime examples. Third, modeling agents, especially the collective agents that

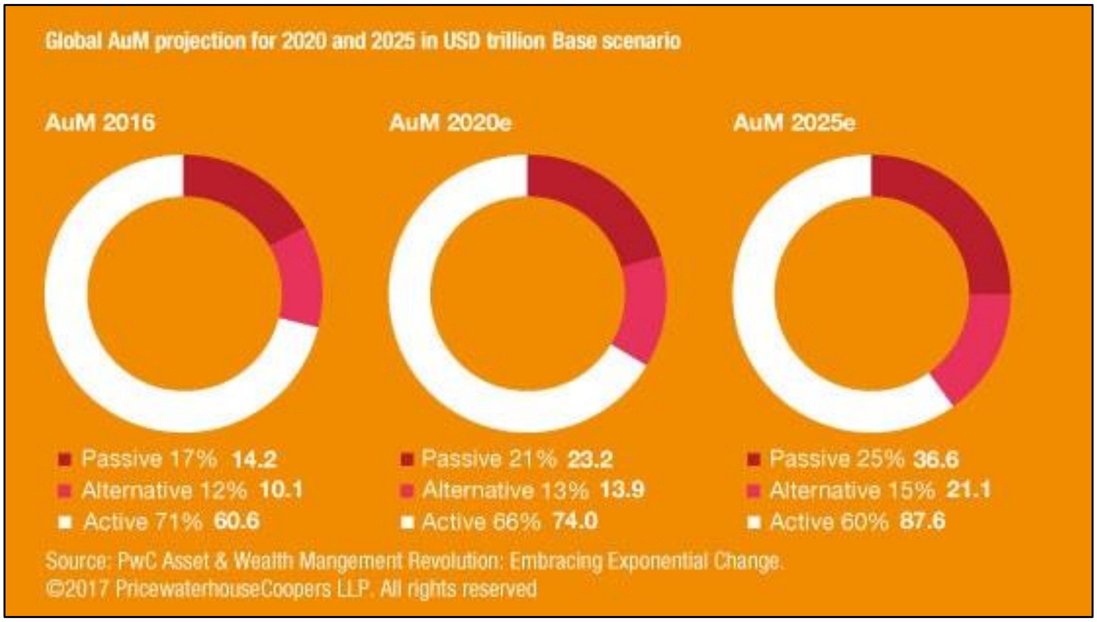


Figure 3: Global AUM Projection for 2020 to 2025

constitute "the market," is an extremely challenging problem that demands specialized techniques." For these reasons, they believe this field is not just the application of data science in finance or minor improvement over econometric models and techniques.

Furthermore, as Arnott et al. [8] provided the Backtesting protocol for machine learning applications in finance, they argue that although machine learning brings a promising set of powerful tools and techniques for investment management research, choosing suitable applications before applying the tools is critical. The authors state that several lessons would help investment researchers have a more realistic approach to using machine learning tools. First, they should be cautious about a false strategy that can work in the cross-validated sample. This ignorance would dangerously result in a single historical path problem. Second, financial data is minimal (compared to other natural sciences), and this small sample is a challenge for most machine-learning applications. It would be considered tiny for advanced approaches such as deep learning. Third, techniques such as unsupervised learning do not necessarily incorporate economic principles and theories in their modeling approach. If such a strategy works, “it works in retrospect, but not necessarily in the future.” So, to successfully assess investment strategies, we need to use financial theories that can help us filter out ideas without an ex-ante economic basis. Such consideration demonstrates the critical role of theory and scientific processes in financial machine learning in investment research.

1.2 Quantitative Investing Primer

The use of machine learning in investment has witnessed its early adopters in the last 30 years in quantitative hedge funds such as AQR, Renaissance Technologies, WorldQuant, D.E. Shaw, Two Sigma, and Bridgewater Associates. However, more systematic and quantitative fund managers have recently started applying machine learning methods in investment research and practice due to the abundance of data from too many sources. Such techniques create more robust and systematic approaches in factor modeling, portfolio analysis and construction, derivative pricing, and optimal hedging and risk management.

Artificial intelligence and machine learning are changing virtually every aspect of the financial services industry. Investment research and practice are benefiting from the rise of machine learning accomplishments that only professional human experts could perform until recently. Price prediction, hedging, portfolio construction and optimization, alpha capture, and sentiment analysis, to name a few, are areas that machine learning has already impacted. [8], [9], [10], [11] as an example, the complex and chaotic nature of financial market price forecasting is a challenging problem in a dynamic environment. Many studies from various research areas have used machine learning to provide some predictions to address this problem. These methods have resulted in promising outcomes. [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23]

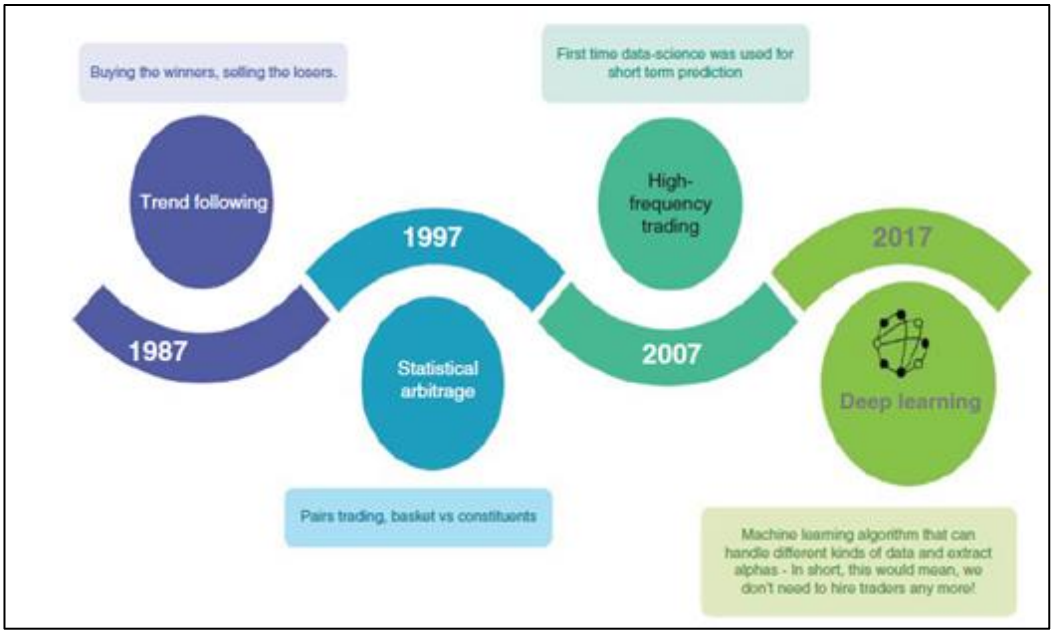


Figure 4: Evolution of Quantitative Investing

These studies show how financial research has started to consider the value of machine learning in price forecasting, given that the data is chaotic and relationships are non-linear.

Technological advances in big data analytics and our increased computational capacity have made it possible to analyze a large volume of structured and unstructured data in a short period. One example is the application of machine learning in portfolio construction and optimization, which requires high computational power. Shen et al., for example, have proposed an orthogonal portfolio framework that represents the combining effects of passive and active investment styles based on a risk-adjusted function. Results demonstrate outperformance in both risk-adjusted return and cumulative wealth. [24] Gu et al. suggested an empirical asset pricing framework for portfolio construction based on the canonical problem of asset risk premia, highlighting the value of machine learning in both empirical studies and financial innovation. [12]

Machine learning demonstrates excellent promise for empirical asset pricing as well. Machine learning has shown the potential to improve empirical testing and understanding of expected asset returns at the holistic level. [12] The capability of crunching a massive amount of big data with a wide variety, velocity, and volume and feeding it into predictive models enables researchers and professionals to dig deeper into empirical analysis beyond traditional econometric models. Rapach et al., for instance, apply Lasso to predict global stock market returns. [25] several papers use artificial neural networks and decision trees to forecast derivative prices and credit card defaults. Other types of machine learning investment applications include studying a cross-section of stock returns, factor pricing models, portfolio sorting, and selection. All recently applied machine learning techniques in such studies represent a promising future. [26], [27], [28], [29], [30]

As Marco Lopez de Prado stated, machine learning provides the opportunity to gain insights from: "(a) new datasets that cannot be modeled with econometric methods; and (b) old datasets that incorporate complex relationships still unexplored. Key strengths of ML methodologies include (i) focus on out-of-sample predictability over variance adjudication; (ii) usage of computational methods to avoid relying on (potentially unrealistic) assumptions; (iii) ability to "learn" complex specifications, including non-linear, hierarchical, and non-continuous interaction effects in high-dimensional space; and (iv) feature importance analysis robust to multicollinearity."

In addition, asset pricing and factor models, which have been a deep area of research in the last four decades, are recognized as potentially another area of research that machine learning and deep learning can shed light on [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41] There is an abundance of empirical research in this area based on econometric analysis. Since machine learning and deep learning excel at absorbing large datasets from various sources and identifying reliable patterns, they are well-suited for the empirical study of asset pricing models. The main challenge in this case is prediction and practical testing, in which machine learning excels in solving prediction problems, and its empirical testing methods have been developing over time.

The changes imposed by new data sets and machine learning techniques will likely take the investment landscape to a higher dimension. As more investors adopt new methodologies and alternative data, today's interconnected and complex capital markets will start reacting faster. In the long term, machine learning techniques and alternative datasets will become a standard approach for quantitative fund managers. This trend will highly likely become normal for systematic asset managers.

However, despite an impressive ML research outcome of recent applications, ML as a standalone research area in finance and asset management practice is still in its early stages. The asset management industry has an emergent understanding of the potential and future of machine learning, which is why research about financial machine learning is so valuable. Notably, the investment landscape's current data science, analytics, and machine learning studies are still sporadic and fragmented. Furthermore, implementing machine learning and applying advanced data science techniques is also the initial step among practitioners. [42], [43] [42], [44], [45], [46]

1.3 Research Motivation

In recent years, we have observed the beginning of a change in the money management industry. Many fund managers face consistent fee pressure and technological innovations, forcing them to adapt or risk losing the competition to more prominent names. The U.S. has experienced the most significant percentage of capital inflow into passive funds. As reported by Morningstar, only three of the top 10 funds worldwide are actively managed funds. [47] Quant funds are no exception in this industry shift and suffered even more in 2019. According to Hedge Fund Research (HFR), an index measuring long-short equity hedge funds that apply "sophisticated quantitative techniques" lost 1.8% in 2019. [48] A sample of machine learning and artificial intelligence funds tracked by Bernstein generated only a 6% absolute return in 2019, compared to 31% for the S&P 500 index. This type of underperformance in active systematic quant funds has redirected investors' money into passively managed "smart beta" exchange-traded funds (ETFs). However, this does not necessarily mean that too many AI-based ETFs are in the investment landscape. Unfortunately, only a few ETFs execute investment decisions using AI [50]. It turns out that both human and machine intelligence find it challenging to beat the market. Nevertheless, AI-based funds need more time to demonstrate their capabilities, as they are mostly short-lived.

A global trend shows that nearly every large asset management company has teams of AI and data scientists focusing on developing machine learning technologies. Many firms have already integrated these technologies into their investment decision-making processes. However, academic research and industry reports have argued that many investment strategies uncovered by practitioners and academics are false discoveries. This can partially explain the high failure rate, especially among quantitative funds. [49] One reason could be that there is too much focus on strategy backtesting rather than the scientific process. In scientific methods, scientists seek to test hypotheses and not run backtests to generate viable rules. We need the same financial research approach when using machine learning tools and technologies. [10] Machine learning and data science technologies could help economic researchers develop more theories rather than find the holy grail. So, the scientific research process, empowered by new techniques and technologies, could be a promising area in financial research.

Following this research and considering the challenges in applying AI technologies in investment research, the study examines financial data science, machine learning projects, and research methods to address current economic and investment problems. One example is the abundance of data. In the age of big data, with millions of tweets published in less than 3 minutes and millions of Google searches completed in less than 20 seconds, how do econometricians respond to this abundance of data? The current machine learning capabilities could be the solution. [50]

Many of today's complex and unstructured datasets are beyond the scope of financial econometrics analysis. Some research studies have shown that financial machine learning is not a pure black box but complementary to econometrics, dispelling the perception that machine learning strategies are opaque, with no transparency regarding how algorithms make decisions. For instance, De Prado has shown a one-to-one correspondence between steps in the econometric and machine learning research processes, making it more transparent how researchers and practitioners can adopt these new technological techniques. [43] The primary steps in the process include goal setting, visualization, outlier detection, feature extraction, regression, classification, feature importance assessment, model selection/prevention of overfitting, and model validation. Hence, these critical steps should be systematically interconnected throughout the research process, demanding more thoughtful consideration.

Brooks et al. [50] stated that “while econometrics and financial data science differ in their intellectual point of departure (i.e., statistical techniques and data sets, respectively), the two fields have many more aspects in common than divide them. Both use econometric concepts and techniques, and both fields develop their hypotheses informed by some form of economic theorizing. Similarly, both will likely use the wealth of newer and bigger data sets from digitalization.” Hence, both fields represent complementary perspectives on the same process. See the figure below.

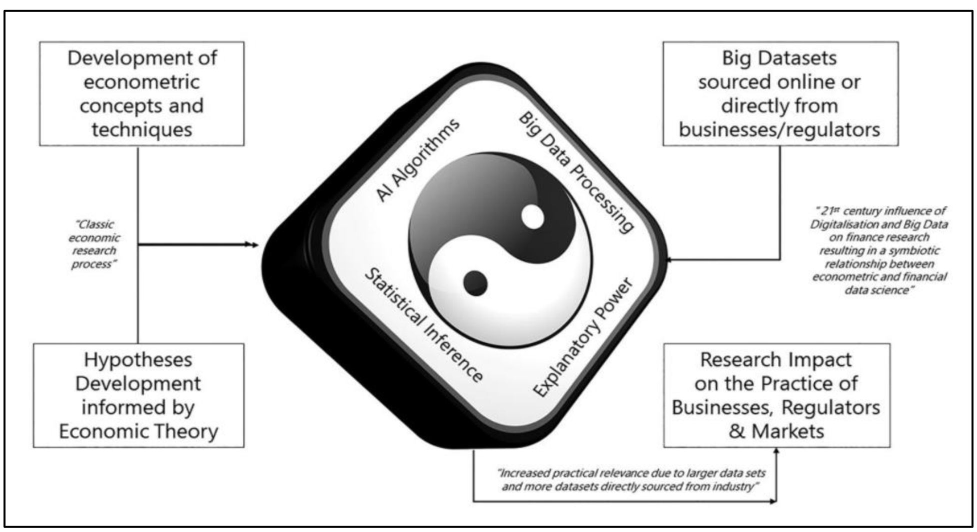


Figure 5: The Complementary Relationship of Econometrics and Financial Data Science

The above figure exhibits the interdisciplinary nature of financial data science and machine learning and the importance of systematic and multidisciplinary research processes.

Furthermore, Khraisha [51] has shed light on different aspects of the financial data science process to provide a holistic approach that considers multiple research process factors. He believes that as financial data science brings more methodological and technological components to the analysis process, there is a need for having a holistic view in the successful management of financial data science projects. Andreas et al. [52] have also emphasized other aspects of financial data science and machine learning, which are essential to more transparency and making conclusions from machine learning methods. They have correspondingly proposed significance in economic forecasting, statistical relevance in risk modeling, and explainability of novel data sets. These aspects of analysis also demonstrate the importance of multiple perspectives in analyzing the decision-making process in financial data science research.

Likewise, some researchers explain the necessity of financial data science and related research processes as a unique and emergent field of research and the need for more systematic methods in developing and evaluating investment strategies. [7], [44], [53], [54], [55], [56], [57], [58], [59], for example, Li et al. have suggested three main principles in the practical usage of financial machine learning in equities. They are instability, interpretability, and interesting model predictions, meaning the model should convincingly outperform simpler models.

To integrate multiple perspectives into the research process of financial data science projects, the theoretical underpinning of this research is depicted below:

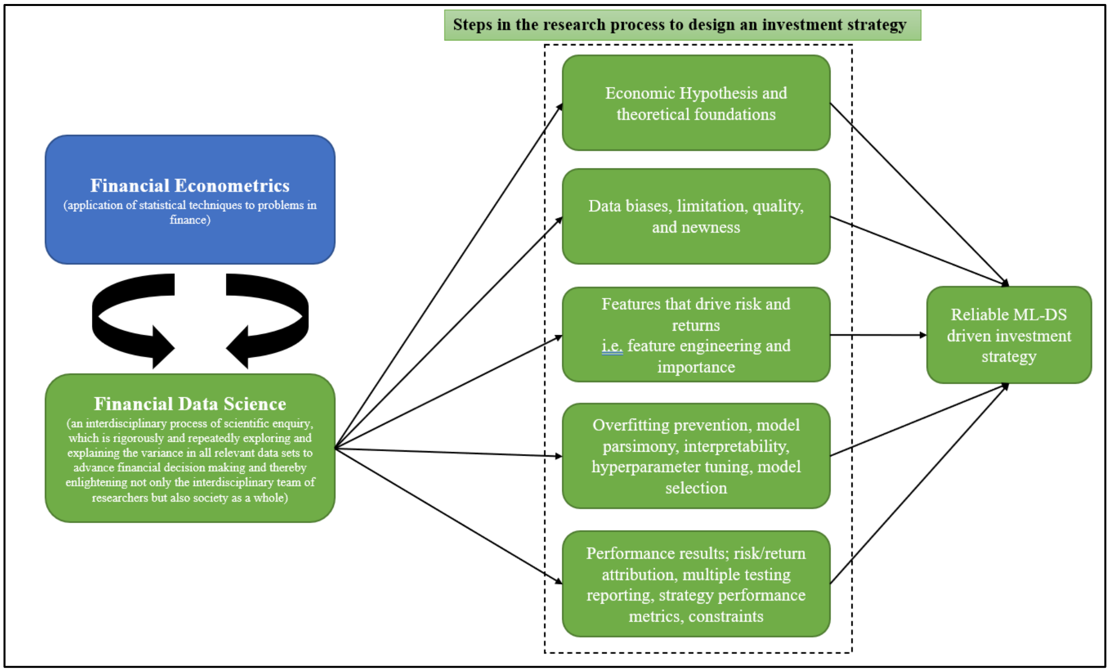


Figure 6: The Theoretical Model

1.4 Problem Statement

There is a growing stream of research on AI and machine learning techniques and technologies in investment research and asset management. [60] However, some studies still indicate the high failure rate in AI-based quantitative funds and the abundance of reporting false discoveries. In fact, “researchers want to minimize false positives, but to do it in a way that does not miss too many good strategies.” With growing computational power and increasing complexity of models – especially in more sophisticated machine learning models – -based techniques, there is a need for standard protocols to improve the outcome of backtesting results. [54] additionally, in today’s complex and interconnected world, many forms of financial data are beyond the grasp of econometric models. Applying data science and machine learning techniques can offer a better understanding of data features such as unstructured and alternative data, non-linear relationships among variables, and the high dimensionality of data. [9], [61]

Furthermore, “econometrics lacks the tools and methods to analyze alternative datasets such as social media streams, geological data, patents, news, and microdata on consumer behavior.” [51] Although this does not mean financial data science will replace traditional econometrics techniques, on the contrary, it will play a complementary role in economic research to empower economic research outcomes. In turn, the adoption of machine learning models in addressing financial problems is in its early stages, which needs robust frameworks.

Additionally, some of the characteristics are unique to the financial domain. In contrast to machine learning applications in other fields, finance could not simply apply these techniques without being cautious about some issues. For example, financial market data are noisy, with a low signal-to-noise ratio. So, the naïve machine learning applications would be hazardous to the drive of financial decisions. Moreover, model interpretability is a significant challenge in AI-based investment strategies. Although more advanced and complex techniques like tree-based or deep learning models might generate more accurate outcomes in their predictions, their interpretability is not straightforward.

Consequently, to cope with such challenges in today's financial industry, "machine learning offers a modern set of tools specifically suited to overcome the challenges of new economic and financial data sources and increasingly complex associations in financial markets. [43] financial data science as an interdisciplinary field and machine learning as an approach to solving economic problems have been starting to address the economic issues and prove the technology capabilities and solutions.

Therefore, there is a need for a model or framework that can help investment research teams and asset management firms to be more confident about the quality of results of strategies that arise from using machine learning and data science in the investment research process. To achieve this goal, such a model should have some features:

- Identifies potential vital factors that significantly impact the success of investment strategies based on machine learning and data science.

- Evaluate the reliability of strategy results from machine learning and data science investment research.
- Follows scientific processes and depicts multiple perspectives that impact the results.

Developers have implemented machine learning and data science methods across various sectors and industries, resulting in diverse applications; building a general model that can serve and address all kinds of applications requires years of testing and validation. So, it is more feasible to make a model as a starting step in this research direction that can focus on a particular area of finance and have the generalization capacity down the road. Hence, this research will develop a model for designing investment strategies employing data science and machine learning. This area is data-intensive, and challenges are evident from the literature and practice. As investment companies adopt and run initiatives regarding financial machine learning, they could face the same challenges and consequences when using such technologies.

According to the above considerations, this research will aim to identify the main hassles that result in futile investment strategies and, consequently, failures; in this context, machine learning and data science investment strategies refer to investment research projects that seek to benefit from such technologies to make financial decisions. More specifically, this research intends to concentrate on the following:

- Identifying the critical success factors in ML/DS-based investment research projects, mainly quantitative ML/DS strategies, which determine success or failure in practice, based on literature review and expert judgment.
- Develop a multi-criteria model that plays a decision-support role in the investment research decision-making process to evaluate such issues and increase the reliability of results and the chance of success in practice.
- To ensure the reliability of the model results, subject-matter experts will validate and quantify it.
- Finally, applying the model to a sample ML/DS investment strategy tests its efficacy.

Therefore, the author believes the rigorous scientific method is the best approach to address investment management challenges and problems.

CHAPTER 2. LITERATURE REVIEW

2.1 AI and Machine Learning

Artificial intelligence (AI), particularly machine learning, is considered one of the disruptive technologies in the fourth wave of the industrial revolution. Recently, we have witnessed its growing impact across most industries. Machine learning has gained significant popularity in research and practice, with many applications, including image classification, text analytics, voice generation and recognition, and natural language processing. Machine learning (ML) is a subset of the artificial intelligence field that aims to build and test systems capable of learning from data without explicit programming. The explosion of data and remarkable advancements in computational technologies have led to new research and practices in this field. There is increasing interest in machine learning in general, and deep learning, as an emerging and robust method, has significantly generated new use cases for such technologies. [17], [28], [62], [63], [64], [65]

One of the leading industries significantly impacted by this trend is financial services, particularly asset management. Many established investment firms have already integrated AI and machine learning into their investment research processes and decision-making. This integration encompasses portfolio management, risk management, trading, asset pricing, and transaction cost analysis. As increasing investment companies embrace a more data-driven approach, machine learning methods present numerous exciting avenues for solving prediction problems and unveiling the underlying data generation processes often hidden in plain sight. Despite the highly successful results achieved by machine learning and deep learning methods across various use cases, their evolution and expansion within asset management are still in their early stages.

There are several classes of machine learning models, and the type of problem determines which learning models should be applied. [6], [66], [67], [68], [69] The following are broad categories in this domain: supervised learning, unsupervised learning, reinforcement learning, and deep learning.

2.1.1 Supervised

The term 'supervised' arises from the modeler guiding or supervising the algorithm by providing labeled data or a training set, along with the output or predicted variables. The objective, therefore, is to establish the association between independent attributes and the designated outcomes. This association manifests as a mathematical or algorithmic structure and pattern that captures the relationship between the predictors and the predictable. Generally, all supervised problems fall into two main categories: regression and classification. Both approaches are employed to predict values. In classification models, the modeler aims to anticipate a discrete or categorical output or response. In contrast, regression seeks to predict a continuous variable. [15], [46], [63], [70], [71] [72], [73], [74] They typically exhibit the following characteristics: utilizing a training dataset to train a model, followed by applying the trained model for validation to test predictions and verify the results.

Consider an example of predicting the mathematical relationship between market returns and the fundamental factors of companies within a specific sector. In traditional econometrics, an analyst would employ multiple linear regressions to calculate the beta of market returns concerning each of these predictable variables. However, by embracing machine learning tools and techniques falling under the category of supervised learning, they can utilize more advanced models capable of capturing nonlinear relationships, accounting for outliers, identifying the most crucial variables, and so forth. Such advancements beyond classical financial econometrics are attainable within financial data science and machine learning. [60], [75], [76], [77]

2.1.2 Unsupervised

In many real-world scenarios, datasets lack labels, and our primary objective is to gain insights into the data. In such instances, unsupervised learning techniques come into play. Unsupervised models are employed to uncover patterns within the data without the guidance provided by input labels. One of the most frequently utilized approaches in this context is clustering. For example, consider a scenario where one wishes to categorize companies within a specific sector into distinct groups based on shared characteristics. Clustering can be instrumental in grouping these companies based on their similarities, thus enabling the development of diverse investment strategies for each specific group. Unsupervised learning algorithms autonomously identify patterns within the feature space without requiring external supervision by identifying similar data points within the dataset. [78], [79], [80], [81]

2.1.3 Reinforcement

Reinforcement learning models constitute one of the current research areas that have piqued the interest of many scholars. The premise of this category involves an agent who endeavors to achieve a goal within an uncertain and complex decision environment through a reinforcing feedback loop of learning. Essentially, the agent iteratively acquires the ability to respond to its environment to maximize rewards, guiding it toward the desired outcome. Simply put, the primary objective is to maximize total compensation to attain the goal. Within a game-like environment, the agent receives either rewards or penalties based on the game's rules. After numerous iterations of trial and error, it develops creative strategies to reach the goal. Various domains have applied these learning models. For instance, they can be employed in finance to build investment strategies that learn from market dynamics to achieve the best risk-adjusted returns while effectively managing transaction costs. [82], [83], [84], [85]

2.1.4 Deep Learning

Although machine learning has received considerable attention recently, deep learning has emerged as a leader in this field. As a subset of machine learning, deep learning emulates biological neurons and abstract representations of their activities. The exponential growth of data and advancements in computational techniques have facilitated computational modeling capabilities previously inaccessible through deep learning. The foundations of deep learning models originate from traditional neural networks, and their superiority in predictive accuracy surpasses that of other machine learning algorithms. The processes and technologies incorporated into deep learning models exhibit vast diversity, encompassing image recognition, text analytics, numerical predictions, and video recommendations. [69], [82], [85], [86] Artificial neural networks with multiple layers are called deep learners. Deep neural networks are structured so that each layer performs sophisticated computation to make sense of the data. Models are typically shallow or deep based on the number of layers used in the architecture. Besides, deep learning systems require much data and are highly computationally intensive to generate credible results.

Furthermore, deep learning has addressed current issues in traditional machine learning problems. Feature engineering, for instance, has been one of the highly researched areas in the domain of machine learning and artificial intelligence. By applying deep learning, the model can automatically extract features and sometimes create more complex nonlinear features internally without human intervention. [87], [88] Many deep learning architectures and models can be selected and applied based on the use case and the problem. (See Figure 7)

The table below illustrates machine learning techniques' most notable features and capabilities. This table is not an exhaustive list of elements but covers the most important ones.

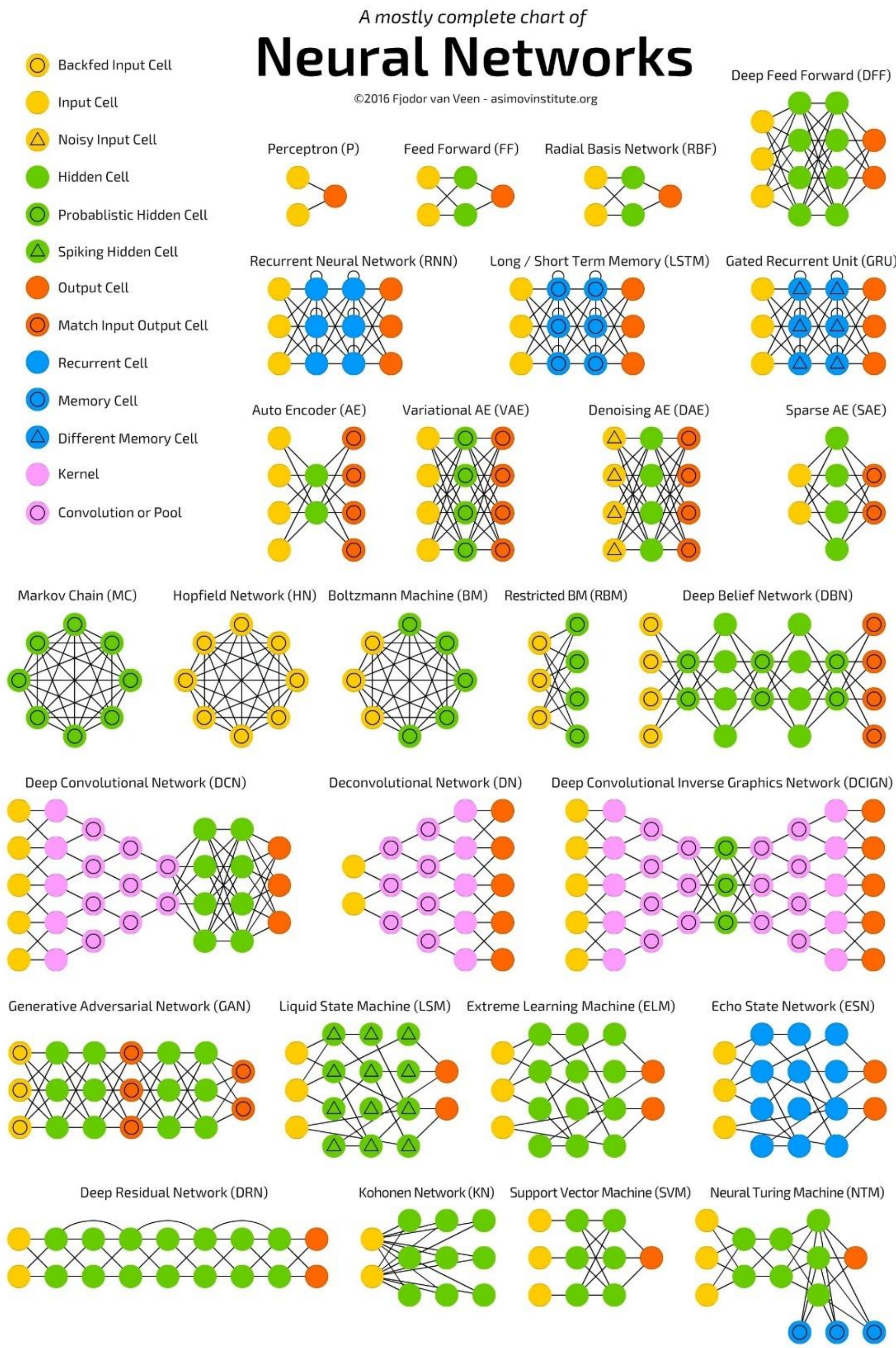


Figure 7: Neural Network Architectures

(adapted from [84])

Table 1: Machine Learning Characteristics

ML Characteristics	
Visualization	This feature is essential, especially for visualizing high-dimensional data in different domains. Visualization techniques can summarize and display data in a form to help researchers and practitioners absorb and make sense of a large amount of data [89], [90].
Feature Extraction	In the era of big data, datasets with hundreds of features are ordinary. This capability reduces the number of features in a dataset by creating new feature representations. This technique helps to summarize most of the information with a lower number of features [91], [92], [93].
Regression	Regression is the most well-known and widely used statistical learning approach in econometrics. It is popular because it is a simple mapping between the inputs and outputs. This technique follows some assumptions. For instance, one hypothesis in this model is that variables typically follow a specific distribution. Additionally, it assumes a linear relationship between input and output [94], [95], [96], [97].
Classification	Like regression, classification aims to map the relationship between input and output variables. However, it predicts a class of data points. The main goal of this approach is to predict the class of outputs based on the input data. Some applications include credit rating and mortgage classification [98], [99], [100], [101].
Feature Importance	These are the techniques that try to rank the input features based on how predictive and valuable they are at estimating the target variable [10], [102], [103], [104], [105].
Sentiment Analysis	Researchers often use it to classify emotions in subjective and textual data, leveraging techniques like Natural Language Processing (NLP). Its applications include analyzing customer messages, gauging sentiment in financial markets, and assessing news sentiment, among others [81], [106], [107], [108], [109].
Clustering	In clustering, models only cluster input data and identify groups within the dimension space, unlike supervised learning, where target variables are utilized [110], [111], [112].
Interpretability	“Interpretability is the degree to which a human can understand the cause of a decision. [Tim Miller] It is the degree to which a human can consistently predict the model’s result” [Been Kim]. “The higher the interpretability of the model, the easier for the decision-makers to understand why certain predictions have been made” [113], [114], [115], [116], [117], [118].

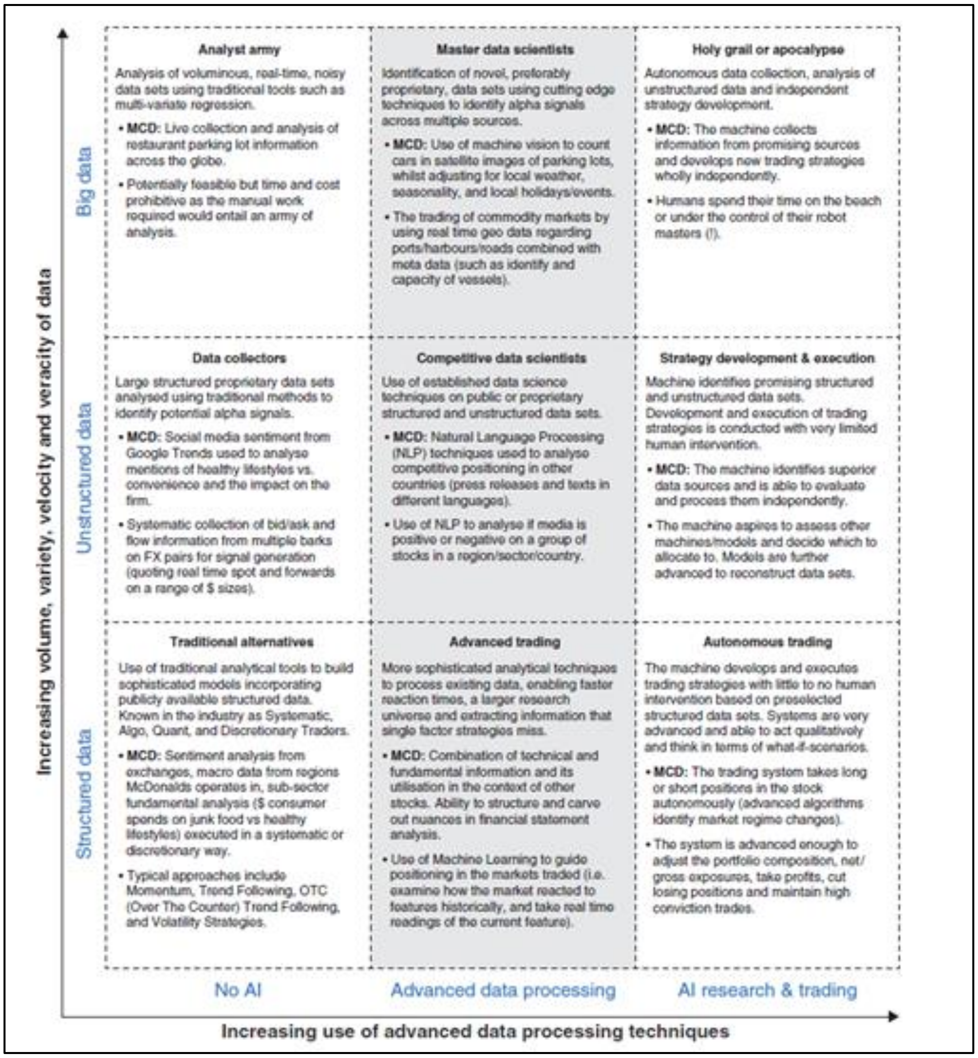
2.2 AI in Finance and Asset Management

As mentioned in the preceding section, AI is among the technologies that have garnered significant interest from scholars and technologists in recent years. As one of the largest industries, asset management is no exception to this trend. We have witnessed a broad spectrum of use cases and applications, ranging from the automation of existing investment processes to specific machine learning applications in alpha generation. Deep learning and its characteristics are currently a prominent topic in many applications. AI methods can contribute to investment research from various perspectives. For instance, discretionary managers can integrate DS/ML techniques into traditional fundamental research to enhance results that cannot be achieved solely through conventional fundamental analysis. Additionally, many problems in finance essentially boil down to estimation, an area in which machine learning algorithms demonstrate superior performance. Examples include return prediction, risk estimation, and portfolio optimization. [2], [60], [119]

The emergence of computerized capital markets has also led to more reliance on machines in trading. High-frequency trading and automatic market-making are products of such automation. Thus, algorithmic trading is another investment realm in complex markets that has already attempted and used AI techniques. Processing large amounts of data in nano-seconds is not something humans can realize just by looking at the data. This area is precisely where machines' capabilities blossom, and the algorithm's performance is beyond human intelligence. As depicted in the following figure, Sirotyuk and Bennett (2017) have classified AI in finance. Therefore, as we race towards the era of big data, the level of complexity and automation increases. [120]

AI has the potential to address many financial problems. In a systematic study by JP Morgan Asset Management in 2017 on the implication of machine learning for the investment community, they have classified the different types of tasks in investment that data science and machine learning methods can solve. As shown below, researchers can use a corresponding list of techniques for every question. Although this list is not exhaustive, it demonstrates the capabilities of AI in solving investment problems. [6]

Moreover, Robo-advisor applications have gained significant popularity in recent years. They are computer programs that provide investment advisory services at a scale that was impossible in the past without powerful computers and big data. We also observe their development across other areas of asset management, such as wealth management and retail trading. The rising value of startups joining the unicorn club is evidence of more democratization of investment and asset management services.



Increasing volume, variety, velocity and veracity of data	No AI	Advanced data processing	AI research & trading
Big data	**Analyst army** Analysis of voluminous, real-time, noisy data sets using traditional tools such as multi-variate regression. • **MCD:** Live collection and analysis of restaurant parking lot information across the globe. • Potentially feasible but time and cost prohibitive as the manual work required would entail an army of analysis.	**Master data scientists** Identification of novel, preferably proprietary, data sets using cutting edge techniques to identify alpha signals across multiple sources. • **MCD:** Use of machine vision to count cars in satellite images of parking lots, whilst adjusting for local weather, seasonality, and local holidays/events. • The trading of commodity markets by using real time geo data regarding ports/harbours/roads combined with meta data (such as identify and capacity of vessels).	**Holy grail or apocalypse** Autonomous data collection, analysis of unstructured data and independent strategy development. • **MCD:** The machine collects information from promising sources and develops new trading strategies wholly independently. • Humans spend their time on the beach or under the control of their robot masters (!).
Unstructured data	**Data collectors** Large structured proprietary data sets analysed using traditional methods to identify potential alpha signals. • **MCD:** Social media sentiment from Google Trends used to analyse mentions of healthy lifestyles vs. convenience and the impact on the firm. • Systematic collection of bid/ask and flow information from multiple barks on FX pairs for signal generation (quoting real time spot and forwards on a range of $ sizes).	**Competitive data scientists** Use of established data science techniques on public or proprietary structured and unstructured data sets. • **MCD:** Natural Language Processing (NLP) techniques used to analyse competitive positioning in other countries (press releases and texts in different languages). • Use of NLP to analyse if media is positive or negative on a group of stocks in a region/sector/country.	**Strategy development & execution** Machine identifies promising structured and unstructured data sets. Development and execution of trading strategies is conducted with very limited human intervention. • **MCD:** The machine identifies superior data sources and is able to evaluate and process them independently. • The machine aspires to assess other machines/models and decide which to allocate to. Models are further advanced to reconstruct data sets.
Structured data	**Traditional alternatives** Use of traditional analytical tools to build sophisticated models incorporating publicly available structured data. Known in the industry as Systematic, Algo, Quant, and Discretionary Traders. • **MCD:** Sentiment analysis from exchanges, macro data from regions McDonalds operates in, sub-sector fundamental analysis ($ consumer spends on junk food vs healthy lifestyles) executed in a systematic or discretionary way. • Typical approaches include Momentum, Trend Following, OTC (Over The Counter) Trend Following, and Volatility Strategies.	**Advanced trading** More sophisticated analytical techniques to process existing data, enabling faster reaction times, a larger research universe and extracting information that single factor strategies miss. • **MCD:** Combination of technical and fundamental information and its utilisation in the context of other stocks. Ability to structure and carve out nuances in financial statement analysis. • Use of Machine Learning to guide positioning in the markets traded (i.e. examine how the market reacted to features historically, and take real time readings of the current feature).	**Autonomous trading** The machine develops and executes trading strategies with little to no human intervention based on preselected structured data sets. Systems are very advanced and able to act qualitatively and think in terms of what-if-scenarios. • **MCD:** The trading system takes long or short positions in the stock autonomously (advanced algorithms identify market regime changes). • The system is advanced enough to adjust the portfolio composition, net/gross exposures, take profits, cut losing positions and maintain high conviction trades.

Increasing use of advanced data processing techniques

Table 2: AI in Finance Classification

2.2.1 Current Trends

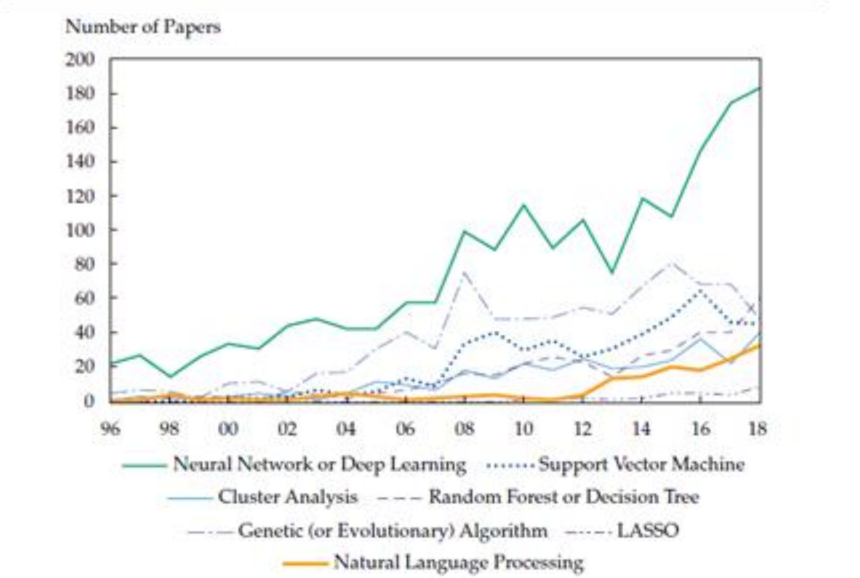
The surge of publications on applications of AI and ML in finance shows its popularity. In a report presently released by a team of researchers affiliated with the CFA Institute, there is clear evidence of growing interest in applying these techniques in asset management. It shows that neural networks or so-called deep learning methods are the most used technique in this stream of research. The authors attributed these trends to three leading developments: the increasing computational and storage capacity in recent years has significantly improved the utilization of AI and ML methods. Second, prominent data

Question	Data Analysis Technique
Given set of inputs, predict asset price direction	Support Vector Classifier, Logistic Regression, Lasso Regression, etc.
How will a sharp move in one asset affect other assets?	Impulse Response Function, Granger Causality
Is an asset diverging from other related assets?	One-vs-rest classification
Which assets move together?	Affinity Propagation, Manifold Embedding
What factors are driving asset price?	Principal Component Analysis, Independent Component Analysis
Is the asset move excessive, and will it revert?	
What is the current market regime?	Soft-max classification, Hidden Markov Model
What is the probability of an event?	Decision Tree, Random Forest
What are the most common signs of market stress?	K-means clustering
Find signals in noisy data	Low-pass filters, SVM
Predict volatility based on a large number of input variables	Restricted Boltzmann Machine, SVM
What is the sentiment of an article / text?	Bag of words
What is the topic of an article/text?	Term/InverseDocument Frequency
Counting objects in an image (satellite, drone, etc)	Convolutional Neural Nets
What should be optimal execution speed?	Reinforcement Learning using Partially Observed Markov Decision Process

Source: J.P.Morgan Macro QDS

Table 3: Investment Questions and Associated Data Analysis Techniques

attributes, including variety, velocity, and data integrity, have substantially led to this trend. Finally, the improved ML algorithms and accessibility have given researchers and practitioners the to apply them in many use cases. [60] That report also stated that machine learning (ML), as a subfield of AI, has received the most applications among researchers. Consequently, one can consolidate these applied machine learning techniques into a concise list of methods extensively utilized in asset management applications. Their study is founded on analyzing AI techniques in financial research, encompassing all working papers posted on SSRN. One noteworthy observation here is that this finding aligns with the research of bibliometric literature review results, which indicate a consistent upward trend in the utilization of ML in general and deep learning in particular. [6], [60]



Notes: The figure presents the number of published papers with specific keywords by year. It is based on the number of published papers listed on Scopus starting in 1996 and ending in 2018. The papers have "finance" and/or "asset management" keywords together with at least one of the following keywords: "cluster analysis," "genetic algorithm" or "evolutionary algorithm," "lasso," "natural language processing," "neural network" or "deep learning," "random forest" or "decision tree," and "support vector machine."

Figure 8: Number of Papers Published Using AI Technology Over Time, 1996 – 2018

Many applications exist to address various investment challenges concerning the use cases of machine learning in asset management. Examples include portfolio construction, risk management, algorithmic trading, sentiment analysis, empirical asset pricing and factor investing, bankruptcy prediction, and FX rate forecasting, to name a few. The techniques typically employed encompass artificial neural networks (both shallow and deep), decision trees and random forests, support vector machines (SVM), LASSO, cluster analysis, evolutionary models, and natural language processing (NLP). The subsequent section delves into these techniques and shows their applications. [121], [122], [123], [124], [125]

2.2.2 Machine Learning Techniques in Asset Management

Although AI applications have steadily grown in asset management, we are far from replacing all investment process steps with automation and machines. Most AI and ML use cases are categorized into commonly used models. (See the following figure) As illustrated, one of the widely used techniques is artificial neural networks. This class of models is very good at capturing nonlinearities and complex patterns that other models cannot. However, model interpretability could be a challenge. There is a broad spectrum of researched architectures in deep learning models, which is worth digging deeper into and having a short review of, as the case study for this research will also use the deep learning model.

2.2.3 Financial Data Science and Machine Learning

Financial data science and machine learning represent an emerging interdisciplinary field of study whose popularity has surged recently. The scientific analysis of financial data has been the domain of financial econometrics over the past few decades. Traditionally, financial econometrics, which relies on statistical methods to address economic problems, has been the cornerstone of financial modeling.

The substantial uncertainty inherent in financial data has necessitated a firm reliance on statistics and tools such as multivariable linear regression, parameter estimation, hypothesis testing, and multiple testing. However, financial econometrics has its limitations, as researchers have highlighted. For instance, econometric models heavily depend on low-dimensional analysis, which is inadequate for modeling high-dimensional datasets. Econometrics primarily focuses on modeling traditional datasets and is less equipped to handle alternative data sources such as user-generated data on the web, social media data, patents, and news. Additionally, multiple testing in strategy backtesting and cherry-picking have led to false discoveries in investment research. [61], [126], [127] [128] Recently, advances in computational and analytics technologies have created new opportunities and provided novel perspectives to solve financial problems. Different terms refer to this unknown trajectory, such as quantitative investment, financial data science, financial machine learning, and so forth. For example, Krishna [51] defines it as follows:

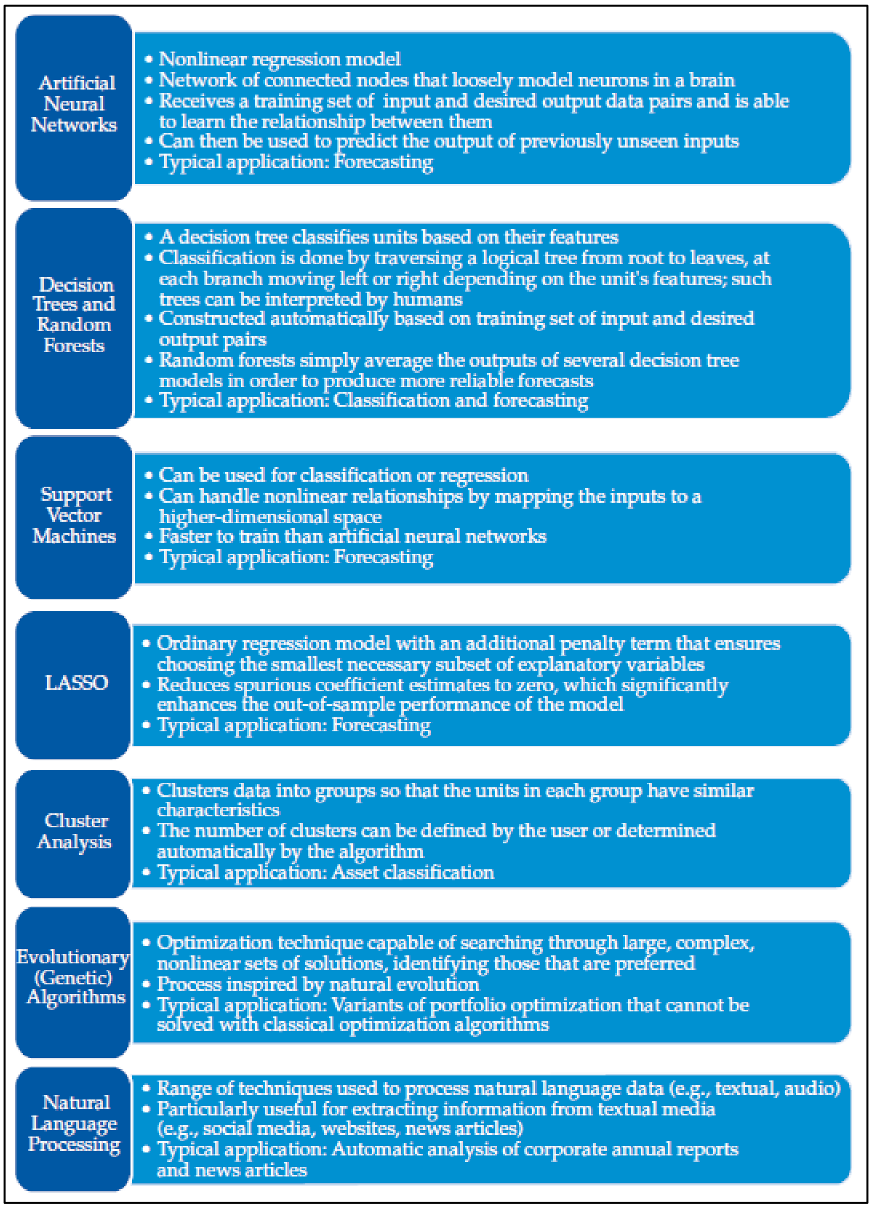


Figure 9: Primary ML Techniques Commonly Used in Asset Management

"Financial data science is a distinct, interdisciplinary area of research and practice that combines tools and methods from financial economics, statistics, computer science, machine learning, and data mining to scientifically analyze and understand a wide variety of datasets to solve existing and new problems in finance."

Other scholars have also suggested different definitions. Financial data science is "an interdisciplinary process of scientific inquiry, which is rigorously and repeatedly exploring and explaining the variance in all relevant data sets to advance financial decision making and thereby enlightening not only the interdisciplinary of researchers but also society as a whole."

There are vital definitions that can help researchers better understand terms used in this area of research. The following covers some of the important ones that increase the clarity of concepts:

- Quantitative investment strategies: A quantitative investment strategy is a "systematic, data – and model-based approach to making investment decisions. The most important characteristic of the quantitative modeling approach is the scientific approach. This approach provides a paradigm that guides and informs empirical work. This approach in quantitative modeling attempts to describe, inquire, and interpret with precision" [44]
- Strategy backtesting: "A *backtest* is a historical simulation of an algorithmic investment strategy. It calculates the profits and losses such an algorithm would have generated if it had been run over that period." [129]

- Financial machine learning is a subset of data science that seeks to draw insights and make predictions using statistical and computational models. It applies machine learning tools and techniques in solving financial problems and making more informed financial decisions.

Therefore, the advances in computational technologies and the abundance of data in recent years resulted in the emergence of financial data science and machine learning. The following examples highlight some of the cases that are out of the grasp for traditional econometrics, and financial DS/ML could have solutions for:

- In the age of big data, with millions of published tweets in less than 3 minutes and millions of Google searches completed in less than 20 seconds, how do econometricians respond to this abundance of data? [130]
- With growing computational power and increasing complexity of models, especially in machine learning-based techniques, researchers aim to minimize false positives while ensuring that they do not overlook too many promising strategies. Therefore, there is a pressing need for standard protocols to enhance the accuracy of backtesting results. [54]
- Many forms of financial data are beyond the grasp of econometric models. Unstructured data, non-linear relationships between variables, and the high dimensionality of data, to name a few, are data features that can be understood better by applying data science and machine learning techniques. [43]

- “Machine learning offers a modern set of tools specifically suited to overcome the challenges of new economic and financial data sources and increasingly complex associations in financial markets. [54]

Moreover, the following two figures represent how financial DS/ML and econometrics can work together and empower the research results. The left figure is about the different terminology used in both fields. Although we have two class terms, both refer to the same reality. Also, the correct figure is how traditional econometrics / statistical and data-driven research steps have too many overlapping phases. Generally, “most quantitative models are based on two approaches of thinking – a hypothesis-based (deductive) and pattern-based (inductive). Each approach requires a different model-building research process. For the hypothesis-based approach, the starting point is some insight into why a trading opportunity exists. It depends on an economic thesis or hypothesis on how the market works or why the opportunity exists. Frequently, the “story” precedes the empirical work. The second approach is inductive or pattern-based. This approach is exploratory, and the discovery of insights emerges from the practical work. A key feature is that learning occurs throughout the process. In this approach, it is critical to be able to distinguish between correlation and causation. Are measured statistical correlations spurious or causal? Understanding underlying economic mechanisms and theory may provide insights into this question. [6]

Term in Statistics	Equivalent Term in Machine Learning
Statistical Learning	Classical Machine Learning
Independent Variable, X	Input Feature, attribute
Dependent Variable, Y	Output Feature, response
In-Sample	Training Set
Out-of-Sample	Test Set
Estimate, Fit a Model	Learn a Model
Model Parameters	Model Weights
Regression	Supervised Learning
Clustering and Dimensionality Reduction	Unsupervised Learning
Classifier	Hypothesis
Response (e.g. 0 or 1)	Label
Data Point	Example, Instance, Training Sample

Source: J.P Morgan Macro QDS

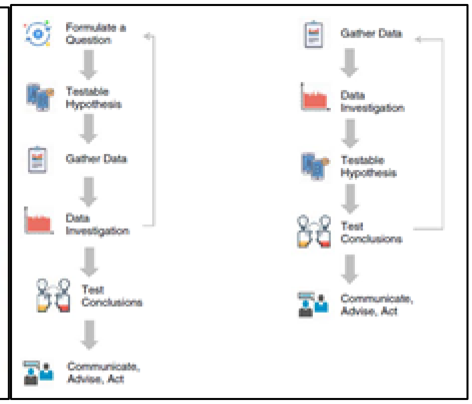


Figure 10: Left: Statistics vs. Machine Learning Terminology, Proper: Hypothesis-Based vs. Data-Driven Analysis

A report released by AQR asset management is a good indication of the current state of financial machine learning:

“Financial machine learning has the potential to be the next leap forward in quantitative investing. Understanding the current state of machine learning in asset management requires grasping two key points. First, research is advancing, leaving many important questions unanswered. Second, early research evidence suggests potential economically and statistically significant improvements in portfolio performance by leveraging machine learning tools. However, these gains represent an evolutionary progression rather than a revolutionary leap.

The ideas behind machine learning – leveraging new data sets to identify robust additive portfolio performance and using methods to extract information systematically – are the modus operandi of quantitative investment processes. For decades, asset managers have used human-intensive, decentralized statistical learning; machine learning offers a systematic approach to investing that mechanizes information from more new sources faster, including unstructured data previously untapped, and provides tools to search through increasingly flexible economic models that seek to capture complex realities of financial markets better. The evolution of machine learning in finance is just beginning." [131]

Contrary to some critics' suggestions that machine learning is only beneficial for short-term predictions and relies primarily on black-box models, scholars have demonstrated the power of ML models in studying systematic equity investment and uncovering hidden market structures. [132] Machine learning can construct benchmarks to test financial theories. It helps develop approaches and explains systematic variations not captured by traditional ideas. Unlike conventional economic research, we deduce rules using machine learning and data science; we let the data tell us which rules are in place and might evolve. [133]

Machine learning also brings its benefits and challenges to solving financial problems. These are summarized in Table 2 and Table 3 below:

Table 4: Benefits of Machine Learning and Data Science in Investment Research

Benefits		Reference
Capturing Nonlinearities in Asset Pricing	Compared to traditional econometric models, ML models can improve the description of price behaviors. Their capabilities in capturing nonlinearities help provide insights and open opportunities to investigate asset pricing models in factor investing research.	[12], [95], [134], [135], [136]

Generation of Synthetic Datasets	Synthetic dataset applications are growing, and their importance in finance with short data histories is critical. ML algorithms can generate synthetic data with the same statistical characteristics as datasets to help test the strategies on new observations and reduce the probability of overfitting.	[10], [137], [138], [139]
Portfolio Construction	ML methods that rely on a few assumptions are applied to capture hierarchical associations between variables, which is impossible for traditional methods. The clustering capabilities of ML can improve the classical mean-variance framework and its subsequent approaches to capturing covariance relationships.	[11], [75], [76], [79], [116]
Outlier (anomaly) Detection	Traditional regression models are susceptible to outliers, which results in biased estimates. Many ML methods are precious to process large amounts of data to detect and identify outliers.	[140], [141], [142], [143]
Bet Sizing	Determining the size of the bets has always been critical for executing investment strategies. "a meta-labeling classification algorithm can learn bet sizing."	[9], [61]
Sentiment Analysis	The amount of textual data is vast, and most of it is unlabeled. In dealing with such datasets, most traditional econometrics methods are silent. ML models can analyze these datasets differently, such as finding the sentiment and categorizing topics.	[81], [106], [107]
Feature Importance	Feature importance in machine learning assigns a score to input features based on how useful their contribution predicts a target output. This capability helps to tie the results to the essential input factors.	[102], [103], [104], [104], [105]
Credit Ratings and Analyst Recommendations	Many machine learning have shown their power in solving the credit rating problem. Regression, classification, and clustering techniques, to name a few, have demonstrated positive results in providing reliable credit ratings and analyzing analyst recommendations.	[144], [145], [146]
Controlling for Effects and Interactions	One of the central powers of machine learning and data science techniques is their capability to detect and address nonlinear interactions. Also, researchers can control and test several variable effects in a completely controlled experimental environment.	[23], [147], [148]

Table 5: Challenges and Considerations of Machine Learning and Data Science in Investment Research

Challenges and Considerations		Reference
Interpretability of Results	The interpretability of ML models is a hot topic nowadays. The lack of equation specification does not necessarily mean ML algorithms are a black box. ML models can provide the building blocks of theory generation, and researchers can utilize them: the model interpretability research and its aspects in a growing field in AI.	[113], [115], [118], [149]
Risk of Overfitting or Underfitting	Overfitting is a common issue encountered in strategy backtesting. When the model's performance varies between in-sample and out-of-sample data, this inconsistency indicates overfitting or underfitting of the data. Nonetheless, numerous researchers have developed valuable tools to mitigate this risk.	[56], [129], [129], [150], [151], [152]
Performance Attribution Difficulty	One of the critical areas in finance and investment is performance attribution to the determining factors. Attributing performance to predictive factors is not always straightforward, especially when using more complex models like neural networks. It is not always easy to map the relationship between the outcome and the factors that resulted in that specific performance outcome. (i.e., the Barra risk factor model might not be suitable for AI strategies)	[37], [153], [154]
Incorrect Inference	AI models can make wrong decisions based on incorrect inferences based on the spurious patterns captured in the data. Sometimes, researchers create a simpler model to produce more understandable inferences than AI models.	[77], [155], [156]
Heavy Reliance on Data Quality	The quality and reliability of data are primary sources of concern. Poor data will easily take the ML process to the well-known “garbage-in, garbage-out.” Some data pre-processing techniques might help but do not guarantee data quality.	[51], [60], [157]
Requirement for Large Amounts of Data	ML and data science are data-intensive fields; their primary raw material is data. To solve real problems using machine learning, we need to have data to get accurate results. Some specific models, like deep NNs, require too much data to generate relevant results.	[30], [82], [158]
Multiple Testing	“probability of obtaining a false positive would increase as a test is repeated multiple times over the same dataset.” Multiple testing is one of the most common occurrences in published financial research. This practice, a cherry-picking approach, involves finding the best strategy and reporting only the winning outcomes. Such issues have been addressed in other fields of science. Data-driven economic researchers must be aware of this	[53], [54], [128], [159]

	phenomenon and ethically report both the successful and unsuccessful outcomes.	
Research Culture	Financial data science and ML success require teamwork and a culture that supports the scientific process rather than finding the winning strategy. If managers encourage scientific rigor and accept that most tests might lead to failure, this culture will survive by adopting this technology as the best investment research method.	[10], [51], [54]
Complexity	One of the most powerful features of ML is its capability to handle high-dimensional datasets, a task typically beyond the reach of traditional models. However, researchers should consistently strive to generate the most precise and practical model specifications. ML offers assertive techniques for dimensionality reduction, and scholars should comprehend the trade-off between inference and utilizing the model merely as a black box.	[54], [140], [160], [161]

There are implications for machine learning applications in asset management. Instead of finding the best strategies by several rounds of backtesting, we need to focus on the scientific experiments that will result in reliable outcomes. “The scientific method is an approach for examining and understanding phenomena, developing new theories, or modifying or integrating existing theories based on the presentation of empirical and measurable evidence subject to specific principles of reasoning.”

The characteristics of the scientific approach as it relates to quantitative equity strategy modeling include the following phases:

- “Development of a thoughtful hypothesis or thesis to be evaluated.
- Use empirical work to attempt to put precision around investment decisions and economic reasoning.
- Reliance on high standards of analytical rigor.

- Use of sensitivity analysis to challenge assumptions and context in which the strategy was developed.

- Incorporation of adjustments to the strategy based on the judgment.

- Ability to explicitly measure results.

- Incorporation of revisions or updates to the model as new information becomes available" [44]

The following table represents a sample of the research that addresses the need for data-driven investment research processes.

Table 6: Related Work to The Financial Data Science Research Process

Study	Challenges/Considerations	References
The Ten Reasons Most Machine Learning Funds Fail	Quantitative finance and financial machine learning often experience a high failure rate. One common mistake is conducting economic research solely through backtesting rather than following the rigorous scientific research process.	[61]
Who Needs Newtonian Finance?	Most empirical finance research still uses traditional econometrics techniques, some unsuitable for addressing today's complex financial problems.	[127]
Triumph of The Empiricists: The Birth of Financial Data Science	Authors argue that financial data science should be considered a stand-alone interdisciplinary field in economic research.	[7]
Machine Learning: An Applied Econometric Approach	Relevant modern big data can be sources of economic analysis in financial research. Machine learning provides a powerful, flexible way of making quality predictions. (i.e., policy prediction, testing theories, data-driven inductive reasoning)	[46]
Best Practices in Research for Quantitative Equity Strategies	Model development in quantitative investment strategies should follow the scientific research process and best practices regardless of the asset class and strategy category.	[44]

A Backtesting Protocol in the Era of Machine Learning	Research questions based on economic theories, multiple testing, data and sample choice, model validation, model dynamics, complexity, and research culture are building blocks of the suggested protocol in financial ML research.	[54]
Financial Data Science: The Birth of a New Financial Research Paradigm Complementing Econometrics?	Need much more engagement with performance management standards to prevent weak performance models. Although there are suggested protocols, pre-registering the research design, and actual out-of-sample results, there is a need for more comprehensive approaches.	[50]
A Holistic Approach to Financial Data Science: Data, Technology, and Analytics	The emerging financial data science and classical econometrics are complementary. The challenge is that most financial ML research only relies on some of the essential elements of research steps, and there is no holistic method. A strategy of 9 interrelated parts is proposed to manage a financial data science project efficiently.	[51]

Financial DS/ML research needs to develop robust research experiments and best practices according to scientific methods. “Best practices involve understanding how to make better decisions in the research design process. It is useful to draw on sciences from other disciplines that study decision making, often in experimental settings; these include psychology, philosophy, And organizational behavior.” [44] “A good question is, “How do we make better decisions in developing quantitative strategies?” … “The research process is the heart of developing successful quantitative strategies.”

2.3 Primer of Empirical Asset Pricing and Factor Investing

Factor investing has become part of the vocabulary of academics and finance professionals in today's world. Factor investing, which selects investments based on specific characteristics, has increased in popularity in the last three decades. The history of research on factors goes back to the 1930s when Graham and Dodd proposed value premiums. Then, two models were established: the foundational theories of modern portfolio theory, namely the Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT) [162], [163], [164], [165], [166] The main objective of these models was to demonstrate that the returns of securities can be modeled as a function of different factors. (Cerniglia & Fabozzi, 2018) The essence of these models is based on understanding the risk and return attributes of various investments. Academics believe that if investors can identify the right cross-sectional attributes of securities called anomalies, they can construct portfolios capable of beating the market.

Grouping investments, in this case, stocks, based on relative cross-sectional performance and building a portfolio that long the top-performing and shorts the bottom-performing factors achieve significant investment performance above the associated cap-weighted benchmark. [167] However, beating the market has never been easy. So, one of the perpetual objectives of researchers and investors is identifying robust, systematic, and repeatable sources of return. One of the pioneers in recognizing factors is Alfred Ross. [166] he was one of the first to note that one approach to understanding the return of stocks is to model them as a function of exposures to various factors that correspond to a set of characteristic attributes of stock returns. Although a broad spectrum of macroeconomic, fundamental, statistical, and technical factors have been proposed, the most widely referred to come from Fama and French's seminal works (1992, 1993) and Carhart (1997). [168], [169], [170] Following this academic field of research, the terms *factor portfolios* or *factor investing* was born in the industry, and one of the well-known strategies has been generated: Smart beta. [36]

Essentially, factors (value, size, momentum, quality, low volatility) are systematic drivers of stock returns – they explain why we see co-movement among some stocks and why certain stocks gain higher expected returns. Over the decades, researchers have identified several distinguishing factors as the primary sources of expected stock returns. [171] although researchers have generated a rich area of research and practitioners have produced trendy products, there are still questions in the field worth digging deeper into: 1. Which factors are independent? 2. Which factors are important? 3. Why do factors move prices? [40] academic researchers have started to examine these questions. In this regard, researchers have employed various approaches to obtain reliable answers. These approaches involve utilizing human judgment, conducting regression analysis, and employing data science techniques, particularly machine learning. [172], [173]

As we systematically look at the number of proposed factors in the literature, it looks like a jungle of factors that explain expected stock returns associated with specific factors. For this reason, researchers have started to research factor combinations. How factors are combined is as essential as which factors are used in modeling. [171], [174] Piotroski’s F-score and Mohanram’s G score, which combines fundamental factors into a holistic score to rank stocks, are well-known models in response to this challenge. This area of vital research has started to grow in recent years.

Making a multi-factor model is an essential building block of quantitative investing. The typical approach to dealing with such models has been running cross-sectional regressions to develop the relationship between future stock returns and some attributes of individual companies. Another strand of research is typically time-series regressions of portfolio returns and macroeconomic variables. The third strand of research is applying data mining techniques to find patterns and nonlinear relationships among factors from the bottom up. [31], [38], [40], [171]

These classical methods have some limitations that advanced tools and techniques in machine learning and data science can help overcome. For instance, many well-documented predictor variables have been in the last 50 years. [173], [175] The traditional methods are ill-suited to address such nonlinearity and high dimensionality issues. The main challenge is recognizing the importance of each predictor variable and assessing its predicting power compared to all other identified factors. Factor investors must prioritize this aspect due to the extensive range of proposed factors. Statistical machine learning tools hold significant potential in enhancing value within this domain.

Investors typically utilize these factor models in two ways. They can be used to enhance returns or to decompose risk for risk control. Investors achieve higher returns in the former by tilting portfolios toward factors with predictive power. This approach is called the alpha model, as portfolio managers aim to boost returns through tilting.

On the other hand, referring to the diversification concept in the portfolio construction and analysis, factors can capture the primary sources of correlation in stock returns, which will diversify away from the specific risk of each stock. The relationship between factors under study is considered linear in return and risk uses. So, considering the assumption of the nonlinear relationship between variables can be assessed.

Academics and practitioners have been seeking factors that can explain the cross-section of stock returns or capture the significant source of correlation risk between stocks. [12], [33], [36], [134], [167], [176], [177] Among many debated areas in factor research, the challenge that researchers have been facing with the increasing number of factors is mainly two folds: first, what factors have high importance and explanatory capability, second, how they should be ranked based on the level of importance.

Despite the growing interest in research on factor models, this research area has been around for several decades. The foundation of this strand of research was the pioneering work of empirical asset pricing via factors by Ross and subsequently by Fama and French in providing the analytical factor portfolio framework. [166], [169], [170] The question about the drivers of stock returns has been the foundation of modern finance. The capital Asset Pricing Model is the most well-known model of stock returns. The reason is that this model was the first model to decompose the sources of risk and create a factor-based approach to explaining stock returns.[163], [165], [178], [179] In CAPM, stocks are driven by two primary sources of risk: systematic and unsystematic (idiosyncratic). Systematic risk comes from the stock exposure to the market captured by beta, which demonstrates how much the stock returns to the market. While the idiosyncratic risk is diversifiable, investors will be compensated only for exposure to market risk.

However, the notion of "factors" was essentially popularized by Ross's Arbitrage Pricing Theory (APT) model in 1976. He suggested that the expected future returns of stocks can be a function of multiple factors. (i.e., macroeconomic). The main difference between CAPM and APT was that APT did not explicitly state those factors. This model opened a new perspective on empirical asset pricing. In the APT world, there was no pre-determined number of factors, and it could be any number and of any nature, which varied across the market. Generally, factors are considered attributes related to a group of stocks and systematically explain their risk and returns. For instance, exposure to the market is the most critical equity factor in the CAPM model.

The literature assesses and studies three alternative factors—total, immutable characteristics that explain stock returns. Value, Size, Growth, Momentum, and volatility are the most well-known factors in this category. [32], [169], [170] (see table below). Macroeconomic factors are the second study category. GNP surprise, inflation, and any other macro criteria could potentially have explaining power in depicting the expected stock returns. The last one is statistical, which typically refers to statistical factors arising from applying statistical methods like Principal Component Analysis (PCA) and other dimensionality reduction techniques. Each of these factors has been a potential source of research and practice.

2.3.1 Machine Learning in Empirical Asset Pricing and Factor Investing

There is an emerging literature on applying data science and machine learning methods in empirical asset pricing. One of the systematic and comparative studies is the work of Gu et al. (2020), in which they applied multiple machine learning methods to predict individual US stock returns and show the benefits and power of such techniques in capturing nonlinearities and addressing the estimation problem in empirical asset pricing. It has been demonstrated in their study that nonlinear models like neural networks and random forests have the best performance in capturing nonlinear interactions and producing more accurate predictions.

Systematic Factors	What It is	Commonly Captured by
Value	➢ Captures excess returns to stocks that have low prices relative to their fundamental value	➢ Book to price, earnings to price, book value, sales, earnings, cash earnings, net profit, dividends, cash flow
Low Size (Small Cap)	➢ Captures excess returns of smaller firms (by market capitalization) relative to their larger counterparts	➢ Market capitalization (full or free float)
Momentum	➢ Reflects excess returns to stocks with stronger past performance	➢ Relative returns (3-mth, 6-mth, 12-mth, sometimes with last 1 mth excluded), historical alpha
Low Volatility	➢ Captures excess returns to stocks with lower than average volatility, beta, and/or idiosyncratic risk	➢ Standard deviation (1-yr, 2-yrs, 3-yrs), Downside standard deviation, standard deviation of idiosyncratic returns, Beta
Dividend Yield	➢ Captures excess returns to stocks that have higher-than-average dividend yields	➢ Dividend yield
Quality	➢ Captures excess returns to stocks that are characterized by low debt, stable earnings growth, and other "quality" metrics	➢ ROE, earnings stability, dividend growth stability, strength of balance sheet, financial leverage, accounting policies, strength of management, accruals, cash flows

Table 7: The Main Risk Factors

(Adapted from [180])

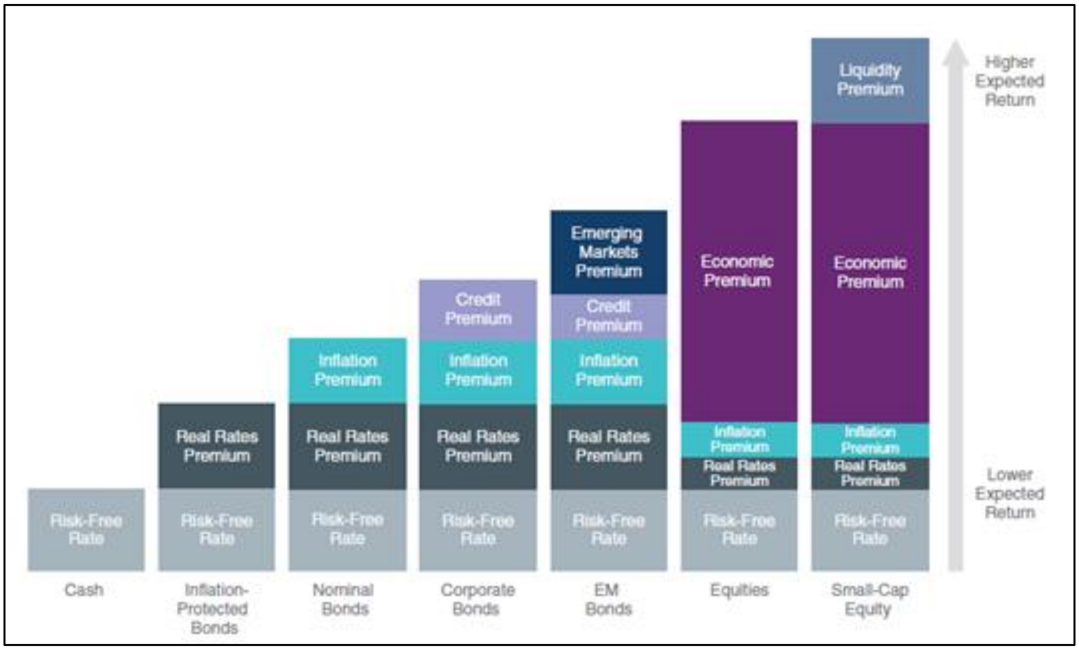


Figure 11: Asset Class Breakdown into Factors

We have witnessed sporadic research on machine learning techniques in recent years in the literature. This research seeks to answer some factor-related questions using machine learning and data science. The sample questions are as follows: What factors influence future equity returns? Is there any new factor to capture alpha? What is the relationship between factors? Do they apply across several asset classes and markets? How do we find false positives in factor analysis? These are just a small set of questions that researchers have tried to answer.

However, sporadic utilization of machine learning applications has been observed in factor investing and empirical asset pricing. Gu et al. [12] show the comparative analysis of machine learning methods in practical asset pricing. The primary purpose of this is the traditional problem of estimating asset risk premiums. Their results demonstrate the promising outcomes for machine learning portfolios. As they have stated, “measurement of asset’s risk premium is fundamentally a problem of prediction … machine learning, whose methods are largely specialized for prediction tasks, is thus ideally suited to the problem of risk premium measurement”. They also indicated that traditional prediction methods (OLS regression, for instance) are not well suited when the number of variables is high and there is a high degree of co-linearity. Machine learning has techniques such as dimension reduction, variable importance identification, and variable selection that can improve the analysis level and address such challenges. Additionally, some machine learning methods are designed to estimate complex nonlinear relationships among variables, a task where traditional methods are almost silent. Moreover, model selection criteria in financial machine learning encompass a broader range of methods that enhance the probability of finding true positive findings while simultaneously reducing the rate of false discoveries.

Noel [62] states the use of machine learning techniques in building systematic strategies that are statistically stable. He demonstrates the use of cases of machine learning in the statistical estimation of covariance and the efficient frontier. Again, this research shows some potential asset allocation improvements compared to traditional methods. Kelly et al. [136]propose one model of deep learning called “Autoencoder” to allow for a flexible nonlinear function of covariates. The resulting outcome of this research is that they were able to produce the asset pricing model with a minor out-of-sample pricing error that is less than that of other factor models. These studies are just a sample representative of the growing trend of machine learning methods in empirical asset pricing and factor investing.

Furthermore, machine learning helps address the factors from different perspectives:

- Selecting and test factors [173]
- Generating latent factors [181]
- Extracting nonlinear signals [182]
- Create an augmented linear factor model [183]

CHAPTER 3. RESEARCH GAPS, GOALS, AND OUTPUTS

3.1 Research Gap

The challenges and issues in the emerging financial data science and machine learning field mandate the more profound research of new decision models and frameworks to highlight and address the research process challenges. One approach involves applying multi-criteria decision models and emphasizing the research process over finding the best investment strategies. Some traditional finance problems that can be addressed via ML/DS methods have been mentioned in previous sections. Therefore, in the following section, we will focus on the gaps in the literature.

3.1.1 Overview

We have conducted a systematic and comprehensive study to identify the existing elements of strategy development based on ML/DS methods. Our research has focused on areas where applications of ML/DS methods, techniques, tools, and practices have been observed. Additionally, we investigate the current challenges in using these methods and explore viable solutions provided by financial data science and machine learning methods. ML/DS is an emerging area in finance and asset management, and its concepts require more clarity to empower and complement traditional econometrics methods.

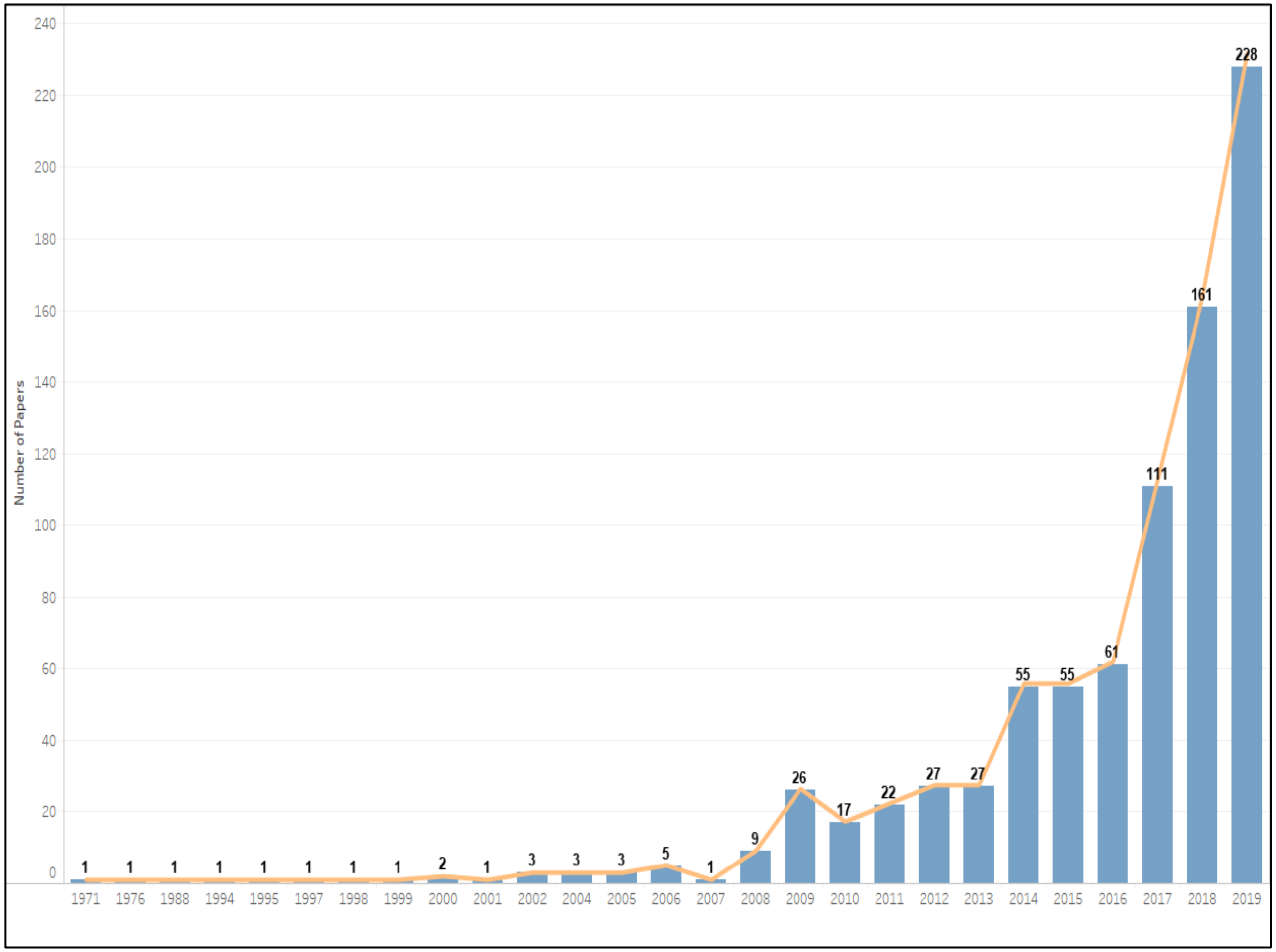
Despite AI and machine learning having long been utilized in financial services, their usage has predominantly relied on outdated techniques and technologies. However, with the abundance of data and advancements in computational technologies in recent years, many companies are adopting new techniques and identifying novel use cases. Also, as Marcos

Lopez de Prado stated, “Most machine learning investment strategies are naïve to backtest results, spurious patterns, and false positives. However, the quantity and quality of financial research projects are increasing.” Furthermore, he mentions that “Econometrics is the application of classical statistical methods to economic and financial series. The essential econometrics tool is multivariate linear regression, an 18th-century technology Gauss had already mastered before 1794 (Stigler [1981]). Standard econometric models do not learn. It is hard to believe that something as complex as 21st-century finance could be grasped by something as simple as inverting a covariance matrix.” [Advances in financial machine learning].

Nevertheless, this field undergoes rapid evolution, with implementations maturing and use cases expanding. Investment firms' adoption of machine learning is increasing, albeit accompanied by persistently high rates of false positives. This dynamic underscores the ongoing evolution of financial machine learning.

Furthermore, a bibliometric analysis of machine learning and deep learning applications in finance and investment was conducted to demonstrate the holistic view of the current research streams, best practices, use cases, and trends. In this direction, using a systematic literature review over the comprehensive Scopus database, the study investigated and mapped the literature at the intersection of machine learning and deep learning as a subset of finance and investment. The findings highlight the most important articles (highly cited), techniques applied, and active research topics by graphing keywords. One of the findings was that the investment landscape's current data science, analytics, and machine learning studies are still sporadic and fragmented.

Figure 12: The Growing Trend of Publications



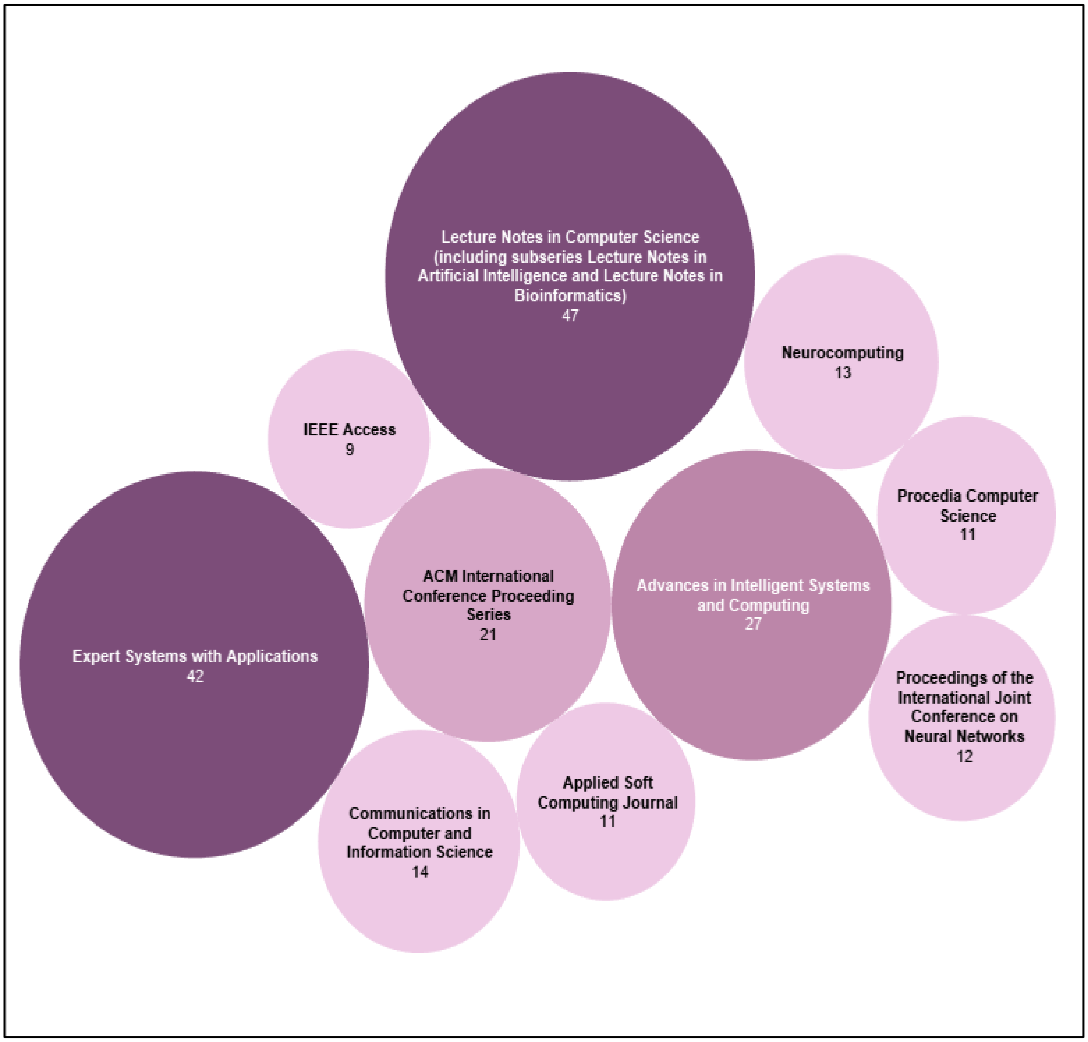


Figure 13: Top 10 Journals by The Number of Publications

Additionally, this research selected 833 papers for the study following a systematic protocol. By visualizing the co-occurrence of keywords, we demonstrated a growing interest in applying deep learning in a financial setting. The artificial neural network is not a new concept in a financial application; however, with a higher amount of data, a vast range of contemporary architectural models, and accessible computational capacity, scholars have produced more creative use cases in this area of research.

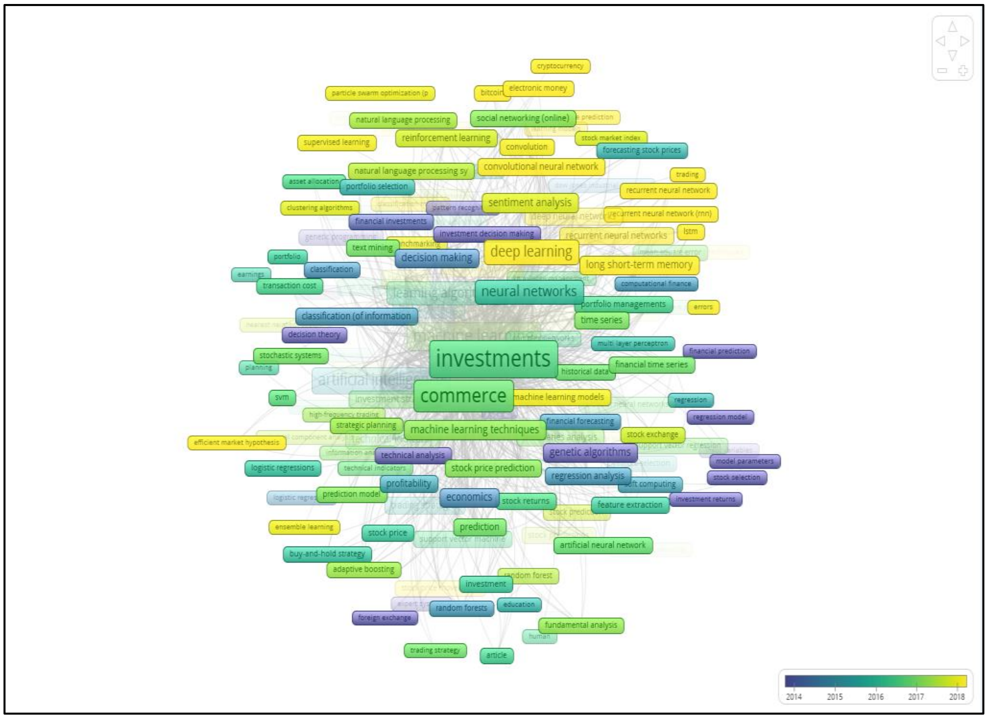
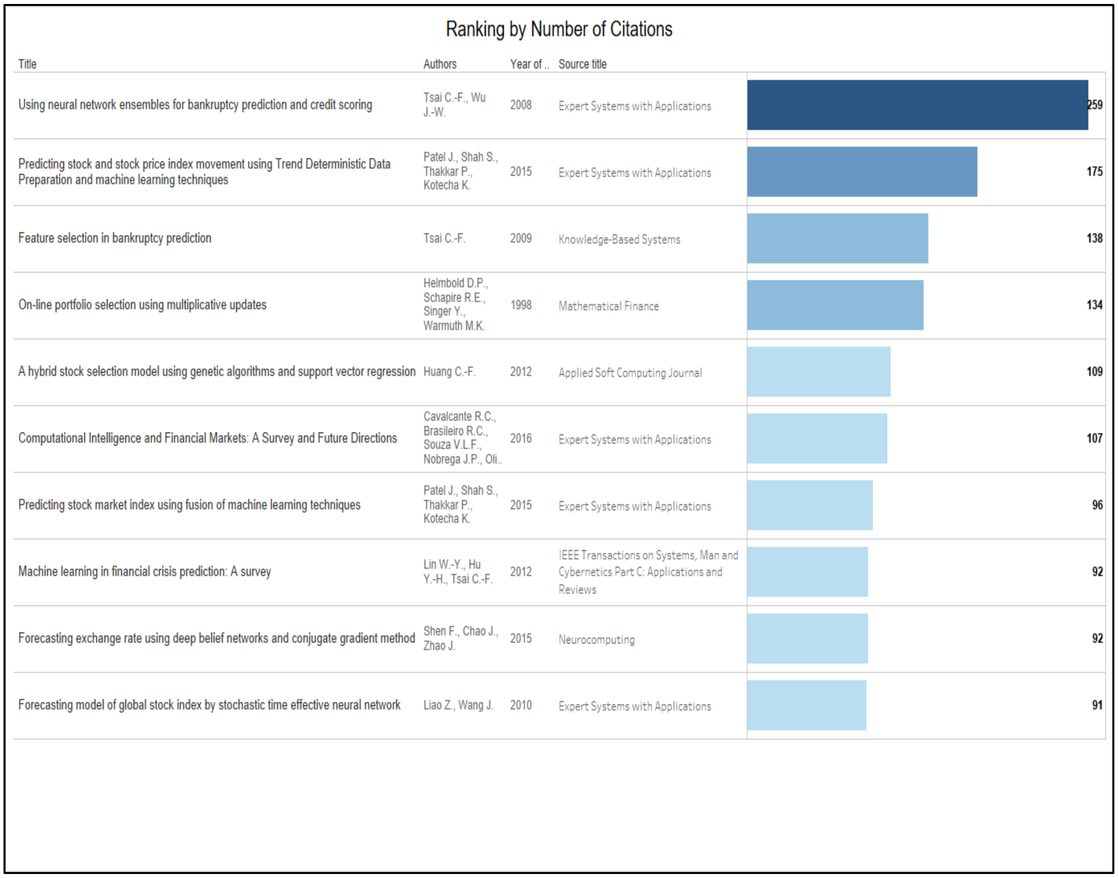


Figure 14: Co-occurrence of Keywords and Growing Trend of Deep Learning Research in Finance

Moreover, as illustrated below, ranking the papers by the number of citations shows that some papers have been extensively cited, and some authors should also be considered the most recognized and influential in this stream of research.

We have identified several gaps according to the literature findings from the general review and systematic bibliographic analysis. We have also established the goals of this research and introduced research questions. The following section presents the research gaps, goals, and questions.



Ranking by Number of Citations

Title	Authors	Year of ..	Source title	
Using neural network ensembles for bankruptcy prediction and credit scoring	Tsai C.-F., Wu J.-W.	2008	Expert Systems with Applications	259
Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques	Patel J., Shah S., Thakkar P., Kotecha K.	2015	Expert Systems with Applications	175
Feature selection in bankruptcy prediction	Tsai C.-F.	2009	Knowledge-Based Systems	138
On-line portfolio selection using multiplicative updates	Helmbold D.P., Schapire R.E., Singer Y., Warmuth M.K.	1998	Mathematical Finance	134
A hybrid stock selection model using genetic algorithms and support vector regression	Huang C.-F.	2012	Applied Soft Computing Journal	109
Computational Intelligence and Financial Markets: A Survey and Future Directions	Cavalcante R.C., Brasileiro R.C., Souza V.L.F., Nobrega J.P., Oli..	2016	Expert Systems with Applications	107
Predicting stock market index using fusion of machine learning techniques	Patel J., Shah S., Thakkar P., Kotecha K.	2015	Expert Systems with Applications	96
Machine learning in financial crisis prediction: A survey	Lin W.-Y., Hu Y.-H., Tsai C.-F.	2012	IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews	92
Forecasting exchange rate using deep belief networks and conjugate gradient method	Shen F., Chao J., Zhao J.	2015	Neurocomputing	92
Forecasting model of global stock index by stochastic time effective neural network	Liao Z., Wang J.	2010	Expert Systems with Applications	91

Table 8: Ranking of Papers by The Number of Citations

3.1.2 Gap Analysis

As aforementioned in the literature review about data science and machine learning applications in financial settings, many techniques have been applied, but there is a lack of research to address best practices and frameworks to ensure the quality of outcomes. (See Figure 20)

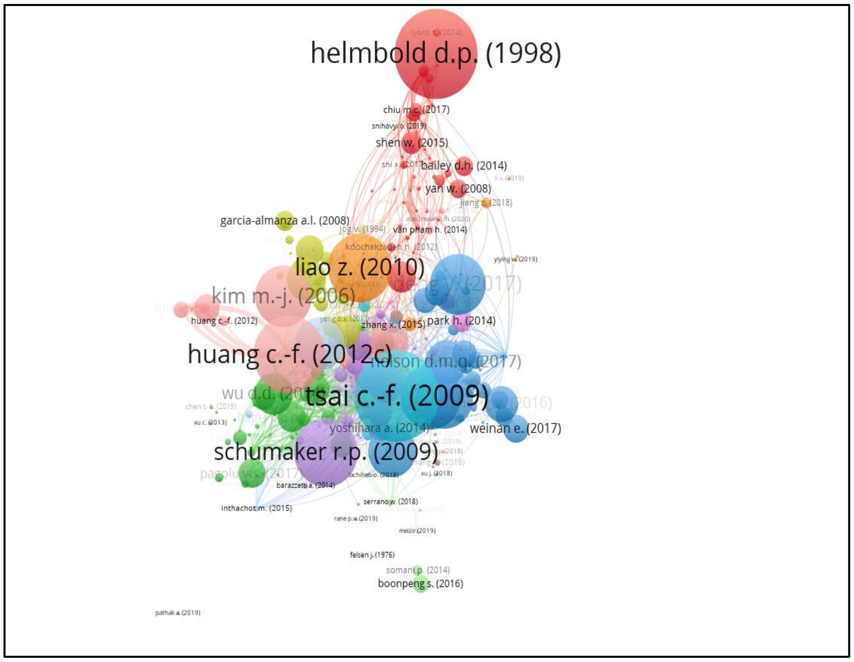


Figure 15: Bibliographic Coupling Network of Papers

Gap 1: Paucity of multi-criteria holistic studies to assess the financial DS/ML research.

One of the critical advantages of multi-attribute decision models is their capability to evaluate the problem from multiple dimensions, helping decision-makers see it more broadly. The DS/ML research process is no exception. After reviewing the literature on financial data science and the applications of machine learning in investment, it is crystal clear that there is a lack of systematic studies to assess the quality and conduct a health check of strategy development research projects, ensuring that the results are scientifically reliable for use in the practice of investment decisions. Therefore, there is a need for an overarching model to consider multiple perspectives in both qualitative and quantitative manners. This kind of study sheds light on the entire research process of these projects and tremendously facilitates researchers in seeing the big picture.

Gap 2: Scarcity of studies systematically evaluating the productivity and quality of financial DS/ML research projects.

Most research publications applying DS/ML methods have focused solely on the technical applications and specific use cases. Many exclusively run DS/ML models on datasets and analyze the results. The gap here is that the rest do not address the health check process and results, with only a handful of papers doing so. Additionally, the current research mainly concentrates on finding the best models (in the case of this study, the best quantitative strategies) with high accuracy rather than designing robust experimental research processes to study the problem scientifically. "Although models are quantitative, the research process is subject to data and model decisions that are more qualitative." [44]

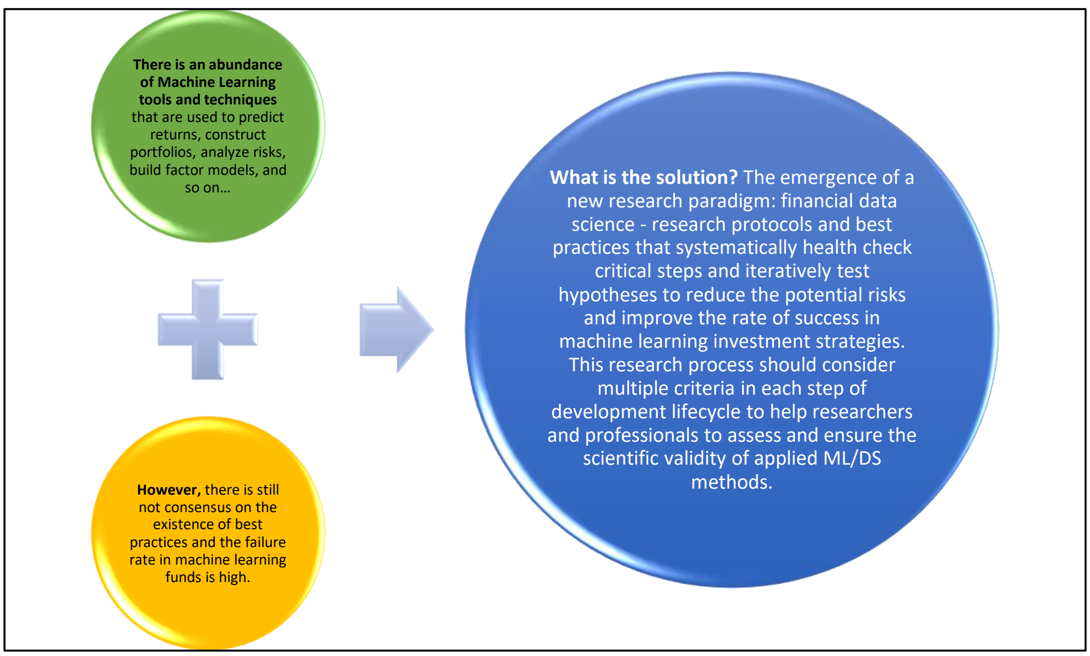


Figure 16: Gap Analysis Logic

There is a need for a holistic model based on the collective intelligence of experts to consolidate the fragmented thoughts in this line of research. This study takes some steps in this direction.

Gap 3: Lack of studies that highlight the most critical factors impacting the reliability and quality of financial DS/ML research projects.

Despite the invaluable applications of financial DS/ML and the promising outcomes, only a few papers in the literature cover some of the factors that every data science and quantitative research team should consider. Those studies are also not comprehensive and serve as just the starting point for a new stream of research. Identifying the critical success factors that help assess the research project and conducting health checks on the points that determine the quality of final results would provide an invaluable framework in the arsenal of investment teams. Thus, finding and defining these criteria in developing the model is essential and could lead to higher adoption of DS/ML techniques in investment research and practice.

Gap 4: Lack of studies based on collective intelligence and the expert's judgments and present the importance level of the factors and perspectives considered in the assessment.

One of the most ignored elements of DS/ML research is collecting and integrating experts' judgments and quantifying the impacting factors in producing the best practices and protocols. No single study utilizes the expert judgments of AI scholars, quantitative researchers, and data scientists to address this issue. The field of DS/ML is multidisciplinary. Therefore, the solution could lie in interdisciplinary fields that cross the boundaries of individual fields. The proposed approach in this research is an attempt to do just that.

Gap 5: There is an extensive literature gap in economic research to address the challenges of the promising financial DS/ML field.

Financial data science is an emerging field, and its capabilities could empower and complement the current econometrics methods. However, naive applications of such powerful tools could result in perilous outcomes. Recent studies have primarily focused on use cases and applications, and there is an evident lack of technology management decision methods to see the bigger picture and provide solutions for the entire research process. Every step in financial DS/ML projects needs special consideration to incorporate all the capacities into current investment practices fully. Exploring the criteria and identifying different aspects are the tasks that the financial DS/ML community should start thinking about and addressing to reduce the number of false results.

3.2 Research Goals

This research aims to develop a multi-criteria framework and, subsequently, a score to evaluate the quality and reliability of ML-based quantitative investment strategies, seeking to prevent the pitfalls of naive financial machine learning applications and consequently report false results. This research generally develops an evaluation framework for ML-based empirical research in finance that relies on research design. Therefore, the main objectives of this research are as follows:

- Develop a framework and a score to evaluate the quality and reliability of financial DS/ML research projects.

- Identify the factors impacting the reliability of results in DS/ML research.

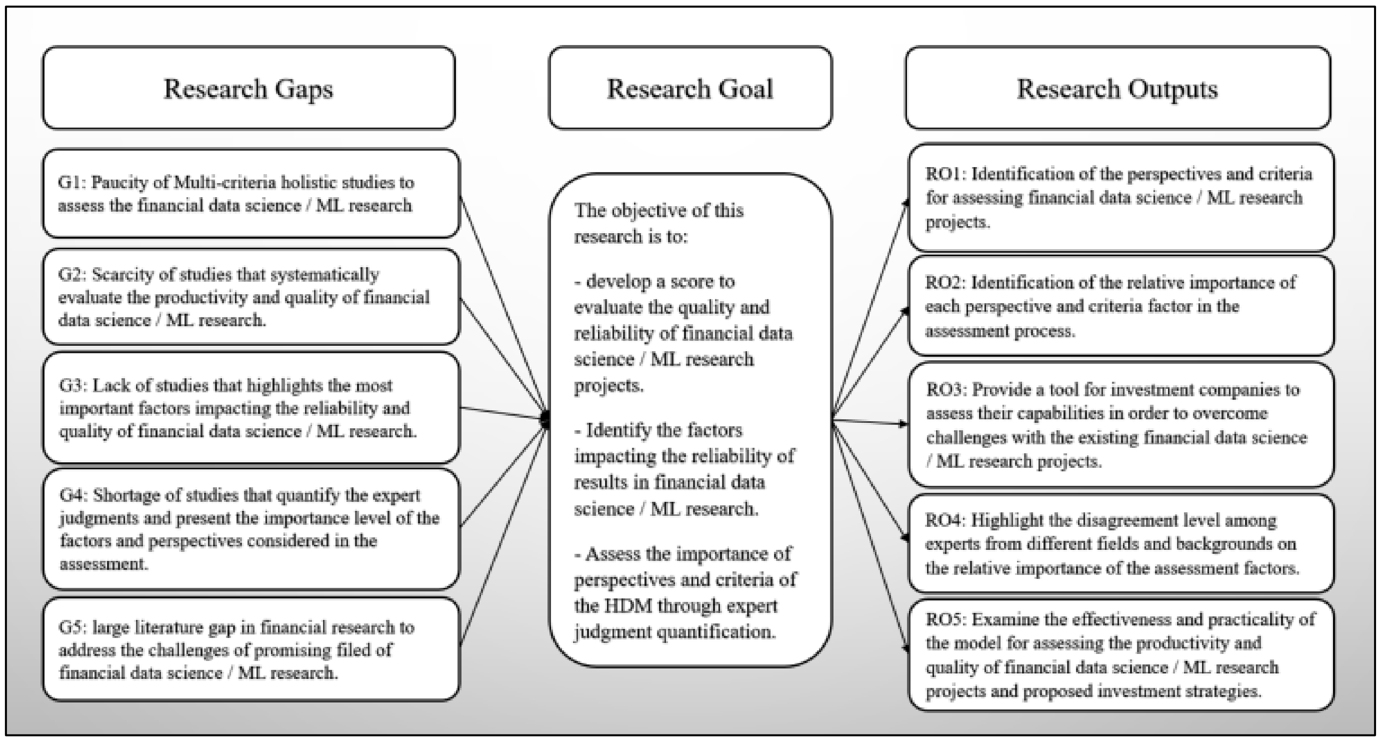
- Assess the importance of perspectives and criteria of the HDM model through expert judgment quantification.

3.3 Research Outputs

Drawing from the research process and obtaining the research goals, the research output is five folds:

- RO1: Identification of the perspectives and criteria for assessing financial data science / ML research projects.
- RO2: Identification of the relative importance of each perspective and criteria factor in the assessment process.
- RO3: Provide a tool for investment companies to assess their capabilities to overcome challenges with the existing financial data science / ML research projects and to be able to systematically evaluate ML/DS-based funds before getting exposed to those funds.
- RO4: Highlight the disagreement level among experts from different fields and backgrounds on the relative importance of the assessment factors.
- RO5: Examine the effectiveness and practicality of the model for assessing the productivity and quality of financial data science / ML research projects and proposed investment strategies.

Figure 17: Research Gaps, Goals, and Outputs



CHAPTER 4. RESEARCH FRAMEWORK

The research framework followed in this study comprises three main phases. As illustrated below, these three main steps are model development, data collection/analysis, and case study/results. Each of these steps has its subset of activities. The first phase begins with a literature review and construction of a multi-criteria model based on the research gap factors. In the second phase, we form a panel of research experts, quantify and validate the model and desirability curves with the panel, and subsequently analyze and evaluate the collected data. Finally, in the last step of this process, we design a case study to test the model and report the results after assessment.

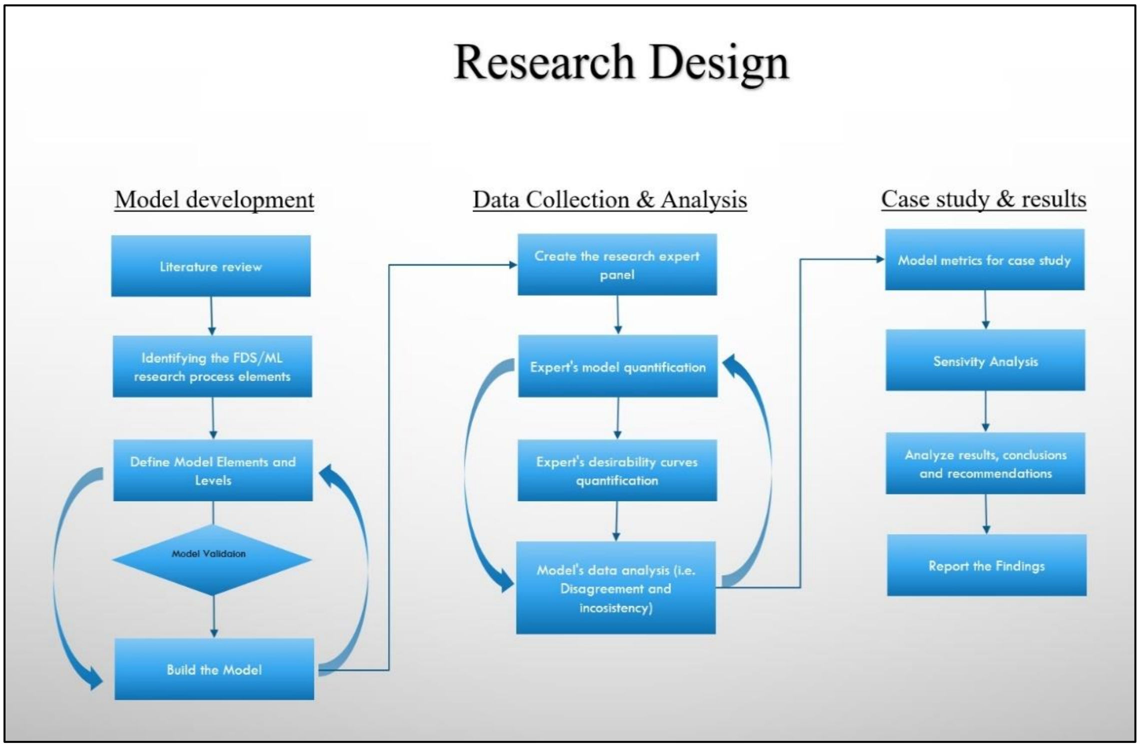


Figure 18: Research Design

4.1 Background and Literature Review

The background and literature review cover the main contributions of machine learning and data science to the field of finance and asset management, including pricing nonlinearities, working with synthetic datasets, portfolio construction, outlier detection, feature importance, sentiment analysis, risk management, systematic backtesting strategies, and challenges and considerations in using such technologies. Most of the literature review content relies on academic publications and some professional reports in the industry. All these findings are consolidated, resulting in the development of the initial HDM model, which is the proposed framework of this research. Recapping, we can break down the results of the literature review into the following primary outcomes:

- Review of the background of quantitative strategies
- Review AI, machine learning, and data science, specifically in investment and asset management.
- A brief review of deep learning applications in finance.
- Gap identification, research objectives, and research outcomes.
- Identify the model components, including high-level objectives, perspectives, and criteria.

4.2 Research Model Development

The initial model is created based on the literature review regarding the applications of ML/DS in developing investment strategies. This model is then built upon multiple factors associated with developing ML/DS investment strategies. The constituent elements that play a vital role in developing quantitative investment strategies are considered as the foundation for assessment and, consequently, the model's factors. The outcome is the initial HDM model based on the literature and subsequently enhanced with expert judgments.

4.3 Panel Formation, Model Validation and Quantification

As one of the cornerstones of the methodology applied in this research, a panel of experts will be formed to validate and quantify all the proposed factors in the model. The rounds of expert judgment feedback will shape the final version of the model, which is a more robust version that incorporates their expertise. Designing two surveys using the Qualtrics survey design platform, the first for validation and the second for quantification, experts provide insights based on pairwise comparisons at various model layers to achieve this goal. The results of this phase include both model validation and model quantification.

4.4 Model Application and Results Analysis

To analyze the reliability of individual and collective pairwise comparisons and expert judgments, we will evaluate inconsistency and disagreement indices to ensure that the judgments have not exceeded an acceptable threshold. After completing validation in this step, we finalize the model and assign factor importance along with their corresponding weights to the model and desirability curves. Subsequently, experts will quantify the desirability curves, and the model will be ready for testing in case studies to assess the quality and reliability of ML/DS investment research projects.

4.5 Discussion and Conclusion

This phase is the final step in the research process. We will conclude and recommend avenues for improvement and future developments based on the analyzed results. The findings of this research should shed light on the most critical factors impacting the quality and reliability of DL/ML investment research projects and provide guidelines to facilitate the application of such frameworks in the investment research process, resulting in more robust and consistent financial outcomes.

CHAPTER 5. RESEARCH METHODOLOGY

After conducting an extensive literature review and investigating the current research avenues and published papers on the evaluation of impacting factors and the ways ML/DS is adopted in financial settings, the researcher selected the Hierarchical Decision Model (HDM) as a capable methodology to address the current challenges facing this area. One of the key findings was that the research process, which relies on adopting ML/DS methods, is not a one-dimensional problem, and the community needs a more systematic and multi-criteria approach to address the issue. The reasoning is that when researchers start using ML/DS tools in the research and development process of investment strategies, they are faced with multiple dimensions that could affect the research results. For instance, when designing an ML/DS-based strategy, solely considering data issues would overlook other perspectives like foundational theory and model development. Consequently, the author believes a multi-perspective decision model would address this problem more reliably.

5.1 HDM Model

5.1.1 Hierarchical Decision Model (HDM) Overview

This research applies one of the multi-criteria decision models (MCDM) called the Hierarchical Decision Model (HDM). HDM and the Analytic Hierarchy Process (AHP) are two well-known and widely used methodologies in MCDM. This method was introduced by two pioneer scholars, Cleland and Kocaoglu [184], as an analytical tool capable of incorporating the expertise of subject-matter experts into the hierarchical process, which results in the ranking of alternatives according to multiple criteria. This hierarchy is a complete mesh network of relationships that systematically evaluates the experts' judgments and generates the best-performing outcomes. [185], [186], [187] This methodology can structure the decision problem into multiple levels; at each level, it seeks to do a pairwise comparison between elements, calculate priorities, check consistency, and develop the best alternative. HDM's power is its ability to systematically process experts' judgments in an absolute and relative manner.

Each level's layers and decision components depend on the studied decision problem. The top of the hierarchy starts with the model's primary objective, and at the second and third layers, we will have high-level perspectives and sub-criteria, respectively. At the bottom of the model, alternatives or outcomes of the decision are organized. Generally, HDM follows a process similar to AHP's but with a different weighting scheme. [98], [188], [189], [190] In HDM, judgments of the subject-matter experts will be converted to numerical numbers (weights), which will be utilized to calculate the importance of each criterion at every specific level of the hierarchy. It means experts evaluate each criterion in a pairwise comparison model to demonstrate the relative weight of that criterion to each final alternative. Following this procedure, each criterion will have local and global weights relative to another criterion in the model. Therefore, HDM explicitly displays the importance of each alternative based on the expert's judgments.

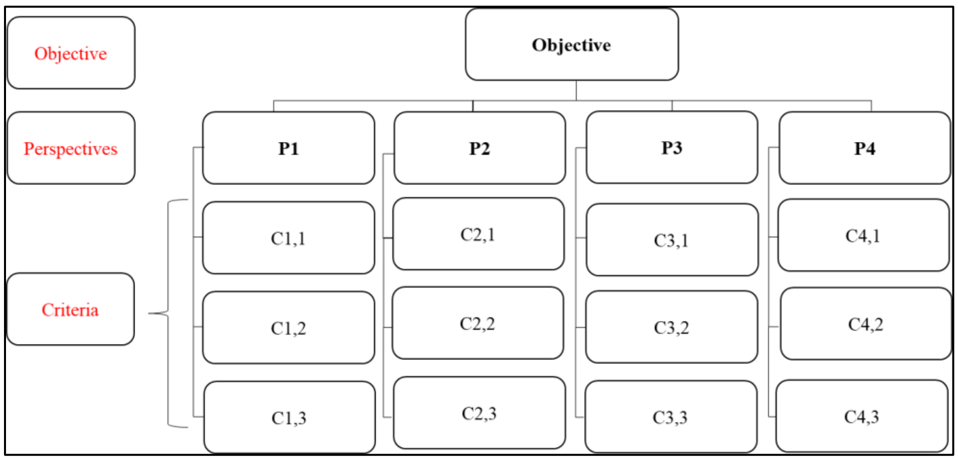


Figure 19: Example of an HDM Hierarchy

Furthermore, we will validate the reliability of the HDM model. We should check multiple metrics for the model results, including inconsistency, disagreement, and sensitivity analysis. At the individual level of an expert's judgment, inconsistency refers to the degree of disagreement among the responses received from an expert. However, disagreement will assess the extent to which there is a wide range of answers for the perspectives and quantifications among all the experts involved in the research. [187]

The following is a list of previous research that has used the same research methodology in their studies:

- A Measurement System for Science and Engineering Research Center Performance Evaluation [191]
- Development of a Readiness Assessment Model for Evaluating Big Data Projects: Case Study of Smart City in Oregon, USA [192]
- Development of a Technology Transfer Score for Evaluating Research Proposals: Case Study of Demand Response Technologies in the Pacific Northwest [193]
- A Scoring Model to Assess Organizations' Technology Transfer Capabilities: The Case of a Power Utility in the Northwest USA [194]
- Innovation Measurement: A Decision Framework to Determine Innovativeness of a Company [195]

HDM pursues a systematic top-down process on different levels that starts from the main objective at the top of the hierarchy and goes down to the final layer representing the chosen alternatives. At each level of criteria and sub-criteria, this methodology follows a systematic process to incorporate the ideas of subject-matter experts (SME) into the model by evaluating each criterion and comparing it to the overall purpose of the decision and the rest of the requirements. The collaborative nature of HDM in collecting and imputing the expert's ideas into the model is one of its most potent perspectives in relying on collective intelligence and finding the best alternative in environments with a high level of uncertainty and ambiguity. This approach helps decision-makers to have a more robust and consistent framework to find the best option for their objective. Hence, in a logical and step-by-step process, this methodology takes the decision-maker toward finding the most critical factors that impact the purpose and finally showing the best alternative.

The breakdown of the HDM model is depicted in the following figure:

As it is crystal clear from the figure, the model's overall structure is hierarchical. One key feature of this hierarchy is that it is a full mesh, and all its components interact at different decision-making layers. This hierarchy helps to have a global and overarching picture of the problem/decision we are trying to solve and its impacting factors. At the top of the hierarchy, we typically have the primary objective or mission and sub-objectives, criteria, sub-criteria, and alternatives at the lowest level. However, another powerful feature of HDM is its flexibility in decision breakdown; given the type of objective or decision, the decision-maker wants to find the best alternative. This leveling flexibility based on the context and problem makes HDM a good decision framework. [191], [192], [193], [195], [196]

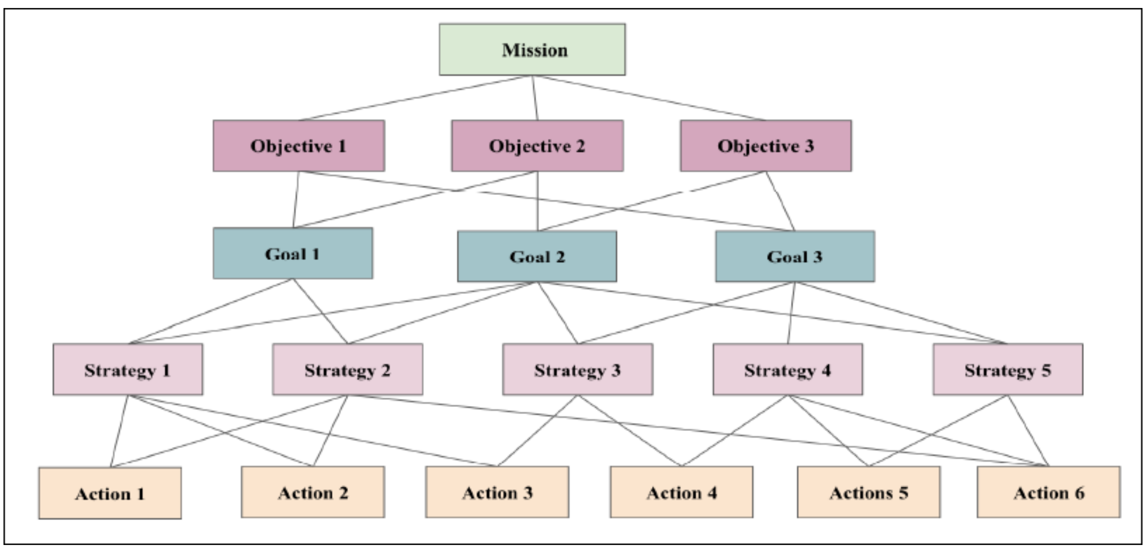


Figure 20: HDM Framework

5.1.2 Desirability Curves

Desirability curves come into play when employing the model on multiple occasions or when dealing with various options. The integration of desirability curves into the HDM framework was first introduced by Phan in 2013. Experts assign numerical values to model parameters and desirability metrics, which remain constant in this context. At the same time, decision-makers assess diverse strategies against these benchmarks based on their performance on the desirability metrics scale. In the present study, desirability curves evaluate the degree of desirability or value attached to a given metric by decision-makers.

For each criterion within the model, experts are presented with specific units of measurement and associated categories. These experts then assign a numerical score ranging from zero to 100 to each category for every criterion, indicating the level of desirability related to each category. Subsequently, the curves are plotted based on the average assessments provided by the experts. Desirability curves capture the nuances and intricacies of each criterion, illustrating the dynamic nature of these components. The chief advantage of incorporating desirability functions lies in the flexibility they offer to the model. A sample desirability curve for one criterion is presented below, with further discussions on the remaining criteria in Chapter 6.

Invited to construct desirability curves, experts discuss, validate, and assign values to each criterion within the model, considering typical scenarios encountered in machine learning and data science investment strategies. They are tasked with assessing the potential outcomes for each criterion that investment strategies may experience. Furthermore, they express their preferences regarding desirable criteria and the degree of desirability associated with each. A survey is administered to all experts to facilitate the exchange of their collective expertise.

In practical applications, investment teams take the following steps: Firstly, they evaluate each strategy's readiness by considering its current status and proficiency within each criterion, identifying how each criterion currently affects its readiness. Subsequently, the team utilizes desirability curve values to select the value level that best represents the strategy's current status during the investigation, determining the value levels and their corresponding scores at this stage. Using these selected value levels, we calculate the readiness score for a data science/machine learning investment strategy by multiplying the weight assigned to each criterion by its corresponding desirability curve value. The subsequent graph provides an illustrative example of a desirability curve for the feature selection and importance criteria within the feature perspective.

5.1.3 Inconsistency, Disagreement, and Sensitivity Analysis

We ascertain the reliability of HDM through a rigorous examination, including assessing inconsistency, analyzing disagreement, and evaluating sensitivity. Inconsistency in expert judgment arises when discrepancies emerge within an expert's assessment in logical comparisons, signifying incongruities in their comparative evaluations.

This factor measures the actions taken and techniques used in feature construction, feature extraction, feature selection, and feature importance. Techniques in feature importance, for instance, can provide insight into the dataset and highlight the most relevant feature to the target variable. Below are the categories:

- No feature engineering and importance.
- Low level of feature engineering and importance.
- Medium level of feature engineering and importance.
- High level of feature engineering and importance.

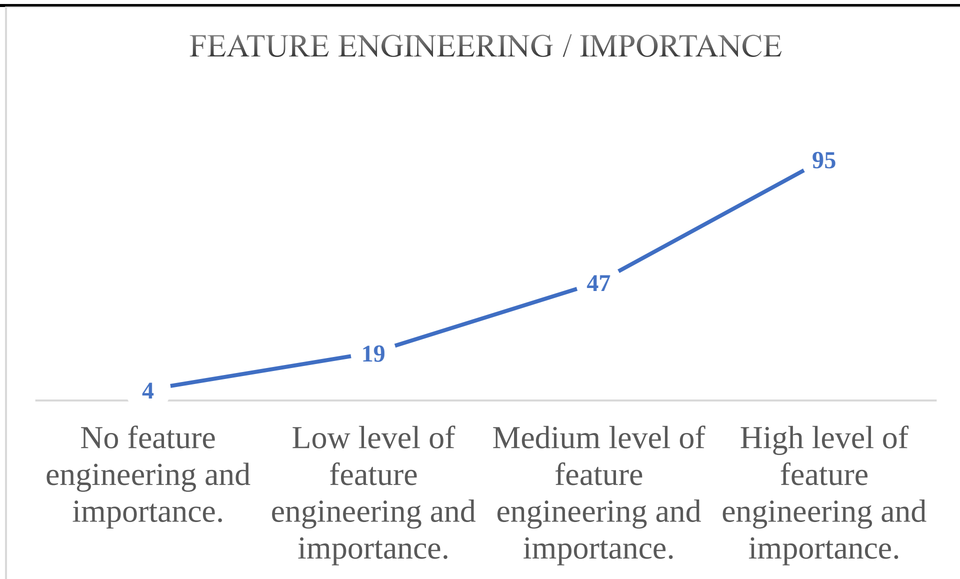


Figure 21: Sample Desirability Curve

Feature engineering/Importance	
Description	Desirability
No feature engineering and importance.	4
Low level of feature engineering and importance.	19
Medium level of feature engineering and importance.	47
High level of feature engineering and importance.	95

Table 9: Measure Description

Expert disagreements can offer diverse quantifications and contrasting perspectives within the same analytical framework. In addition, sensitivity analysis gauges the model's adaptability to alterations [186] [191].

Several preceding research endeavors have adopted a comparable approach to the one outlined in this study, addressing and applying these metrics, including:

- ➢ Development of a Technology Transfer Score for Evaluating Research Proposals: Case Study of Demand Response Technologies in the Pacific Northwest [193].
- ➢ A Scoring Model to Assess Organizations' Technology Transfer Capabilities: The Case of a Power Utility in the Northwest USA [194].
- ➢ Development of a Readiness Assessment Model for Evaluating Big Data Projects: Case Study of Smart City in Oregon, USA [192].
- ➢ A Measurement System for Science and Engineering Research Center Performance Evaluation [191].

The forthcoming section will comprehensively explore inconsistency, disagreement, and sensitivity analysis. This discussion will explore the methods for identifying, quantifying, and addressing inconsistencies and disagreements within expert judgments.

5.1.4 HDM Benefits and Limitations

Strengths and limitations exist for each methodology used, and HDM is no exception. The following are considered as pros and cons for the HDM methodology in the context of this research:

Strengths:

- HDM is undoubtedly one of the best decision support frameworks for complex and multi-criteria decisions.

- HDM is a mixture of qualitative/quantitative methodology in its essence. Incorporating experts' judgments inside the model and converting them to quantitative measures makes it a powerful tool for decision-makers.
- HDM efficiently evaluates multiple criteria through the hierarchical structure through systematic analysis and collective intelligence of expert judgments.
- Robustness and inconsistency in the small group of expert judgments can be effectively managed and controlled.
- The level of aggregation in HDM takes advantage of the wisdom of the crowd (expert's judgment in HDM)
- The reusability of the model could help users apply it multiple times as decision support without starting from scratch.
- The marriage of expert judgments with domain knowledge with the structured mesh representing a procedural logic makes HDM beneficial in different contexts.

Limitations:

- Research from various fields has shown that even experts have behavioral biases in making decisions. Although they are experts in specific fields, this does not necessarily mean they make the most optimum decisions, even in their expertise. Therefore, one critical step in this research is how experts will be selected and formed in different panels to address this issue.
- Furthermore, a degree of disagreement consistently exists among decision-makers, especially in complex decisions with heightened uncertainty. Experts are not exempt from this phenomenon.

- Defining an “expert” in a specific field is controversial and fuzzy, especially when the study area is interdisciplinary.
- HDM model could be susceptible to radical changes in the value of variables. The lack of robustness in the model raises doubt about its application in practice.

5.1.5 Justification of The Method

In order to fulfill the research objectives and address the defined research questions, the method should be capable of considering multiple perspectives and criteria and employing a systematic process to handle the complex decision of selecting and ranking investment strategies in asset/investment management companies. Hence, the proposed methodology should be able to incorporate the researched attributes into a well-structured framework. To achieve this goal, we must ensure the following characteristics are met.

This research aims to assess and evaluate the readiness of ML/DS-based investment strategies in asset management companies for practical implementation. We assume this complex and multi-attribute decision will be evaluated through scientific and systematic decision-making processes. As depicted in the literature review of this research, several requirements in different steps of ML/DS-based investment strategy development vividly demonstrate the need for a multi-perspective decision model.

Additionally, deciding which strategies are ready to be selected and implemented for an investment research team in an asset management company is critical. Investment strategy readiness can be systematically addressed through multiple steps: first, an extensive literature review discovered the most critical factors. Second, the model represents the criteria for improving investment strategies. Finally, the model can indicate why some strategies might have underperformed or overperformed. Furthermore, desirability helps decision-makers see the level of desirability for each criterion and how far a sample strategy is from those values. In essence, desirability values will work as benchmark levels against which researched ML/DS investment strategies can be compared.

Moreover, end-users should easily use the model (i.e., investment research teams) and be explicitly understandable. In addition, the reusability of the developed model would tremendously save time and money for strategy research and development processes, which would be invaluable for investment companies.

Finally, this study's research area and problem are interdisciplinary, and making simple assumptions in decision-making could not be a suitable approach. Such emerging research areas require an interdisciplinary decision process to generate consistent outcomes based on systematic scientific methods. Methodologies like HDM can shine in these settings as they need experts with different backgrounds and areas of expertise. These multi-perspective methods can capture diverse viewpoints and embed them into the reusable model, increasing the probability of finding the best strategies for asset allocation and practical implementation.

As mentioned, there are several methods in MCDM with specific strengths and weaknesses, which might be potential candidates for solving complex decision problems. Regarding suitability, HDM is considered one of the most appropriate methodologies to address the research problem in this study and achieve its goals. The following underlying reasons have been the driver of selecting HDM as a research methodology:

- HDM hierarchical structure allows the researcher to decompose a complex problem into sub-problems that are easier to solve.
- HDM method is suitable as it can incorporate multiple criteria into the model.
- Expert judgment's quantification, depicting the relative importance of criteria, and assessing the expert's individual and collective discrepancies through inconsistency and disagreement concepts are enabling factors that increase the reliability of results.
- It provides a step-by-step process to evaluate and select the best alternative, which is then validated by the outcome of the expert's judgments.
- Desirability curves bring more flexibility to identifying criteria, and the level of investment strategies is far from the benchmark values. This feature is a powerful characteristic that helps investment teams rely on robust performance evaluation methods.
- One of the main cornerstones of academic research is its reusability and replication by other researchers. The essence of desirability curves allows users to accomplish results based on some aggregated expertise captured via values.
- Furthermore, sensitivity analysis can help decision-makers test the model against extreme scenarios and evaluate how sensitive the model outcomes are to specific parameters.

5.1.6 Model Generalizability

The comprehensive literature review is conducted regarding the implementation of DS/ML-driven investment strategy development and its readiness within the investment management field, alongside prior independent studies, and the reviewing of assessment methods has significantly contributed to identifying key determinants influencing the incorporation of systematic and robust research processes. We categorize these determinants into distinct perspectives, specifically Economic foundations and research underpinnings, Data, Features, Modeling, and Performance. The proposed model in this research is designed to evaluate DS/ML-driven investment strategy readiness and assess a wide range of investment strategies. However, it is imperative to acknowledge that the generalizability of the research findings depends on contextual and temporal variables. The recognized factors and perspectives may evolve and not reflect the new conditions under study.

It is essential to take the model factors through a validation process to develop an overarching objective of achieving generalizability. This process has a dual purpose: to ensure the inclusion of the most critical factors within the model and to enhance its fidelity to real-world scenarios. Experts proficient in quantitative investment and financial data science are solicited for their input to ascertain the model's application beyond the specific use cases. They are engaged in verifying the model’s perspectives, corresponding criteria, the suitability of desirability curves and their levels, and evaluating the outcomes derived from the model quantification process.

When selecting experts to validate the model, choosing individuals with a profound understanding of this interdisciplinary domain is crucial, including those with diverse professional backgrounds and varied experiential insights. A comprehensive delineation of the criteria for expert selection and panel formation is demonstrated in Section 5.2.2. Please pay special attention to the model validation process to ensure dependable outcomes and develop a model that can be used confidently across diverse contexts while substantiating its generalizability.

We use several validity measures to ensure the reliability of research results and the model's generalizability. These measures include assessing expert judgments, evaluating disagreement among experts in a panel, and conducting sensitivity analysis. Sensitivity analysis evaluates the impact of potential variations in the values assigned to different levels of the Hierarchical Decision Model (HDM) to gauge the model's robustness. In addition, this approach sheds light on any ranking changes that might occur in the outcomes under extreme conditions. Conducting sensitivity analysis along with the HDM results is instrumental in formulating a comprehensive strategy capable of addressing a spectrum of contingencies. This analytical approach offers a lucid depiction of the interrelationships between various levels and their constituent components.

5.2 Experts Judgement

5.2.1 Expert's Judgement Overview

Panel formation and expert judgment collection require proper consideration regarding perspectives and criteria. Drawing a line or boundary within expert selection is not often straightforward in interdisciplinary research domains. Expert definition based on one's background can pose a complex challenge, as it frequently involves converging multiple knowledge areas. Consequently, it is imperative to establish precise selection criteria and firmly ground the research problem and goals on a robust literature review foundation. Effectively communicating these criteria and objectives to the experts is paramount, as it enhances the likelihood of achieving successful outcomes and bolsters the reliability of results. Nonetheless, it is essential to acknowledge that inherent biases in human judgment are an intrinsic aspect of a human's cognitive nature. Even experts are not entirely immune to these cognitive biases, which can potentially influence and skew the research findings.

Furthermore, in today's rapidly evolving world, with too many things that keep people busy, engaging individuals and securing their active participation in research activities is not trivial. Many experts find themselves entangled in many commitments and responsibilities, necessitating careful consideration of their willingness and availability in advance to align with the project's timeline. It is crucial to recognize that a lack of willingness from experts can pose a substantial issue, even if they reluctantly agree to participate in the research. This reluctance may result in a diminished capacity to contribute their expertise and insights to the study, which is particularly problematic in cases where the reliability of results requires stringent verification and validation.

Another potential challenge is maintaining a high communication standard with experts and precisely articulating the research goals and expected outcomes. Any inadvertent bias introduced into the communication process by the interviewer can lead to unintended and erroneous research outcomes. Therefore, it is imperative to establish a well-defined and consistent communication protocol to mitigate such risks.

Lastly, the analysis of the gathered responses should consider several critical aspects. These include assessing the consistency in the experts' responses, identifying disagreements among participating experts, and evaluating the model's sensitivity to specific variables at an aggregated level. These considerations play a pivotal role in analyzing the collected data and ensuring the robustness and integrity of the research findings.

Considering these considerations, we invite experts with the following backgrounds to participate in each panel: financial data scientists, quantitative finance researchers, investment professionals, academic researchers, AI finance project managers, and quantitative investment teams comprised of public employees. By making this diverse and interdisciplinary cohort, the researcher aims to augment the likelihood of incorporating multiple perspectives into the study. However, it is essential to note that effectively managing potential disagreements within the group is also a critical aspect of this endeavor.

5.2.1.1 Expert Characteristics

Expertise can be conceptualized as a multidimensional prototype characterized by seven key attributes [195], [196], [197]:

Advanced Problem-Solving Processes: The ability to engage in complex and sophisticated problem-solving methodologies.

Significant Knowledge Base: Possessing an extensive reservoir of knowledge in a particular domain.

Advanced Knowledge Organization: Proficiency in structuring and organizing knowledge effectively within the chosen domain.

Effective Knowledge Utilization: The skill to apply knowledge efficiently and effectively in practical contexts.

Creative Aptitude: The capacity to generate novel insights and knowledge by building upon existing information.

Automated Actions: The capability to perform tasks and actions within the domain almost instinctively, without conscious effort.

Practical Mastery: A deep understanding of how to excel and thrive within one's field.

Expertise is a multifaceted concept containing several dimensions. It involves formal knowledge, also known as declarative knowledge, which originates from structured education and a deep comprehension of theoretical principles. In addition, expertise incorporates practical knowledge, often termed procedural knowledge, which is acquired through hands-on experience, fostering the development of practical skills and a profound sense of "knowing-how." Furthermore, expert knowledge extends to self-regulative knowledge, which includes reflective skills employed by individuals to evaluate and assess their actions and decision-making processes critically. These three components collectively contribute to the rich tapestry of expertise, enhancing individuals' ability to excel and innovate within their chosen field. [195]

5.2.1.2 Levels of Expertise

Experts progress through developmental stages as they evolve into seasoned professionals. Dreyfus and Dreyfus [198] introduced a five-stage model elucidating the acquisition of expertise:

Novice Stage: At this initial phase, individuals lack prior experience or knowledge in the relevant subject or context, yet they possess a rudimentary understanding of basic rules, although their practical skills are limited.

Advanced Beginner Stage: Advancing from the novice stage, individuals begin to grasp the contextual nuances and discern distinctions within the domain. They exhibit acceptable performance but remain in the early stages of skill development.

Competence Stage: In this phase, individuals further cultivate their skills, accumulate experience and knowledge, and develop a comprehensive understanding of the subject or situation's intricacies. They can proficiently manage various elements and procedures, demonstrating efficiency and confidence in their actions.

Proficiency Stage: Proficient individuals perceive a situation holistically, interpreting its significance in the context of long-term objectives. Proficiency is achieved through integrating experiential learning in an intuitive, non-theoretical manner. Responses to various situations become automatic and intuitive rather than solely reasoned.

Expertise Stage: At the pinnacle of expertise, individuals possess an in-depth comprehension of the situation, enabling them to make subtle and nuanced distinctions. They exhibit heightened flexibility and exceptional proficiency, enabling rapid and intuitively informed responses to complex scenarios.

5.2.1.3 Expert Identification and Selection

After establishing the criteria for selecting experts, the subsequent critical step pertains to the methodology employed by the researcher to locate and enlist these experts. Identifying experts in the realm of interdisciplinary fields such as financial data science and machine learning, especially in investment practice, is notably challenging, given the emergent nature of this technology and the relative scarcity of widespread expertise. However, several practical and recommended approaches exist that researchers can employ to identify and engage the requisite experts. Tran advocates for an approach that commences with personal connections, followed by a snowball sampling technique and social network analysis [201].

Within the context of this interdisciplinary research, the most suitable candidate organizations would encompass investment and asset management companies or any investment teams within other organizations that incorporate machine learning and data science into their research and development for investment strategies. Drawing from both existing literature and the author's own experience, we can source prospective experts from three primary channels:

Quantitative Asset Managers: These individuals may hail from prominent quantitative asset management firms such as DE Shaw, Two Sigma, Point72, Millennium, Citadel, AQR Capital, or from quantitative investment units within larger investment management companies and financial institutions like Goldman Sachs, JP Morgan Chase, Morgan Stanley, Bank of America Merrill Lynch, UBS, BlackRock, Vanguard, Fidelity Investments, and State Street Global Advisors.

Academic and Professional Researchers: Experts from this category typically straddle the realms of quantitative finance and investing, AI and ML research, financial data science, and factor investing. They often maintain affiliations with academic institutions and investment entities in research within these specialized areas.

Social Network Analysis (SNA): Bibliometrics and Social Network Analysis, often called SNA, constitute a strategic approach to illuminate the intricate landscape of publication ecosystems. These methods are valuable tools for identifying authors and published works that exert significant influence, predicated on various criteria, including centrality, betweenness, and citation counts. In the context of this research, these analytical techniques can be applied to pinpoint experts who have made the most pronounced impact and substantial contributions to the relevant field. As per the guidance provided by Garces et al. (2017), these methodologies can effectively aid in identifying and selecting such experts. [197]

In identifying experts, the steps outlined in Figure 25 serve as the guiding framework for generating a list of experts pertinent to this study. As previously mentioned, various metrics are available to discern the most significant contributors within the field. The primary selection criterion in this endeavor is the expert's co-citation count in published academic papers. The interconnectedness of these published works is ascertained based on the frequency with which they are cited together. To facilitate this analysis, we employ the widely recognized VOSviewer software as valuable software for visualizing the relationships among these publications [198]. Figure 25 below shows the clusters of significant financial data science and machine learning publications. This illustration demonstrates four principal clusters of authors who have made notable contributions to this burgeoning field.

To identify experts, we meticulously follow the steps outlined in Figure 25. The Web of Science is the primary scientific database for this research endeavor. We systematically search multiple relevant keywords within this database and select pertinent papers for visualization using the VOSViewer tool. Experts whose research papers have garnered numerous co-citations across many publications emerge as central figures, signifying their

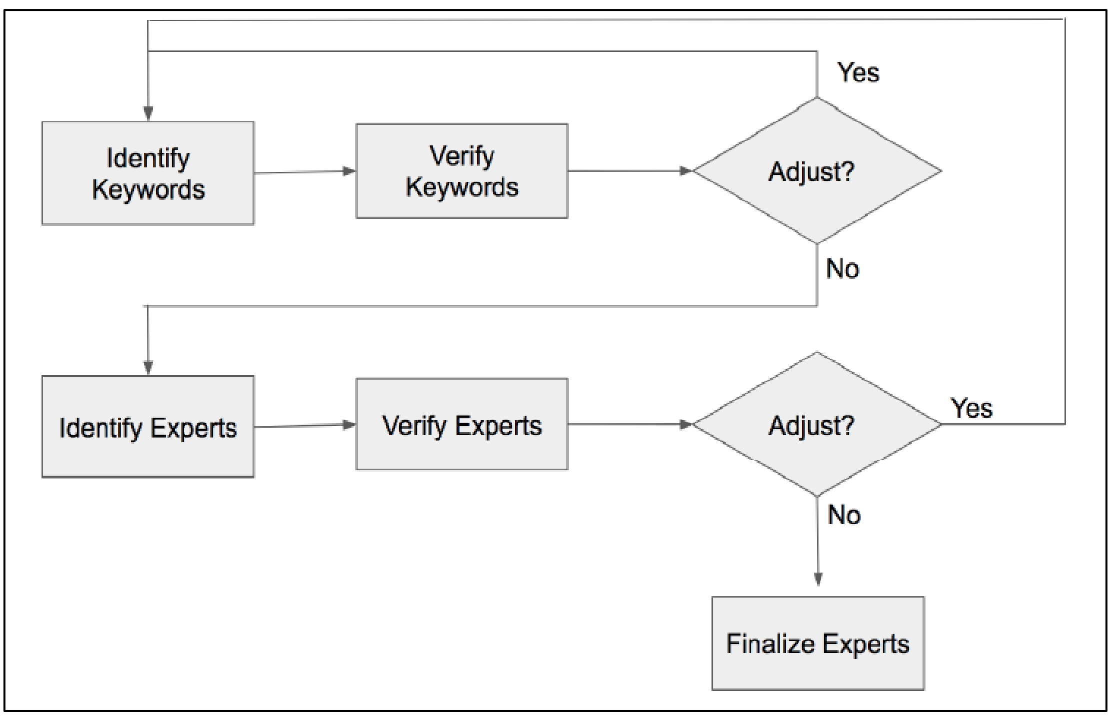


Figure 22: Social Network Analysis (SNA) Steps

prominence and expertise within the subject matter. Some prominent figures are also included among the 50 experts listed in the subsequent section (refer to Figures 26 and 27). A discernible pattern emerges from the network of publications, revealing the presence of four principal clusters of experts actively engaged in this specific realm of research. These clusters constitute a valuable resource for selecting experts drawn from the literature.

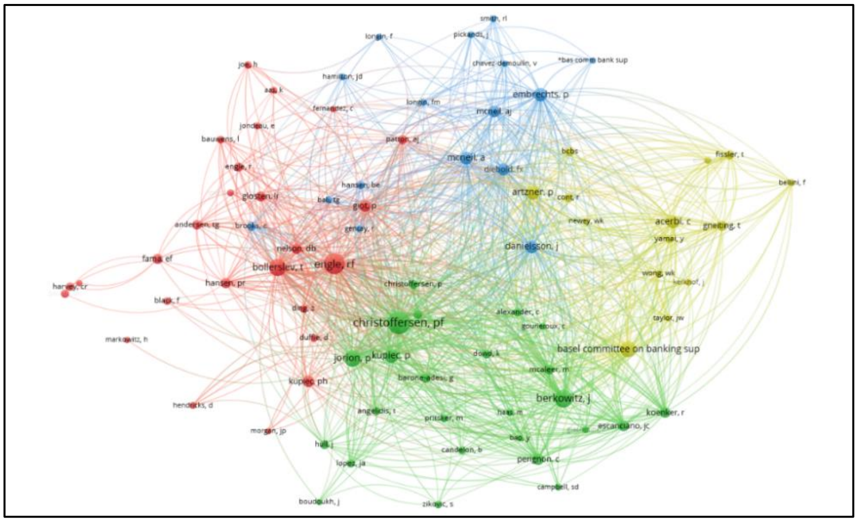


Figure 23: Network Visualization of Scientific Papers and Their Citation Importance

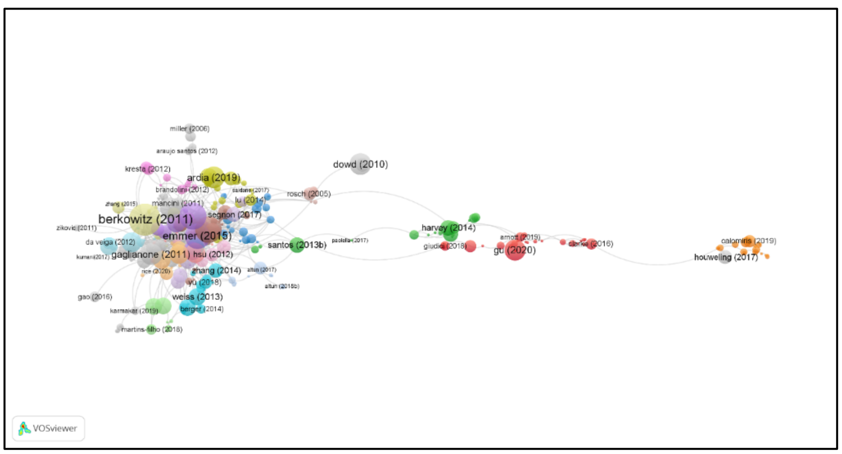


Figure 24: Citation Network Visualization of Papers on Web of Science

Therefore, the identification of experts is underpinned by multiple sources. Firstly, the researcher leverages the connections established over the past several years. Secondly, he actively tapped into and engaged with people within the LinkedIn community in the field. Additionally, insights are drawn from the results of the Social Network Analysis (SNA). This diverse pool of contacts constitutes a valuable resource for expert identification, given that many of these individuals enjoy widespread recognition and have a history of notable achievements within the field.

Moreover, the email script presented in Appendix A, titled "Letter of Invitation to Experts, " facilitated inviting experts." This correspondence involved the distribution of three distinct links to experts, each addressing specific requirements about model validation, model quantification, and the quantification of desirability curves. The model validation link directed experts to a tailored Qualtrics survey designed explicitly to validate model factors by experts. An interface was crafted within the ETM HDM software tool for model quantification, allowing experts to use pairwise comparisons directly within a controlled development environment. Finally, we meticulously designed another Qualtrics Survey to enable experts to quantify Desirability Curves and assign importance levels to them.

5.2.2 Expert Panel Formation

The practice of collecting expert judgments is a prevalent approach in both academic and professional research studies. Subject matter experts have emerged as essential players in validating and assessing research projects, offering invaluable expertise that plays a central role in appraising research quality and charting pathways for future investigations. This significance is particularly pronounced in qualitative research, where their insights assume a critical role due to the inherent uncertainty and vagueness often associated with model variables and characteristics. Furthermore, regardless of the specific contribution, an expert's judgment serves as a significant input across various facets of scientific research, including idea generation, hypothesis testing, model development, and the provision of context-specific expertise. [184], [186], [199], [200], [201], [202], [203], [204], [205]

In certain scientific studies, the acquisition of data or the validation of models can present inherent challenges. In such contexts, when conventional methods may prove elusive, the insights and expertise offered by experts play a crucial role, facilitating insights and inferences that would otherwise remain unattainable. The Hierarchical Decision Model (HDM) modeling approach is a prime example. Incorporating expert judgments into various process stages, including drawing model boundaries, selecting criteria and perspectives, validating and quantifying criteria, and addressing inherent model uncertainties, contributes to generating robust outcomes. We may execute data collection through established methods such as surveys and interviews. Furthermore, we can evaluate experts using various team-based methodologies like the Delphi method, pair-wise comparison, and focus groups, enhancing the comprehensiveness and reliability of the research process. [184], [206], [207], [208], [209], [210]

Developing expert panels and gathering their assessments requires scholars to follow a systematic approach. This approach ensures the adoption of a consistent and resilient systematic method throughout the panel formation and elicitation process: [193], [201]

• Before forming an expert panel and initiating contact with potential participants, it is imperative for the researcher to attain a lucid comprehension of the research problem and the nature of data required to address said problem. This preliminary phase is the foundation upon which the decision to utilize expert judgments as a viable method or opt for alternative approaches is predicated.

• After framing the research problem with precision and determining that an expert panel is the chosen methodology, the researcher must establish the criteria that delineate an expert and deliberate on the composition and quantity of panels to be constituted. For instance, when confronted with a specific research issue, considerations encompassing attributes such as educational background, age, gender, and diversity may necessitate contemplation to construct an apt panel for collecting judgments.

• Once the research problem has been unambiguously defined and the appropriate panel(s) have been formed, the researcher should adhere to a systematically structured scientific methodology while acquiring and eliciting expert judgments. This phase aligns with the foundational principles of scientific philosophy, wherein the gathering of judgments from the expert panel follows the established norms of scientific inquiry.

• At this juncture, the research problem is well-defined, panels are constituted following predefined criteria, and a structured scientific process is in place. Consequently, the scholar may proceed to initiate contact with experts who have been selected based on the rigorously defined criteria established during the panel formation phase.

• In the final stage, the gathered judgments are systematically collected, securely stored, and meticulously organized to facilitate their application in the subsequent stages of the research endeavor.

Henceforth, the identification and selection of experts include pivotal phases in the constitution of expert panels. The quality and dependability of an expert's expertise within the context of the research problem wield substantial influence over the credibility of the resultant findings. While the quantity of experts bears significance in the pursuit of robust conclusions, it is equally essential that the panels accurately represent the related field of inquiry. If the selected panel members fail to mirror the larger population of experts within the field adequately, it raises concerns regarding the reliability of the research outcomes. This emphasis on expert quality is not a novel concept, as the peer-review process has long served as a cornerstone of scientific investigation. This factor enhances the authenticity of research findings, as peer reviews within a specific domain corroborate the degree to which a given study is well-established and its underlying problem's significance in the field. [187], [201], [211]

5.2.2.1 Expert Panel Definition

An expert panel refers to a collective of individuals who possess specialized knowledge and insights and are enlisted when a project necessitates highly technical input and expert opinions [199]. They are selected with diverse fields of expertise to engage in and ultimately offer recommendations to facilitate well-informed decision-making. The expert panel members must maintain current knowledge and exhibit impartiality regarding the research findings [200]. The literature underscores the significance of constructing expert panels that exhibit balance by incorporating experts with diverse knowledge and expertise and ensuring an unbiased approach to addressing the specific decision or problem under examination [175]. This section will delve into the crucial aspects surrounding the selection and formation of expert panels.

5.2.2.2 Expert Panel Size

The optimal number of experts within an expert panel to achieve its intended objectives has been extensively discussed in the literature. Determining the appropriate panel size poses a notable challenge [175]. A minimal number of experts in a panel can undermine the study's reliability, while an extensive panel can introduce process complexity and complicate panel management. Nevertheless, the number of experts required within each panel varies depending on the requisite expertise level and research objectives. Successful studies have been conducted with as few as three to five experts [201] and [194]. Victoria recommends a panel size ranging from 2 to 8 experts [199], whereas Mitchell suggests a minimum of 8 to 10 experts is essential [202]. Using the Delphi method, Phan involves 10 to 15 experts for each panel [168]. Consequently, in alignment with similar dissertations, this research will encompass 13 panels, each comprising 6 to 12 experts, to validate and quantify the research model [177], [178], [176], [187], [172].

The subsequent tables illustrate the composition of the expert panels and elucidate how individuals with varying backgrounds and areas of expertise will be allocated to each respective panel. This strategic approach intends to harness broad insights and knowledge, ultimately enriching the research process and its outcomes.

Table 10: Expert Panels

Panel	Role	Research Tool	Panel Size	Participated
Panel 1	Perspectives Validation	Qualtrics survey	≥ 6	20
Panel 2	Factor validation of economic foundations and research perspective	Qualtrics survey	≥ 6	20
Panel 3	Factor validation of data perspective	Qualtrics survey	≥ 6	20
Panel 4	Factor validation of feature perspectives	Qualtrics survey	≥ 6	20
Panel 5	Factor validation of modeling perspective	Qualtrics survey	≥ 6	20
Panel 6	Factor validation of performance perspective	Qualtrics survey	≥ 6	20
Panel 7	Quantification of perspectives	ETM HDM software+ Qualtrics survey	≥ 6	11
Panel 8	Quantification of economic foundations factors and related desirability curves	ETM HDM software + Qualtrics survey	≥ 6	10
Panel 9	Quantification of data factors and corresponding desirability curves	ETM HDM software + Qualtrics survey	≥ 6	10
Panel 10	Quantification of feature factors and related desirability curves	ETM HDM software + Qualtrics survey	≥ 6	13
Panel 11	Quantification of modeling factors and	ETM HDM software + Qualtrics survey	≥ 6	13

	corresponding desirability curves			
Panel 12	Quantification of performance factors and related desirability curves	ETM HDM software + Qualtrics survey	≥ 6	11

5.2.3 Experts Inconsistencies

Humans are biologically wired, so our brains heavily rely on heuristics in facing complex and multi-faceted decisions. Experts and their subjective opinions are no exception. At the same time, despite the robust nature of the human mind, we are not staggeringly good at evaluating complicated situations that require complex logical evaluations. This feature will most likely end up with inconsistent results in such situations.

Table 11: Expert Panels by Background

Panels	Financial data scientist	Quantitative finance researcher	Investment professional	Academic scientist	AI finance project managers	Quant investment teams - public employees
Panel 1	✓	✓	✓	✓	✓	✓

Panel 2	✓	✓	✓	✓	✓	✓
Panel 3	✓	✓	✓	✓	✓	✓
Panel 4	✓	✓	✓	✓	✓	✓
Panel 5	✓	✓	✓	✓	✓	✓
Panel 6	✓	✓	✓	✓	✓	✓
Panel 7	✓	✓	✓	✓	•	•
Panel 8	✓	✓	✓	✓	•	•
Panel 9	✓	✓	✓	✓	•	•
Panel 10	✓	✓	✓	✓	•	•
Panel 11	✓	✓	✓	✓	•	•
Panel 12	✓	✓	✓	✓	•	•

✓: will be included in a panel without any further consideration of criteria.

●: will only be considered in a panel if the background fits well.

There are different definitions of "inconsistency" in the literature. For instance, Estep (2017) has defined it as a "disagreement within an individual's evaluation." [193] This research will ask experts to conduct pairwise comparisons to evaluate the HDM model. Each expert will use judgment to choose the criteria that will impact the decision. In the context of this research, if we suppose there are five factors in the HDM model and the experts are asked to determine the relative importance of each one on the decision by using pairwise comparisons, given a scale of 0 to 100, each expert conducts this step of research as follows:

Factor 1: Economic foundations and research

Factor 2: Data

Factor 3: Features

Factor 4: Modeling

Factor 5: Performance

Pseudo expert, hypothetically, might evaluate the relative importance of these factors like this:

{Economic foundations and research 75:25 Data}, {Economic foundations and research 75:25 Features}, {Economic foundations and research 80:20 Modeling}, {Economic foundations and research 75:25 Performance}, {Data 75:25 Features}, {Data 50:50

Modeling}, {Data 60:40 Performance}, {Features 50:60 Modeling}, {Features 50:60 Performance}, {Modeling 50:50 Performance}

If we summarize the relative importance of these factors based on the given weights, this is what is logically selected:

Factor 1 > Factor 2 > Factor 4 > Factor 5 > Factor 3

For instance, chosen Economic foundations and research 75:25 Data means F1 is three times more important than F2. The above relationship between factors shows the rank or relative importance of factors and their degree of importance, which follows rational logic. To detect an expert's judgment inconsistency, it is worth looking at Abbas's (2016) definition of inconsistency:

"Inconsistency is a slight or gross, deliberate or unintentional error in the elicited pairwise judgment related to the rank order and mutual preference proportionality of alternatives." So, given this definition, the researcher is to assess the expert's judgments for inconsistency through two main perspectives. First, ensure logical maintenance of the chosen factors' order. Second, ensure logical consistency in the weighting used to represent each criterion's relative importance. Logical consistency in weighting means that if Factor 1 is three times more important than Factor 2, and Factor 2 is three times more important than Factor 3, then if the expert selects Factor 1 as only one time more important than Factor 3, it results in evident logical inconsistency in the chosen weights.

This inconsistency might appear relatively intuitive. However, experts are highly likely to make inconsistent judgments in the face of complex decision problems. Hence, measurement and assessment of expert judgment inconsistencies are vital to validate the credibility of the decision model. One might also ask how this evaluation should be measured and checked to ensure no inconsistency in expert responses. There are acceptable thresholds mentioned in the literature in which the consistency level can get the pass/no pass score to ensure the model is robust enough to use. [186], [193], [201], [212], [213] In the context of HDM modeling, the concept of standard deviation plays a critical role, and inconsistency calculation is essentially based on that. Inconsistency is the sum of the standard deviation of pairwise comparisons. Phan (2013) [195] has elegantly illustrated its equations as follows:

$r$ij: relative value of the ith, element in the jth, orientation for an expert.

i: mean relative value of the i+, element for that expert

Equation 1: Inconsistency Formula

$$\text{Inconsistency} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{n!}\sum_{j=1}^{n!}(\bar{r}_i - r_{ij})^2}$$

As it is crystal clear from the above formula, it is the same as a well-known and widely used statistical formula for calculating variance applied in HDM to measure the degree of variation or inconsistency for each expert's judgment. One logical question that might arise when using the inconsistency formula is the acceptable threshold or inconsistency value. According to some research papers, various methods have proposed acceptable values ranging from 0.0 to 0.1. However, the acceptable range depends on the mechanism used and the level of criticality of the decision being made. [199], [200] Another consideration is when there are some observed inconsistencies among collected responses from experts. If inconsistency surpasses the threshold, the researcher should instruct the expert to redo the assessment and ensure they fully understand what is asked for in the model evaluation. If the inconsistency persists, the specific expert must be removed from the expert population.

5.2.4 Experts Disagreement

Several methods have been proposed in the literature to measure and treat the disagreement among experts. However, before going through them, it is worth knowing how "disagreement" is generally defined. According to the Cambridge Dictionary, disagreement is "an argument or a situation in which people do not have the same opinion." In other terms, it is defined as "the act of disagreeing or the state of being at variance" in the Merriam-Webster dictionary. The variation is the concept that these definitions implicitly imply. This concept sheds light on the measurement and calculation of disagreement and notes that it is highly probable that it would be, in mathematical terms, a formula or equation associated with the variation.

One of the seminal research studies about the essence of expert disagreements is the work of Hammond entitled "Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice." [214], [215] he raised unique perspectives about the cognitive processes under which human judgments are shaped. He asked a fundamental question: what does half a century of research tell us about these processes and human judgment? Then he stated, “It tells us that there is little doubt that, at least in certain circumstances, (1) human cognition is not under our control, (2) we are not aware of our judgment and decision processes, and (3) our reports about those processes are not to be trusted.” He later asserted that classic explanations of expert disagreement, including incompetence, venality, and ideology, are inadequate.

Additionally, preferences or priorities aggregation in group decision-making methods has been extensively studied to characterize disagreement. In group decision-making, experts with different backgrounds provide subjective judgments that should be systematically aggregated to develop the overall picture of the objective and alternative priorities. However, as Li and Lu (2012) [205] have shown in their research, the subjects of disagreement dynamics among decision-makers, consensus measures, and the homogeneity of group preferences are new areas that have been started to be investigated. They also argued that experts’ disagreement (due to differences in viewpoints) can be tested rigorously by applying well-known statistical methods, including regression, multiplicative models, and variance component models.

Despite many proposed methods and techniques to identify and reduce disagreement among experts, they do not necessarily guarantee that disputes can be eliminated. Moreover, although a better decision environment and the means and sources of information can alleviate the level of disagreement, they do not provide any means to ensure that existing disagreement, although at a reduced level, will be gone. Furthermore, if the level of disagreement is beyond the acceptable threshold, there would be doubt about the robustness of the model and the credibility of its results. Hence, there is a need to put some protocols in the modeling process to ensure disagreement identification, measurement, and reduction. [187], [191], [192], [193], [214]

The HDM methodology used in this research is no exception. As group decision-making heavily relies on subject matter experts (SMEs) and their judgments, the researcher needs to address the disagreement level that might arise in the model. As disagreement essentially happens at the aggregated level, the primary source will be when experts conduct pairwise comparisons to determine the importance and weight of factors in the HDM model. Their subjective judgment will determine the degree of importance of each factor/criterion. In the context of HDM model disagreement measurement, several methods have been proposed in the literature to calculate the level of disagreement and its acceptable threshold. [192], [193], [205] Also, disagreement measures have been extensively researched, and many statistical methods that apply to HDM modeling have been suggested. However, this research calculates disagreement based on the Hierarchical Agglomerative Clustering (HAC) method that has been developed in the form of HDM software by the Engineering and Technology Management (ETM) Department at Portland State University, which is recognized as one of the Technology schools of thought. [191], [192] It is defined as below:

m: The number of experts, k= 1 … m

n: The number of decision elements, i=1 … n

$\overline{r}_{ik}$: The mean relative value of the ith element for kth expert

R_i: The group relative value of the ith element for m experts is

$$R_i = \frac{1}{m}\sum_{k=1}^{m} \bar{r}_{ik} \ for\ i = 1, 2, \ldots, n$$

Therefore, we calculate the level of disagreement among experts given the above-defined elements as follows:

Equation 2: Disagreement Formula

$$d = \sqrt{\frac{1}{n.m}\sum_{k=1}^{m}\sum_{i=1}^{n}(R_i - \bar{r}_{ik})^2}$$

In mathematical terms, this formula essentially calculates the distance or, in the context of this research, the level of dissimilarity or disagreement among the objects. As noted in the literature, d = 0.1 or 10% threshold is acceptable, meaning the model disagreement value cannot pass this acceptance level. If it exceeds this threshold, the researcher should address it to ensure the results are validated and, therefore, reliable.

The modeler can take different approaches to achieve the required level of disagreement in the model. At the individual level, the inconsistency value of an expert can potentially represent those experts who are causing the disagreement at an aggregated level. Once identified, they can be eliminated, or preventive actions can be taken to reduce inconsistency. Furthermore, shedding light on the reasons for their response differences through more in-depth discussion might result in rational reasonings that can be cascaded to the whole group. Further assessment of such reasons could also open perspectives for another expert in the group. So, these deviations should not always be treated as outliers and need more evaluation. Finally, as previous research shows and also reported in the ETM HDM tool in this research, statistical tests like the F-test can be applied to statistically test whether disagreement over 0.1 is acceptable. Such hypothesis testing and the considered significance level will determine the credibility of the disagreement value.

5.2.5 Sensitivity Analysis

One well-known and widely used method to evaluate the impact of potential changes in input values on the model outcomes is called “Sensitivity Analysis” (SA). Saltelli [216] defines SA as a “study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input.” Sensitivity analysis plays a critical role in model development and reporting its results. SA generates essential insights on model behavior at the aggregated level and shows the model's overall structure and how different input parameters impact the resulting outcomes. Therefore, this research uses this technique to analyze the potential impacts of HDM inputs on its results. [216], [217], [218], [219]

In HDM, decision components at each level have a local contribution to other levels, and hence, final decisions in ranking the decision alternatives rely on these local contributions. However, as Chen and Kocaoglu [219] argued, “values of the local contributions are seldom known at a 100% confidence level and are subject to variations as the environment changes. Besides, HDM's various pairwise comparison scales and judgment quantification techniques usually yield different local contribution values. Thus, different results for the same problem and various group-opinion combining methods may change the current decision.” Therefore, to ensure a robust model and understand how uncertainty might impact the final solution in the HDM decision model, we cannot merely consider the rank order of alternatives by experts as the definitive and complete solution.

We can highlight several benefits of conducting the SA for HDM as follows: (1) it helps depict the impact of changes of higher levels on lower levels of the hierarchy; (2) it helps identify the most impacting factors in the model; (3) check decision model robustness; (4) conduct what-if analysis to analyze different scenarios of the model. However, although there is an abundance of SA methods applied in operations research and management science models in the literature, most of them in the context of HDM has been the incremental change of input parameters and showing how the corresponding results change. This approach is namely called “numerical incremental analysis.” Additionally, a handful of studies employ simulation methods to evaluate HDM sensitivity. In this approach, researchers incorporate probability distributions into the model and determine the expected value of alternative ranks after hundreds of simulation iterations. This probabilistic approach introduces stochasticity into the model outputs, fundamentally altering the deterministic nature of the model and shedding light on more complex scenarios that could occur. Finally, mathematical deduction methods are utilized, in which closed-form solutions can explain the input-output relationships in the decision model. This technique has also demonstrated better performance and less computational complexity.

In this study, the researcher will use SA to analyze the impact of potential changes at higher levels of HDM on the lower levels and the decision alternatives. It will show how sensitive the model is to priority perturbations. Furthermore, SA helps to demonstrate the reasonable range of values for perspectives and criteria that will keep the assessment score at its original level. The HDM SA algorithm developed by Chen and Kocaoglu is used to achieve the SA assessment goals in this research. The following elements represent how this method is defined to address the overall contribution of each alternative to the mission in the HDM model:

Equation 3: Global and Local Criterion Contribution in a Sensitivity Analysis

C_l^{O-M}: Local contribution of the Lth objective to the mission

C_{kl}^{G-O}: Local contribution of the kth goal to the Lth objective

C_i^{A-M}: Overall contribution of ith alternative to the mission

C_{ik}^{A-G}: Local contribution of ith alternative to the Kth goal

C_{il}^{A-O}: Global contribution of ith alternative to the Lth objective

$$C_i^{A-M} = \sum_{l=1}^{L} \sum_{k=1}^{K} C_l^{O-M} \cdot C_{kl}^{G-O} \cdot C_{ik}^{A-G}$$

In the context of this research, the components are associated as follows:

- Mission: Readiness assessment score to implement ML/DS-based investment strategies
- Objectives: they are represented as “Perspectives” on the HDM model in this research
- Goals: they are demonstrated as factors/criteria impacting ML/DS-based investment strategies
- Alternatives: A Desirability Curve is used in this study; therefore, there will not be alternative ML/DS-based investment strategies, and they will be scored independently.

Overall, different scenarios will be analyzed using numerical incremental analysis to understand better the nature and impact of uncertainty in the model and the influence of each perspective on the scores of the ML/DS-based investment strategies.

5.3 Methodology Comparison with Other Methods

There are a wide variety of multi-criteria decision-making methods with their pros and cons. We have discussed below the main ones that are most relevant in the context of this research. They are all multi-criteria decision models (MCDM) that can evaluate many multi-faceted decision problems and provide invaluable insights.

TOPSIS: Technique for Order of Preference by Similarity, abbreviated as TOPSIS, is one of the methods classified in the multi-criteria decision aid (MCDA) or multi-criteria decision-making (MCDM) methodologies. In this method, a set of alternatives or choices is weighted for each criterion. Then, the normalized score for each criterion is compared with the ideal choice based on a mathematical distance formula. The perfect alternative has the best score in a specific criterion. So, the aggregated comparisons demonstrate how each criterion is close to the ideal solution and far from the worst alternative. Therefore, if we imagine a spectrum from a negative perfect solution to a positive ideal solution, this algorithm selects the solution with the shortest distance from the perfect solution and the longest distance from the negative perfect solution. In terms of applications, TOPSIS has been widely used in various areas, including business and marketing management, supply chain management, energy management, manufacturing, and human resources management. Overall, this algorithm/method searches the finite alternatives space to identify a solution extracted from the simultaneous minimization of the ideal solution and maximization of the negative ideal solution. [206], [220], [221], [222], [223]

PROMETHEE: Preference Ranking Organization Method for Enrichment Evaluations or PROMETHEE is a multi-criteria decision method developed in the early 1980s and has been widely studied and applied since then. In this method, the focus is not on finding the "right" decision; instead, it helps find the choice that best suits the goal in the context in which the decision problem is defined. This method has two main versions: PROMETHEE I and PROMETHEE II. The decision-maker receives a partial ranking; the latter is a complete ranking of the alternatives. The core concept in this method is called the decision or evaluation matrix. It is a matrix that shows all options and the criteria used to compare them against each other. The first step in normalizing the evaluation matrix is based on the Beneficial and Non-Beneficial criteria and their maximum and minimum values. Next, the decision-maker should calculate the difference between the alternatives relative to others. Then, the preference function and the aggregated preference should be calculated. Finally, the leaving, entering, and net outranking flows are calculated, and the alternatives are ranked based on net outranking flow. Overall. This method works under the concept of net flow indifference and preference thresholds. [224], [225], [226], [227] This method also has been extensively used in different settings.

AHP: Analytical Hierarchy Process (AHP) is a systematic and structural methodology for assessing and evaluating complex decision problems. Thomas L. Saaty developed this method in the 1970s. This technique is based on decision criteria quantification and expert judgments using pair-wise comparison to establish the relative importance of those criteria. It decomposes a complex decision through a hierarchical structure consisting of a primary goal/objective on top of the hierarchy, a set of factors or criteria that associate the alternatives to the overall goal, and a set of options/alternatives at the lowest level. If the decision problem is complex, one can drill down the hierarchy and create sub-levels to address the problem requirements. The AHP hierarchy design should consider the context of the decision problem and the expertise and knowledge of the people involved in designing the model. This method has been extensively researched and applied in Group Decision Making. It is structurally like HDM; however, AHP and HDM apply different weighting mechanisms in assigning and calculating weights. This process is the distinctive feature and the main point of departure from HDM methodology. This technique is used globally in various contexts, such as business, technology, government, and education, in terms of applications. [98], [227], [228], [229], [230]

ANP: The Analytic Network Process is a rather general format of the AHP method in multi-criteria decision-making. ANP and AHP break down the decision problem into objectives, criteria, and alternatives. The main difference between the two methods is the structure they use in this decomposition. AHP uses a hierarchy, while the ANP method structures the problem in a network form. One more difference between these methods is the concept of "independence." One requirement in the AHP process is that all components should be independent. For instance, alternatives are independent relative to each other and their criteria on a higher level. However, in ANP, elements can have interdependence, which is more suitable for real-world decision problems with relationships between elements. The literature shows that this method has been widely used in manufacturing, transportation, energy, healthcare, finance, and banking, to name a few. [199], [231], [232]

ELECTRE: elimination and choice translating reality or simply ELECTRE is a family of methods in multi-criteria decision analysis proposed by B. Roy in 1965, used in solving MCDM problems. This family of techniques focuses on MDCM problems with independent criteria. This method is famous for its outranking associations to rank a set of choices. Sometimes, it is only used to eliminate those undesired criteria and then apply other MCDM methods to organize the alternatives. However, one of the challenges with some versions of this method is dealing with interdependence and prioritization of criteria. In some real-world decision problems, we cannot simply ignore the concept of prioritization and its importance in weighting the criteria. [233], [234]

MAUT: Multi-Attribute Utility Theory is used in decision analysis and selection. This method's fundamental assumption is that a decision-maker chooses an alternative with the highest utility compared to other alternatives. An alternative is considered as a set of attributes. The decision-maker is supposed to evaluate each alternative based on its characteristics and relative importance. Finally, each alternative's aggregated utility is assigned based on attributes' values and importance weighting. The alternative with the highest utility score will be selected as the best choice. In general, MAUT is a beneficial method for quantifying and selecting multi-attribute alternatives based on their relative attractiveness. [235], [236], [237], [238], [239], [240]

RAND Corporation initially developed Delphi in the 1950s as a forecasting methodology to forecast the effect of technology on warfare. This technique is a process to collect group judgments and opinions through several rounds of questionnaires sent to a panel of experts. It aims to achieve the best-aggregated results based on the expert panel consensus. It is essentially a structured and iterative process to solve complex problems through group communication and accomplish an outcome that is better and more significant than individual ones. It is an effective technique, especially when there is incomplete information and knowledge about the decision problem. It is well suited to improve the problem understanding, find potential solutions and opportunities, or develop forecasts. It has been widely used in information systems research to study managerial decision-making. Overall, the main objective is to reduce the range of responses and develop an expert consensus, which is considered the best outcome. [207], [208], [209], [210], [241]

The following table briefly overviews the pros and cons of each illuminated method above.

CHAPTER 6. RESEARCH MODEL DEVELOPMENT AND RESULTS

6.1 The Initial HDM Model

Based on the literature review, this study's proposed HDM model builds upon an initial structure. The HDM comprises five high-level perspectives, each encompassing severalcriteria designated to assess the reliability of DS/ML-based investment strategies. These perspectives are economic background/foundations, data, features, modeling, and performance, as delineated below. Under each perspective, multiple criteria are directly linked to that high-level perspective. The figure below illustrates the entire model, which is underpinned by factors that could impact the results of investment strategies.

Table 12: Comparison of Methods

Name	Pros	Cons	Sample Reference
TOPSIS	Application of mathematical programming to aggregate the expert judgements	Lack of robust mechanism for weighting	[26]–[30]
	It is a heuristic method with less number of comparisons comparing to HDM	It is not evident what causes one criterion to be close to each side of the spectrum	
	Flexible on the number of criteria		
PROMETHEE	Applicable in a wide variety of settings	Concept of net flow is not always easy to follow for those who execute the method.	[31]–[34]
		No clear criteria contribution to the outcome	
		Preference thresholds makes the process more complicated	
AHP	Similar to HDM	The main point of departure from HDM is scoring schema of AHP which could be confusing for expert judgements in their evaluations.	[34]–[38]
ANP	Similar to HDM and AHP	Tracking dynamics is not always straightforward	[39]–[41]
	Elements can have interdependence	No much applicable to independent components	
ELECTRE	Able in outranking associations to rank a set of choices	challenge in dealing with interdependence and prioritization of criteria	[42], [43]
MAUT	Flexible in relative comparisons	Not able to properly address decision problems that contain the high level of uncertainity	[44]–[49]
	able to quantify and select multi-attribute alternatives based on their relative attractiveness		
Delphi	Divers group of people with different perspective add value to capture different insights.	Reaching consensus is not always straightforwad.	[50]–[54]
	The decision is based on the collective contribution of participants which reduces the level of bias.	Multiple rounds of iteration might be exhausting for participants which might lead to biases.	
	All people are involved in the process of decision making and coming up with consensus.	Lack of proper quantification mechanism	
		Requires a professional person to write and communicate clearly.	

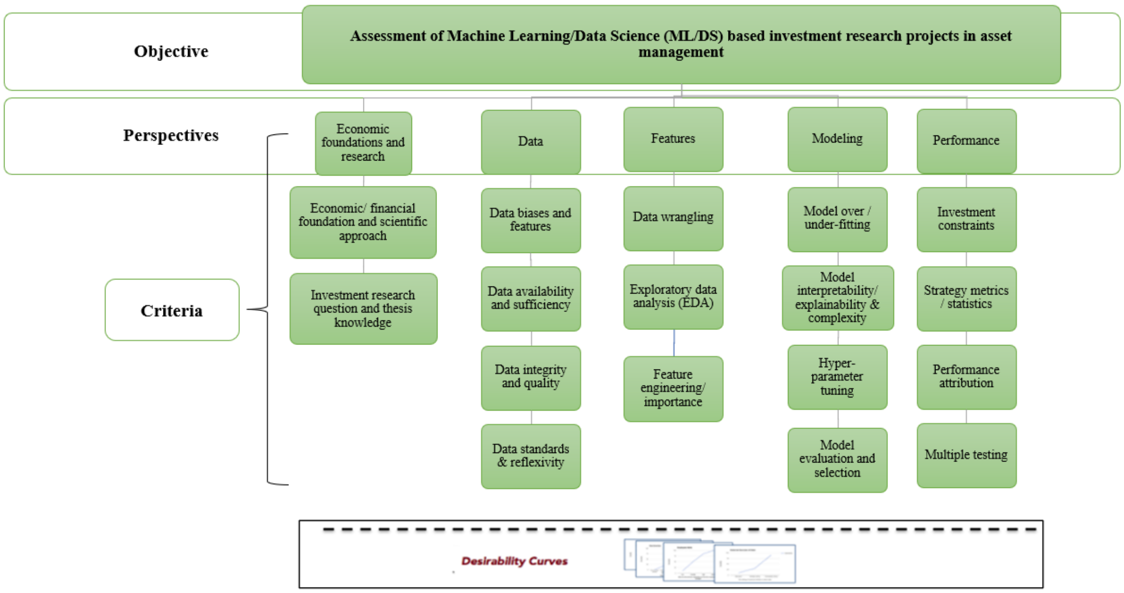


Figure 25: The Initial HDM

6.1.1 Model Perspectives

Table 13: HDM Model Perspectives

Perspective	Details
Economic Foundations Perspective	This perspective captures the basis for the financial and economic foundation of the strategy design. Factors such as the scientific basis of the strategy and investment thesis are under this perspective.
Data Perspective	This perspective includes data biases, availability and sufficiency, integrity and quality, and standards and reflexivity.
Features Perspective	The features perspective concentrates on how strategy addresses features in the process, including wrangling, exploratory data analysis, feature engineering, and importance.

Modeling Perspective	This perspective covers the modeling aspect of strategy research and development to shed light on factors such as model over and underfitting, interpretability of the model, hyper-parameter tuning, and model evaluation and selection.
Performance Perspective	This perspective includes the performance-related side of strategy development. It employs investment constraints, performance metrics and stats, attribution and decomposition of risk/return, and multiple testing.

6.1.2 Model Criteria

Table 14: HDM Model Criteria

Criteria	Details	References
Economic Foundations and Research Perspective		
Economic/ Financial Foundation and Scientific Approach	This factor carefully measures the structure of the machine learning problem to ensure that it is guided by scientific and economic thinking before conducting the research. Most quantitative models are based on three lines of thought: deductive, inductive, and simulation. Each approach requires a different model-building process. Deductive or hypothesis-based thinking starts with an initial idea or insight about an investment opportunity. Those initial thoughts come from an economic theory or mechanism about market functionality. The hypothesis precedes the empirical study in this approach. Secondly, the pattern-oriented or inductive approach is essentially exploratory and discovery-driven. In this data-driven approach, insights emerge from studying the data. Here, empirical analysis precedes the hypothesis and could be the source for generating theories inductively. As we dig deeper into the data, we learn more, and learning occurs throughout the process. Thirdly, simulation is a method for conducting virtual experiments. It can be used for both exploration and theory development. This approach utilizes a simulation model, a simplified and computational representation of the real-world phenomenon. Researchers can also compare the simulation results with empirical data.	[8],[11], [45], [55], [225]

Investment Research Question and Thesis Knowledge	This factor measures the investment team's level of clarity and awareness in terms of the research question, objectives, and investment thesis. Essential and intriguing research questions with economic/financial foundations and based on the scientific background are critical. In any scientific endeavor, the research project starts with underlying assumptions, questions with foundations in the literature, and strategies for the applied methodologies. The investment thesis is essential as it will be the basis for the data representing it and building signals.	[7], [45], [53], [157]

Criteria	Details	References
Data Perspective		
Data Biases and Features	Financial datasets could have some biases that investment teams should be aware of and check for the reliability of results. It includes survivorship bias, look-ahead bias, pre-processed alternative data, dividends and splits, and ticker updates; with increasing alternative data applications in investment management, controlling for such characteristics is becoming more critical. This includes identifying these biases and checking their existence to implement more robust backtesting results. This factor measures and contains the approaches taken to ensure they are considered in the investment strategy development process. This measure confirms that researchers have taken steps to ensure the usability and reliability of the dataset for further exploration.	[2], [9], [60], [61]
Data Availability and Sufficiency	Compared to other domains of knowledge, The data is very limited in scope in finance and investment. Given the limited standard financial data, the researchers and investment teams should address the small sample size problem when running machine learning models. "Today, we have about 55 years of high-quality equity data (or less than 700 monthly observations) for many of the metrics in each of the stocks we may wish to consider." This is a tiny sample, especially for deep learning models that are highly data intensive. This issue does not allow for many cross-validation runs on the data to minimize the risk of overfitting. This factor measures how the research team has approached the problem of data availability and limited data and which techniques,	[54], [60], [75], [76], [91], [157], [242]

	such as synthetic datasets data augmentation, have been used to address this problem.	
Data Integrity and Quality	High-quality data are essential for developing machine learning models. The increasing amount of data from multiple sources (for example, mobile and web applications) and vendors exacerbated the problem of data quality and integrity. Data quality and integrity should be addressed in developing investment strategies, as AI models heavily rely on the integrity and quality of data and can easily result in garbage-in and garbage-out problems. Furthermore, data quality and integrity are essential since most ML models are considered black-box, and interpreting results at face value is crucial. Developing machine learning models relies heavily on high-quality data. The increasing amount of data from multiple sources, such as mobile and web applications and vendors, has exacerbated the problem of data quality and integrity. This factor sheds light on this problem and ensures that the investment research team is aware of it and has implemented preventive protocols in the strategy development process. Various examples address this problem, including checking for missing data, data duplicates, and data heterogeneity.	[2], [51], [60], [243]
Data Standards Reflexivity	In the era of big data with high volume, variety, and velocity, the necessity of data standards is undeniable. Big data is a promising area for providing insights and discovery. However, researchers and practitioners should understand and standardize datasets based on the widely accepted and used financial data standard protocols to achieve this goal. Data standards and identifiers have significantly impacted the data collection, validation, and analysis steps. This factor measures the financial data standard health check level in the investment development process to ensure the associated data standard aspects are based on accepted global financial data regulations and standards. In addition, If the data is openly available and too many investors have access to it, the research team needs to consider it in the development process. The reason is that prices/returns in financial markets	[50], [51], [60], [244], [245] [6], [10], [54] [54], [246], [247], [248], [249]

	are reflective, which means if too many people are following and using the same data or strategy, then the value of that data in generating excess returns is doubtful. Because its value is already reflected in prices and unlike other natural phenomena that we collect the data and analyze, reality changes itself based on investors' actions in financial markets. This factor measures the ability and actions taken in the research process to anticipate the market's reflexivity and the value of the data used in the strategy. Moreover, the critical characteristic of financial markets is that prices quickly (returns) adapt to new information and make a profit from an overcrowded strategy or model harder. This means prices are reactive to the actions of market players, and these dynamics arbitrage away opportunities for newcomers to already crowded strategies. Also, the market is prone to structural changes that might break the models or strategies that have been performing well in the past. This criterion aims to demonstrate that investment researchers know of such unique phenomena and have proper approaches to address them. This factor is then a lens that shows and checks the probability of overcrowding and reflexivity and decreases the likelihood of failure.	

Criteria	Details	References
Features Perspective		
Data Wrangling	Data wrangling or munging has mostly been thought of as simply data cleaning in data science projects. However, other aspects in this phase, including combining data sources, reproducible processes, and controlling data provenance, should be considered. Also, given that data cleansing accounts for 50 to 80 percent of the time of analytics projects, the recognition of practical importance and applicable workflows are critical throughout the process. The whole point is that high-quality and reliable data wrangling can be accomplished systematically (i.e., data acquisition, data	[250], [251], [252], [253]

	unification, and data cleansing) and simultaneously be flexible enough to address the problem in different circumstances. An efficient process in data wrangling helps to know what data was gathered, when it was collected, what the location is, and why it was gathered in the first place. These 4Ws (what, when, where, why) help bring more clarifications to the data science project. Having such workflows and health checks provides more efficiency and helps reduce many sources of error contributing to different kinds of data problems. Therefore, these protocols result in higher standards and reliability in strategy development.	
Exploratory Data Analysis (EDA)	One of the critical steps in any data science/machine learning project is EDA. The main aim of EDA is to help researchers understand the data before making any formal assumptions. It also helps investment teams to ensure the results are produced and validated after EDA, which can be more reliable for further machine learning modeling. EDA is primarily a step before formal modeling and hypothesis testing. It sheds some light on the dataset characteristics, such as missing data, outlier detection, understanding variables and their types, and finding exciting associations among variables. This factor measures the level of completed EDA on the dataset before starting any machine learning modeling or hypothesis testing. This helps ensure the EDA steps in the strategy development process, which increases the probability of reliable results and team awareness about its importance.	[50], [51], [133], [192], [254]

Feature Engineering/Importance	Feature engineering and importance is a crucial step in the lifecycle of ML/DS projects. This step helps make the dataset proper for machine learning modeling and improves the model performance. Better features mean more flexibility, simpler models, and more reliable results. The feature engineering process transforms the raw dataset into applicable features that better represent the problem. Although some ML models, like deep learning algorithms, automatically manage this step, this step can help improve the predictive results for most machine learning models. This factor measures the actions taken and techniques used in feature construction, feature extraction, feature selection, and feature importance. Techniques in feature importance, for instance, can provide insight into the dataset and highlight the most relevant feature to the target variable.	[88], [91], [93], [102], [139]

Criteria	Details	References
Modeling Perspective		
Model Over/Underfitting	Model overfitting and underfitting are fundamental problems in using machine learning models. Researchers can train a wide range of models on the dataset with high accuracy, but the models fail when facing unseen data. This means these models do not generalize well from training data to test data. This is a critical issue, especially in financial markets with a low signal-to-noise ratio, and such overfitting leads to fitting the model to the noise rather than the signal. Hence, the trade-off between bias-variance (too simple vs. too complex) is a fundamental concept that should be addressed in running machine learning models on financial data. Overfitting detection	[56], [58], [129], [129], [152], [255]

	and prevention are two crucial building blocks in this step. Investment researchers who use machine learning models must provide reasonable solutions to ensure that the model is not overfitted. There are several techniques, to name a few, that they can use to decrease the risk of model overfitting: cross-validation, feature removal, early stopping, and regularization. This factor aims to demonstrate that this issue is considered in the investment strategy design and development, and appropriate techniques are implemented to ensure the model is not overfitted or underfitting. Additionally, researchers understand that valid out-of-sample testing is only possible when the model is tested in real-world and live trading.	
Model Interpretability/Explainability & Complexity	One of the main barriers to adopting machine learning models, especially in investments, is that algorithms do not explain their predictions. This has become incredibly important in deep learning models, considered black boxes. “interpretability is the degree to which a human can understand the cause of a decision… it is the degree to which a human can consistently predict the model’s result.” This is a critical feature for investors attributing specific strategies' performance to the constituting inputs. That shows how the model has made the decision that resulted in the particular outcome. This way, we can trust the model without asking why it has made a specific decision. This important factor measures the level of interpretability techniques used to	[113], [114], [115], [118], [149], [256] [30], [54], [60], [82]

	explain and interpret the model. In turn, it seeks to ensure researchers are aware of this critical factor and have applied suitable interpretation methods to shed light on the model performance outcomes. Some techniques are partial dependence plots (PDP), individual condition expectations, feature interaction, Permutation feature importance, Global surrogate, LIME, and Shapley values. According to Occam's razor, the more straightforward explanations (models) are better; among many models that make identical predictions, the simplest is preferred. This means researchers should have taken into consideration the principle of parsimony. Therefore, researchers should have taken steps to develop more parsimonious specifications in modeling. Given different modeling criteria, the appropriability of model parsimony consideration in designing the investment strategy should be considered.	
Hyper-Parameter Tuning	In designing machine learning models, researchers typically face design choices regarding the leading architecture of the model. This is particularly evident in the visual representation of deep learning models. The parameters defining the model architecture are called hyper-parameters. Deciding on the optimal architecture is crucial. Although there is no straightforward way to determine the optimal architecture for the model, some techniques and methods can help. Grid search, Random search, and Bayesian optimization are some examples. This factor evaluates how researchers	[10], [257], [258], [259]

	clarify hyper-parameter tuning and which systematic methods have been used to find the optimal model structure.	
Model Evaluation and Selection	Given ever-increasing machine learning models and accessible libraries, choosing a range of models best suited for the problem is challenging. With the emergence of AutoML platforms, researchers can choose from an extensive collection of models. The process of selecting a final model is called model selection. There are several methods that researchers can apply to have a better model selection approach for a specific problem. (examples: Random splitting, Bootstrap) Also, researchers must evaluate the model results based on well-established performance measurements. Such metrics should address the correctness of the model on test data. Depending on the type of problem, whether it is regression, classification, or clustering, researchers need to use the proper set of output metrics to evaluate the model results. Such metrics include confusion matrix, accuracy, precision, recall, F1 score, AUC-ROC, MSE, RMSE, R-squared, MAE, adjusted R-squared, and learning curves. This factor measures the level at which researchers have paid attention to this step and the approaches and methods they have used in selecting and evaluating strategy models.	[138], [155], [255], [257], [260], [261]

Criteria	Details	References
Performance Perspective		
Investment Constraints	Even with the best-trained model and a promising strategy in hand, researchers should still consider other investment considerations in their analysis and evaluate their impact on the performance of the model or strategy. These considerations include transaction costs, market impact, liquidity, short selling, leverage, and strategy turnover. These constraints could negatively impact the strategy performance in live trading. For instance, an investment team may find a profitable ML strategy based only on trading over a few stocks with liquidity problems. This illiquidity could have a very negative effect on the strategy results as they are not readily tradable. This criterion measures how researchers know about such practical issues and which systematic actions exist.	[10], [262], [263]
Strategy Metrics/Statistics	Regardless of the strategy type that the researcher chooses, there are investment strategy performance metrics that he should use in reporting the results. These measurements will provide a systematic way for investors to compare competing strategies within and across kinds of strategies. These metrics are mostly coming from the investment field. Sample metrics are turnover, holding period, drawdown, and annualized return. The standard investment performance framework is GIPS (Global Investment Performance Standards), which includes some of these metrics. However, these measurements should be considered with model measures mentioned in model selection and evaluation criteria to provide a more comprehensive picture of strategy performance. This factor measures the level and quality of acquired and reported strategy performance results.	[10], [44], [53], [54], [132], [132], [264]
Performance Attribution	Performance attribution aims to explain and identify the sources of excess return. It is essentially decomposing the performance based on the well-documented risk factors. Attribution techniques break down the performance into several components, a good representation of how much performance is attributed to systematic and	[153], [170], [265], [266]

	unsystematic risks. Factor attribution shows the amount of performance that comes from exposure to common risk factors and that component from the manager's skills (alpha). This factor measures and evaluates the performance that is explainable by exposure to common risk factors, generated alpha, and stock selection capabilities. Three generally considered forms of attribution are multi-factor analysis, style analysis, and return decomposition analysis.	
Multiple Testing	Running the backtest often and selecting and reporting good results are the main reasons for fund failures. This procedure is called selection bias under multiple testing. If we only run one hypothesis, there is a slight chance of getting a significant result. However, if we run that test a thousand times, we dramatically increase the chance of getting false discoveries. This approach has a long history in statistics, and many scientists have already raised the flag in reporting false discoveries. It has been a critical problem in reported investment strategies, and journals are full of such strategies. This factor tries to address and measure this issue and ensure that researchers are aware of such an issue, scientific methods are used, and all trials are reported.	[53], [53], [54], [126], [128]

6.2 MODEL VALIDATION AND QUANTIFICATION

6.2.1 HDM Model Validation

The experts validated the initial HDM model using a Qualtrics survey in Appendix A. In this analysis, experts went through all criteria identified by the literature review as the most critical factors affecting the success or failure of ML/DS-based investment strategies. Furthermore, experts were allowed to suggest additional factors based on their knowledge and experience. Based on the survey results, all criteria were recognized as essential and approved by participants who collaborated with the expert panel.

Table 15: The Expert Panels' Roles in The Validation Phase

Panel	Role	Tool	Size
Panel 1	Perspectives Validation	Qualtrics Survey	20
Panel 2	Factor validation of economic foundations and research perspective	Qualtrics Survey	20
Panel 3	Factor validation of data perspective	Qualtrics Survey	20
Panel 4	Factor validation of feature perspectives	Qualtrics Survey	20
Panel 5	Factor validation of modeling perspective	Qualtrics Survey	20
Panel 6	Factor validation of performance perspective	Qualtrics Survey	20

6.2.1.1 Pre-Validation HDM Model

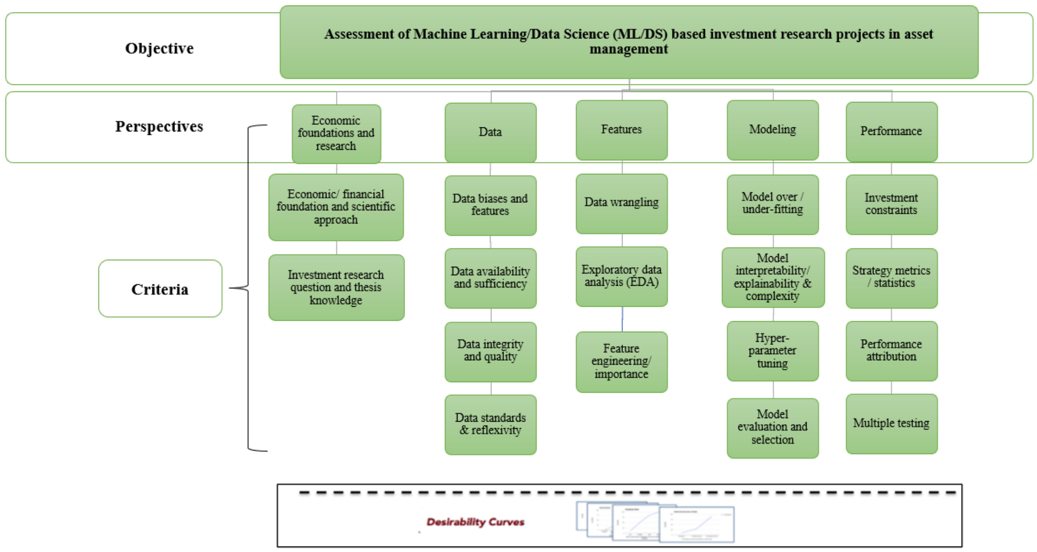


Figure 26: The Pre-Validation HDM Model

6.2.1.2 Perspective Level Validation

Panel 1 comprised 20 experts, all unanimously concurred on the significance of data, features, and modeling. In comparison, 19 experts endorsed the importance of economic foundations and performance perspectives in the research and development of DS/ML-driven investment strategies. To assess the validity of these perspectives, experts in Panel P1 uniformly applied a threshold of 67% for approval, validating all perspectives with a validation rate surpassing the 67% benchmark level. Figure 32 and Tables 9 and 10 below depict the validation results and their corresponding details.

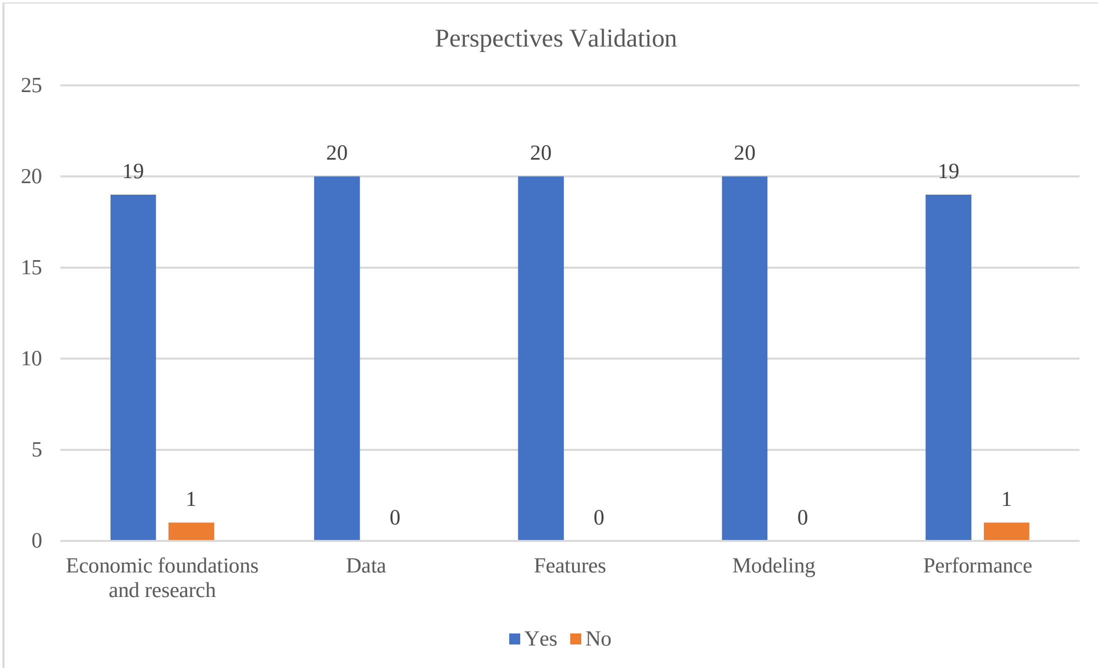


Figure 27: Perspectives Validation

Table 16: Perspective's Validation Summary

Perspectives	# Experts	Answered?	Yes	No	Validation %
Economic foundations and research	20	20	19	1	95%
Data	20	20	20	0	100%
Features	20	20	20	0	100%
Modeling	20	20	20	0	100%
Performance	20	20	19	1	95%

Table 17: Detailed Perspectives Validation

Panel	Economic foundations and research	Data	Features	Modeling	Performance
Expert 1	Yes	Yes	Yes	Yes	Yes
Expert 2	Yes	Yes	Yes	Yes	Yes
Expert 3	Yes	Yes	Yes	Yes	Yes
Expert 4	Yes	Yes	Yes	Yes	Yes
Expert 5	Yes	Yes	Yes	Yes	Yes
Expert 6	Yes	Yes	Yes	Yes	Yes

Expert 7	Yes	Yes	Yes	Yes	No
Expert 8	Yes	Yes	Yes	Yes	Yes
Expert 9	Yes	Yes	Yes	Yes	Yes
Expert 10	Yes	Yes	Yes	Yes	Yes
Expert 11	No	Yes	Yes	Yes	Yes
Expert 12	Yes	Yes	Yes	Yes	Yes
Expert 13	Yes	Yes	Yes	Yes	Yes
Expert 14	Yes	Yes	Yes	Yes	Yes
Expert 15	Yes	Yes	Yes	Yes	Yes
Expert 16	Yes	Yes	Yes	Yes	Yes
Expert 17	Yes	Yes	Yes	Yes	Yes
Expert 18	Yes	Yes	Yes	Yes	Yes
Expert 19	Yes	Yes	Yes	Yes	Yes
Expert 20	Yes	Yes	Yes	Yes	Yes

6.2.1.3 Economic Foundations and Research Validation

Most experts approved all criteria under the Economic foundations and research. The following two tables illustrate the summary of responses and their details.

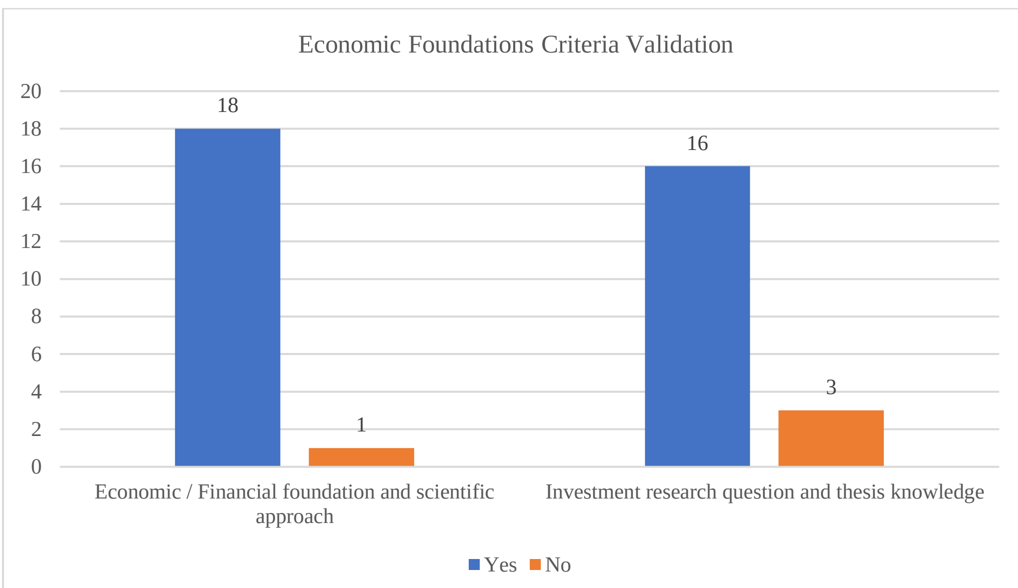


Figure 28: Economic Foundations and Research Validation

Table 18: Economic Foundations and Research Criteria Summary

Perspective	Criterion	# Experts	Answered?	Yes	No	Validation %
Economic foundations and research	Economic / Financial foundation and scientific approach	20	19	18	1	95%
	Investment research question and thesis knowledge	20	19	16	3	84%

Table 19: Detailed Summary of Economic Foundations and Research Criteria

Panel	Economic / Financial Foundation and Scientific Approach	Investment Research Question and Thesis Knowledge
Expert 1	Yes	Yes
Expert 2	Yes	Yes
Expert 3	Yes	Yes
Expert 4	Yes	Yes
Expert 6	Yes	Yes
Expert 7	Yes	No
Expert 8	Yes	Yes
Expert 9	Yes	Yes
Expert 10	Yes	Yes
Expert 11	No	No
Expert 12	Yes	No
Expert 13	Yes	Yes
Expert 14	Yes	Yes
Expert 15	Yes	Yes
Expert 16	Yes	Yes
Expert 17	Yes	Yes
Expert 18	Yes	Yes
Expert 19	Yes	Yes
Expert 20	Yes	Yes

6.2.1.4 Data Validation

Most experts in the panel approved all factors falling under the Data perspective. A summary of the results is presented in the following tables.

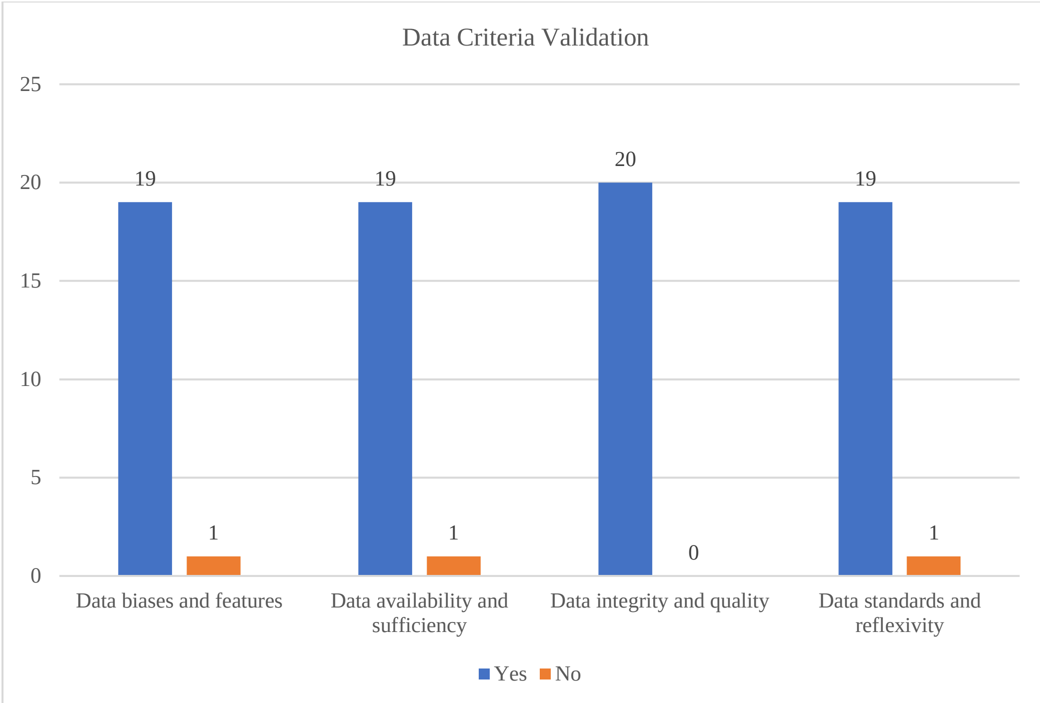


Figure 29: DataValidation

Table 20: Data Criteria Summary

Perspective	Criterion	# Experts	Answered?	Yes	No	Validation %
Data	Data biases and features	20	20	19	1	95%
	Data availability and sufficiency	20	20	19	1	95%
	Data integrity and quality	20	20	20	0	100%
	Data standards and reflexivity	20	20	19	1	95%

Table 21: Detailed Summary of Data Criteria

Panel	Data Bases and Features	Data Availability and Sufficiency	Data Integrity and Quality	Data Standards and Reflexivity
Expert 1	Yes	Yes	Yes	Yes
Expert 2	Yes	Yes	Yes	Yes
Expert 3	Yes	Yes	Yes	Yes
Expert 4	Yes	No	Yes	Yes
Expert 5	Yes	Yes	Yes	Yes
Expert 6	Yes	Yes	Yes	Yes
Expert 7	Yes	Yes	Yes	Yes

Expert 8	Yes	Yes	Yes	Yes
Expert 9	Yes	Yes	Yes	No
Expert 10	Yes	Yes	Yes	Yes
Expert 11	No	Yes	Yes	Yes
Expert 12	Yes	Yes	Yes	Yes
Expert 13	Yes	Yes	Yes	Yes
Expert 14	Yes	Yes	Yes	Yes
Expert 15	Yes	Yes	Yes	Yes
Expert 16	Yes	Yes	Yes	Yes
Expert 17	Yes	Yes	Yes	Yes
Expert 18	Yes	Yes	Yes	Yes
Expert 19	Yes	Yes	Yes	Yes
Expert 20	Yes	Yes	Yes	Yes

6.2.1.5 Features Validation

The experts in the panel approved all factors under the Features perspective, as demonstrated in the tables below:

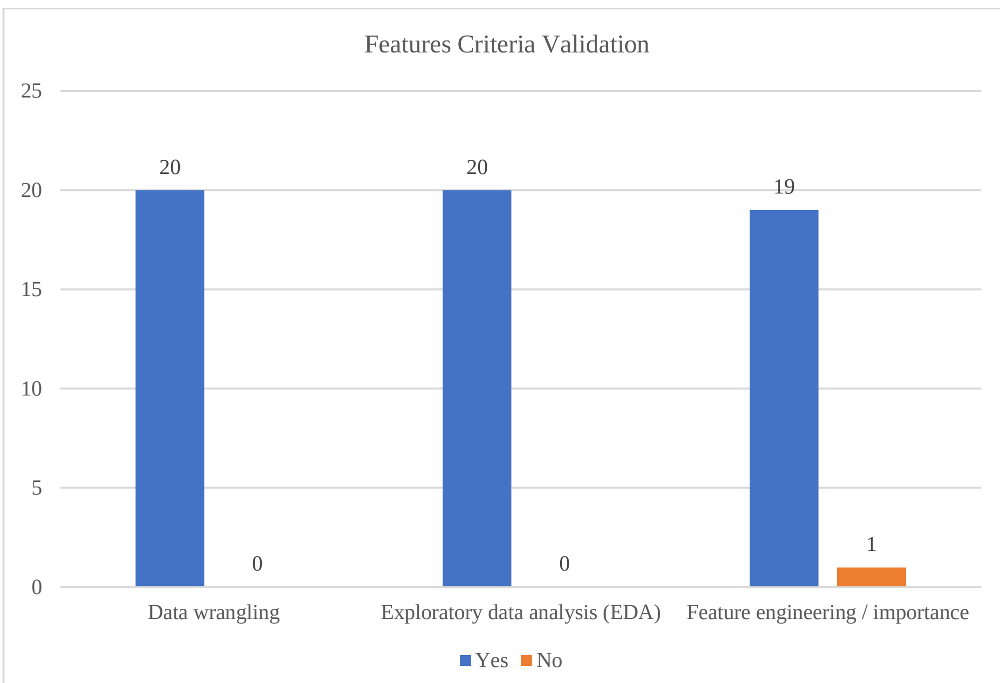


Figure 30: Feature Validation

Table 22: Feature Criteria Summary

Perspective	Criterion	# Experts	Answered?	Yes	No	Validation %
Features	Data wrangling	20	20	20	0	100%
	Exploratory data analysis (EDA)	20	20	20	0	100%
	Feature engineering/Importance	20	20	19	1	95%

Table 23: Detailed Summary of Features Criteria

Panel	Data Wrangling	Exploratory Data Analysis (EDA)	Feature Engineering/Importance
Expert 1	Yes	Yes	Yes
Expert 2	Yes	Yes	Yes
Expert 3	Yes	Yes	Yes
Expert 4	Yes	Yes	Yes
Expert 5	Yes	Yes	Yes
Expert 6	Yes	Yes	Yes
Expert 7	Yes	Yes	No
Expert 8	Yes	Yes	Yes
Expert 9	Yes	Yes	Yes
Expert 10	Yes	Yes	Yes
Expert 11	Yes	Yes	Yes
Expert 12	Yes	Yes	Yes
Expert 13	Yes	Yes	Yes
Expert 14	Yes	Yes	Yes
Expert 15	Yes	Yes	Yes
Expert 16	Yes	Yes	Yes
Expert 17	Yes	Yes	Yes
Expert 18	Yes	Yes	Yes
Expert 19	Yes	Yes	Yes
Expert 20	Yes	Yes	Yes

6.2.1.6 Modeling Validation

Most experts participating in the model validation phase approved all factors under the Modeling perspective.

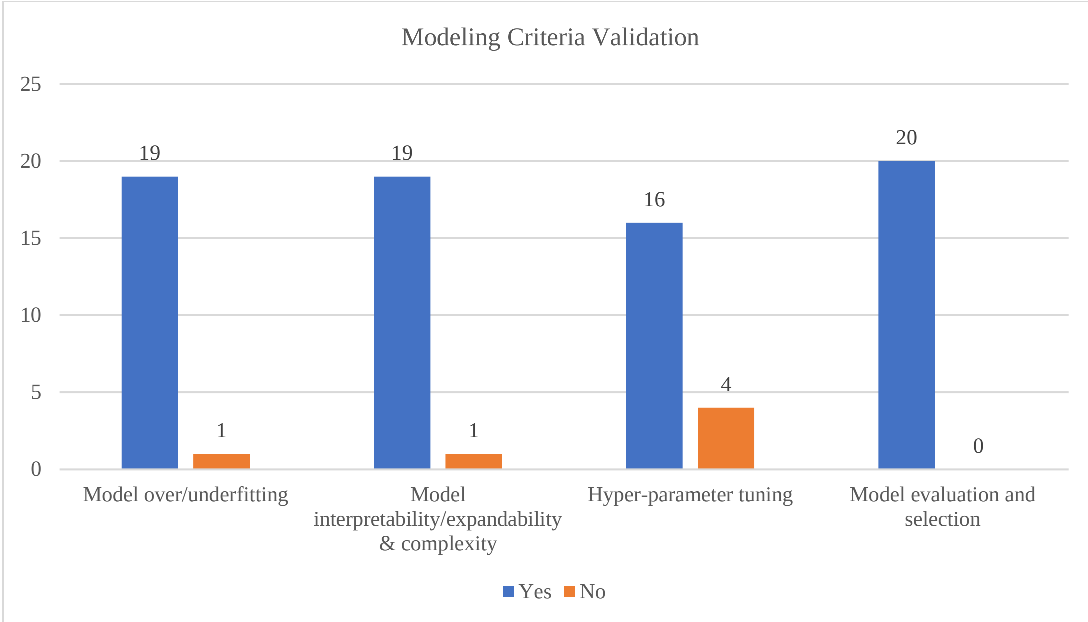


Figure 31: Modeling Validation

Table 24: Modeling Criteria Summary

Perspective	Criterion	# Experts	Answered?	Yes	No	Validation %
Modeling	Model over/underfitting	20	20	19	1	95%
	Model interpretability/expandability & complexity	20	20	19	1	95%
	Hyper-parameter tuning	20	20	16	4	80%
	Model evaluation and selection	20	20	20	0	100%

Table 25: Detailed Summary of Modeling Criteria

Panel	Model Over/Underfitting	Model Interpretability/Expandability & Complexity	Hyper-Parameter Tuning	Model Evaluation and Selection
Expert 1	Yes	Yes	Yes	Yes
Expert 2	Yes	Yes	Yes	Yes
Expert 3	Yes	Yes	Yes	Yes
Expert 4	Yes	Yes	Yes	Yes
Expert 5	Yes	Yes	Yes	Yes

Expert 6	Yes	Yes	Yes	Yes
Expert 7	Yes	Yes	No	Yes
Expert 8	Yes	Yes	Yes	Yes
Expert 9	Yes	No	Yes	Yes
Expert 10	Yes	Yes	Yes	Yes
Expert 11	Yes	Yes	Yes	Yes
Expert 12	Yes	Yes	Yes	Yes
Expert 13	Yes	Yes	No	Yes
Expert 14	Yes	Yes	Yes	Yes
Expert 15	Yes	Yes	Yes	Yes
Expert 16	Yes	Yes	Yes	Yes
Expert 17	No	Yes	Yes	Yes
Expert 18	Yes	Yes	Yes	Yes
Expert 19	Yes	Yes	No	Yes
Expert 20	Yes	Yes	No	Yes

6.2.1.7 Performance Validation

Most experts in the group approved all factors under the Performance perspective. The following tables show their judgment on HDM Performance criteria.

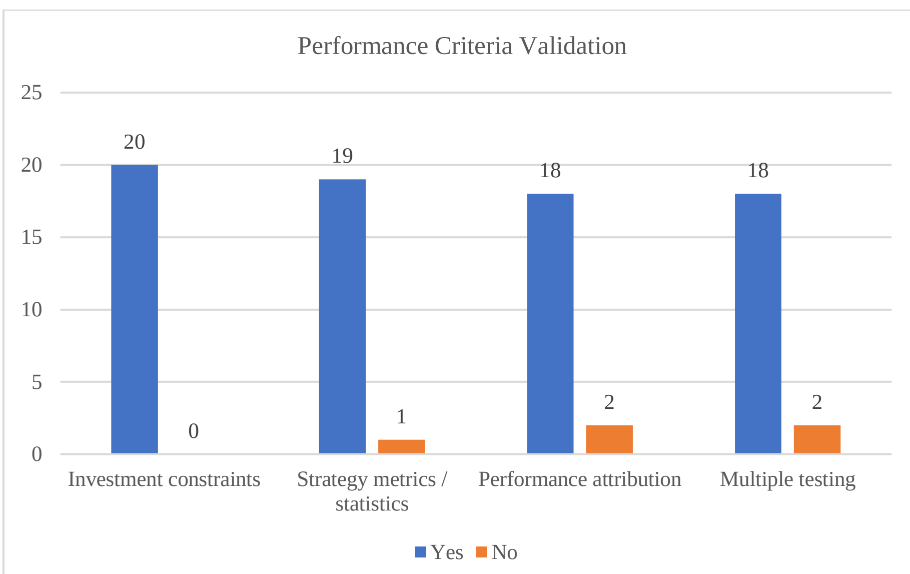


Figure 32: Performance Validation

Table 26: Performance Criteria Summary

Perspective	Criterion	# Experts	Answered?	Yes	No	Validation %
Performance	Investment constraints	20	20	20	0	100%
	Strategy metrics/statistics	20	20	19	1	95%
	Performance attribution	20	20	18	2	90%
	Multiple testing	20	20	18	2	90%

Table 27: Detailed Summary of Performance Criteria

Panel	Investment Constraints	Strategy Metrics/Statistics	Performance Attribution	Multiple Testing
Expert 1	Yes	Yes	Yes	Yes
Expert 2	Yes	Yes	Yes	Yes
Expert 3	Yes	Yes	Yes	Yes
Expert 4	Yes	Yes	Yes	Yes
Expert 5	Yes	Yes	Yes	Yes
Expert 6	Yes	Yes	Yes	Yes
Expert 7	Yes	No	No	No
Expert 8	Yes	Yes	Yes	Yes
Expert 9	Yes	Yes	Yes	No
Expert 10	Yes	Yes	Yes	Yes
Expert 11	Yes	Yes	Yes	Yes
Expert 12	Yes	Yes	Yes	Yes
Expert 13	Yes	Yes	Yes	Yes
Expert 14	Yes	Yes	Yes	Yes
Expert 15	Yes	Yes	Yes	Yes
Expert 16	Yes	Yes	Yes	Yes
Expert 17	Yes	Yes	Yes	Yes
Expert 18	Yes	Yes	Yes	Yes
Expert 19	Yes	Yes	No	Yes
Expert 20	Yes	Yes	Yes	Yes

Table 28: The Final HDM Model Validation

Perspectives and Criteria				
Perspectives		Criterion	Validation %	67% Threshold Check
Economic Foundations and Research	95%	Economic / Financial foundation and scientific approach	95%	Pass
		Investment research question and thesis knowledge	84%	Pass
Data	100%	Data biases and features	95%	Pass
		Data availability and sufficiency	95%	Pass
		Data integrity and quality	100%	Pass
		Data standards and reflexivity	95%	Pass
Features	100%	Data wrangling	100%	Pass
		Exploratory data analysis (EDA)	100%	Pass
		Feature engineering/Importance	95%	Pass
Modeling	100%	Model over/underfitting	95%	Pass
		Model interpretability/expandability & complexity	95%	Pass
		Hyper-parameter tuning	80%	Pass
		Model evaluation and selection	100%	Pass
Performance	95%	Investment constraints	100%	Pass
		Strategy metrics/statistics	95%	Pass
		Performance attribution	90%	Pass
		Multiple testing	90%	Pass

Since most experts have approved all perspectives, surpassing the 67% threshold, there is no need for further model refinements. Consequently, the post-validation HDM model presented will be utilized in the subsequent quantification steps in this research.

6.2.1.8 Post-Validation HDM Model

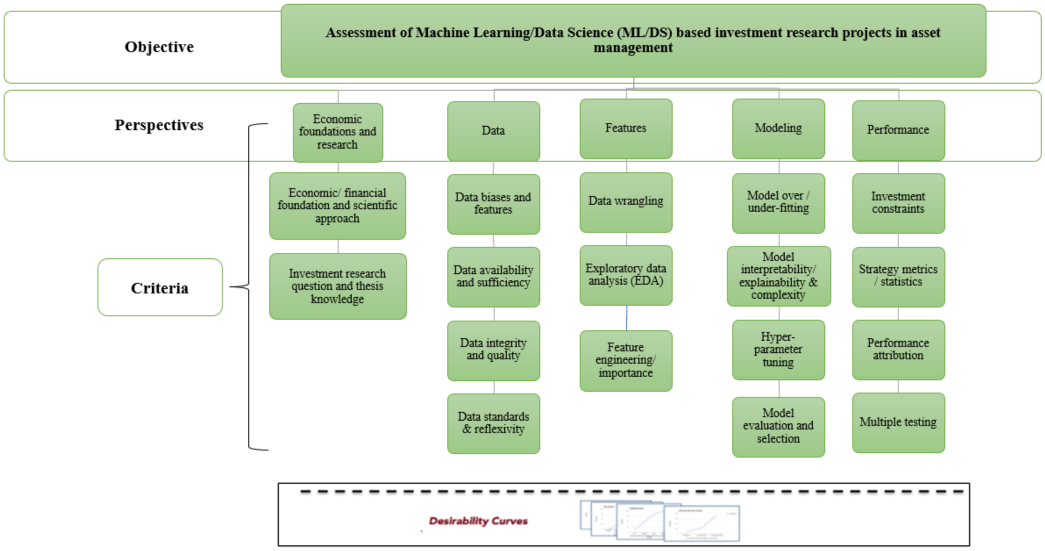


Figure 33: The Post-Validation HDM Model

6.2.2 HDM Model Quantification

During the quantification phase, 20 experts actively participated, distributed across six panels as delineated below. The data was collected and analyzed via HDM software. Below, we will find two tables elucidating the roles of the various expert panels and a list of experts who participated in this phase. In this implementation phase, panel experts quantify the HDM model perspectives and criteria.

Table 29: The Expert Panels' Roles in The Quantification Phase

Panel	Role	Tool	Size
P7	Quantification of perspectives	ETM HDM software + Qualtrics survey	11
P8	Quantification of economic foundations factors and related desirability curves	ETM HDM software + Qualtrics survey	10

P9	Quantification of data factors and related desirability curves	ETM HDM software + Qualtrics survey	10
P10	Quantification of feature factors and related desirability curves	ETM HDM software + Qualtrics survey	13
P 11	Quantification of modeling factors and related desirability curves	ETM HDM software + Qualtrics survey	13
P12	Quantification of performance factors and related desirability curves	ETM HDM software + Qualtrics survey	11

6.2.2.1 Perspective Level Quantification

A panel of 11 experts ranked the relative importance of each perspective in comparison to other perspectives in the model. Figure 39 is a graphical depiction of the pairwise comparison results, and Table 24 shows the details.

Table 30: Experts' Distribution Across The Quantification Panels

No.	Title	P7	P8	P9	P10	P11	P12
Expert 1	Founder Hedge Fund Manager	X	X	X	X	X	
Expert 3	Director of Investment Manager Research		X				X
Expert 4	Associate Professor of Finance and Data Science	X	X	X	X	X	X
Expert 6	Principal Data Scientist	X	X		X	X	X
Expert 7	Hedge Fund Manager	X				X	X
Expert 8	Qualitative Model Validation Researcher	X	X		X	X	X
Expert 9	Systematic Trader	X		X	X	X	X
Expert 10	PhD Data Scientist			X	X	X	
Expert 11	PhD Data Scientist			X			

Expert 12	Financial Data Scientist	X	X		X	X	X
Expert 13	Data scientist/QR - Investment Research	X	X	X	X	X	X
Expert 14	PhD Data Scientist			X	X	X	
Expert 15	Machine Learning Scientist			X	X	X	
Expert 16	Computer Science Professor	X	X	X	X	X	X
Expert 17	PhD Portfolio Optimization & Machine Learning Researcher	X	X		X	X	X
Expert 19	Founder and Quantitative Researcher	X	X	X	X		X
Total		11	10	10	13	13	11

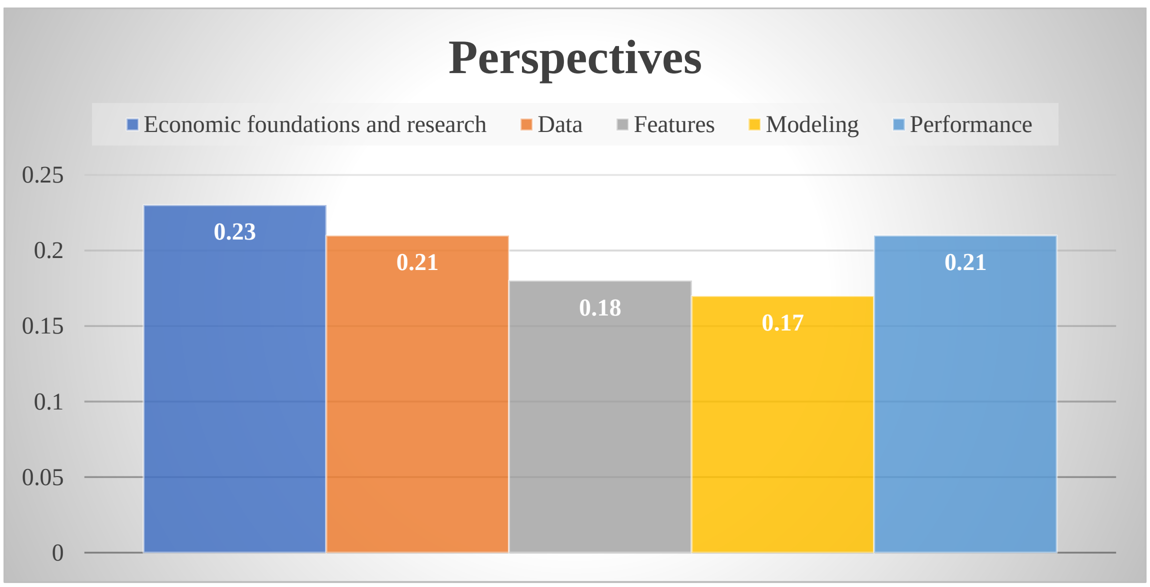


Figure 34: Relative Weight of Perspectives in Descending Order

Table 31: Perspectives Detailed Summary

Perspectives						
Panel	Economic Foundations and Research	Data	Features	Modeling	Performance	Inconsistency

Expert 6	0.2	0.2	0.2	0.2	0.2	0
Expert 1	0.2	0.2	0.2	0.2	0.2	0
Expert 4	0.23	0.25	0.23	0.18	0.11	0.01
Expert 17	0.33	0.14	0.09	0.14	0.3	0.05
Expert 16	0.27	0.2	0.14	0.1	0.29	0.02
Expert 8	0.29	0.23	0.15	0.14	0.19	0.03
Expert 12	0.11	0.21	0.23	0.19	0.25	0.01
Expert 19	0.27	0.28	0.15	0.18	0.11	0.01
Expert 13	0.21	0.23	0.21	0.18	0.17	0.01
Expert 9	0.24	0.15	0.24	0.14	0.23	0.02
Expert 7	0.15	0.24	0.16	0.23	0.22	0.01
Mean	0.23	0.21	0.18	0.17	0.21	
Minimum	0.11	0.14	0.09	0.1	0.11	
Maximum	0.33	0.28	0.24	0.23	0.3	
Std. Deviation	0.06	0.04	0.04	0.04	0.06	
Disagreement						0.045

6.2.2.2 Economic Foundations and Research Quantification

The panel of 10 experts ranked the relative importance of criteria under the economic foundations and research perspective. Figure 40 is a graphical illustration of the pairwise comparison results, and Table 25 shows the details.

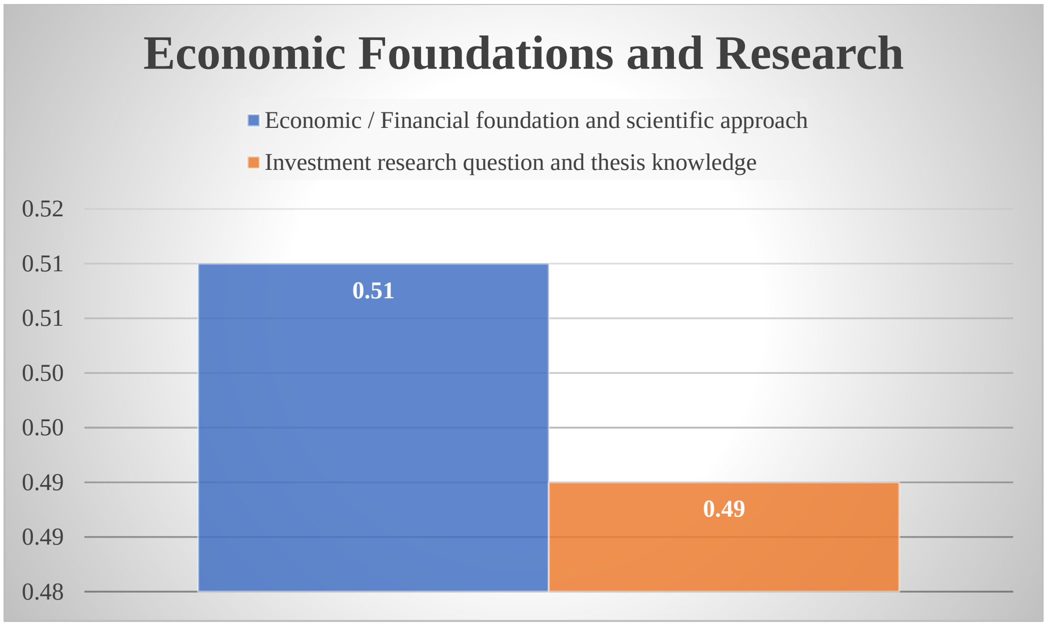


Figure 35: Relative Weight of Factors in Descending Order

Table 32: Economic Foundations and Research Factors Detailed Summary

Economic foundations and research			
Panel	Economic/ financial foundation and scientific approach	Investment research question and thesis knowledge	Inconsistency
Expert 6	0.5	0.5	0
Expert 1	0.5	0.5	0
Expert 3	0.55	0.45	0
Expert 4	0.4	0.6	0
Expert 17	0.5	0.5	0
Expert 16	0.5	0.5	0
Expert 3	0.5	0.5	0
Expert 8	0.72	0.28	0
Expert 12	0.4	0.6	0
Expert 19	0.5	0.5	0
Expert 13	0.51	0.49	0
Mean	0.4	0.28	
Minimum	0.72	0.6	
Maximum	0.08	0.08	
Std. Deviation	0.5	0.5	
Disagreement			0.051

6.2.2.3 Data Quantification

The panel of 10 experts ranked the relative importance of criteria under the Data perspective. Figure 41 is a graphical illustration of the pairwise comparison results, and Table 26 shows the details.

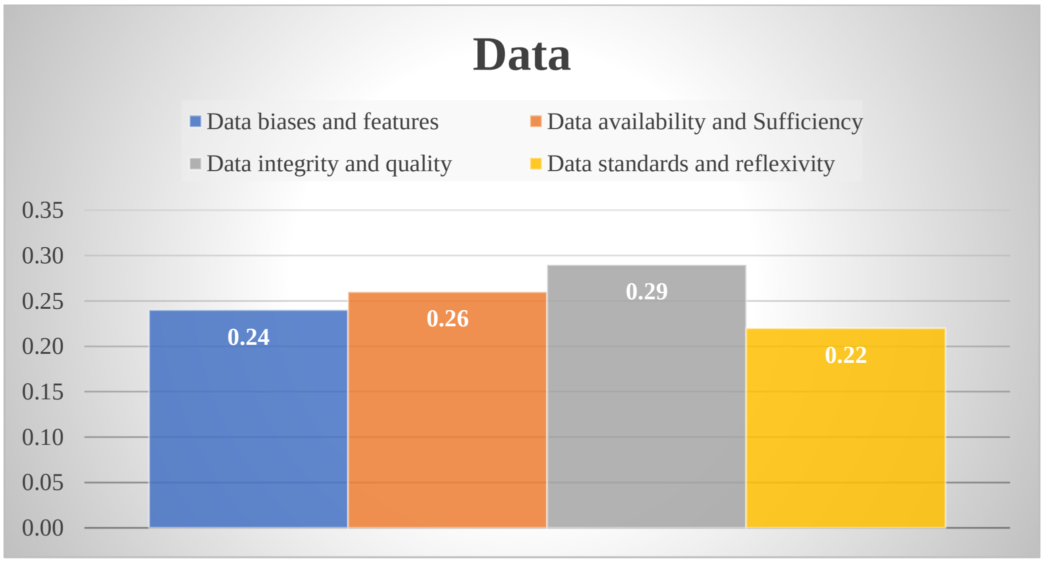


Figure 36: Relative Weight of Factors in Descending Order

Table 33: Data factors detailed summary

Panel	Data biases and features	Data availability and Sufficiency	Data integrity and quality	Data standards and reflexivity	Inconsistency
Expert 14	0.21	0.23	0.34	0.22	0
Expert 1	0.25	0.25	0.25	0.25	0
Expert 4	0.26	0.25	0.28	0.21	0
Expert 10	0.27	0.21	0.35	0.18	0
Expert 16	0.18	0.33	0.27	0.22	0.01
Expert 11	0.16	0.34	0.3	0.21	0
Expert 19	0.25	0.25	0.25	0.25	0
Expert 13	0.22	0.18	0.37	0.24	0.02
Expert 9	0.36	0.27	0.22	0.15	0.03
Expert 15	0.25	0.25	0.25	0.25	0
Mean	0.24	0.26	0.29	0.22	
Minimum	0.16	0.18	0.22	0.15	
Maximum	0.36	0.34	0.37	0.25	

Std. Deviation	0.05	0.05	0.05	0.03	
Disagreement					0.041

6.2.2.4 Features Quantification

Thirteen experts ranked the relative importance of criteria under the Features perspective. Figure 42 is a graphical presentation of the pairwise comparison, and Table 27 shows the details.

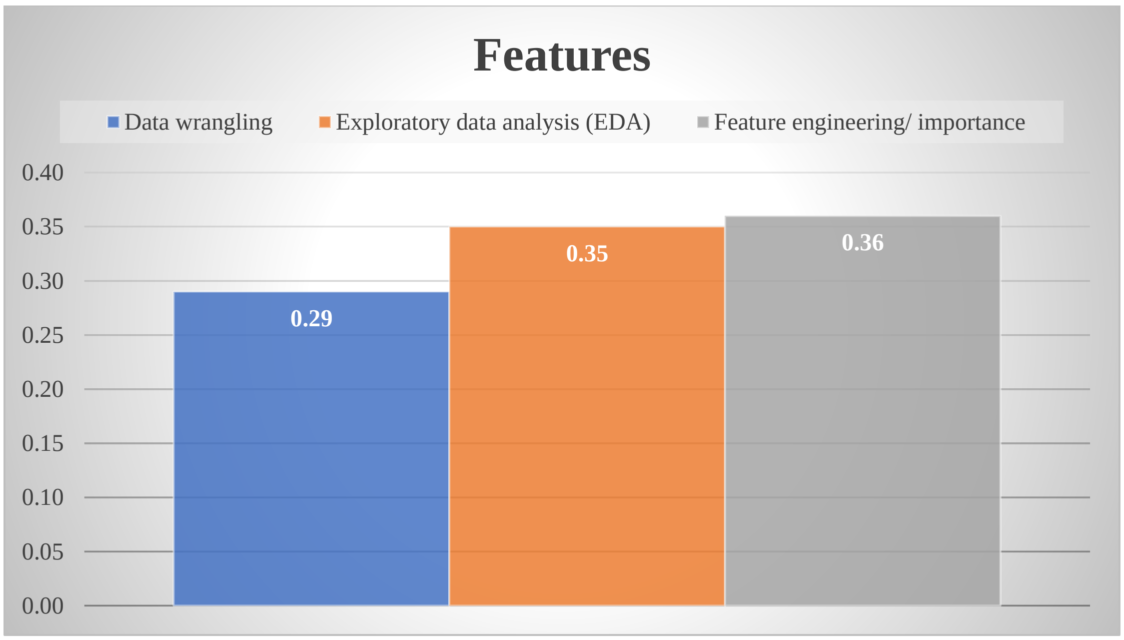


Figure 37: Relative Weight of Factors in Descending Order

Table 34: Features Factors Detailed Summary

Features				
Panel	Data Wrangling	Exploratory Data Analysis (EDA)	Feature Engineering/ Importance	Inconsistency
Expert 14	0.33	0.33	0.33	0
Expert 6	0.33	0.33	0.33	0
Expert 1	0.33	0.33	0.33	0
Expert 4	0.33	0.33	0.33	0
Expert 17	0.15	0.19	0.66	0
Expert 10	0.27	0.33	0.4	0

Expert 16	0.25	0.43	0.33	0
Expert 8	0.29	0.38	0.33	0
Expert 12	0.24	0.49	0.27	0
Expert 19	0.44	0.44	0.12	0
Expert 13	0.33	0.33	0.33	0
Expert 9	0.18	0.29	0.54	0.05
Expert 15	0.33	0.33	0.33	0
Mean	0.29	0.35	0.36	
Minimum	0.15	0.19	0.12	
Maximum	0.44	0.49	0.66	
Std. Deviation	0.07	0.07	0.12	
Disagreement				0.069

6.2.2.5 Modeling Quantification

Panel experts evaluated and ranked all factors' relative weight or importance under the Modeling perspective. Figure 43 is the graphical representation of the results of these pairwise comparisons, and Table 28 illustrates the detailed summary.

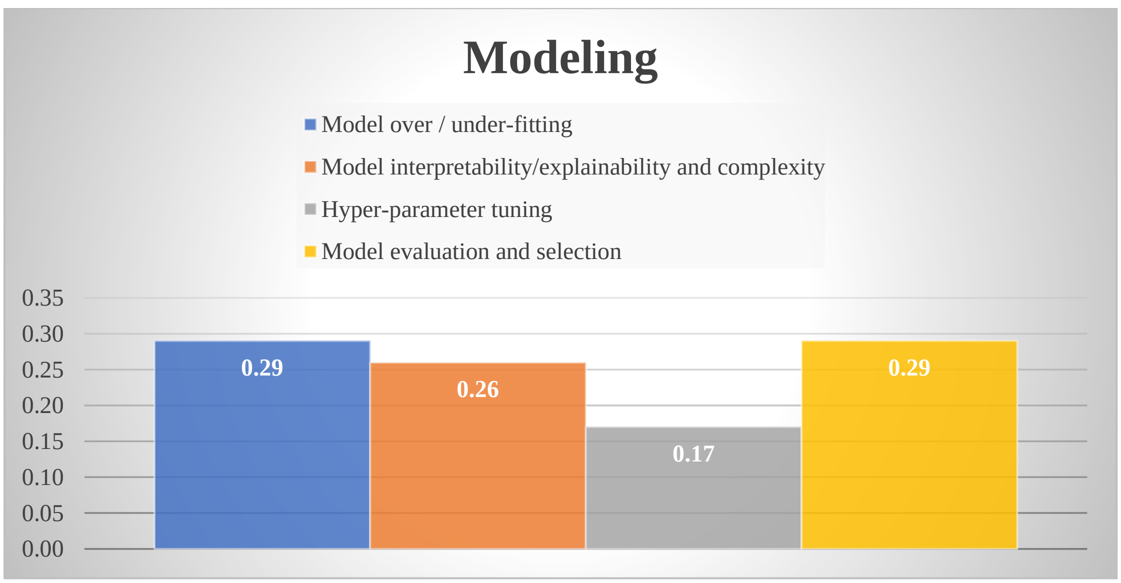


Figure 38: Relative Weight of Factors in Descending Order

Table 35: Modeling Factors Detailed Summary

Panel	Model over / under-fitting	Model interpretability/explainability and complexity	Hyper-parameter tuning	Model evaluation and selection	Inconsistency
Expert 14	0.25	0.25	0.25	0.25	0
Expert 6	0.17	0.17	0.17	0.5	0
Expert 1	0.47	0.16	0.09	0.28	0.05
Expert 4	0.32	0.18	0.21	0.3	0
Expert 17	0.35	0.45	0.04	0.16	0.02
Expert 10	0.27	0.32	0.18	0.23	0
Expert 16	0.22	0.27	0.2	0.3	0
Expert 8	0.3	0.24	0.15	0.32	0.02
Expert 12	0.42	0.17	0.13	0.29	0.01
Expert 13	0.29	0.24	0.17	0.3	0.01
Expert 9	0.27	0.28	0.17	0.28	0.06
Expert 15	0.25	0.25	0.25	0.25	0
Expert 7	0.17	0.34	0.15	0.34	0.05
Mean	0.29	0.26	0.17	0.29	
Minimum	0.17	0.16	0.04	0.16	
Maximum	0.47	0.45	0.25	0.5	
Std. Deviation	0.08	0.08	0.06	0.07	
Disagreement					0.062

6.2.2.6 Performance Quantification

The Panel of 11 experts evaluated and ranked the relative weight or importance of all factors under the Performance perspective. Figure 44 is the graphical representation of the results of these pairwise comparisons, and Table 29 illustrates the detailed summary.

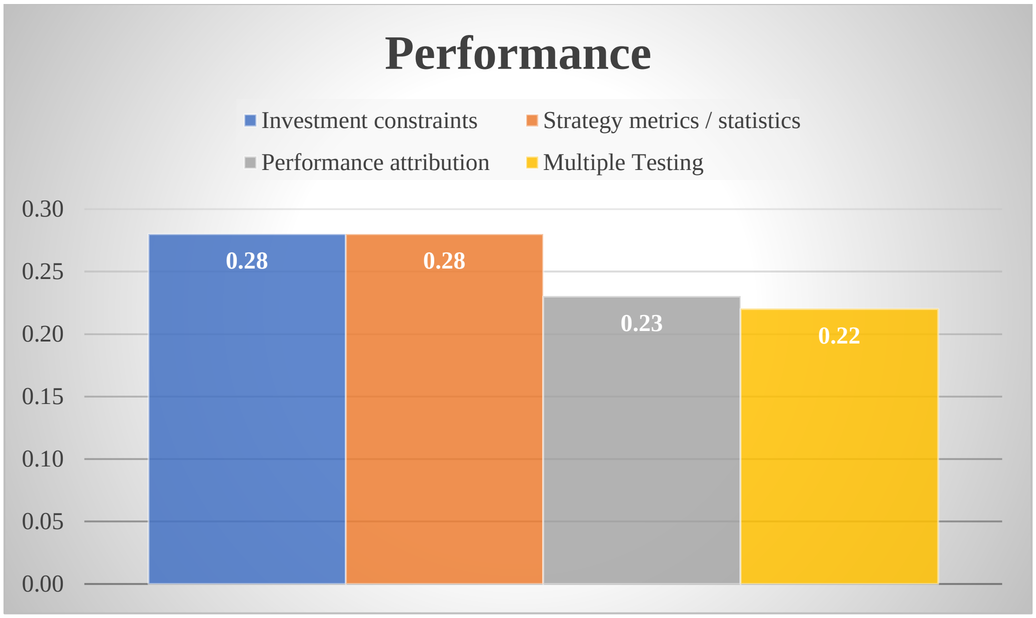


Figure 39: Relative Weight of Factors in Descending Order

Table 36: Performance Factors Detailed Summary

Performance					
Panel	Investment Constraints	Strategy Metrics/Statistics	Performance Attribution	Multiple Testing	Inconsistency
Expert 6	0.12	0.5	0.12	0.25	0
Expert 4	0.21	0.21	0.21	0.38	0
Expert 17	0.26	0.26	0.35	0.12	0
Expert 16	0.2	0.33	0.27	0.2	0
Expert 3	0.33	0.21	0.25	0.21	0
Expert 8	0.33	0.15	0.29	0.24	0.03
Expert 12	0.42	0.13	0.27	0.18	0.02
Expert 19	0.43	0.18	0.22	0.17	0.02
Expert 13	0.19	0.42	0.19	0.2	0
Expert 9	0.22	0.41	0.16	0.21	0.02
Expert 7	0.38	0.24	0.17	0.21	0.01
Mean	0.28	0.28	0.23	0.22	
Minimum	0.12	0.13	0.12	0.12	
Maximum	0.43	0.5	0.35	0.38	
Std. Deviation	0.1	0.12	0.06	0.06	
Disagreement					0.084

6.3 Results Analysis

6.3.1 Inconsistency and Disagreement Analysis

As previously discussed in this research, inconsistencies in expert judgment and disagreements among experts serve the dual objective of validating the model’s outcomes and pinpointing areas ripe for improvement. In the context of HDM modeling, it is imperative to analyze and quantify the experts' judgments, as well as qualify them, to consider them as a reliable source of insight in the modeling process. As previously indicated, experts were asked to perform model validation and pairwise comparisons using Qualtrics surveys and the HDM software tool. Even though the researcher has tried to follow an ordered protocol to reduce the amount of confusion and potential sources of error, human judgments are intrinsically susceptible to bias, which could result in inconsistencies and disagreements.

Therefore, measures for inconsistency and disagreement are calculated using the formulas in their respective sections. For inconsistency and disagreement, a threshold of 10% is defined as an acceptable threshold. If either of these measures exceeds the acceptable threshold, the researcher should identify the root causes and suggest solutions to lessen or eliminate such issues. As illustrated in the tables above, all expert inconsistencies are below the 10% threshold in responses captured at the perspectives level. Upon examining the overall inconsistency in expert judgments, it is evident that some experts have inconsistency scores exceeding the acceptable 10% threshold. In such situations, consultative approaches should be the starting point to ensure that the expert understands the model's context and purpose. In the worst case, if the researcher continues to receive inconsistent results above the acceptable level, the expert will be eliminated from the group and replaced by another expert.

From a disagreement standpoint, the findings reveal that none of the perspectives exhibit a disagreement level surpassing 10%. Notably, the Performance perspective demonstrates the highest degree of discord among experts, at 8%. Conversely, the data perspective stands out with the lowest level of disagreement, garnering a consensus of 4% among experts and underscoring a significant alignment of viewpoints.

6.3.2 Final Model Weights

After model finalization, informed by the validation of factors derived from the comprehensive literature review, domain experts assigned quantifiable values to these factors. This quantification was undertaken to determine the relative significance of these factors concerning the overarching goal of evaluating the DS-ML-driven investment strategy development in investment management firms. Figures X and Y below illustrate a visual representation of the model, complete with the respective weightings assigned to the identified factors.

The table below illustrates the final model weights. Global Weights are the result of Local Weight and Perspective Weight multiplication.

Table 37: HDM Model Final Weights

Perspectives	Factors	Local Weight	Global Weight
Economic Foundations and Research (23%)	Economic/ financial foundation and scientific approach	51.0%	11.7%
	Investment research question and thesis knowledge	49.0%	11.3%
Data (21%)	Data biases and features	24.0%	5.0%
	Data availability and sufficiency	26.0%	5.5%
	Data integrity and quality	29.0%	6.1%
	Data standards and reflexivity	22.0%	4.6%
Features (18%)	Data wrangling	29.0%	5.2%
	Exploratory Data Analysis (EDA)	35.0%	6.3%
	Feature engineering and importance	36.0%	6.5%
Modeling (17%)	Model over / under-fitting	29.0%	4.9%
	Model interpretability/explainability and complexity	26.0%	4.4%
	Hyper-parameter tuning	17.0%	2.9%
	Model evaluation and selection	29.0%	4.9%
Performance (21%)	Investment constraints	28.0%	5.9%
	Strategy metrics/statistics	28.0%	5.9%
	Performance attribution	23.0%	4.8%
	Multiple Testing	22.0%	4.6%

The following figure represents the global weight of factors in descending order.

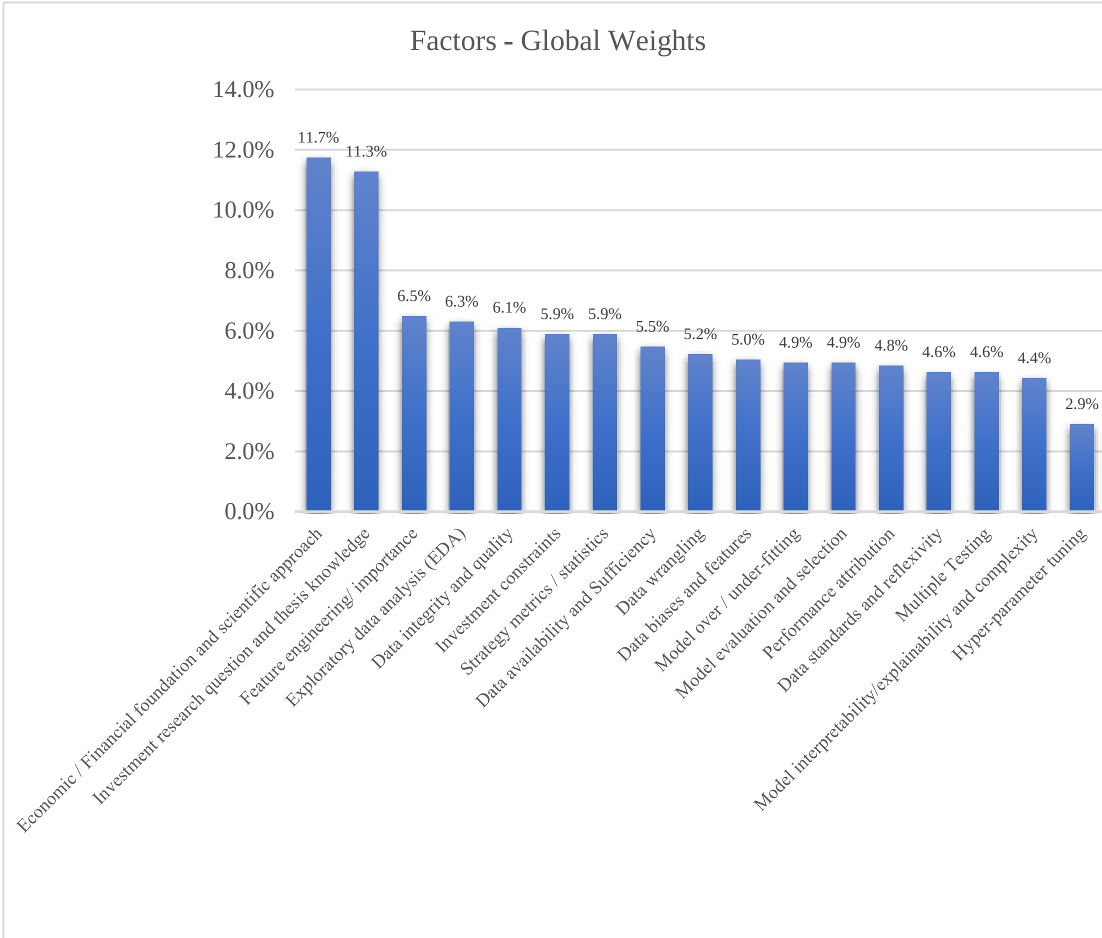


Figure 40: Factors With Their Impact Weight on ML-DS-based Investment Strategies

Figure 41: Local Criteria Weights

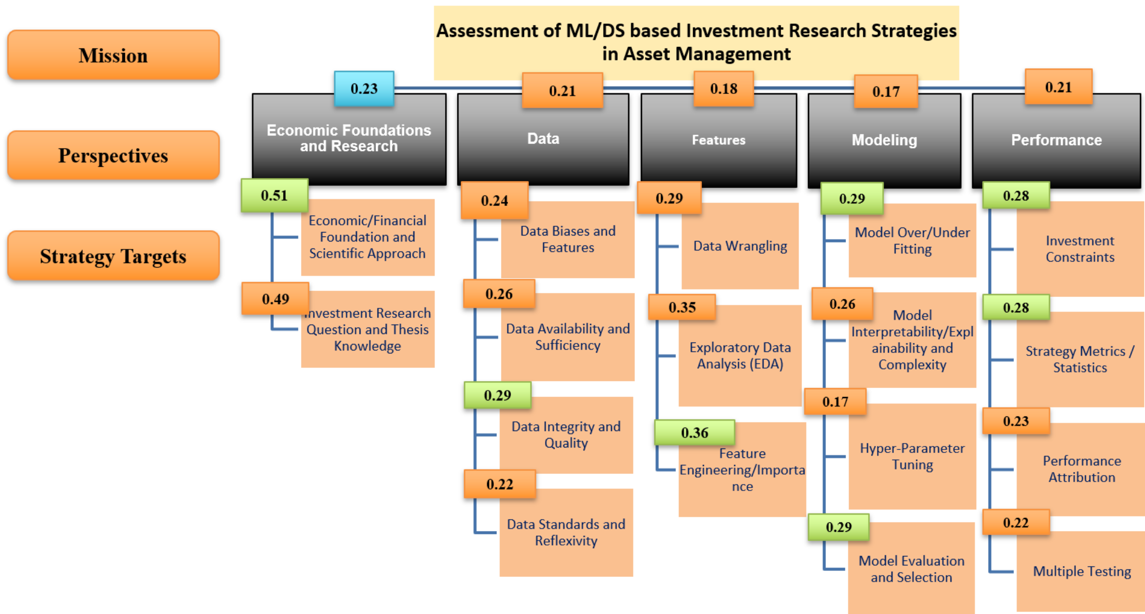
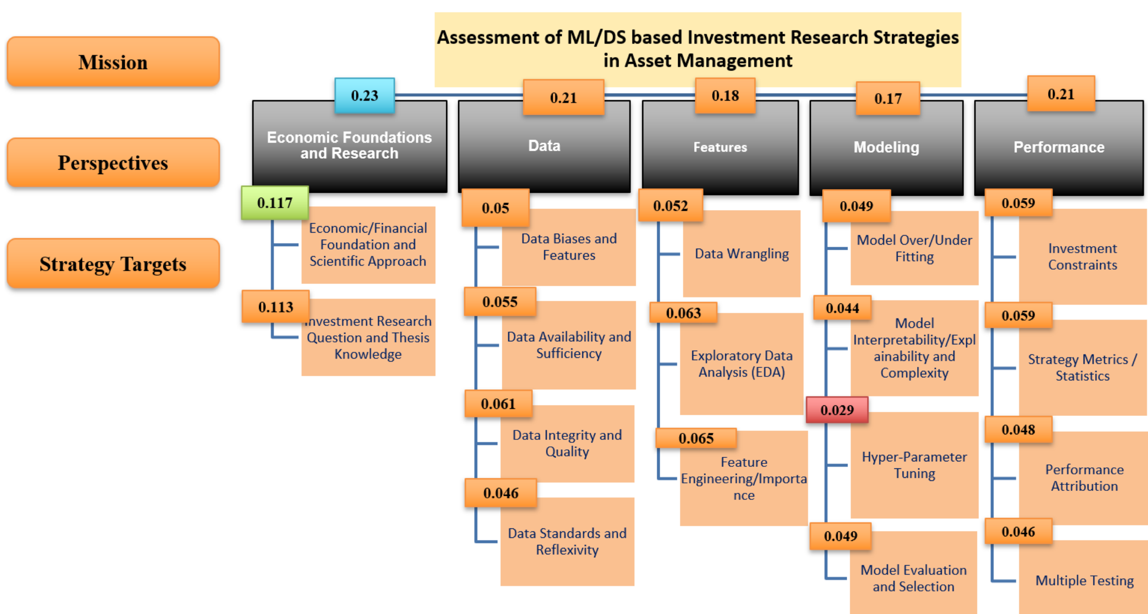


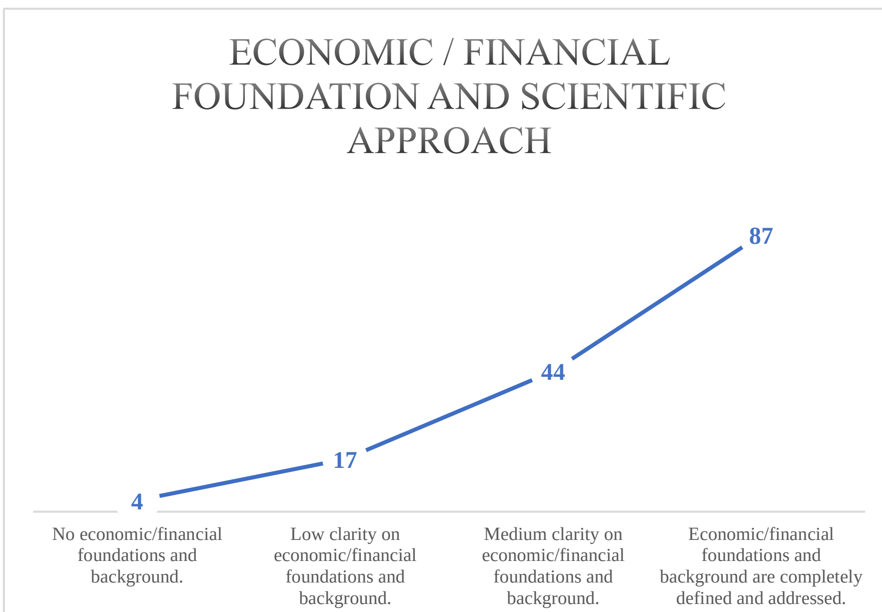
Figure 42: Global Criteria Weights



6.4 Desirability Curves Validation and Quantification

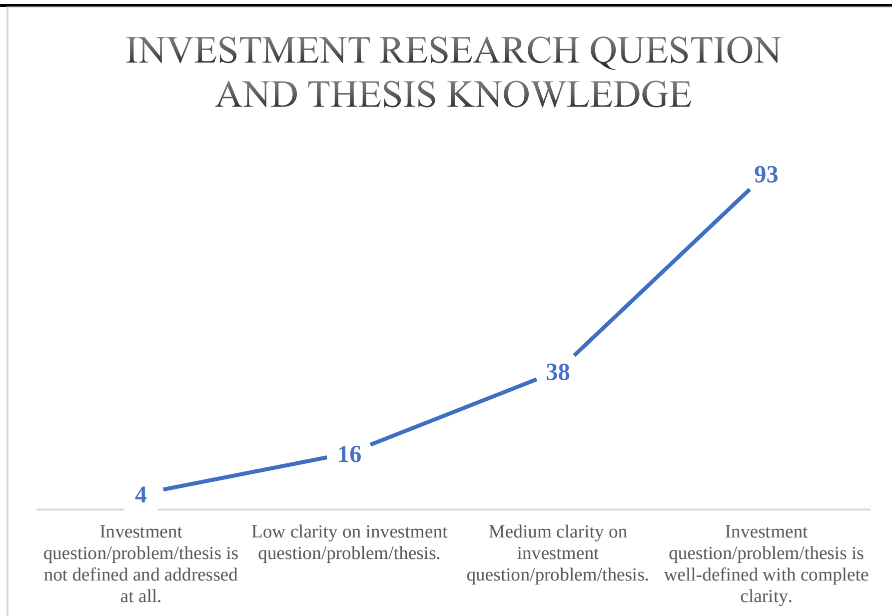
The following charts show value curves and their scores based on the results of the Qualtrics Survey that experts completed for all factors. The desirability level of each factor is calculated as a mean of scores across all experts who completed the survey.

Table 38: Desirability Curves for The Model Criteria

Criteria	Desirability Curve
Economic Foundations and Research Perspective	
Economic/ Financial Foundation and Scientific Approach	This factor measures the structure of the machine learning problem to ensure that the problem is guided by scientific and economic thinking before the research is conducted. Below are the categories: • No economic/financial foundations and background. • Low clarity on economic/financial foundations and background. • Medium clarity on economic/financial foundations and background. • Economic/financial foundations and background are entirely defined and 

<table>
<tr><td></td><td>
<table>
<tr><td>Economic/ Financial
Foundation and Scientific Approach</td><td></td></tr>
<tr><td>Description</td><td>Desirability</td></tr>
<tr><td>No economic/financial foundations and background.</td><td>4</td></tr>
<tr><td>Low clarity on economic/financial foundations and background.</td><td>17</td></tr>
<tr><td>Medium clarity on economic/financial foundations and background.</td><td>44</td></tr>
<tr><td>Economic/financial foundations and background are entirely defined and addressed.</td><td>87</td></tr>
</table>
</td></tr>
<tr><td>Investment Research Question and Thesis Knowledge</td><td>This factor measures the level of clarity and robustness of the strategy in terms of a research question, objectives, and investment thesis. An investment thesis is vital as it will be the basis for the data representing it. Hence, in this step, the investment strategy should be based on economic foundations and the research question or problem it is trying to solve. Below are the categories:

• The investment question/problem/thesis is not defined or addressed.
• Low clarity on investment question/problem/thesis.
• Medium clarity on investment question/problem/thesis.
• The investment question/problem/thesis is well-defined with complete clarity.</td></tr>
</table>



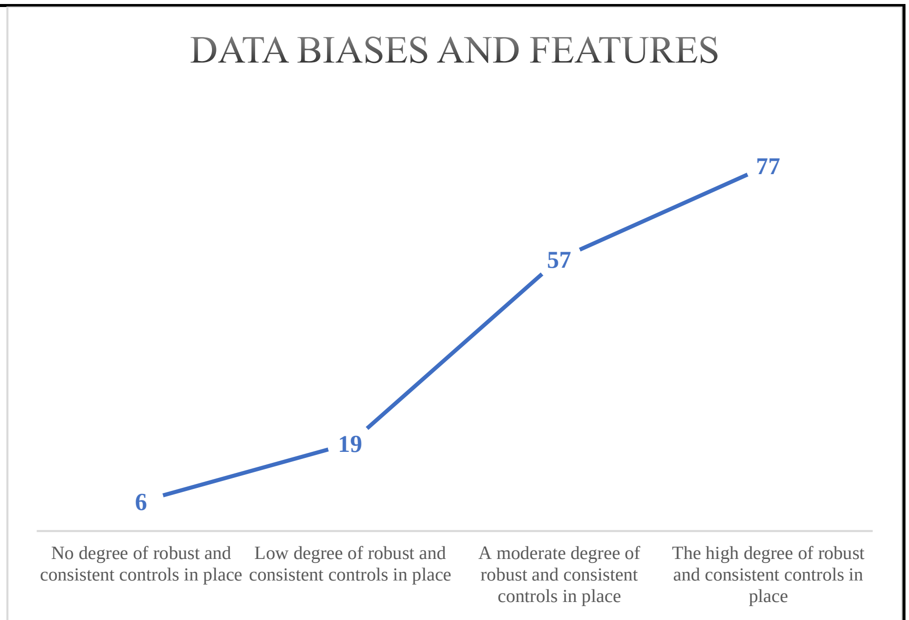
Investment Research Question and Thesis Knowledge	
Description	Desirability
The investment question/problem/thesis is not defined or addressed.	4
Low clarity on investment question/problem/thesis.	16
Medium clarity on investment question/problem/thesis.	38
The investment question/problem/thesis is well-defined with complete clarity.	93

Data Perspective

Data Biases and Features

This factor measures the taken approaches and controls to ensure data biases are considered in the investment strategy development process. This measure confirms that the data used for strategy development is checked to ensure the usability and reliability of the dataset for further exploration. Below are the categories:

- No degree of robust and consistent controls in place
- Low degree of robust and consistent controls in place
- A moderate degree of robust and consistent controls in place
- The high degree of robust and consistent controls in place

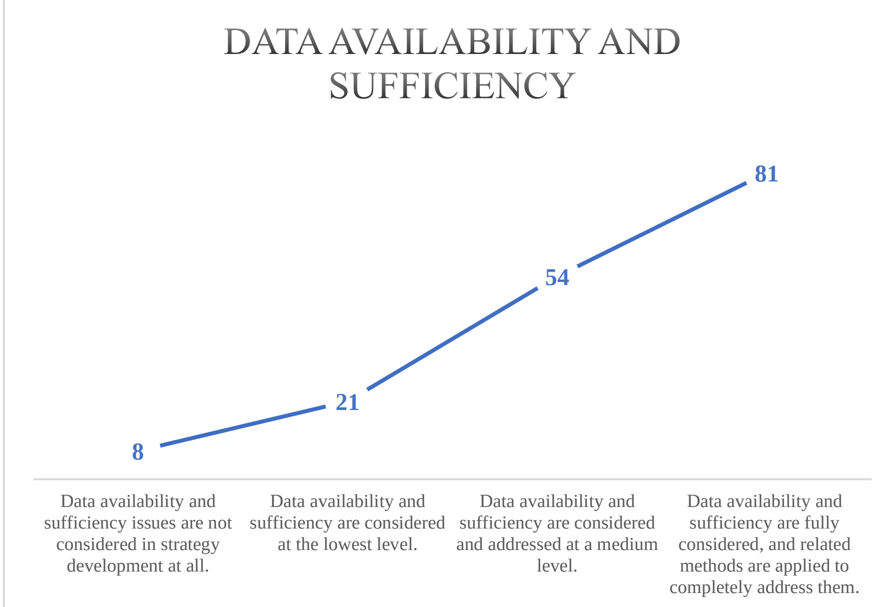


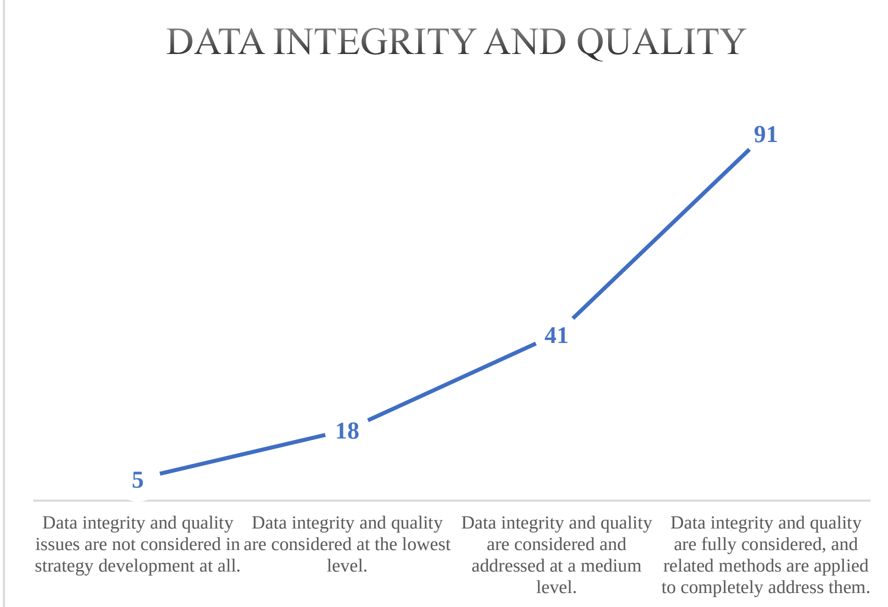
Data Biases and Features	
Description	Desirability
No degree of robust and consistent controls in place	6
Low degree of robust and consistent controls in place	19
A moderate degree of robust and consistent controls in place	57
The high degree of robust and consistent controls in place	77

Data Availability and Sufficiency

This factor measures how data availability and limited data issues are addressed and which techniques, such as synthetic datasets and data augmentation, have been used to address this problem. Below are the categories:

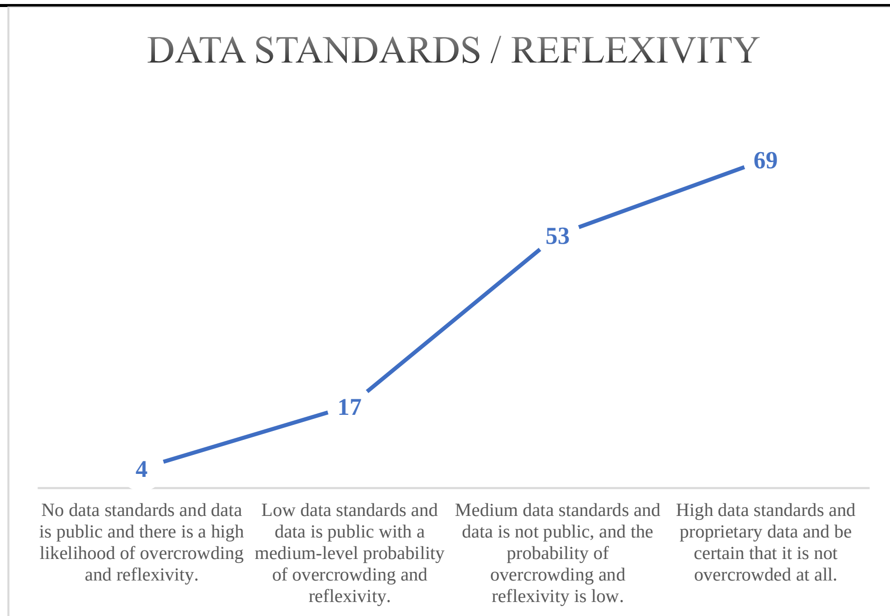
- Data availability and sufficiency issues are not considered in strategy development at all.
- Data availability and sufficiency are considered at the lowest level.
- Data availability and sufficiency are considered and addressed at a medium level.
- Data availability and sufficiency are fully considered, and related methods are applied to address them thoroughly.

	 ata Availability and Sufficiency Description | Desirability Data availability and sufficiency issues are not considered in strategy development at all. | 8 Data availability and sufficiency are considered at the lowest level. | 21 Data availability and sufficiency are considered and addressed at a medium level. | 54 Data availability and sufficiency are fully considered, and related methods are applied to address them. | 81
Data Integrity and Quality	AI models heavily rely on the integrity and quality of data and can easily result in garbage-in and garbage-out problems. Also, since most ML models are considered black-box and the results are interpreted at face value, the importance of data quality and integrity goes higher. This factor measures this aspect of data. Below are the categories • Data integrity and quality issues are not considered in strategy development at all. • Data integrity and quality are considered at the lowest level. • Data integrity and quality are considered and addressed at a medium level. • Data integrity and quality are fully considered, and related methods are applied to completely address them.

<table>
<tr>
<td></td>
<td>


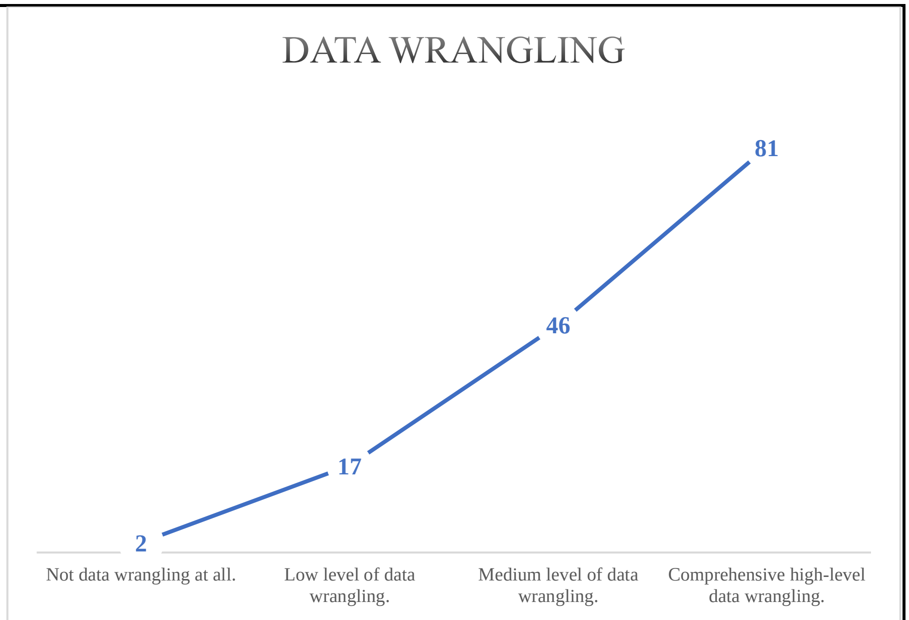
<table>
<tr><th colspan="2">Data Integrity and Quality</th></tr>
<tr><th>Description</th><th>Desirability</th></tr>
<tr><td>Data integrity and quality issues are not considered in strategy development at all.</td><td>5</td></tr>
<tr><td>Data integrity and quality are considered at the lowest level.</td><td>18</td></tr>
<tr><td>Data integrity and quality are considered and addressed at a medium level.</td><td>41</td></tr>
<tr><td>Data integrity and quality are fully considered, and related methods are applied to completely address them.</td><td>91</td></tr>
</table>
</td>
</tr>
<tr>
<td>Data Standards and Reflexivity</td>
<td>This factor measures the financial data standard health check level in the investment development process to ensure the associated data standard aspects are based on accepted global financial data regulations and standards.

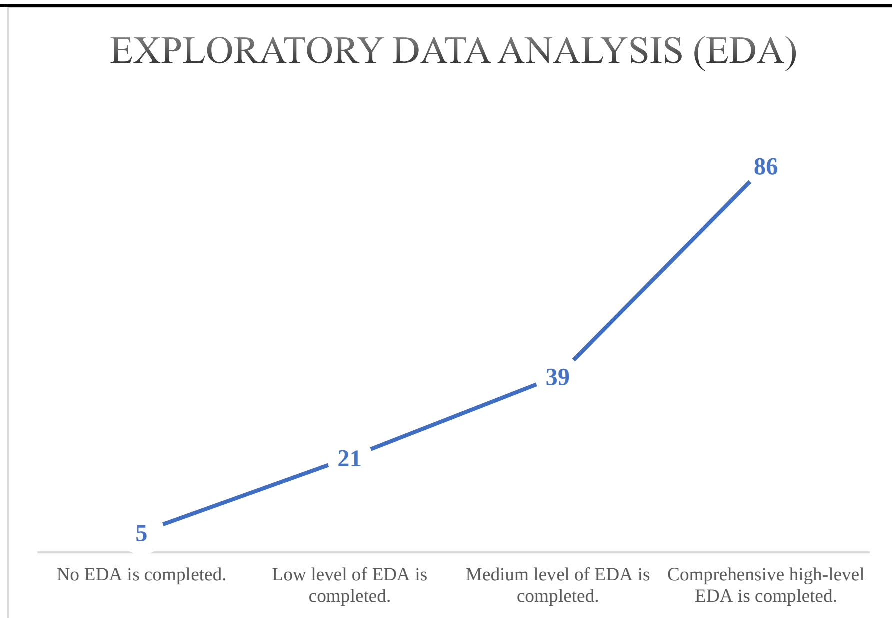
• No data standards data is public, and there is a high likelihood of overcrowding and reflexivity.
• Low data standards and data are public with a medium-level probability of overcrowding and reflexivity.
• Medium data standards and data are not public, and the probability of overcrowding and reflexivity is low.
• High data standards and proprietary data, and be specific that it is not overcrowded at all.</td>
</tr>
</table>



Data Standards and Reflexivity	
Description	Desirability
No data standards data is public, and there is a high likelihood of overcrowding and reflexivity.	4
Low data standards and data are public with a medium-level probability of overcrowding and reflexivity.	17
Medium data standards and data are not public, and the probability of overcrowding and reflexivity is low.	53
High data standards and proprietary data, and be sure that it is not overcrowded at all.	69

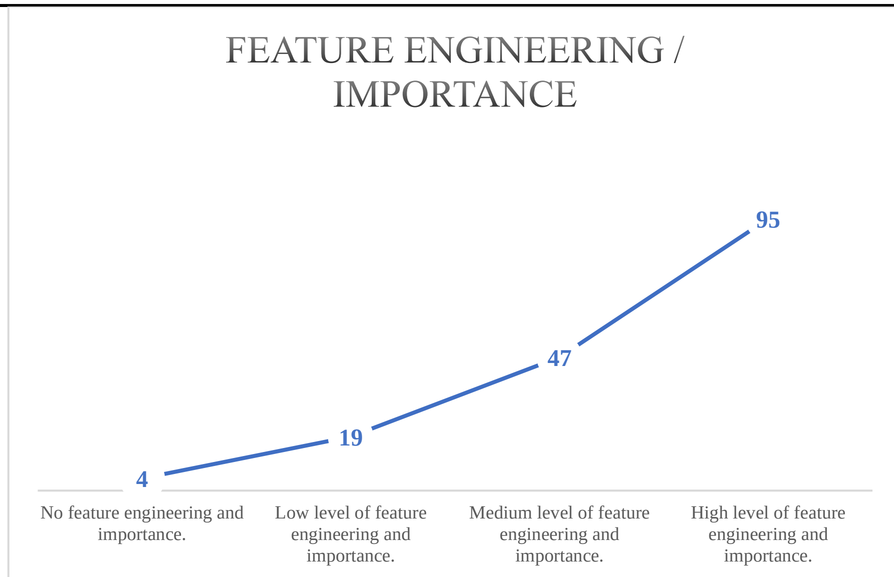
Feature Perspective	
Data Wrangling	This fracture measures the level of data wrangling that is considered before using the dataset further down the research process. By cleaning, transforming, and mapping data, this step sheds light on any potential holes in the dataset, which is crucial in strategy development. • No data wrangling at all. • Low level of data wrangling. • Medium level of data wrangling. • Comprehensive high-level data wrangling.

	 **Data Wrangling** | Description | Desirability | |---|---| | No data wrangling at all. | 2 | | Low level of data wrangling. | 17 | | Medium level of data wrangling. | 46 | | Comprehensive high-level data wrangling. | 81 |
Exploratory Data Analysis (EDA)	This factor measures the level of completed EDA on the dataset before starting any machine learning modeling or hypothesis testing. This helps ensure that the EDA steps have been completed in the strategy development process, increasing the probability of reliable results. • No EDA is completed. • A low level of EDA is completed. • Medium level of EDA is completed. • Comprehensive high-level EDA is completed.

<table>
<tr>
<td></td>
<td>


<table>
<tr><th colspan="2">Exploratory Data Analysis (EDA)</th></tr>
<tr><th>Description</th><th>Desirability</th></tr>
<tr><td>No EDA is completed.</td><td>5</td></tr>
<tr><td>A low level of EDA is completed.</td><td>21</td></tr>
<tr><td>Medium level of EDA is completed.</td><td>39</td></tr>
<tr><td>Comprehensive high-level EDA is completed.</td><td>86</td></tr>
</table>
</td>
</tr>
<tr>
<td>Feature Engineering/Importance</td>
<td>This factor measures the actions taken and techniques used in feature construction, feature extraction, feature selection, and feature importance. Techniques in feature importance, for instance, can provide insight into the dataset and highlight the most relevant feature to the target variable. Below are the categories:

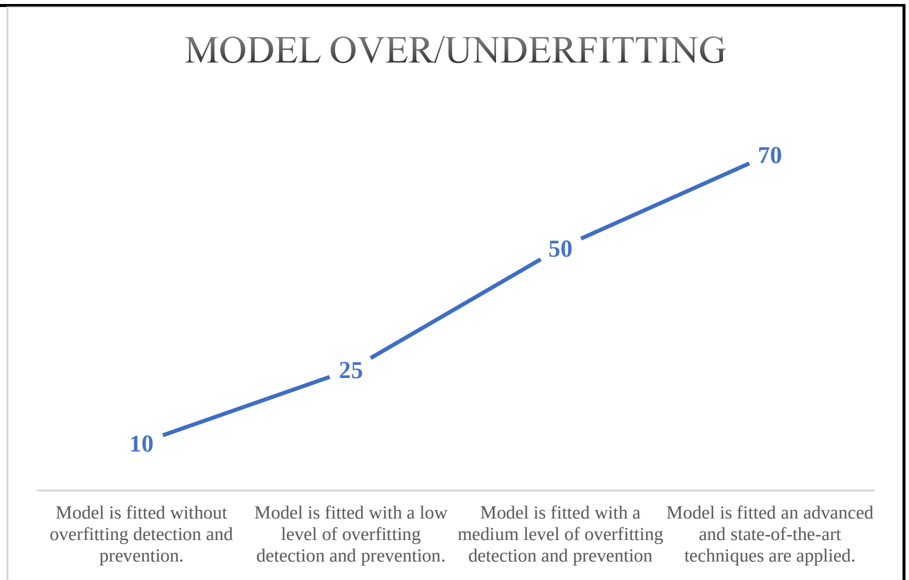
• No feature engineering and importance.
• Low level of feature engineering and importance.
• Medium level of feature engineering and importance.
• High level of feature engineering and importance.</td>
</tr>
</table>



Feature Engineering/Importance	
Description	Desirability
No feature engineering and importance.	4
Low level of feature engineering and importance.	19
Medium level of feature engineering and importance.	47
High level of feature engineering and importance.	95

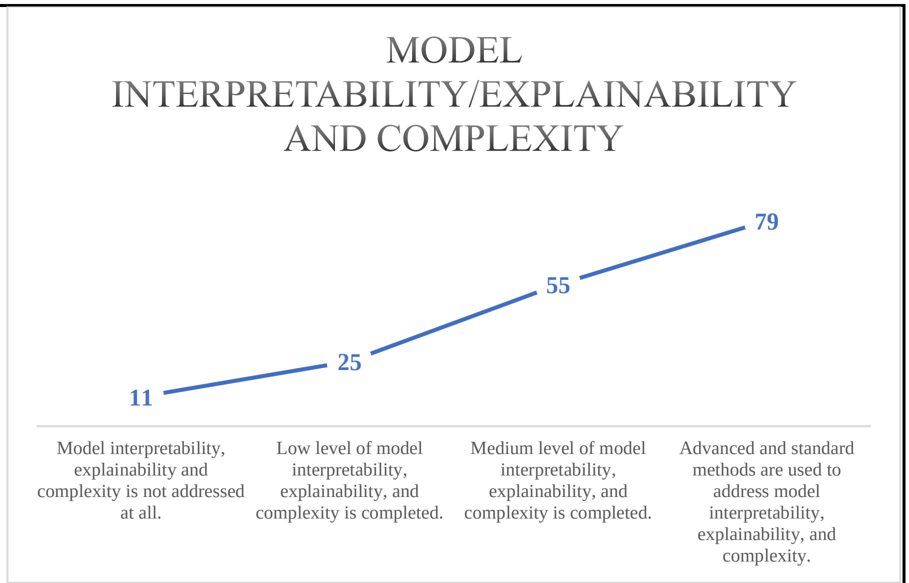
Modeling Perspective

Model Over/Underfitting	This factor intends to show that this issue is considered in the investment strategy design and development, and suitable techniques are in place to ensure the model is not overfitted or underfitting. Also, to illustrate that true out-of-sample testing is only possible when the model is tested in real-world and live trading. Below are the categories: • The model is fitted without overfitting detection and prevention. • The model is fitted with a low level of overfitting detection and prevention. • The model is fitted with a medium level of overfitting detection and prevention. • The model is fitted, and advanced and state-of-the-art techniques are applied.



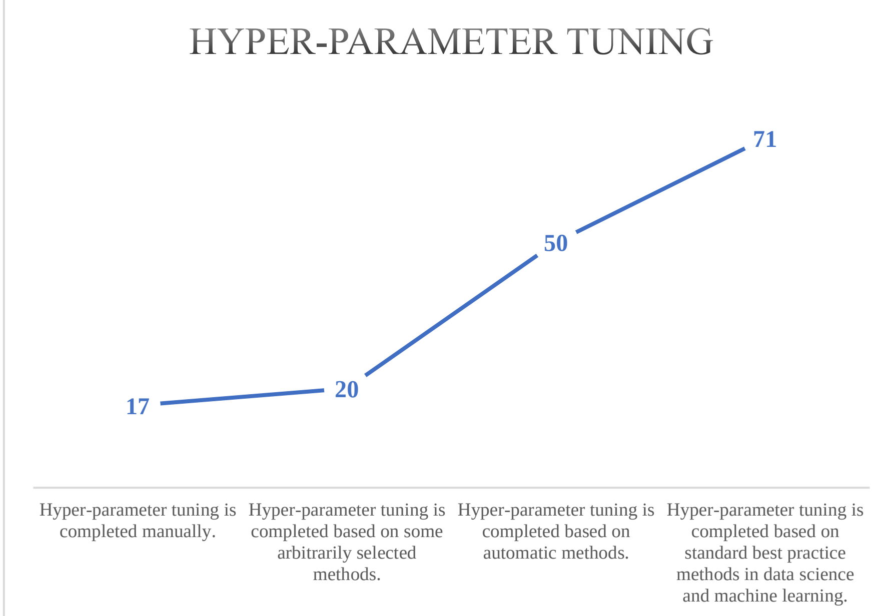
Model Over/Underfitting	
Description	Desirability
The model is fitted without overfitting detection and prevention.	10
The model is fitted with a low level of overfitting detection and prevention.	25
The model is fitted with a medium level of overfitting detection and prevention	50
The model is fitted with advanced and state-of-the-art techniques.	70

Model Interpretability/Explainability and Complexity	This factor evaluates whether steps are taken to develop more parsimonious specifications in modeling. Given different modeling criteria, this factor measures the appropriability of model parsimony consideration in designing the investment strategy. Below are the categories: • Model interpretability, explainability, and complexity are not addressed at all. • A low level of model interpretability, explainability, and complexity is completed. • Medium level of model interpretability, explainability, and complexity is completed. • Advanced and standard methods address model interpretability, explainability, and complexity.



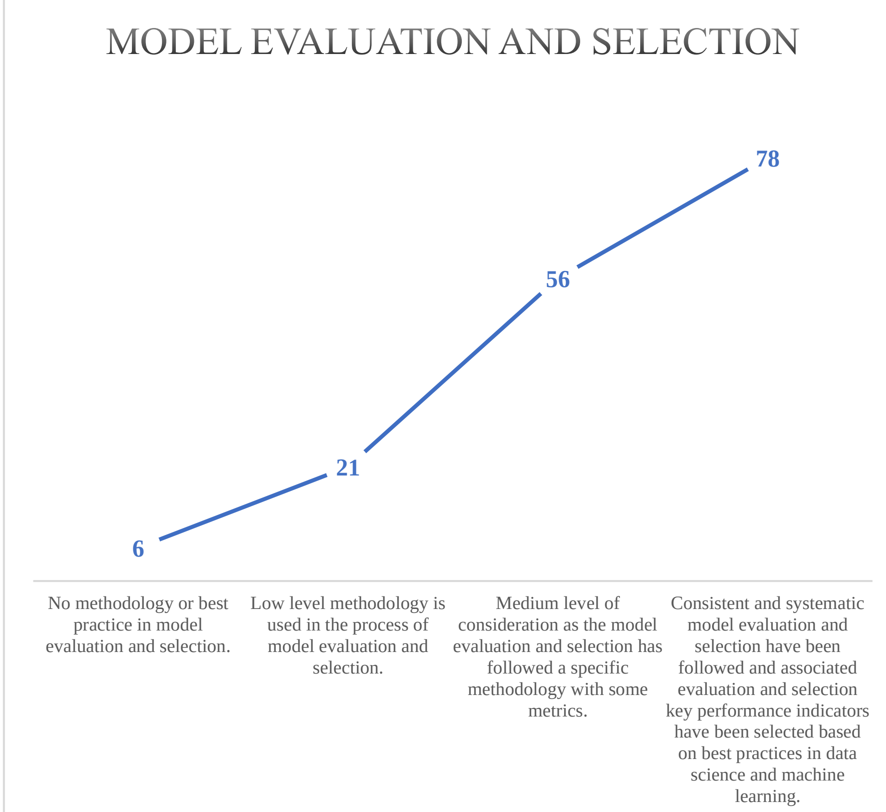
Model Complexity	
Description	Desirability
Model interpretability, explainability, and complexity are not addressed at all.	11
A low level of model interpretability, explainability, and complexity is completed.	25
Medium level of model interpretability, explainability, and complexity is completed.	55
Advanced and standard methods address model interpretability, explainability, and complexity.	79

Hyper-Parameter Tuning	This factor measures how strategy hyper-parameters are tuned and which systematic methods have been used to find the optimal model structure. • Hyper-parameter tuning is completed manually. • Hyper-parameter tuning is completed based on some arbitrarily selected methods. • Hyper-parameter tuning is completed based on automatic methods. • Hyper-parameter tuning is completed based on standard best practice data science and machine learning methods.



Hyper-Parameter Tuning	
Description	Desirability
Hyper-parameter tuning is completed manually.	17
Hyper-parameter tuning is completed based on some arbitrarily selected methods.	20
Hyper-parameter tuning is completed based on automatic methods.	50
Hyper-parameter tuning is completed based on standard best practice data science and machine learning methods.	71

Model Evaluation and Selection	This factor measures the level of the model evaluation and selection approaches and methods used in selecting and evaluating strategy models. The desirability of these criteria is measured through the following categories: • There is no methodology or best practice in model evaluation and selection. • The low-level methodology is used in the process of model evaluation and selection. • Medium level of consideration as the model evaluation and selection has followed a specific methodology with some metrics. • Consistent and systematic model evaluation and selection have been followed. The associated assessment and selection of key performance indicators have been selected based on best practices in data science and machine learning.



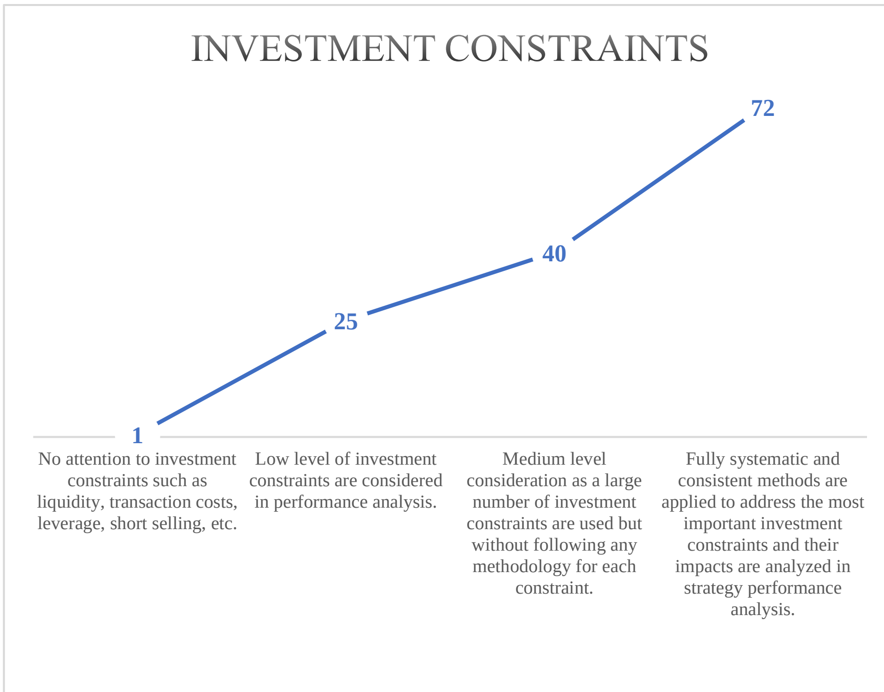
Model Evaluation and Selection	
Description	Desirability
There is no methodology or best practice in model evaluation and selection.	6
Low-level methodology is used in the process of model evaluation and selection.	21
Medium level of consideration as the model evaluation and selection has followed a specific methodology with some metrics.	56
Consistent and systematic model evaluation and selection have been followed. The associated assessment and selection of key performance indicators have been selected based on best practices in data science and machine learning.	78

Performance Perspective	
Investment Constraints	This factor measures the level at which investment constraints are addressed, and systematic actions are in place. The desirability curve of this factor will be measured through the following categories:

- No attention to investment constraints such as liquidity, transaction costs, leverage, and short selling.

- Low levels of investment constraints are considered in performance analysis.

- Medium-level consideration as many investment constraints are used but without following any methodology for each constraint.
- Fully systematic and consistent methods are applied to address the most essential investment constraints, and their impacts are analyzed in strategy performance analysis.



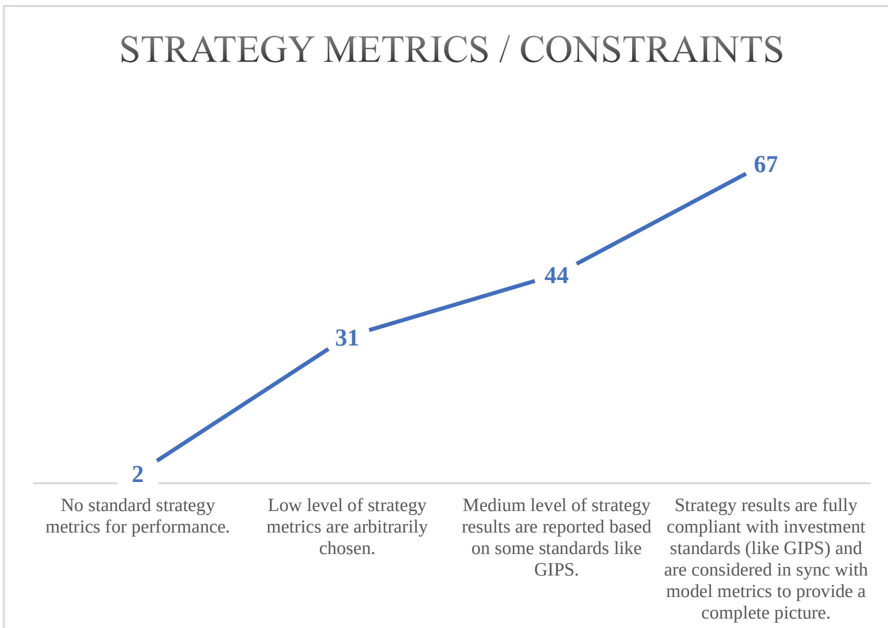
Investment Constraints	
Description	Desirability
No attention to investment constraints such as liquidity, transaction costs, leverage, and short selling.	1
A low level of investment constraints is considered in performance analysis.	25
Medium-level consideration, as many investment constraints are used without following any methodology for each constraint.	40

	Fully systematic and consistent methods are applied to address the most important investment constraints, and their impacts are analyzed in strategy performance analysis.	72

Strategy Metrics/Statistics

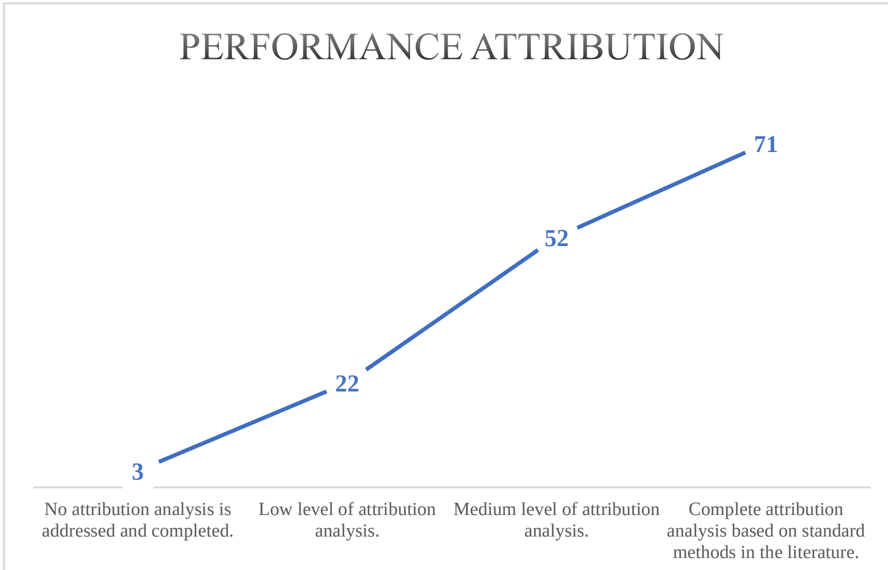
These measurements should be considered with model measures mentioned in model selection and evaluation criteria to provide a more comprehensive picture of strategy performance. This factor measures the level and quality of acquired and reported strategy performance results. Below are the categories:

- There are no standard strategy metrics for performance.
- A low level of strategy metrics is arbitrarily chosen.
- Medium-level strategy results are reported based on some standards like
- Strategy results fully comply with investment standards (like GIPS) and a considered in sync with model metrics to provide a complete picture.



Strategy Metrics/Statistics	
Description	Desirability
There are no standard strategy metrics for performance.	2
A low level of strategy metrics is arbitrarily chosen.	31
Medium-level strategy results are reported based on some standards like GIPS.	44
Strategy results fully comply with investment standards (like GIPS) and are considered in sync with model metrics to provide a complete picture.	67

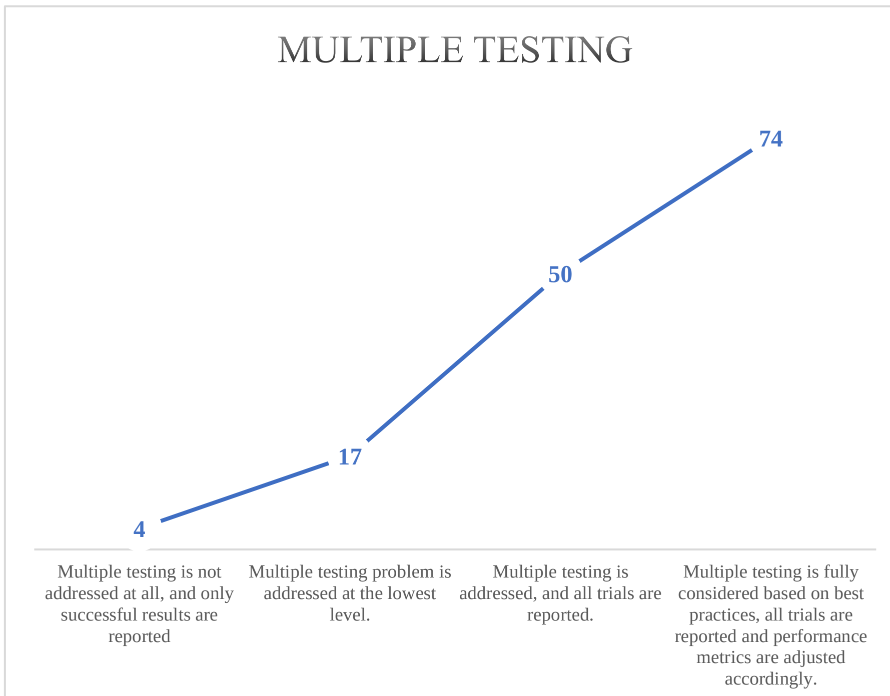
Performance Attribution	This factor measures and evaluates the performance that is explainable by exposure to common risk factors, generated alpha, and stock selection capabilities. Three generally considered forms of attribution are multi-factor analysis, style analysis, and return decomposition analysis. Below are the categories that show the level of attribution analysis: • No attribution analysis is addressed and completed. • Low level of attribution analysis. • Medium level of attribution analysis. • Complete attribution analysis based on standard methods in the literature.



Performance Attribution	
Description	Desirability
No attribution analysis is addressed and completed.	3
Low level of attribution analysis.	22
Medium level of attribution analysis.	52
Complete attribution analysis based on standard methods in the literature.	71

<table>
<tr>
<td>Multiple Testing (Selection Bias)</td>
<td>
Running the backtest often and selecting and reporting good results are the main reasons for fund failures. This factor tries to address and measure this issue and ensure that strategy is checked for this fundamental issue and suitable actions are taken to report all trials. Below are the categories:

Multiple testing is not addressed at all, and only successful results are rep
Multiple testing problems are addressed at the lowest level.
Multiple testing is addressed, and all trials are reported.
Multiple testing is fully considered based on best practices, all trials are re metrics are adjusted accordingly.



<table>
<tr><th colspan="2">Multiple Testing</th></tr>
<tr><th>Description</th><th>Desirability</th></tr>
<tr><td>Multiple testing is not addressed at all, and only successful results are reported</td><td>4</td></tr>
<tr><td>Multiple testing problems are addressed at the lowest level.</td><td>17</td></tr>
<tr><td>Multiple testing is addressed, and all trials are reported.</td><td>50</td></tr>
<tr><td>Multiple testing is fully considered based on best practices, all trials are reported, and performance metrics are adjusted accordingly.</td><td>74</td></tr>
</table>
</td>
</tr>
</table>

CHAPTER 7. RESEARCH MODEL APPLICATION

Case Study 1: The US Hedge Fund Investing in European Equity Market

The top-level items on any company's income statement will ultimately drive its long-term success. Of course, this obvious point is often obscured by the share price fluctuations of publicly traded companies.

The point at which the link between fundamental profitability and share price is most vital will be when financial results are reported. To capitalize on this simple reasoning, the strategy relies on forecasting quarterly or semi-annual surprises more accurately than other market participants.

Since the strategy predicts earnings announcements ahead of time, it can benefit from three associated phenomena associated with earnings releases. Firstly, it aims to catch the pre-announcement drift a few months before the published results. This is when prices move up before positive surprises and down before negative surprises. Secondly, it can capture the announcement effect on the release day. The announcement contains news and is often accompanied by abnormal volatility and volume. On the announcement, the price effect generally reflects the surprise's magnitude and direction. Thirdly, the strategy can participate in the post-announcement drift, where the stock prices continue to drift after the announcement in response to the surprise.

The key differentiating factor of the strategy is the use of various sources of non-conventional economic data to predict the near-term fundamental performance of specific companies.

The conventional data (fundamentals, price, and reporting dates.) have been sourced from several high-quality providers, such as Refinitiv and Bloomberg, since 2000. Despite the premium nature of these data, cleaning is necessary to correct myriad discrepancies in reporting dates and other entries relevant to specific aspects of the strategy.

The non-conventional data consists of two primary types. The first type is time series data obtained from several sources. The second type of data connects the raw data to specific companies based on historical relationships.

The size and nature of the data require a substantial amount of proprietary pre-processing techniques, cleaning, and some human economic analyses before any numerical processing can be employed.

The number of potentially valuable signals in non-conventional data sources is almost infinitesimally small relative to massive databases, and the time series of quarterly or semi-annual income statements is limited even over twenty years.

Spurious results and over-fitted models can overwhelm valuable signals, even with the most advanced data-mining techniques. Thus, the strategy necessitates rigorous validation of any potential model features. In some cases, this validation can be done quantitatively. The strategy incorporates a second layer of checks through human analysis of every likely portfolio company’s business model and markets. The final number of features is kept as minimal as possible to obtain meaningful predictions. Overall, the features are derived through an iterative approach using regression techniques and human insight.

As mentioned, the signal-to-noise ratio is minimal for the strategy's data sources. For the validation of linking non-conventional data to specific companies, the potentially helpful training data are too small relative to the number of portfolio companies for machine learning to be effective, so more deterministic techniques were devised.

For the time series correlations, linear regression variants such as LASSO and Ridge were employed with the caveat that human validation of models is required (or beneficial). Standard statistics (such as adjusted R-squared) and some proprietary metrics enable the models to be evaluated by their accuracy in forecasting items relevant for surprises by the respective portfolio companies.

The models are updated on an ongoing basis as new time-series data are available, as well as when fundamental analysis of a portfolio company reveals a change in their business model, markets, or other factor relevant to the model.

Backtests have been performed on the models, but due to the relatively sparse time series and requirements for human analysis of each model's economic viability, it is challenging to eliminate look-ahead bias from the results.

The best performance metrics for the models are comparing the difference between the forecasted and reported results for each respective model and portfolio company.

The financial performance of the trading system is currently partially dependent on human decisions "in the loop" on position sizing and timing, so it is impossible to analyze the results from a machine-learning perspective.

Case Study 2: The Mexican Quant Firm Investing in the US Equity Market

For a robot-advisory strategy, the team implemented different optimization targets depending on the 10Y treasuries rate, with restrictions based on the risk profile (designed for five profiles). When the last value exceeds/falls below the previous 3-year average, the approach shifts to a Min Vol approach for higher rates and Max Sharpe. The concept behind this strategy is to rebalance the portfolio and gain alpha through these adjustments. The data was the historical prices of the mutual fund's universe that could be included in the robo advisors. We calculated their expected return using consensus estimates for equity and current rates for fixed income adjusted with the expected inflation. We also needed that historical 10Y US rate, historical inflation, expectations, holdings of the mutual funds, and all ETFs.

Regarding features, the most essential part of the process was calculating the fund's expected returns; we developed a code that calculated an anticipated return for ETFs and shares and analyzed fixed income rate and duration. Once we had those numbers, we adjusted them to avoid bias from consensus sell-side numbers. Regarding the optimization constraints for each profile, we identified how we can blend the exposure to 7 types of funds, and this helped to adjust risk by asset class. We also included a risk score using clustering to establish those restrictions for another project. Once we created the historical expected returns for the model, we built a back tester code to replicate quarterly rebalances and assign the optimization depending on the rate.

We are still trying to find another signal that triggers the risk-on/off change. Maybe a PCA of more economic variables could make sense, but then we should be able to rebalance more often and include trading costs. Once we had all these calculations, we backtested the strategy to see if we could beat the benchmark. We got different results by profile; for the most conservative, we could not beat the benchmark due to the short duration of the strategy and the costs of the mutual funds. For the other four strategies, we were able to beat the benchmark that was composed of ACWI and CETES.

CHAPTER 8. CASE STUDIES AND SENSITIVITY ANALYSIS

In this chapter, we will employ the developed model to evaluate the overall readiness ratings of the two cases introduced in Chapter 7. In-depth consultations were conducted with experts associated with each case to assign value curve scores to each investment strategy based on various factors. These consultations were conducted via Zoom meetings. The computation of the ultimate readiness score is carried out utilizing the mathematical equations expounded in Chapter 5. Subsequently, scenario analysis will be employed to gauge the model's sensitivity and the implications for each case under varying scenarios. Lastly, we will discuss how the model can be leveraged to augment the readiness score for each strategy, providing comprehensive insights and deliberations on the matter.

8.1 Readiness Assessment Scores

The quantification of the model factors and desirability metrics will remain consistent. Still, various investment strategies will be evaluated against these results by considering their performance levels on the desirability metrics scale. The readiness levels of different investment strategies on this metric scale will vary, depending on their preparedness concerning each specific criterion. For instance, one investment strategy may exhibit a high degree of economic foundations and research background based on financial literature, while another may lag. The latter strategy must enhance its capabilities to elevate its readiness level. For more details, please refer to the discussion on desirability curves and the computation of readiness scores in Section 5.1.2.

The following tables demonstrate each investment strategy's final readiness assessment and score.

Table 39: Strategy (1) Assessment Score

Perspectives	Factors	Global Weight	Value Curve (VC) Score	Final Score (Weighs * VC)	Perspectives Global Score
Economic Foundations and Research (23%)	Economic/ financial foundation and scientific approach	11.7%	87	10.18	20.69
	Investment research question and thesis knowledge	11.3%	93	10.51	
Data (21%)	Data biases and features	5.0%	77	3.85	17.03
	Data availability and sufficiency	5.5%	81	4.46	
	Data integrity and quality	6.1%	91	5.55	
	Data standards and reflexivity	4.6%	69	3.17	
Features (18%)	Data wrangling	5.2%	46	2.39	6.77
	Exploratory Data Analysis (EDA)	6.3%	21	1.32	
	Feature engineering and importance	6.5%	47	3.06	
Modeling (17%)	Model over / under-fitting	4.9%	50	2.45	6.79
	Model interpretability/explainability and complexity	4.4%	25	1.10	
	Hyper-parameter tuning	2.9%	17	0.49	
	Model evaluation and selection	4.9%	56	2.74	
Performance (21%)	Investment constraints	5.9%	25	1.48	4.54
	Strategy metrics/statistics	5.9%	31	1.83	
	Performance attribution	4.8%	22	1.06	
	Multiple Testing	4.6%	4	0.18	
Overall Score				55.82	

Table 40: Strategy (2) Assessment Score

Perspectives	Factors	Global Weight	Value Curve (VC) Score	Final Score (Weighs * VC)	Perspectives Global Score
Economic Foundations and Research (23%)	Economic/ financial foundation and scientific approach	11.7%	17	1.99	12.50
	Investment research question and thesis knowledge	11.3%	93	10.51	
Data (21%)	Data biases and features	5.0%	57	2.85	10.76
	Data availability and sufficiency	5.5%	54	2.97	
	Data integrity and quality	6.1%	41	2.50	
	Data standards and reflexivity	4.6%	53	2.44	
Features (18%)	Data wrangling	5.2%	81	4.21	5.80
	Exploratory Data Analysis (EDA)	6.3%	21	1.32	
	Feature engineering and importance	6.5%	4	0.26	
Modeling (17%)	Model over / under-fitting	4.9%	25	1.23	5.17
	Model interpretability/explainability and complexity	4.4%	55	2.42	
	Hyper-parameter tuning	2.9%	17	0.49	
	Model evaluation and selection	4.9%	21	1.03	
Performance (21%)	Investment constraints	5.9%	40	2.36	5.12
	Strategy metrics/statistics	5.9%	31	1.83	
	Performance attribution	4.8%	3	0.14	
	Multiple Testing	4.6%	17	0.78	
Overall Score				39.33	

Tables 34 and 35 demonstrate the results of the case study research in which it is crystal clear that strategy (1) has achieved a higher assessment score than strategy (2). Although both strategies are in the same asset class, the total assessment score of strategy (1) is 55.82, while the total assessment score of strategy (2) is 39.33. This sample case study analysis shows that seeing a quantitative investment strategy from multiple perspectives would increase the probability of success for that specific strategy. Investors can be more confident in selecting the strategies with higher assessment scores from a pool of strategies.

Table 41: Strategy Application Overall Assessment Score

Perspectives	Case 1	Case 2
Economic foundations and research	20.69	12.50
Data	17.03	10.76
Modeling	6.77	5.80
Performance	6.79	5.17
Features	4.54	5.12
Overall Score	55.82	39.33

It is evident from the case study results that strategy (1) has outperformed strategy (2) across all assessment perspectives. Both strategies assign The highest scores to economic foundations/research and data perspectives. However, compared to strategy (1), strategy (2) is very close in modeling. The following section will discuss areas of improvement for both strategies.

8.2 Strengths, Weaknesses, and Improvement Simulation

The following table shows the strengths and weaknesses of each case. It is a demonstrative point that the proposed framework has been able to assess each strategy from multiple perspectives and decompose how each one of the attributes contributed to the overall assessment score of the strategy.

Table 42: Strengths and Weaknesses of Case 1

Case 1		Criterion Definition	Assessment Score
Strengths	Economic/ Financial Foundation and Scientific Approach	The economic foundation of the strategy is directly related to business fundamentals and not to any statistical models. As such, the underlying strategy will continue to be valid even if elements of the models need to be adjusted.	87
	Investment Research Question and Thesis Knowledge	The investment thesis relies on well-established principles of financial performance linked to official corporate reporting and communication. By developing superior financial models, the strategy can exploit changes in short- and medium-term price movements resulting from such communications.	93
	Data Integrity and Quality	The data sources are from varied sources and require significant cleaning, evaluation, and processing. However, confidence in their integrity is pretty high. Multiple validation steps ensure that the data remains relevant and accurate.	91

Weaknesses	Performance Attribution	The performance of the actual strategy is difficult to separate between the quantitative signals and qualitative human judgment. Individual factors could be evaluated within the quantitative portion to determine their effects on the observed performance.	22
	Multiple Testing	Because human evaluation of the data and judgment of investment decisions are necessary, creating a bias-free backtest is challenging. While backtests have been conducted, confidence in their accuracy is low.	4
	Investment Constraints	Many of the typical constraints (such as position sizing, transaction costs, and capacity) have not been thoroughly considered due to the relatively low frequency of our trades. However, they will become more critical as the strategy evolves.	25

Table 43: Strengths and Weaknesses of Case 2

Case 2		Criterion Definition	Assessment Score
Strengths	Investment Research Question and Thesis Knowledge	Strategy is grounded on the research and practice of robo-advisor methods and is driven by specific research questions.	93
	Investment Constraints	Regarding portfolio constraints, the strategy undergoes testing and assessment to ensure it demonstrates more realistic outcomes.	40

	Model Interpretability/Explainability & Complexity	The developed model is highly explainable, and the complexity level is low, making it easier to understand the drivers of risk/return.	55
Weaknesses	Feature Engineering/Importance	The development of the strategy completely overlooks feature engineering and its significance, which inevitably leads to the generation of highly unreliable outcomes.	4
	Performance Attribution	The analysis of the model's outcomes lacks the incorporation of performance attribution, a fundamental process for meticulously dissecting and comprehending the returns generated by the investment strategy. This omission represents a significant gap in the evaluation methodology and hinders a comprehensive understanding of the strategy's overall performance and underlying drivers.	3
	Model Evaluation and Selection	The model undergoes partial evaluation without a systematic method for selecting the best model. Consequently, it overlooks a range of potential models, elevating the risk of selection bias in the analysis.	21

In both cases under examination, notable areas exist where these strategies exhibit commendable readiness levels and competencies about the readiness to be applied in practice. Nevertheless, there remain ample opportunities for enhancing their preparedness in this regard.

In the context of Case 1, several noteworthy strengths come to the forefront of the US Hedge Fund. Firstly, the strategy demonstrates a sturdy economic underpinning grounded in company fundamentals, bolstering its overall rationale and viability. However, it is crucial to note a significant shortcoming in neglecting multiple testing, a practice that can substantially inflate return performance. If left unaddressed, this oversight can introduce bias into the results following the execution of numerous backtests.

Furthermore, performance assessment of the actual strategy proves challenging, primarily due to the intricate interplay between quantitative signals and qualitative human judgment. Within the quantitative realm, individual factors warrant evaluation to discern their respective impacts on the observed performance. It is essential to highlight that performance attribution, a critical aspect of strategy evaluation, has not been addressed in this context in Case 1, and it is one of its weaknesses. Similarly, a parallel limitation is evident in Case 2, where the formulated strategy exhibits a noteworthy deficiency by entirely omitting any consideration of performance attribution.

In addition, a notable shortcoming within the strategy in Case 1 lies in the incomplete consideration of several conventional constraints, including position sizing, transaction costs, and capacity. This oversight can be attributed to the relatively low frequency of trades conducted. Acknowledging that these constraints will likely assume greater significance and require comprehensive assessment as the strategy evolves is imperative.

In Case 2, there is also a weakness in model evaluation and selection. The model undergoes a partial review, and a rigorous and systematic approach is conspicuously absent when selecting the optimal model. Consequently, disregarding a comprehensive array of models fails to account for potential selection bias and elevates the associated risk.

8.3 Sensitivity Analysis

The study performed sensitivity analysis to evaluate the effects of changes in input parameters on decision outcomes and to explore nonlinear relationships between inputs and outputs. Five scenarios were analyzed to investigate how pushing these perspectives to extreme levels impacts the overall strategy score, thus achieving the study's goal. Table 34 outlines these five scenarios employed to evaluate the model's resilience.

This sensitivity analysis was explicitly applied to the two case studies within this research. Each scenario involved elevating the value of one perspective to 96% while reducing the values of all other perspectives to 1% to observe how responsive the decision score is to changes in input levels.

Table 44: Implemented Scenarios

Perspectives	Perspective Base Case	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Economic foundations and research	23.00%	96.00%	1.00%	1.00%	1.00%	1.00%
Data	21.00%	1.00%	96.00%	1.00%	1.00%	1.00%
Features	18.00%	1.00%	1.00%	96.00%	1.00%	1.00%
Modeling	17.00%	1.00%	1.00%	1.00%	96.00%	1.00%
Performance	21.00%	1.00%	1.00%	1.00%	1.00%	96.00%

The following tables summarize the results, and corresponding score changes are reported.

In the initial scenario, the economic foundation's aspect was augmented to its highest attainable value of 96%. The outcome reveals a significant increase in the overall score for Case 1, increasing from 55.82 to 88.15, and Case 2 exhibited an increase from 39.33 to

53.45. The favorable alteration observed in Case 2 suggests that in instances where empirical evidence designates the economic foundations perspective as the preeminent determinant, it can be addressed with enhanced confidence. The subsequent table elucidates the modifications in the overall scores for both cases and the corresponding adjustments in the economic foundation’s perspective scores.

In scenario 2, the Data dimension has been elevated to its extreme value of 96%. The resultant findings indicate a surge in the overall score for Case 1, increasing from 55.82 to 79.76, and a similar jump is observed for Case 2, surging from 39.33 to 50.63. Both cases experienced a straight impact. These alterations suggest that the readiness score will be positively affected when empirical evidence underscores the pivotal role of Data perspective factors. Therefore, there is a need for specific measures to enhance strategies in these domains. The subsequent table elucidates the variations in the overall scores for both cases, alongside the corresponding adjustments in the Data perspective scores.

In scenario 3, the Features facet was elevated to its maximum threshold of 96%. The outcomes reveal a reduction in the overall score for Case 1, diminishing significantly from 55.82 to 38.43, and Case 2 incurred a less negative impact, resulting in a decline from 39.33 to 32.59. The adverse shift observed in Cases 1 and 2 underscores that, in instances where empirical evidence shows the paramount importance of Features perspective factors, adjustments can be made with heightened confidence. Furthermore, Case 1 has demonstrated a more significant Feature impact on its score than Case 2. The subsequent table elucidates the variations in the overall scores for both cases and the corresponding adjustments in the Features perspective scores.

Table 45: Scenario 1 Outcomes for Case 1 and Case 2

Scenario 1						
Perspectives	**Factors**	**Perspective value**	**Local Weight**	**Global Weight**	**Case 1**	**Case 2**
Economic foundations and research (23%)	Economic/ financial foundation and scientific approach	0.96	51.0%	0.4896	42.60	8.32
	Investment research question and thesis knowledge		49.0%	0.4704	43.75	43.75
Data (21%)	Data biases and features	0.01	24.0%	0.0024	0.18	0.14
	Data availability and sufficiency		26.0%	0.0026	0.21	0.14
	Data integrity and quality		29.0%	0.0029	0.26	0.12
	Data standards and reflexivity		22.0%	0.0022	0.15	0.12
Features (18%)	Data wrangling	0.01	29.0%	0.0029	0.13	0.23
	Exploratory Data Analysis (EDA)		35.0%	0.0035	0.07	0.07
	Feature engineering and importance		36.0%	0.0036	0.17	0.01
Modeling (17%)	Model over / under-fitting	0.01	29.0%	0.0029	0.15	0.07
	Model interpretability/explainability and complexity		26.0%	0.0026	0.07	0.14
	Hyper-parameter tuning		17.0%	0.0017	0.03	0.03
	Model evaluation and selection		29.0%	0.0029	0.16	0.06
Performance (21%)	Investment constraints	0.01	28.0%	0.0028	0.07	0.11
	Strategy metrics / statistics		28.0%	0.0028	0.09	0.09
	Performance attribution		23.0%	0.0023	0.05	0.01
	Multiple Testing		22.0%	0.0022	0.01	0.04
				Scenario 1 Score	**88.15**	**53.45**
				Base Case	**55.82**	**39.33**
				% Change	**57.92%**	**35.90%**

Table 46: Scenario 2 Outcomes for Case 1 and Case *2*

Scenario 2						
Perspectives	**Factors**	**Perspective value**	**Local Weight**	**Global Weight**	**Case 1**	**Case 2**
Economic foundations and research (23%)	Economic/ financial foundation and scientific approach	0.01	51.0%	0.0051	0.44	0.09
	Investment research question and thesis knowledge		49.0%	0.0049	0.46	0.46
Data (21%)	Data biases and features	0.96	24.0%	0.2304	17.74	13.13
	Data availability and sufficiency		26.0%	0.2496	20.22	13.48
	Data integrity and quality		29.0%	0.2784	25.33	11.41
	Data standards and reflexivity		22.0%	0.2112	14.57	11.19
Features (18%)	Data wrangling	0.01	29.0%	0.0029	0.13	0.23
	Exploratory Data Analysis (EDA)		35.0%	0.0035	0.07	0.07
	Feature engineering and importance		36.0%	0.0036	0.17	0.01
Modeling (17%)	Model over / under-fitting	0.01	29.0%	0.0029	0.15	0.07
	Model interpretability/explainability and complexity		26.0%	0.0026	0.07	0.14
	Hyper-parameter tuning		17.0%	0.0017	0.03	0.03
	Model evaluation and selection		29.0%	0.0029	0.16	0.06
Performance (21%)	Investment constraints	0.01	28.0%	0.0028	0.07	0.11
	Strategy metrics / statistics		28.0%	0.0028	0.09	0.09
	Performance attribution		23.0%	0.0023	0.05	0.01
	Multiple Testing		22.0%	0.0022	0.01	0.04
				Scenario 2 Score	**79.76**	**50.63**
				Base Case	**55.82**	**39.33**
				% Change	**42.89%**	**28.73%**

Table 47: Scenario 3 Outcomes for Case 1 and Case *2*

Scenario 3						
Perspectives	**Factors**	**Perspective value**	**Local Weight**	**Global Weight**	**Case 1**	**Case 2**
Economic foundations and research (23%)	Economic/ financial foundation and scientific approach	0.01	51.0%	0.0051	0.44	0.09
	Investment research question and thesis knowledge		49.0%	0.0049	0.46	0.46
Data (21%)	Data biases and features	0.01	24.0%	0.0024	0.18	0.14
	Data availability and sufficiency		26.0%	0.0026	0.21	0.14
	Data integrity and quality		29.0%	0.0029	0.26	0.12
	Data standards and reflexivity		22.0%	0.0022	0.15	0.12
Features (18%)	Data wrangling	0.96	29.0%	0.2784	12.81	22.55
	Exploratory Data Analysis (EDA)		35.0%	0.3360	7.06	7.06
	Feature engineering and importance		36.0%	0.3456	16.24	1.38
Modeling (17%)	Model over / under-fitting	0.01	29.0%	0.0029	0.15	0.07
	Model interpretability/explainability and complexity		26.0%	0.0026	0.07	0.14
	Hyper-parameter tuning		17.0%	0.0017	0.03	0.03
	Model evaluation and selection		29.0%	0.0029	0.16	0.06
Performance (21%)	Investment constraints	0.01	28.0%	0.0028	0.07	0.11
	Strategy metrics / statistics		28.0%	0.0028	0.09	0.09
	Performance attribution		23.0%	0.0023	0.05	0.01
	Multiple Testing		22.0%	0.0022	0.01	0.04
				Scenario 2 Score	**38.43**	**32.59**
				Base Case	**55.82**	**39.33**
				% Change	**-31.15%**	**-17.14%**

In scenario 4, the Modeling dimension was elevated to its maximum attainable value of 96%. The outcomes indicate a downswing in the overall score for both Case 1 and Case 2, reducing from 55.82 to 40.83, and a parallel decrease is observed for Case 2, descending from 39.33 to 30.93. In this scenario, both Case 1 and Case 2 demonstrated very close negative downward shifts of -26% and -21% in the scores, respectively. The subsequent table illustrates the alterations in the overall scores for both cases and the corresponding shifts in the Modeling perspective scores. In scenario 4, the Performance dimension was elevated to its maximum value of 96%. The results indicate an upturn in the overall score for Case 1, rising from 55.82 to 61.37, and a more significant increase is observed for Case 2, escalating from 39.33 to 54.02. The favorable adjustments experienced by both Case 1 and Case 2 suggest that, in cases where empirical evidence underscores the paramount importance of Performance factors, endeavors can be pursued with enhanced confidence.

Furthermore, Case 2 can gain more favorable outcomes by considering this factor as a focal point for improvement. The subsequent table elucidates the alterations in the overall scores for both cases and the corresponding shifts in the Performance perspective scores.

As the table below shows, Case 1 reaches its highest point when the Economic Foundations and Research perspective are enhanced, resulting in a substantial increase of 57.92%. Conversely, its lowest performance is observed when the Modeling perspective is emphasized, leading to a notable decrease of -31.15%. In contrast, Case 2 exhibits sensitivity primarily to the Features and Performance perspectives. When the Features perspective is amplified, Case 2's score experiences a significant improvement to 54.02, representing a noteworthy increase of 37.35%. Conversely, the emphasis on the

Table 48: Scenario 4 Outcomes for Case 1 and Case *2*

Scenario 4						
Perspectives	**Factors**	**Perspective value**	**Local Weight**	**Global Weight**	**Case 1**	**Case 2**
Economic foundations and research (23%)	Economic/ financial foundation and scientific approach	0.01	51.0%	0.0051	0.44	0.09
	Investment research question and thesis knowledge		49.0%	0.0049	0.46	0.46
Data (21%)	Data biases and features	0.01	24.0%	0.0024	0.18	0.14
	Data availability and sufficiency		26.0%	0.0026	0.21	0.14
	Data integrity and quality		29.0%	0.0029	0.26	0.12
	Data standards and reflexivity		22.0%	0.0022	0.15	0.12
Features (18%)	Data wrangling	0.01	29.0%	0.0029	0.13	0.23
	Exploratory Data Analysis (EDA)		35.0%	0.0035	0.07	0.07
	Feature engineering and importance		36.0%	0.0036	0.17	0.01
Modeling (17%)	Model over / under-fitting	0.96	29.0%	0.2784	13.92	6.96
	Model interpretability/explainability and complexity		26.0%	0.2496	6.24	13.73
	Hyper-parameter tuning		17.0%	0.1632	2.77	2.77
	Model evaluation and selection		29.0%	0.2784	15.59	5.85
Performance (21%)	Investment constraints	0.01	28.0%	0.0028	0.07	0.11
	Strategy metrics / statistics		28.0%	0.0028	0.09	0.09
	Performance attribution		23.0%	0.0023	0.05	0.01
	Multiple Testing		22.0%	0.0022	0.01	0.04
				Scenario 2 Score	**40.83**	**30.93**
				Base Case	**55.82**	**39.33**
				% Change	**-26.86%**	**-21.37%**

Table 49: Scenario 5 Outcomes for Case 1 and Case *2*

Scenario 5						
Perspectives	**Factors**	**Perspective value**	**Local Weight**	**Global Weight**	**Case 1**	**Case 2**
Economic foundations and research (23%)	Economic/ financial foundation and scientific approach	0.01	51.0%	0.0051	0.44	0.09
	Investment research question and thesis knowledge		49.0%	0.0049	0.46	0.46
Data (21%)	Data biases and features	0.01	24.0%	0.0024	0.18	0.14
	Data availability and sufficiency		26.0%	0.0026	0.21	0.14
	Data integrity and quality		29.0%	0.0029	0.26	0.12
	Data standards and reflexivity		22.0%	0.0022	0.15	0.12
Features (18%)	Data wrangling	0.01	29.0%	0.0029	0.13	0.23
	Exploratory Data Analysis (EDA)		35.0%	0.0035	0.07	0.07
	Feature engineering and importance		36.0%	0.0036	0.17	0.01
Modeling (17%)	Model over / under-fitting	0.96	29.0%	0.2784	13.92	6.96
	Model interpretability/explainability and complexity		26.0%	0.2496	6.24	13.73
	Hyper-parameter tuning		17.0%	0.1632	2.77	2.77
	Model evaluation and selection		29.0%	0.2784	15.59	5.85
Performance (21%)	Investment constraints	0.96	28.0%	0.2688	6.72	10.75
	Strategy metrics / statistics		28.0%	0.2688	8.33	8.33
	Performance attribution		23.0%	0.2208	4.86	0.66
	Multiple Testing		22.0%	0.2112	0.84	3.59
				Scenario 2 Score	**61.37**	**54.02**
				Base Case	**55.82**	**39.33**
				% Change	**9.94%**	**37.35%**

Performance perspective leads to a sharp decline of -21.37%, reducing Case 2's score to 30.93. These fluctuations offer valuable insights for quantitative investment teams, facilitating more informed and systematic decision-making in developing investment strategies that rely exclusively on data science and machine learning techniques. Moreover, the adaptability of such frameworks enables investment professionals to apply them to different asset classes and geographic regions while tailoring the perspectives to specific needs. Significantly, these systematic approaches diminish the reliance on arbitrarily chosen methods when designing quantitative investment strategies that leverage artificial intelligence.

Table 50: Summary of Scenario Analysis

Scenario	Boosted Perspective	Case 1 Score	Case 1 Score Change %	Case 2 Score	Case 2 Score Change %
Base Case Scenario	None	55.82	None	39.33	None
Scenario 1	Economic foundations and research	88.15	57.92%	53.45	35.90%
Scenario 2	Data	79.76	42.89%	50.63	28.73%
Scenario 3	Modeling	38.43	-31.15%	32.59	-17.14%
Scenario 4	Performance	40.83	-26.86%	30.93	-21.37%
Scenario 5	Features	61.37	9.94%	54.02	37.35%

8.4 Recommended Improvements

This research aims to develop a model to assist investment and asset management firms evaluate their DS/ML investment strategy readiness. This model seeks to identify and prioritize the critical factors influencing the strategy development process and to pinpoint any vulnerabilities that may impede project success. Doing so enables investment firms to implement improvements and corrective measures based on the identified weaknesses.

The strengths and weaknesses section within this chapter thoroughly examines each case study, highlighting areas of both strength and weakness. Within this section, there is a demonstration of how the research model can provide added value and enhance readiness scores, ultimately improving the likelihood of success. The research not only seeks to identify weaknesses but also strives to offer comprehensive guidelines and recommendations for their amelioration. These enhancements are explicitly targeted at areas where strategies have received low scores, and appropriate recommendations are provided.

The subsequent tables outline potential enhancements for both case studies, considering their respective scores. Investment teams can approach these enhancements with varying degrees of conservatism, moderation or by implementing more significant changes. Teams can refer to value curves to determine their current position and ascertain the next enhancement level for each model element and the optimal level for a particular factor. This process can be iterative, starting with conservative changes and progressing step by step until the desired readiness score and confidence level are attained.

Tables 51 and 52 show the tactical actions that can be taken for each case to improve their scores and performance. (both tables 51 and 52 are available as supplemental files in the Appendix D.) The central insight from this assessment procedure is that tactical actions will be identified and executed after initial evaluation and achieving the scores.

Furthermore, value curves will be pivotal as a reference point for further enhancements. They will be directed towards setting objectives at each iteration for various attributes, facilitating an assessment of the current state versus the anticipated future state of the strategy. This approach will illuminate specific facets of the investment strategy that warrant focused attention and necessitate substantial improvement efforts.

Adhering to this systematic approach will enable investment teams to monitor diverse facets of the evolution of the investment strategy over time. It ensures the adoption of a rigorous and scientifically grounded methodology in developing a quantitative strategy.

Table 51: Improvement Simulation – Case 1

Improvement Simulation							
Perspectives	Factors	Global Weight	Value Curve (VC) Score	Final Score (Weighs * VC)	New Value Curve (VC) Score	New Final Score (Weighs * VC)	Tactical Action
Economic foundations and research (23%)	Economic/ financial foundation and scientific approach	11.7%	87	10.18	87	10.18	-
	Investment research question and thesis knowledge	11.3%	93	10.51	93	10.51	-
Data (21%)	Data biases and features	5.0%	77	3.85	77	3.85	-
	Data availability and sufficiency	5.5%	81	4.46	81	4.46	-
	Data integrity and quality	6.1%	91	5.55	91	5.55	-
	Data standards and reflexivity	4.6%	69	3.17	69	3.17	-
Features (18%)	Data wrangling	5.2%	46	2.39	46	2.39	-
	Exploratory Data Analysis (EDA)	6.3%	21	1.32	21	1.32	-
	Feature engineering and importance	6.5%	47	3.06	47	3.06	-
Modeling (17%)	Model over / under-fitting	4.9%	50	2.45	50	2.45	-
	Model interpretability/explainability and complexity	4.4%	25	1.10	25	1.10	-
	Hyper-parameter tuning	2.9%	17	0.49	17	0.49	-
	Model evaluation and selection	4.9%	56	2.74	56	2.74	-
Performance (21%)	Investment constraints	5.9%	25	1.48	40	2.36	Consideration of investment constraints could be done in parallel with the hypothetical quantitative trading model.
	Strategy metrics / statistics	5.9%	31	1.83	31	1.83	-
	Performance attribution	4.8%	22	1.06	52	2.50	Breaking down some of the quantitative factors as well as increased automation of the trading decisions will allow the performance to be attributed to some extent.
	Multiple Testing	4.6%	4	0.18	10	0.46	Creating a strategy with purely quantitative trading decisions would allow comparison with the human performance and could give some insight into how the strategy might perform broadly, though specfic tuning of parameters would probably not be justifiable.
Overall Score				55.82	→	58.42	-

Table 52: Improvement Simulation – Case 2

Improvement Simulation							
Perspectives	**Factors**	**Global Weight**	**Value Curve (VC) Score**	**Final Score (Weighs * VC)**	**New Value Curve (VC) Score**	**New Final Score (Weighs * VC)**	**Tactical Action**
Economic foundations and research (23%)	Economic/ financial foundation and scientific approach	11.7%	17	1.99	17	1.99	-
	Investment research question and thesis knowledge	11.3%	93	10.51	93	10.51	-
Data (21%)	Data biases and features	5.0%	57	2.85	57	2.85	-
	Data availability and sufficiency	5.5%	54	2.97	54	2.97	-
	Data integrity and quality	6.1%	41	2.50	41	2.50	-
	Data standards and reflexivity	4.6%	53	2.44	53	2.44	-
Features (18%)	Data wrangling	5.2%	81	4.21	81	4.21	-
	Exploratory Data Analysis (EDA)	6.3%	21	1.32	21	1.32	-
	Feature engineering and importance	6.5%	**4**	0.26	**19**	1.24	Run feature engineering and importance to see which features are creating value
Modeling (17%)	Model over / under-fitting	4.9%	25	1.23	25	1.23	-
	Model interpretability/explainability and complexity	4.4%	55	2.42	55	2.42	-
	Hyper-parameter tuning	2.9%	**17**	0.49	**20**	0.58	Tune the parameters of the model to improve the overall model performance
	Model evaluation and selection	4.9%	21	1.03	21	1.03	-
Performance (21%)	Investment constraints	5.9%	40	2.36	40	2.36	
	Strategy metrics / statistics	5.9%	31	1.83	31	1.83	
	Performance attribution	4.8%	**3**	0.14	**22**	1.06	Decompose risk/return to shed light on drivers of value
	Multiple Testing	4.6%	17	0.78	17	0.78	
Overall Score				**39.33**	→	**41.31**	-

CHAPTER 9. RESEARCH VALIDATION

In order to uphold the integrity and adhere to preceding doctoral dissertations, this study comprehensively addresses three research validity dimensions: Content, Construct, and Criterion [267].

As the primary facet of research validity, content validity necessitates meticulous consideration throughout the research process. In this study, we validated content by forming expert panels. These panels were convened to ascertain the suitability and relevance of the perspectives and criteria identified in the literature concerning the research's overarching objectives. Furthermore, experts were allowed to augment the content by proposing novel factors, thereby enhancing the content validity of the study.

Construct validity assesses the capability and aptness of the developed model to address the research's focal subject matter. Throughout this investigation, input and recommendations from many subject matter experts, academic faculty members, and doctoral students were collected to validate the construct of the research model. Ultimately, the final construct was validated through a disagreement analysis process, ensuring the model did not exhibit substantial disparities in perspectives among various experts.

Criterion validity was addressed both during and after the analysis of research findings. Within this study, multiple academic faculty members and subject matter experts offered feedback concerning the accuracy of research outcomes and validating the results and recommendations. Moreover, during the case study analysis phase, hypothetical companies were devised to test the research model's viability rigorously.

CHAPTER 10. DISCUSSION AND CONCLUSION

10.1 Discussion and Recent Research

The problem statement in Chapter 1 deliberated upon the prevalent issues of substantial failure of data science and machine learning-driven quantitative investment strategies. Hence, readiness assessment before allocating capital and live trading has emerged as a viable strategy for mitigating the risk of failure. Nevertheless, the gap analysis conducted in Chapter 3 revealed that research on backtest protocols, experiments, and systematic processes for evaluating readiness for ML-DS-driven strategies is scarce. Furthermore, the investigation conducted on the existing literature showed that this gap is even wider in the investment decision-making process of developing ML-DS strategies.

For instance, Blitz et al. investigated the emerging literature of machine learning in a wide range of asset pricing applications to demonstrate its capability in this domain. They evaluated the promises and pitfalls of applying machine learning from a practical standpoint by focusing on methodological design choices that impact predictive outcomes. They stated that although machine learning models are data-driven, the users still need to make essential choices in the strategy development process, and these design choices significantly impact the model's outcomes. For example, they argued that rigorous research governance and protocol are unavoidable to successfully navigate the promising results of machine learning applications in asset management without falling into the danger of data mining. They highlighted seven pillars of a healthy research protocol from Arnott et al. [54]: motivation for research, multiple testing, data quality and sample choice, cross-validation, model dynamics, model complexity, and research culture. [268]

Moreover, Mirete-Ferrer et al. conducted a comprehensive review of machine learning effects in asset management. [269] They highlighted the importance of considering financial markets' exceptional nonlinear and dynamic characteristics when using machine learning to invest. Unlike other scientific disciplines, ML in finance faces many unique challenges that quantitative finance experts have constantly faced. They also showed the broad spectrum of applied ML methods, target markets, and performance criteria in different use cases, such as factor investing, portfolio management, algorithmic trading, and price forecasting. On the other hand, they argued about the unique challenges that must be addressed, including standard datasets, reproducibility, multimodal data, and heterogeneous architectures. For instance, they show that a broad range of new models is applied in papers and only a few build upon a solid groundwork to improve them. The state is that finding common patterns and unifying those diverse architectures could have a beneficial effect. In addition, regarding reproducibility, they state that there is no standard methodology or framework for model training and benchmarking, which further highlights the importance of such frameworks. This hurts reproducibility since more models are difficult, if not impossible, to compare against each other, which necessitates the core role of having a systematic approach.

Furthermore, Shukla et al. [270] surveyed data science and artificial intelligence applications in financial decisions. They argued that to utilize the strengths of financial econometrics and data science techniques, in many cases, the hybrid models outperformed traditional econometric models, representing a vast opportunity for applying data science models in financial decisions. This aligns with the earlier argument in this research, as the author indicated the importance of using these methods in conjunction with more traditional econometric approaches to achieve more excellent economic results out of developed ML-DS-driven investment strategies.

Additionally, Avramove et al. examined the capabilities of machine learning to extract signals from hard-to-arbitrage stocks. They indicated that ML-based performance deteriorates in the presence of trading costs due to high turnover and extreme positions in tangency portfolios. This is another example of considering the proper set of constraints and performance measures once using such mechanisms in developing investment strategies. With the recent developments in financial technologies, their findings support the concept that ML-DS-driven investments could hold considerable promise for asset management. Their research provides evidence to show the importance of machine learning applications in asset management and proposes a list of back-testing protocols for academic research. This paper enriches the academic and policy discussions surrounding the adoption of machine learning in asset management, including economic considerations and restrictions, interpretability, sustainability of new trading signals, and the potential regulatory and supervisory implications of applying these methods. [271]

Similarly, there are recent research papers surrounding this topic which include Tang et al. [272], Kaczmarczyk et al. [273], Olorunnimbe and Viktor [274], and Nazareth et al. [275]. Furthermore, an emerging field of inquiry started from the dawn of large language models and one of its most well-known models, “chatgpt,” in 2023. It worth highlighting the works of Wang et al. (2023), Ko and Lee (2023), Dowling and Lucey (2023), Umer and Khan

(2023), Lu et al. (2023), Zaremba and Demir (2023), Aldridge (2023), and Feng et al. (2023). [276], [277], [278], [279], [280], [281], [282]

10.2 Conclusion

This study's distinctive feature lies in its rigorous integration of an extensive literature review connecting data science, machine learning, and investment decision-making in developing quantitative investment strategies. This comprehensive examination has resulted in identifying and quantifying perspectives and their corresponding criteria, which are considered significant. This research has successfully constructed a holistic and systematic framework that facilitates understanding the assessment process by combining multiple and diverse perspectives. This model can generate valuable insights for enhancing DS-ML-driven investment strategy development, quantitative finance teams, asset management companies, and policymakers.

The principal objective of this research is to discern pivotal factors to evaluate the readiness to develop strategies in which data science and machine learning play vital roles. Thus, the proposed framework serves as a valuable tool for investment management firms by which they can evaluate the readiness of their investment strategies in which they have applied machine learning and data science to develop them.

Many such strategies and projects have encountered significant setbacks and even failure. Many of these initiatives have faltered in their pursuit of envisioned objectives, ultimately resulting in the termination or shutdown of such strategies. However, there remains a notable shortage of comprehensive research studies examining the multifaceted nature of investment decision-making that impacts the success of such projects rather than finding the most profitable strategies. This gap includes a lack of robust frameworks designed to augment the success rate of DS-ML-driven investment strategies. The scoring model developed in this study meticulously considers strategy readiness through a multi-perspective lens. It addresses economic foundations, data, features, modeling, and performance aspects of strategy development. In turn, this approach makes investment firms capable of more systematic decision-making while designing new strategies and consequently gaining insights into the areas to rethink the investment process and improve the outcomes.

Furthermore, this research provides valuable insights from case applications, encompassing the following observations.

Although cases have developed different strategies and invested in other markets with specific characteristics, this research demonstrates how they can benefit from such frameworks in their investment decision-making and consequently improve their strategy. Both firms involved in this research confirmed that they had found the model and insights generated by that useful, which provided them with a bird's s-eye view of their strategy development process.

Overall, Case 1 performed better in multiple perspectives than Case 2. However, both cases vividly demonstrate economic foundations and data as the most significant factors in strategy development. Another commonality between the two cases is the ignorance of investment teams about the performance attribution to shed light on the drives of risk and return. This specific aspect of strategy development becomes even more significant in DS-ML-driven strategies as most people still see them as black box models and are keen on seeing how they make decisions and generate results.

These case studies offer insights into nuanced perspectives on the multifaceted landscape of financial data science and machine learning, an interdisciplinary research and practice domain.

10.3 Recommendations

One crucial research recommendation is to emphasize the adoption of rigorous and systematic assessment frameworks to enhance the performance of DS/ML-driven investment strategies. Incorporating structured evaluation processes can provide a robust foundation for decision-making and ensure that strategies are consistently optimized and aligned with evolving market dynamics.

According to empirical evidence, it is recommended to strongly emphasize establishing robust economic foundations and applying high-quality data in developing investment strategies. This recommendation stems from the observation that neglecting these factors can lead to suboptimal investment and even adverse financial outcomes. Prioritizing acquiring reliable economic data and a solid analytical foundation can significantly improve the effectiveness of DS/ML-driven investment approaches.

Another research recommendation centers around the refinement of research design and the meticulous implementation of protocols within investment decision-making. This emphasis on methodological rigor holds the potential to yield substantial value rather than just pursuing purely profitable strategies. A well-structured research design can lead to more accurate and reliable investment decisions, reducing the likelihood of costly errors.

For those applying DS/ML methodologies to investment strategy development, it is advisable to incorporate rigorous scientific validation checkpoints. This recommendation is based on the premise that systematic validation can enhance the reliability and credibility of the developed strategies. By applying rigorous validation procedures and shifting from finding profitable investment strategies, researchers and practitioners can gain greater confidence in their applicability to real-world financial scenarios.

Finally, it is essential to recognize that the financial domain of DS/ML is distinct from other sectors using similar techniques. Therefore, this study recommends carefully considering the unique characteristics of financial markets when deploying these technologies. Failure to do so may result in outcomes vastly different from those encountered during the developmental phase. Understanding and adapting to these intricacies is essential to maximize the effectiveness of DS/ML in the financial sector.

10.4 Expected Contributions

This research generates significant consideration for the newborn field of financial data science and machine learning. The contributions of this research are two-fold: academic and professional.

10.4.1 Academic Contributions

From an academic perspective, this study significantly contributes to the evolving domains of fintech and ML/DS research management, primarily by addressing the challenges related to research process quality and result reliability by utilizing a systematic, multi-criteria decision model. The intricate layers inherent in the Hierarchical Decision Model (HDM) and its associated factors establish a robust framework for consolidating and validating expert judgments to fill a crucial gap in ML/DS research projects.

Despite the expanding applications of ML/DS in asset management and the emergence of numerous successful use cases, the absence of robust and consistent frameworks for ensuring research quality and result reliability is glaring. This research tries to enrich the understanding of how investment firms can assess and implement best practices by introducing a decision support system. As corroborated by the literature review and evident in the gap analysis phase, a lack of systematic and comprehensive protocols guarantees the integrity of research outcomes in this domain. In response, the proposed model in this study serves as a valuable tool for academia and practitioners to enable them to identify and prioritize critical factors in financial data science research applications in investment strategy development. Moreover, this research advances investment companies' knowledge and comprehension, reducing failure rates in DS/ML research projects and investment strategies. In essence, this study provides a coherent framework derived from a comprehensive literature review and the collective wisdom of domain experts.

This research yielded a model developed for assessing the readiness of ML-DS-driven investment strategies. This investigation extensively incorporated current scholarly publications and expert judgments to bridge the gaps and provide a solution to research inquiries in this specific domain. The HDM was applied as the methodology to create a hierarchical representation of the extracted and validated perspectives and criteria, and it was used to elicit expert judgments to identify the relative importance of each criterion. Furthermore, two case studies were implemented to demonstrate the practical aspects of the proposed framework. The following two tables summarize how this research achieves its objectives by filling the gaps and showing research outputs and associated contributions.

Table 53: Summary of The Research Gaps and The Research Contributions

<table>
<tr><th>Research Gaps</th><th>Contributions</th></tr>
<tr><td>There is a lack of multi-criteria holistic studies to assess the financial DS/ML research.</td><td rowspan="2">The HDM framework proposed in this research is a systematic and structured methodology to evaluate the quality and readiness of ML/DS-based investment strategies.</td></tr>
<tr><td>Scarcity of studies systematically evaluating the productivity and quality of financial DS/ML research projects.</td></tr>
<tr><td>There is a lack of studies that highlight the most critical factors impacting the reliability and quality of financial DS/ML research projects.</td><td rowspan="2">This is conducted based on a comprehensive literature review and elicitation of experts' judgments on identifying, validating, and quantifying the most critical factors impacting the reliability and readiness of an ML/DS-driven investment research project and its outcome, which is an investment strategy. Furthermore, the relative importance of criteria is determined by a diverse group of experts, which is essential as this problem is essentially interdisciplinary.</td></tr>
<tr><td>Lack of studies based on the collective intelligence and the expert's judgments and present the importance level of the factors and perspectives considered in the assessment.</td></tr>
</table>

There is an extensive literature gap in economic research to address the challenges of the promising financial DS/ML field.	This research contributes to financial data science and machine learning with a specific focus on investment strategy development by developing a readiness assessment tool in investment management using a robust decision-making model framework.

Table 54: Summary of the research outputs and the research contributions

Research Outputs	Research Contributions
Identifying the perspectives and criteria for assessing financial data science / ML research projects.	This research is constructed upon an exhaustive examination of contemporary academic literature, complemented by integrating insights provided by subject matter experts.
Identify the relative importance of each perspective and criteria factor in the assessment process.	The research discerned the paramount factors influencing the adoption of machine learning and data science techniques in the investment process and subsequently integrated experts' judgments to ascertain their respective significance and priority levels.
Provide a tool for investment companies to assess their capabilities to overcome challenges with the existing financial data science / ML research projects and to be able to systematically evaluate ML/DS-based funds before getting exposure to those funds.	This research introduced a robust decision-making model tool, namely the HDM framework, for readiness assessment of investment strategies to embrace machine learning technology.
Highlight the disagreement level among experts from different fields and backgrounds on the relative importance of the assessment factors.	The experts' disagreement level has been demonstrated to fall within acceptable bounds. The invited experts for participation possess a broad spectrum of expertise and exhibit varying degrees of exposure to the subject matter. (See chapter 5)
Examine the effectiveness and practicality of the model for assessing the productivity and quality of financial data science / ML research projects and proposed investment strategies.	This model was applied to two hedge funds with entirely different strategies as case studies for the study and has proven its capability to assess their readiness (See Chapter 87)

10.4.2 Professional Contributions

From a professional standpoint, this research offers a valuable assessment support system tailored for investment management firms that evaluate the outcomes of quantitative strategies employing DS/ML methodologies. This framework presents a dependable mechanism that instills greater confidence within quantitative investment and financial data science teams regarding the efficacy of their developed strategies. It adeptly tackles the prevalent challenges encountered by these teams by aiding in identifying potential pitfalls that might otherwise elevate the risk of erroneous discoveries.

As previously stated, the triumph of ML/DS-based quantitative strategies hinges not solely on data quality or model complexity but on a holistic view of the entire research process, which includes multiple influencing factors. Consequently, decision support mechanisms, such as the one proposed in this research, empower investment teams to conceive more coherent financial experiments, enhancing the likelihood of success. These models equip investment companies with the tools and perspectives necessary for a more comprehensive analysis of various factors and their respective contributions to the resulting outcomes.

This research holds paramount significance due to its systematic approach to illuminating the investment research process. Moreover, it underscores the critical yet often overlooked role of technology management tools, exemplified by the decision model proposed herein, in ML/DS investment research studies. Consequently, this study pioneers a holistic, multi-perspective approach to equipping investment management professionals with the benefits of decision science methodologies.

In summary, the overarching contributions can be demonstrated as follows:

Academically:

- This study enhances knowledge in technology management and financial data science by furnishing a systematic framework for assessing the quality and reliability of ML-based investment strategies.

- The proposed assessment tool facilitates the expanding understanding of how quantitative investment teams evaluate ML-based investment strategy research and apply scientific processes to develop more robust strategies.
- This research fills the gap by offering a methodical and comprehensive study of the pivotal factors and their impact on the investment research process, which uses data science and ML.

Professionally:

- We provide a framework researchers and practitioners can use as a decision support system.
- We establish a multi-criteria evaluation approach for managing financial data science/machine learning research projects.
- We empower investment research teams to address problems comprehensively and systematically, and examining them from diverse perspectives enhances the evaluation of ML-based investment strategies.

In addition, it is notable to mention the feedback this research investigator received from experts involved in the case studies. The investment team in Case 1 stated their findings and experience working on this project and applying the framework in their investment process. They said, "Participating in your research project was a valuable experience for our team (Snowstorm Capital, LLC). It significantly enhanced our understanding of how our strategy aligns with the larger context. The template view, which summarized the strengths and weaknesses of the cases, effectively highlighted the challenges in our strategy. While certain aspects of our investment strategy are fixed, this project informed us on which components we have the potential to enhance. Working on a data science-based strategy with a small team is resource-intensive, and as a result, we often need to make tough choices about where to allocate our efforts for improvement. We found the Case Study scoring methodology useful in highlighting where the most substantial impact could be achieved. As a result, we plan to incorporate this methodology into our decision-making process when evaluating larger development projects in the future. In summary, this research project not only heightened our awareness of our strategy's strengths and weaknesses but also provided a practical tool for prioritizing improvements and enhancing the impact of our development work. We are excited to leverage this newfound knowledge to drive our strategy forward effectively."

Similarly, the investor of Case 2 stated how this research helped him and his team to make more informed decisions in their investment research and development. He said: "During my career in finance, I have always read only about metrics to measure in different ways the performance and risks of an investment strategy, but thinking about a score regarding not only in attribution metrics but rather in the strategy methodology and design framework helped me add these considerations for some projects I have been working on. I believe having frameworks like the ones proposed by Farshad could help us solve the replicability crisis that many systematic papers claim to achieve and lead us to a more scientific era for investment decisions. Appreciate the opportunity to exchange ideas on this topic and hope we can participate in the evolution of markets." This statement clearly shows that the method proposed in this research has been helpful for the investment team of Case 2 in their practical endeavors.

10.5 Limitations

One limitation of this research revolves around the inherent behavioral biases that decision-makers, including experts, may exhibit when making critical decisions. Even though these individuals possess specialized knowledge in their respective fields, it is essential to acknowledge that expertise does not necessarily guarantee the most optimal decisions, even within their areas of specialization. The impact of these biases on the decision-making process can introduce a layer of subjectivity and potential deviations from what might be considered the ideal choice. To counter this limitation, a pivotal aspect of this research involves the meticulous selection and composition of expert panels. The objective is to ensure that the chosen experts exhibit expertise and can make rational, unbiased decisions within the study context. By doing so, the research aims to address the challenges associated with decision biases and enhance the reliability of the outcomes.

In addition, disagreements among decision-makers are a recurrent challenge, especially when dealing with complex decisions in contexts characterized by high levels of uncertainty. This issue also applies to experts, who may hold different opinions or perspectives on complex matters. These disagreements can introduce variability and subjectivity into the decision-making process and potentially affect the consistency and robustness of the model's results. It is essential to recognize that expert consensus may not always be attainable, and strategies to manage and incorporate these divergent views into the decision-making framework must be established. This inherent variability shows a limitation and underscores the need for comprehensive sensitivity analyses and an exploration of uncertainty in the research.

Another limitation lies in the ambiguous and contentious definition of an "expert," which even stands out more in interdisciplinary fields of study. Defining what constitutes expertise can be challenging, as it may vary across disciplines and evolve. Including experts from diverse backgrounds in an interdisciplinary study can complicate matters further because it may be challenging to establish a clear, universally applicable criterion for expertise. Consequently, there is a degree of subjectivity and uncertainty associated with the selection and categorization of experts. This limitation necessitates a rigorous and transparent process for expert identification and selection to ensure that the chosen individuals genuinely possess the required expertise and can contribute meaningfully to the research.

Lastly, the susceptibility of the HDM model to significant fluctuations in the values of variables is a noteworthy limitation that warrants consideration. Extreme changes in variable values can significantly disrupt the stability and reliability of the model, which raises doubts about its practical applicability. The lack of robustness in the model can undermine its utility in real-world decision-making scenarios, where consistency and dependability are paramount. Therefore, this limitation demonstrates the importance of conducting sensitivity analyses and stress testing to assess the model's resilience to extreme variations in variable inputs. These analyses can inform researchers and practitioners about the model's limitations and guide efforts to enhance its robustness for practical use.

10.6 Future Research

AI and machine learning continue to evolve rapidly, so it is essential to consider re-evaluating the existing investment model in response to new and emerging factors. These ever-advancing technologies bring forth novel data sources, algorithms, and market dynamics that could significantly impact investment strategies. Continuous assessment and adaptation of the model are crucial to ensure its ongoing relevance and effectiveness in navigating the ever-changing financial landscape.

Expanding the scope of research by applying similar methodologies in different regions and asset classes represents a valuable avenue for future research. Comparative studies across diverse geographical regions and asset categories can provide insights into the challenges and opportunities in various investment contexts. This comparative approach enriches our understanding of investment strategy development and fosters the development of versatile models capable of adapting to different market conditions and dynamics.

Further research should focus on identifying and investigating the unique factors specific to various asset classes and designing tailored models to accommodate these nuances. Different asset classes, such as equities, fixed income, or alternative investments, often exhibit distinct characteristics and risk profiles. Developing specialized models that account for these idiosyncrasies can lead to more precise and effective investment strategies within each asset class and enhance overall portfolio performance.

Efforts should be directed towards evaluating the impact of systematic decision models on formulating new investment policies to harness their full potential. Understanding how these models inform investment decisions, shape risk management practices, and guide portfolio construction is pivotal for policymakers and industry practitioners. Insights gained from such research can facilitate the development of policies that harness AI and machine learning to optimize investment strategies and align them with evolving market conditions.

Finally, researchers should explore the value creation by these models as collaborative frameworks for investment companies. Investigating their potential as central hubs for fostering interdisciplinary collaboration within organizations can drive innovation and synergy between different areas of expertise. By building diverse teams around these models, investment companies can break down silos and promote cross-functional collaboration, which leads to novel investment strategies and a more competitive edge in the dynamic financial industry landscape.

REFERENCES

[1] “The world’s largest fund managers - 2019,” Willis Towers Watson. Accessed: Apr. 12, 2020. [Online]. Available: https://www.thinkingaheadinstitute.org/en/Library/Public/Research-and-Ideas/2019/10/P_I_500_2019_Survey

[2] “AI Pioneers in Investment Management,” CFA Institute. Accessed: Apr. 12, 2020. [Online]. Available: /en/research/industry-research/ai-pioneers-in-investment-management

[3] PricewaterhouseCoopers, “Asset & Wealth Management Revolution: Embracing Exponential Change,” PwC. Accessed: Apr. 12, 2020. [Online]. Available: https://www.pwc.com/gx/en/industries/financial-services/asset-management/publications/asset-wealth-management-revolution.html

[4] PricewaterhouseCoopers, “Asset Management 2020: A Brave New World,” PwC. Accessed: Apr. 12, 2020. [Online]. Available: https://www.pwc.com/gx/en/industries/financial-services/asset-management/publications/asset-management-2020-a-brave-new-world.html

[5] “Global AI in Asset Management Market By Technology, By Deployment Mode, By Application, By End User, By Region, Industry Analysis and Forecast, 2020 - 2026.” Accessed: Dec. 01, 2020. [Online]. Available: https://www.reportlinker.com/p05975407/Global-AI-in-Asset-Management-Market-By-Technology-By-Deployment-Mode-By-Application-By-End-User-By-Region-Industry-Analysis-and-Forecast.html?utm_source=GNW

[6] M. Kolanovic and R. T. Krishnamachari, “Machine Learning and Alternative Data Approach to Investing,” p. 280, 2017.

[7] J. Simonian and F. J. Fabozzi, “Triumph of the Empiricists: The Birth of Financial Data Science,” *J. Financ. Data Sci.*, Jan. 2019, doi: 10.3905/jfds.2019.1.010.

[8] M. Lopez de Prado, “Ten Financial Applications of Machine Learning (Presentation Slides),” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3197726, Jun. 2018. Accessed: Jan. 08, 2020. [Online]. Available: https://papers.ssrn.com/abstract=3197726

[9] M. Lopez de Prado, “The 7 Reasons Most Machine Learning Funds Fail (Presentation Slides),” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3031282, Sep. 2017. Accessed: Jan. 08, 2020. [Online]. Available: https://papers.ssrn.com/abstract=3031282

[10] M. Lopez de Prado, “Advances in Financial Machine Learning (Chapter 1),” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3104847, Jan. 2018. Accessed: Jan. 08, 2020. [Online]. Available: https://papers.ssrn.com/abstract=3104847

[11] M. Lopez de Prado, “Building Diversified Portfolios that Outperform Out-of-Sample,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2708678, May 2016. Accessed: Jan. 08, 2020. [Online]. Available: https://papers.ssrn.com/abstract=2708678

[12] S. Gu, B. T. Kelly, and D. Xiu, “Empirical Asset Pricing via Machine Learning,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3159577, Sep. 2019. Accessed: Jan. 08, 2020. [Online]. Available: https://papers.ssrn.com/abstract=3159577

[13] D. M. Q. Nelson, A. C. M. Pereira, and R. A. de Oliveira, “Stock market’s price movement prediction with LSTM neural networks,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, May 2017, pp. 1419–1426. doi: 10.1109/IJCNN.2017.7966019.

[14] K. Chen, Y. Zhou, and F. Dai, “A LSTM-based method for stock returns prediction: A case study of China stock market,” in *2015 IEEE International Conference on Big Data (Big Data)*, Oct. 2015, pp. 2823–2824. doi: 10.1109/BigData.2015.7364089.

[15] R. Aadith Narayan, B. D. Dayana, B. Yagneshwaran, M. R. Vignesh Babu, and K. A. V. Krishna, “Stock value prediction using machine learning,” *Int. J. Recent Technol. Eng.*, vol. 7, no. 6, pp. 839–843, 2019.

[16] A. N. Akansu, S. R. Kulkarni, and D. Malioutov, *Overview: Financial Signal Processing and Machine Learning*. Wiley-IEEE Press, 2016. doi: 10.1002/9781118745540.ch1.

[17] A. Baldominos, I. Blanco, A. J. Moreno, R. Iturrarte, Ó. Bernárdez, and C. Afonso, “Identifying real estate opportunities using machine learning,” *Appl. Sci. Switz.*, vol. 8, no. 11, 2018, doi: 10.3390/app8112321.

[18] H. Y. Kim and C. H. Won, “Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models,” *Expert Syst. Appl.*, vol. 103, pp. 25–37, 2018, doi: 10.1016/j.eswa.2018.03.002.

[19] A. Hasan, O. Kalipsiz, and S. Akyokuş, “Predicting financial market in big data: Deep learning [Büyük Verilerde Finansal Piyasa Tahmini: Derin Öğrenme],” in *2nd International Conference on Computer Science and Engineering, UBMK 2017*, Institute of Electrical and Electronics Engineers Inc., 2017, pp. 510–515. doi: 10.1109/UBMK.2017.8093449.

[20] O. Lachiheb and M. S. Gouider, “A hierarchical Deep neural network design for stock returns prediction,” in *Procedia Computer Science*, J. L. C. Toro C. Hicks Y. ,. Howlett R. J. ,. Jain L. C., Ed., Elsevier B.V., 2018, pp. 264–272. doi: 10.1016/j.procS.2018.07.260.

[21] A. H. Manurung, W. Budiharto, and H. Prabowo, “Algorithm and modeling of stock prices forecasting based on long short-term memory (LSTM),” *ICIC Express Lett.*, vol. 12, no. 12, pp. 1277–1283, 2018, doi: 10.24507/icicel.12.12.1277.

[22] L. Ni, Y. Li, X. Wang, J. Zhang, J. Yu, and C. Qi, “Forecasting of Forex Time Series Data Based on Deep Learning,” in *Procedia Computer Science*, B. R. Yu J. Sun Y., Ed., Elsevier B.V., 2019, pp. 647–652. doi: 10.1016/j.procs.2019.01.189.

[23] F. Rodrigues, I. Markou, and F. C. Pereira, “Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach,” *Inf. Fusion*, vol. 49, pp. 120–129, 2019, doi: 10.1016/j.inffus.2018.07.007.

[24] W. Shen, J. Wang, Y.-G. Jiang, and H. Zha, “Portfolio Choices with Orthogonal Bandit Learning,” in *Twenty-Fourth International Joint Conference on Artificial*

Intelligence, Jun. 2015. Accessed: Jan. 09, 2020. [Online]. Available: https://www.aaai.org/ocs/index.php/IJCAI/IJCAI15/paper/view/10972

[25] D. Rapach and G. Zhou, “Chapter 6 - Forecasting Stock Returns,” in *Handbook of Economic Forecasting*, vol. 2, G. Elliott and A. Timmermann, Eds., in Handbook of Economic Forecasting, vol. 2. , Elsevier, 2013, pp. 328–383. doi: 10.1016/B978-0-444-53683-9.00006-2.

[26] J. M. Hutchinson, A. W. Lo, and T. Poggio, “A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks,” *J. Finance*, vol. 49, no. 3, pp. 851–889, 1994, doi: 10.1111/j.1540-6261.1994.tb00081.x.

[27] J. Yao, Y. Li, and C. L. Tan, “Option price forecasting using neural networks,” *Omega*, vol. 28, no. 4, pp. 455–466, Aug. 2000, doi: 10.1016/S0305-0483(99)00066-3.

[28] A. E. Khandani, A. J. Kim, and A. W. Lo, “Consumer credit-risk models via machine-learning algorithms,” *J. Bank. Finance*, vol. 34, no. 11, pp. 2767–2787, Nov. 2010, doi: 10.1016/j.jbankfin.2010.06.001.

[29] F. Butaru, Q. Chen, B. Clark, S. Das, A. W. Lo, and A. Siddique, “Risk and risk management in the credit card industry,” *J. Bank. Finance*, vol. 72, pp. 218–239, Nov. 2016, doi: 10.1016/j.jbankfin.2016.07.015.

[30] “[1602.06561] Deep Learning in Finance.” Accessed: Jan. 09, 2020. [Online]. Available: https://arxiv.org/abs/1602.06561

[31] A. Ang, “Factor Investing,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2277397, Jun. 2013. Accessed: May 27, 2019. [Online]. Available: https://papers.ssrn.com/abstract=2277397

[32] E. F. Fama and K. R. French, “Dissecting Anomalies with a Five-Factor Model,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2503174, Jun. 2015. Accessed: Jun. 08, 2019. [Online]. Available: https://papers.ssrn.com/abstract=2503174

[33] C. S. Asness, “The Siren Song of Factor Timing,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2763956, Apr. 2016. Accessed: May 27, 2019. [Online]. Available: https://papers.ssrn.com/abstract=2763956

[34] C. S. Asness, S. Chandra, A. Ilmanen, and R. Israel, “Contrarian Factor Timing is Deceptively Difficult,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2928945, Mar. 2017. Accessed: May 27, 2019. [Online]. Available: https://papers.ssrn.com/abstract=2928945

[35] A. J. Bean and A. C. Singer, “Factor graphs for universal portfolios,” in *Conference Record - Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2009, pp. 1375–1379. doi: 10.1109/ACSSC.2009.5469881.

[36] J. Bender, T. Blackburn, and X. Sun, “Clash of the Titans: Factor Portfolios versus Alternative Weighting Schemes,” *J. Portf. Manag.*, vol. 45, no. 3, pp. 38–49, Feb. 2019, doi: 10.3905/jpm.2019.45.3.038.

[37] D. Bektić, J.-S. Wenzler, M. Wegener, D. Schiereck, and T. Spielmann, “Extending Fama–French Factors to Corporate Bond Markets,” *J. Portf. Manag.*, vol. 45, no. 3, pp. 141–158, Feb. 2019, doi: 10.3905/jpm.2019.45.3.141.

[38] D. Blitz, “Factor Investing Revisited,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2626336, Jul. 2015. Accessed: May 27, 2019. [Online]. Available: https://papers.ssrn.com/abstract=2626336
[39] D. Blitz and M. Vidojevic, “The Characteristics of Factor Investing,” *J. Portf. Manag.*, vol. 45, no. 3, pp. 69–86, Feb. 2019, doi: 10.3905/jpm.2019.45.3.069.
[40] J. Cerniglia and F. J. Fabozzi, “Academic, Practitioner, and Investor Perspectives on Factor Investing,” *J. Portf. Manag.*, vol. 44, no. 4, pp. 10–16, Mar. 2018, doi: 10.3905/jpm.2018.44.4.010.
[41] L. W. Cong and D. Xu, “Rise of Factor Investing: Asset Prices, Informational Efficiency, and Security Design,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2800590, Nov. 2016. Accessed: May 27, 2019. [Online]. Available: https://papers.ssrn.com/abstract=2800590
[42] M.-W. Hsu, S. Lessmann, M.-C. Sung, T. Ma, and J. E. V. Johnson, “Bridging the divide in financial market forecasting: machine learners vs. financial economists,” *Expert Syst. Appl.*, vol. 61, pp. 215–234, Nov. 2016, doi: 10.1016/j.eswa.2016.05.033.
[43] M. Lopez de Prado, “Beyond Econometrics: A Roadmap Towards Financial Machine Learning,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3365282, Sep. 2019. Accessed: Jan. 09, 2020. [Online]. Available: https://papers.ssrn.com/abstract=3365282
[44] J. A. Cerniglia, F. J. Fabozzi, and P. N. Kolm, “Best Practices in Research for Quantitative Equity Strategies,” *J. Portf. Manag.*, vol. 42, no. 5, pp. 135–143, Jul. 2016, doi: 10.3905/jpm.2016.42.5.135.
[45] H. Jang and J. Lee, “Machine learning versus econometric jump models in predictability and domain adaptability of index options,” *Phys. Stat. Mech. Its Appl.*, vol. 513, pp. 74–86, 2019, doi: 10.1016/j.physa.2018.08.091.
[46] S. Mullainathan and J. Spiess, “Machine Learning: An Applied Econometric Approach,” *J. Econ. Perspect.*, vol. 31, no. 2, pp. 87–106, May 2017, doi: 10.1257/jep.31.2.87.
[47] S. Waite, A. Massa, and C. Cannon, “Asset Managers With $74 Trillion on Brink of Historic Shakeout,” *Bloomberg.com*. Accessed: Dec. 01, 2020. [Online]. Available: https://www.bloomberg.com/graphics/2019-asset-management-in-decline/
[48] E. Liu, “Quant Funds Struggled in 2019. The Outlook for This Year Is More of the Same.” Accessed: Dec. 01, 2020. [Online]. Available: https://www.barrons.com/articles/quant-funds-struggled-in-2019-the-outlook-for-this-year-is-more-of-the-same-51579782601
[49] M. López de Prado and M. J. Lewis, “Detection of false investment strategies using unsupervised learning methods,” *Quant. Finance*, vol. 19, no. 9, pp. 1555–1565, 2019, doi: 10.1080/14697688.2019.1622311.
[50] C. Brooks, A. G. F. Hoepner, D. McMillan, A. Vivian, and C. W. Simen, “Financial data science: the birth of a new financial research paradigm complementing econometrics?,” *Eur. J. Finance*, vol. 25, no. 17, pp. 1627–1636, Nov. 2019, doi: 10.1080/1351847X.2019.1662822.

[51] T. Khraisha, “A Holistic Approach to Financial Data Science: Data, Technology, and Analytics,” *J. Financ. Data Sci.*, Mar. 2020, doi: 10.3905/jfds.2020.1.031.
[52] A. G. F. Hoepner, D. McMillan, A. Vivian, and C. W. Simen, “Significance, relevance and explainability in the machine learning age: an econometrics and financial data science perspective,” *Eur. J. Finance*, vol. 0, no. 0, pp. 1–7, Dec. 2020, doi: 10.1080/1351847X.2020.1847725.
[53] C. R. Harvey and Y. Liu, “Evaluating Trading Strategies,” *J. Portf. Manag.*, vol. 40, no. 5, pp. 108–118, Sep. 2014, doi: 10.3905/jpm.2014.40.5.108.
[54] R. D. Arnott, C. R. Harvey, and H. Markowitz, “A Backtesting Protocol in the Era of Machine Learning,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3275654, Nov. 2018. Accessed: May 27, 2019. [Online]. Available: https://papers.ssrn.com/abstract=3275654
[55] Y. Li, Z. Simon, and D. Turkington, “Investable and Interpretable Machine Learning for Equities,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3747203, Dec. 2020. doi: 10.2139/ssrn.3747203.
[56] D. H. Bailey, J. M. Borwein, M. López de Prado, and Q. J. Zhu, “Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance,” *Not. Am. Math. Soc.*, vol. 61, no. 5, p. 458, May 2014, doi: 10.1090/noti1105.
[57] D. H. Bailey and M. Lopez de Prado, “Finance is Not Excused: Why Finance Should Not Flout Basic Principles of Statistics,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3895330, Jul. 2021. doi: 10.2139/ssrn.3895330.
[58] D. H. Bailey, S. Ger, M. L. de Prado, A. Sim, and K. Wu, “Statistical Overfitting and Backtest Performance,” p. 10.
[59] K. Lommers, O. E. Harzli, and J. Kim, “Confronting Machine Learning With Financial Research,” p. 34.
[60] “Artificial Intelligence in Asset Management,” CFA Institute. Accessed: Dec. 27, 2020. [Online]. Available: /en/research/foundation/2020/rflr-artificial-intelligence-in-asset-management
[61] M. L. de Prado, “THE 10 REASONS MOST MACHINE LEARNING FUNDS FAIL,” p. 21.
[62] K. Noel, “Application of Machine Learning to Systematic Strategies,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2837664, Jun. 2016. Accessed: May 27, 2019. [Online]. Available: https://papers.ssrn.com/abstract=2837664
[63] R. Aadith Narayan, B. D. Dayana, B. Yagneshwaran, M. R. Vignesh Babu, and K. A. V. Krishna, “Stock value prediction using machine learning,” *Int. J. Recent Technol. Eng.*, vol. 7, no. 6, pp. 839–843, 2019.
[64] S. Aktan, “Application of machine learning algorithms for business failure prediction,” *Invest. Manag. Financ. Innov.*, vol. 8, no. 2, pp. 52–65, 2011.
[65] G. Arutjothi and C. Senthamarai, “Prediction of loan status in commercial bank using machine learning classifier,” in *Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2017*, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 416–419. doi: 10.1109/ISS1.2017.8389442.

[66] F. Rundo, F. Trenta, A. Stallo, and S. Battiato, “Machine Learning for Quantitative Finance Applications: A Survey,” *Appl. Sci.*, vol. 9, pp. 1–20, Dec. 2019, doi: 10.3390/app9245574.
[67] B. M. Henrique, V. A. Sobreiro, and H. Kimura, “Literature review: Machine learning techniques applied to financial market prediction,” *Expert Syst. Appl.*, vol. 124, pp. 226–251, Jun. 2019, doi: 10.1016/j.eswa.2019.01.012.
[68] B. Krollner, B. Vanstone, and G. Finnie, “Financial Time Series Forecasting with Machine Learning Techniques: A Survey,” Jan. 2010.
[69] S. Pouyanfar *et al.*, “A Survey on Deep Learning: Algorithms, Techniques, and Applications,” *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–36, Jan. 2019, doi: 10.1145/3234150.
[70] S. Athmaja, M. Hanumanthappa, and V. Kavitha, “A survey of machine learning algorithms for big data analytics,” in *Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS 2017*, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 1–4. doi: 10.1109/ICIIECS.2017.8276028.
[71] K. Das and R. N. Behera, “A Survey on Machine Learning: Concept,Algorithms and Applications,” undefined. Accessed: May 17, 2020. [Online]. Available: https://www.semanticscholar.org/paper/A-Survey-on-Machine-Learning%3A-Concept%2CAlgorithms-Das-Behera/4d7855b8e5ef36acd4ac41deef596e67ac899e76
[72] R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms,” in *Proceedings of the 23rd international conference on Machine learning*, in ICML ’06. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, Jun. 2006, pp. 161–168. doi: 10.1145/1143844.1143865.
[73] K. Golmohammadi, O. R. Zaiane, and D. Diaz, “Detecting stock market manipulation using supervised learning algorithms,” in *DSAA 2014 - Proceedings of the 2014 IEEE International Conference on Data Science and Advanced Analytics*, K. I. Karypis G. Cao L. ,. Wang W., Ed., Institute of Electrical and Electronics Engineers Inc., 2014, pp. 435–441. doi: 10.1109/DSAA.2014.7058109.
[74] S. Jyothirmayee, V. Dilip Kumar, C. Someswara Rao, and R. Shiva Shankar, “Predicting stock exchange using supervised learning algorithms,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 1, pp. 4081–4090, 2019, doi: 10.35940/ijitee.A4144.119119.
[75] D. Snow, “Machine Learning in Asset Management—Part 1: Portfolio Construction—Trading Strategies,” *J. Financ. Data Sci.*, vol. 2, no. 1, pp. 10–23, Jan. 2020, doi: 10.3905/jfds.2019.1.021.
[76] D. Snow, “Machine Learning in Asset Management: Part 2: Portfolio Construction—Weight Optimization,” *J. Financ. Data Sci.*, Mar. 2020, doi: 10.3905/jfds.2020.1.029.
[77] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*, 2nd edition. New York, NY: Springer, 2016.

[78] R. K. Dubey and P. K. Nizar Banu, “Analysis of supervised and unsupervised technique for authentication dataset,” *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 2867–2873, 2018, doi: 10.14419/ijet.v7i4.9564.

[79] V. Kedia, Z. Khalid, S. Goswami, N. Sharma, and K. Suryawanshi, “Portfolio Generation for Indian Stock Markets Using Unsupervised Machine Learning,” in *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*, Institute of Electrical and Electronics Engineers Inc., 2018. doi: 10.1109/ICCUBEA.2018.8697771.

[80] P. R. Mohapatra, S. K. Swain, and S. S. Basa, “Unsupervised learning based stock price recommendation using collaborative filtering,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 11, pp. 2051–2055, 2019, doi: 10.35940/ijitee.K1932.0981119.

[81] A. Yadav, C. K. Jha, A. Sharan, and V. Vaish, “Sentiment Analysis of Financial News Using Unsupervised and Supervised Approach,” *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, vol. 11942 LNCS, pp. 311–319, 2019, doi: 10.1007/978-3-030-34872-4_35.

[82] M. Z. Alom *et al.*, “A State-of-the-Art Survey on Deep Learning Theory and Architectures,” *Electronics*, vol. 8, no. 3, Art. no. 3, Mar. 2019, doi: 10.3390/electronics8030292.

[83] “Deep Learning: Opportunities and Best Practice,” Peltarion.com. Accessed: May 17, 2020. [Online]. Available: https://peltarion.com/resources/deep-learning-opportunities-and-best-practice-report

[84] S. Leijnen and F. van Veen, “The Neural Network Zoo,” *Proceedings*, vol. 47, no. 1, Art. no. 1, 2020, doi: 10.3390/proceedings47010009.

[85] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, “A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning,” *Arch. Comput. Methods Eng.*, Jun. 2019, doi: 10.1007/s11831-019-09344-w.

[86] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, “State-of-the-art in artificial neural network applications: A survey,” *Heliyon*, vol. 4, no. 11, p. e00938, Nov. 2018, doi: 10.1016/j.heliyon.2018.e00938.

[87] A. Aloraini, “Penalized ensemble feature selection methods for hidden associations in time series environments case study: equities companies in Saudi Stock Exchange Market,” *Evol. Syst.*, vol. 6, no. 2, pp. 93–100, 2015, doi: 10.1007/s12530-014-9124-y.

[88] M. Choi, “Feature generation using machine learning from large spare financial data,” in *25th Americas Conference on Information Systems, AMCIS 2019*, Association for Information Systems, 2019. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85073526630&partnerID=40&md5=590fd39b3837b10dc15fd71c550e2912

[89] F. Antunes, B. Ribeiro, and F. Pereira, “Probabilistic modeling and visualization for bankruptcy prediction,” *Appl. Soft Comput. J.*, vol. 60, pp. 831–843, 2017, doi: 10.1016/j.asoc.2017.06.043.

[90] Z. Bai, N. Jain, Y. Wang, and D. Haughton, “Using SOM-Based Visualization to Analyze the Financial Performance of Consumer Discretionary Firms,” *Adv. Intell. Syst. Comput.*, vol. 976, pp. 90–99, 2020, doi: 10.1007/978-3-030-19642-4_9.

[91] J. Heaton, “An empirical analysis of feature engineering for predictive modeling,” in *SoutheastCon 2016*, Mar. 2016, pp. 1–6. doi: 10.1109/SECON.2016.7506650.
[92] G. Dong and H. Liu, *Feature Engineering for Machine Learning and Data Analytics*. CRC Press, 2018.
[93] S. Khalid, T. Khalil, and S. Nasreen, “A survey of feature selection and feature extraction techniques in machine learning,” in *2014 Science and Information Conference*, Aug. 2014, pp. 372–378. doi: 10.1109/SAI.2014.6918213.
[94] R. Abirami and M. S. Vijaya, “Stock price prediction using support vector regression,” *Commun. Comput. Inf. Sci.*, vol. 269 CCIS, no. PART I, pp. 588–597, 2012, doi: 10.1007/978-3-642-29219-4_67.
[95] T. Adrian, R. K. Crump, and E. Moench, “Regression-based estimation of dynamic asset pricing models,” *J. Financ. Econ.*, vol. 118, no. 2, pp. 211–244, Nov. 2015, doi: 10.1016/j.jfineco.2015.07.004.
[96] T. Al Rashid and V. K. Goyal, “Prediction of market behavior for short term stock prices using regression techniques,” *Int. J. Recent Technol. Eng.*, vol. 8, no. 1, pp. 1176–1183, 2019.
[97] P. C. S. Bezerra and P. H. M. Albuquerque, “Volatility forecasting: The support vector regression can beat the random walk,” *Econ. Comput. Econ. Cybern. Stud. Res.*, vol. 53, no. 4, pp. 115–126, 2019, doi: 10.24818/18423264/53.4.19.07.
[98] S. Khanmohammadi and M. Rezaeiahari, “AHP based Classification Algorithm Selection for Clinical Decision Support System Development,” *Procedia Comput. Sci.*, vol. 36, pp. 328–334, Jan. 2014, doi: 10.1016/j.procs.2014.09.101.
[99] R. N. Ainon, A. Lahsasna, and T. Y. Wah, “A transparent classification model using a hybrid soft computing method,” in *Proceedings - 2009 3rd Asia International Conference on Modelling and Simulation, AMS 2009*, Bandung, Bali, 2009, pp. 146–151. doi: 10.1109/AMS.2009.105.
[100] M. Bayraktar, M. S. Aktas, O. Kalipsiz, O. Susuz, and S. Bayraci, “Credit risk analysis with classification Restricted Boltzmann Machine [Siniflandirilmis Kisitli Boltzmann Makinesi ile Kredi Risk Analizi],” in *26th IEEE Signal Processing and Communications Applications Conference, SIU 2018*, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 1–4. doi: 10.1109/SIU.2018.8404397.
[101] M. S. Basarslan and I. D. Argun, “Classification of a bank data set on various data mining platforms [Bir Banka Müşteri Verilerinin Farkli Veri Madenciliği Platformlarinda Siniflandirilmasi],” in *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting, EBBT 2018*, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 1–4. doi: 10.1109/EBBT.2018.8391441.
[102] G. Casalicchio, C. Molnar, and B. Bischl, “Visualizing the Feature Importance for Black Box Models,” in *Machine Learning and Knowledge Discovery in Databases*, M. Berlingerio, F. Bonchi, T. Gärtner, N. Hurley, and G. Ifrim, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 655–670. doi: 10.1007/978-3-030-10925-7_40.
[103] A. Fisher, C. Rudin, and F. Dominici, “All Models are Wrong, but Many are Useful: Learning a Variable’s Importance by Studying an Entire Class of Prediction Models Simultaneously,” *ArXiv180101489 Stat*, Dec. 2019, Accessed: Dec. 26, 2020. [Online]. Available: http://arxiv.org/abs/1801.01489

[104] U. Grömping, “Variable Importance Assessment in Regression: Linear Regression versus Random Forest,” *Am. Stat.*, vol. 63, no. 4, pp. 308–319, Nov. 2009, doi: 10.1198/tast.2009.08199.

[105] B. Gültekin and B. E. Sakar, “Variable Importance Analysis in Default Prediction using Machine Learning Techniques,” in *DATA 2018 - Proceedings of the 7th International Conference on Data Science, Technology and Applications*, Q. C. Bernardino J., Ed., SciTePress, 2018, pp. 56–62. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85071500700&partnerID=40&md5=42a5bb115c8748102c557dadff7352d3

[106] A. M. Abirami and V. Gayathri, “A survey on sentiment analysis methods and approach,” in *2016 8th International Conference on Advanced Computing, ICoAC 2016*, Institute of Electrical and Electronics Engineers Inc., 2017, pp. 72–76. doi: 10.1109/ICoAC.2017.7951748.

[107] Z. Ahmadi, P. Martens, C. Koch, T. Gottron, and S. Kramer, “Towards bankruptcy prediction: Deep sentiment mining to detect financial distress from business management reports,” in *Proceedings - 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, DSAA 2018*, B. F. Eliassi-Rad T. Wang W. ,. Cattuto C. ,. Provost F. ,. Ghani R., Ed., Institute of Electrical and Electronics Engineers Inc., 2019, pp. 293–302. doi: 10.1109/DSAA.2018.00040.

[108] R. Batra and S. M. Daudpota, “Integrating StockTwits with sentiment analysis for better prediction of stock price movement,” in *2018 International Conference on Computing, Mathematics and Engineering Technologies: Invent, Innovate and Integrate for Socioeconomic Development, iCoMET 2018 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 1–5. doi: 10.1109/ICOMET.2018.8346382.

[109] J. Braun, J. Hausler, and W. Schäfers, “Artificial intelligence, news sentiment, and property market liquidity,” *J. Prop. Invest. Finance*, 2019, doi: 10.1108/JPIF-08-2019-0100.

[110] C.-C. Chang, Z.-T. Lin, W.-W. Koc, C. Chou, and S.-H. Huang, “Affinity propagation clustering for intelligent portfolio diversification and investment risk reduction,” in *Proceedings - 2016 7th International Conference on Cloud Computing and Big Data, CCBD 2016*, Institute of Electrical and Electronics Engineers Inc., 2017, pp. 145–150. doi: 10.1109/CCBD.2016.037.

[111] W.-C. Chen and R. Maitra, “Model-based clustering of regression time series data via APECM-an AECM algorithm sung to an even faster beat,” *Stat. Anal. Data Min.*, vol. 4, no. 6, pp. 567–578, 2011, doi: 10.1002/sam.10143.

[112] F. G. Duarte and L. N. De Castro, “A fuzzy clustering algorithm for portfolio selection,” in *Proceedings - 21st IEEE Conference on Business Informatics, CBI 2019*, N. D. Becker J., Ed., Institute of Electrical and Electronics Engineers Inc., 2019, pp. 414–418. doi: 10.1109/CBI.2019.00054.

[113] S. Chakraborty *et al.*, “Interpretability of deep learning models: A survey of results,” in *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*

(SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), San Francisco, CA: IEEE, Aug. 2017, pp. 1–6. doi: 10.1109/UIC-ATC.2017.8397411.

[114] J. Chen, L. Song, M. J. Wainwright, and M. I. Jordan, “L-Shapley and C-Shapley: Efficient Model Interpretation for Structured Data,” *ArXiv180802610 Cs Stat*, Aug. 2018, Accessed: Dec. 26, 2020. [Online]. Available: http://arxiv.org/abs/1808.02610

[115] L. Cong, K. Tang, J. Wang, and Y. Zhang, “AlphaPortfolio for Investment and Economically Interpretable AI,” 2020, doi: 10.2139/ssrn.3554486.

[116] L. W. Cong, K. Tang, J. Wang, and Y. Zhang, “AlphaPortfolio: Direct Construction through Reinforcement Learning and Interpretable AI,” p. 64.

[117] R. Florez-Lopez and J. M. Ramon-Jeronimo, “Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment. A correlated-adjusted decision forest proposal,” *Expert Syst. Appl.*, vol. 42, no. 13, pp. 5737–5753, 2015, doi: 10.1016/j.eswa.2015.02.042.

[118] A. A. Ismail, M. Gunady, H. C. Bravo, and S. Feizi, “Benchmarking Deep Learning Interpretability in Time Series Predictions,” *ArXiv201013924 Cs Stat*, Oct. 2020, Accessed: Dec. 26, 2020. [Online]. Available: http://arxiv.org/abs/2010.13924

[119] “AI, What Have You Done for Us Lately?,” CFA Institute Enterprising Investor. Accessed: Dec. 01, 2020. [Online]. Available: https://blogs.cfainstitute.org/investor/2019/10/14/ai-what-have-you-done-for-us-lately/

[120] M.-Y. Day, T.-K. Cheng, and J.-G. Li, “Ai robo-advisor with big data analytics for financial services,” in *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018*, B. U. Tagarelli A. Reddy C., Ed., Institute of Electrical and Electronics Engineers Inc., 2018, pp. 1027–1031. doi: 10.1109/ASONAM.2018.8508854.

[121] O. Bustos, A. Pomares, and E. Gonzalez, “A comparison between SVM and multilayer perceptron in predicting an emerging financial market: Colombian stock market,” in *2017 Congreso Internacional de Innovacion y Tendencias en Ingenieria, CONIITI 2017 - Conference Proceedings*, L.-G. C.A, Ed., Institute of Electrical and Electronics Engineers Inc., 2018, pp. 1–6. doi: 10.1109/CONIITI.2017.8273335.

[122] J. Chen, “SVM application of financial time series forecasting using empirical technical indicators,” in *ICINA 2010 - 2010 International Conference on Information, Networking and Automation, Proceedings*, Kunming, 2010, pp. V177–V181. doi: 10.1109/ICINA.2010.5636430.

[123] S. S. Abdullah, M. S. Rahaman, and M. S. Rahman, “Analysis of stock market using text mining and natural language processing,” in *2013 International Conference on Informatics, Electronics and Vision, ICIEV 2013*, Dhaka, 2013. doi: 10.1109/ICIEV.2013.6572673.

[124] J. Chebeir, H. Asala, V. Manee, I. Gupta, and J. Romagnoli, “Comprehensive Investment Framework for Optimal Exploitation of Liquid-Rich Unconventional Reservoirs and Natural Gas Distribution,” *Comput. Aided Chem. Eng.*, vol. 44, pp. 961–966, 2018, doi: 10.1016/B978-0-444-64241-7.50155-5.

[125] J. S. Virdi, W. Peng, and A. Sata, “Feature selection with LASSO and VSURF to model mechanical properties for investment casting,” in *ICCIDS 2019 - 2nd International Conference on Computational Intelligence in Data Science, Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2019. doi: 10.1109/ICCIDS.2019.8862141.
[126] F. J. Fabozzi and M. L. de Prado, “Being Honest in Backtest Reporting: A Template for Disclosing Multiple Tests,” *J. Portf. Manag.*, vol. 45, no. 1, pp. 141–147, Oct. 2018, doi: 10.3905/jpm.2018.45.1.141.
[127] M. Lopez de Prado and F. J. Fabozzi, “Who Needs a Newtonian Finance?,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3104826, Jan. 2018. Accessed: Jun. 08, 2019. [Online]. Available: https://papers.ssrn.com/abstract=3104826
[128] M. L. de Prado, “A Data Science Solution to the Multiple-Testing Crisis in Financial Research,” *J. Financ. Data Sci.*, vol. 1, no. 1, pp. 99–110, Jan. 2019, doi: 10.3905/jfds.2019.1.099.
[129] D. H. Bailey, J. Borwein, M. Lopez de Prado, and Q. J. Zhu, “The Probability of Backtest Overfitting,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2840838, Sep. 2016. Accessed: May 17, 2020. [Online]. Available: https://papers.ssrn.com/abstract=2840838
[130] C. Brooks, A. G. F. Hoepner, D. McMillan, A. Vivian, and C. W. Simen, “Financial data science: the birth of a new financial research paradigm complementing econometrics?,” *Eur. J. Finance*, vol. 25, no. 17, pp. 1627–1636, Nov. 2019, doi: 10.1080/1351847X.2019.1662822.
[131] “Can Machines ‘Learn’ Finance?” Accessed: Jun. 07, 2020. [Online]. Available: https://www.aqr.com/Insights/Research/Alternative-Thinking/Can-Machines-Learn-Finance
[132] T. Guida and G. Coqueret, “Machine Learning in Systematic Equity Allocation: A Model Comparison,” *Wilmott*, vol. 2018, no. 98, pp. 24–33, 2018, doi: 10.1002/wilm.10719.
[133] A. K. Kar and Y. K. Dwivedi, “Theory building with big data-driven research – Moving away from the ‘What’ towards the ‘Why,’” *Int. J. Inf. Manag.*, vol. 54, p. 102205, Oct. 2020, doi: 10.1016/j.ijinfomgt.2020.102205.
[134] A. Weigand, “Machine learning in empirical asset pricing,” *Financ. Mark. Portf. Manag.*, vol. 33, no. 1, pp. 93–104, Mar. 2019, doi: 10.1007/s11408-019-00326-3.
[135] L. Chen, M. Pelger, and J. Zhu, “Deep Learning in Asset Pricing,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3350138, Sep. 2020. doi: 10.2139/ssrn.3350138.
[136] S. Gu, B. T. Kelly, and D. Xiu, “Autoencoder Asset Pricing Models,” *SSRN Electron. J.*, 2019, doi: 10.2139/ssrn.3335536.
[137] X. Lan, “A synthetic dataset for deep learning,” *ArXiv190611905 Cs Stat*, Jun. 2019, Accessed: Feb. 16, 2021. [Online]. Available: http://arxiv.org/abs/1906.11905
[138] M. Hittmeir, A. Ekelhart, and R. Mayer, “On the Utility of Synthetic Data: An Empirical Evaluation on Machine Learning Tasks,” in *Proceedings of the 14th International Conference on Availability, Reliability and Security*, in ARES ’19.

New York, NY, USA: Association for Computing Machinery, Aug. 2019, pp. 1–6. doi: 10.1145/3339252.3339281.
[139] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, “A review of feature selection methods on synthetic data,” *Knowl. Inf. Syst.*, vol. 34, no. 3, pp. 483–519, Mar. 2013, doi: 10.1007/s10115-012-0487-8.
[140] F. Fomin, P. Golovach, F. Panolan, and K. Simonov, “Refined complexity of PCA with outliers,” in *36th International Conference on Machine Learning, ICML 2019*, International Machine Learning Society (IMLS), 2019, pp. 10204–10213. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85078283845&partnerID=40&md5=41b4dcebdf5aa9c33dde4727dad19267
[141] A. A. M. Nurunnabi, A. B. M. S. Ali, A. H. M. R. Imon, and M. Nasser, *Outlier detection in logistic regression*. IGI Global, 2012. doi: 10.4018/978-1-4666-1830-5.ch016.
[142] P. Eichmann, J. Tan, S. Zdonik, F. Solleza, and N. Tatbul, “Metro-Viz: Black-box analysis of time series anomaly detectors,” in *Conference on Human Factors in Computing Systems - Proceedings*, Association for Computing Machinery, 2019. doi: 10.1145/3290607.3312912.
[143] J. Wang *et al.*, “An anomaly prediction framework for financial IT systems using hybrid machine learning methods,” *J. Ambient Intell. Humaniz. Comput.*, 2019, doi: 10.1007/s12652-019-01645-z.
[144] P. Hájek and V. Olej, “Predicting firms’ credit ratings using ensembles of artificial immune systems and machine learning – an over-sampling approach,” *IFIP Adv. Inf. Commun. Technol.*, vol. 436, pp. 29–38, 2014.
[145] Y. Ye, S. Liu, and J. Li, “A multiclass machine learning approach to credit rating prediction,” in *Proceedings - International Symposium on Information Processing, ISIP 2008 and International Pacific Workshop on Web Mining and Web-Based Application, WMWA 2008*, Moscow, 2008, pp. 57–61. doi: 10.1109/ISIP.2008.37.
[146] S. Jadhav, H. He, and K. Jenkins, “Information gain directed genetic algorithm wrapper feature selection for credit rating,” *Appl. Soft Comput. J.*, vol. 69, pp. 541–553, 2018, doi: 10.1016/j.asoc.2018.04.033.
[147] S. L. Brunton, J. L. Proctor, J. N. Kutz, and W. Bialek, “Discovering governing equations from data by sparse identification of nonlinear dynamical systems,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 15, pp. 3932–3937, 2016, doi: 10.1073/pnas.1517384113.
[148] J. K. Jaiswal and R. Das, “Artificial neural network algorithms based nonlinear data analysis for forecasting in the finance sector,” *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 169–176, 2018, doi: 10.14419/ijet.v7i4.10.20829.
[149] Q. Zhang and S.-C. Zhu, “Visual Interpretability for Deep Learning: a Survey,” *ArXiv180200614 Cs*, Feb. 2018, Accessed: Sep. 07, 2020. [Online]. Available: http://arxiv.org/abs/1802.00614
[150] D. Aparicio and M. López De Prado, “How hard is it to pick the right model? MCS and backtest overfitting,” *Algorithmic Finance*, vol. 7, no. 1–2, pp. 53–61, 2018, doi: 10.3233/AF-180231.
[151] D. H. Bailey and M. L. de Prado, “The deflated Sharpe Ratio: Correcting for selection bias, backtest overfitting and non-normality,” p. 22.

[152] F. D. M. Pardo and R. C. López, “Mitigating Overfitting on Financial Datasets with Generative Adversarial Networks,” *J. Financ. Data Sci.*, vol. 2, no. 1, pp. 76–85, Jan. 2020, doi: 10.3905/jfds.2019.1.019.
[153] C. R. Bacon, “Performance Attribution: History and Progress,” *SSRN Electron. J.*, 2019, doi: 10.2139/ssrn.3668966.
[154] A. Ang, “Factor Investing,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2277397, Jun. 2013. Accessed: Jan. 08, 2020. [Online]. Available: https://papers.ssrn.com/abstract=2277397
[155] M. Bal, M. F. Amasyali, H. Sever, G. Kose, and A. Demirhan, “Performance evaluation of the machine learning algorithms used in inference mechanism of a medical decision support system,” *Sci. World J.*, vol. 2014, 2014, doi: 10.1155/2014/137896.
[156] A. Coad, D. Janzing, and P. Nightingale, “Tools for causal inference from cross-sectional innovation surveys with continuous or discrete variables: Theory and applications [Outils pour l’inférence causale d’enquêtes d’innovation de bilan transversal avec des variables continues ou discrètes: Théorie et applications] [Ferramentas para a inferência causal de pesquisas de inovação de corte transversal com variáveis contínuas ou discretas: Teoria e aplicações] [Herramientas para la inferencia causal de encuestas de innovación de corte transversal con variables continuas o discretas: Teoría y aplicaciones],” *Cuad. Econ. Colomb.*, vol. 37, no. 75, pp. 779–808, 2018, doi: 10.15446/cuad.econ.v37n75.69832.
[157] C. Brooks, A. G. F. Hoepner, D. McMillan, A. Vivian, and C. Wese Simen, “Financial data science: the birth of a new financial research paradigm complementing econometrics?,” *Eur. J. Finance*, vol. 25, no. 17, pp. 1627–1636, Nov. 2019, doi: 10.1080/1351847X.2019.1662822.
[158] C. Chen, P. Zhang, Y. Liu, and J. Liu, “Financial quantitative investment using convolutional neural network and deep learning technology,” *Neurocomputing*, 2019, doi: 10.1016/j.neucom.2019.09.092.
[159] “A Data Science Solution to the Multiple-Testing Crisis in Financial Research.” Accessed: May 17, 2020. [Online]. Available: https://www.aqr.com/Insights/Research/Journal-Article/A-Data-Science-Solution-to-the-Multiple-Testing-Crisis-in-Financial-Research
[160] A. Genender-Feltheimer, “Visualizing high dimensional and big data,” in *Procedia Computer Science*, Elsevier B.V., 2018, pp. 112–121. doi: 10.1016/j.procs.2018.10.308.
[161] Y. Li, P. Ni, and V. Chang, “An Empirical Research on the Investment Strategy of Stock Market based on Deep Reinforcement Learning model,” in *COMPLEXIS 2019 - Proceedings of the 4th International Conference on Complexity, Future Information Systems and Risk*, C. V. Firouzi F. Estrada E. ,. Munoz V. M., Ed., SciTePress, 2019, pp. 52–58. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85067524270&partnerID=40&md5=7aeae09effa5b496b860124f6f740628
[162] J. Lintner, “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets,” *Rev. Econ. Stat.*, vol. 47, no. 1, pp. 13–37, 1965, doi: 10.2307/1924119.

[163] J. Mossin, “Equilibrium in a Capital Asset Market,” *Econometrica*, vol. 34, no. 4, pp. 768–783, 1966, doi: 10.2307/1910098.
[164] W. F. Sharpe, “THE CAPITAL ASSET PRICING MODEL: A ‘Multi-Beta’ Interpretation,” in *Financial Dec Making Under Uncertainty*, H. Levy and M. Sarnat, Eds., Academic Press, 1977, pp. 127–135. doi: 10.1016/B978-0-12-445850-5.50011-8.
[165] J. L. Treynor, “Market Value, Time, and Risk,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2600356, Aug. 1961. doi: 10.2139/ssrn.2600356.
[166] S. A. Ross, “The arbitrage theory of capital asset pricing,” *J. Econ. Theory*, vol. 13, no. 3, pp. 341–360, Dec. 1976, doi: 10.1016/0022-0531(76)90046-6.
[167] T. Gupta and B. T. Kelly, “Factor Momentum Everywhere,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3300728, Nov. 2018. Accessed: May 27, 2019. [Online]. Available: https://papers.ssrn.com/abstract=3300728
[168] M. M. Carhart, “On Persistence in Mutual Fund Performance,” *J. Finance*, vol. 52, no. 1, pp. 57–82, 1997, doi: 10.2307/2329556.
[169] E. F. Fama and K. R. French, “The Cross-Section of Expected Stock Returns,” *J. Finance*, vol. 47, no. 2, pp. 427–465, 1992, doi: 10.1111/j.1540-6261.1992.tb04398.x.
[170] E. F. Fama and K. R. French, “Common risk factors in the returns on stocks and bonds,” *J. Financ. Econ.*, vol. 33, no. 1, pp. 3–56, Feb. 1993, doi: 10.1016/0304-405X(93)90023-5.
[171] D. Blitz and M. Vidojevic, “The Characteristics of Factor Investing,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3206798, Jul. 2018. Accessed: May 27, 2019. [Online]. Available: https://papers.ssrn.com/abstract=3206798
[172] C. S. Asness, A. Frazzini, and L. H. Pedersen, “Quality minus junk,” *Rev. Account. Stud.*, vol. 24, no. 1, pp. 34–112, Mar. 2019, doi: 10.1007/s11142-018-9470-2.
[173] G. Feng, S. Giglio, and D. Xiu, “Taming the Factor Zoo: A Test of New Factors,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2934020, Jan. 2019. Accessed: Jun. 10, 2019. [Online]. Available: https://papers.ssrn.com/abstract=2934020
[174] A. Lester, “On the Theory and Practice of Multifactor Portfolios,” *J. Portf. Manag.*, vol. 45, no. 3, pp. 87–100, Feb. 2019, doi: 10.3905/jpm.2019.45.3.087.
[175] C. R. Harvey and Y. Liu, “A Census of the Factor Zoo,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3341728, Feb. 2019. doi: 10.2139/ssrn.3341728.
[176] C. de Franco and B. Monnier, “Multi-Factor Portfolios: A New Factor? Limits of the Static Approach,” *J. Invest.*, vol. 28, no. 1, pp. 97–111, Jan. 2019, doi: 10.3905/joi.2019.28.1.097.
[177] M. F. Dixon and N. G. Polson, “Deep Fundamental Factor Models,” *ArXiv190307677 Cs Stat*, Apr. 2020, Accessed: May 17, 2020. [Online]. Available: http://arxiv.org/abs/1903.07677

[178] J. Lintner, “Security Prices, Risk, and Maximal Gains from Diversification*,” *J. Finance*, vol. 20, no. 4, pp. 587–615, 1965, doi: 10.1111/j.1540-6261.1965.tb02930.x.

[179] “CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK* - Sharpe - 1964 - The Journal of Finance - Wiley Online Library.” Accessed: Jun. 01, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6261.1964.tb02865.x

[180] J. Bender, R. Briand, D. Melas, and R. A. Subramanian, “1. Foundations of Factor Investing,” *SSRN Electron. J.*, 2013, doi: 10.2139/ssrn.2543990.

[181] B. T. Kelly, S. Pruitt, and Y. Su, “Characteristics Are Covariances: A Unified Model of Risk and Return,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3032013, Oct. 2018. doi: 10.2139/ssrn.3032013.

[182] G. Feng, N. Polson, and J. Xu, “Deep Factor Alpha,” p. 28.

[183] T. Guida and G. Coqueret, “Ensemble Learning Applied to Quant Equity: Gradient Boosting in a Multifactor Framework,” in *Big Data and Machine Learning in Quantitative Investment*, John Wiley & Sons, Ltd, 2018, pp. 129–148. doi: 10.1002/9781119522225.ch7.

[184] “Cleland, David I. and Dundar F. Kocaoglu; Engineering Management, McGraw-Hill, 1981 (CK) | Engineering | System | Free 30-day Trial,” Scribd. Accessed: May 27, 2020. [Online]. Available: https://www.scribd.com/document/308713942/Cleland-David-I-and-Dundar-F-Kocaoglu-Engineering-Management-McGraw-Hill-1981-CK

[185] D. F. Kocaoglu, “Identify the Best Alternatives to Help the Diffusion of Teleconsultation by Using the Hierarchical Decision Model (HDM)”, Accessed: May 27, 2020. [Online]. Available: https://core.ac.uk/reader/37775950

[186] T. U. Daim, Ed., *Hierarchical Decision Modeling: Essays in Honor of Dundar F. Kocaoglu*. in Innovation, Technology, and Knowledge Management. Springer International Publishing, 2016. doi: 10.1007/978-3-319-18558-3.

[187] L. Hogaboam, B. Ragel, and T. Daim, “Development of a Hierarchical Decision Model (HDM) for health technology assessment (HTA) to design and implement a new patient care database for low back pain,” in *Proceedings of PICMET ’14 Conference: Portland International Center for Management of Engineering and Technology; Infrastructure and Service Integration*, Jul. 2014, pp. 3511–3517.

[188] R. Ali, S. Lee, and T. C. Chung, “Accurate multi-criteria decision making methodology for recommending machine learning algorithm,” *Expert Syst. Appl.*, vol. 71, pp. 257–278, Apr. 2017, doi: 10.1016/j.eswa.2016.11.034.

[189] P. Xidonas, G. Mavrotas, T. Krintas, J. Psarras, and C. Zopounidis, *Multicriteria Portfolio Management*. in Springer Optimization and Its Applications. New York: Springer-Verlag, 2012. doi: 10.1007/978-1-4614-3670-6.

[190] Y. Beulah Jeba Jaya and J. Jebamalar Tamilselvi, “Simplified MCDM analytical weighted model for ranking classifiers in financial risk datasets,” in *Proceedings - 2014 International Conference on Intelligent Computing Applications, ICICA 2014*, Institute of Electrical and Electronics Engineers Inc., 2014, pp. 158–161. doi: 10.1109/ICICA.2014.42.

[191] E. Gibson and T. U. Daim, “A measurement system for science and engineering research center performance evaluation,” in *2016 Portland International Conference on Management of Engineering and Technology (PICMET)*, Sep. 2016, pp. 2782–2792. doi: 10.1109/PICMET.2016.7806624.
[192] H. Barham, “Development of a Readiness Assessment Model for Evaluating Big Data Projects: Case Study of Smart City in Oregon, USA,” Jan. 2000. doi: 10.15760/etd.6872.
[193] J. Estep, “Development of a Technology Transfer Score for Evaluating Research Proposals: Case Study of Demand Response Technologies in the Pacific Northwest”, Accessed: Jul. 12, 2020. [Online]. Available: https://core.ac.uk/reader/81253812
[194] J. Lavoie, “A Scoring Model to Assess Organizations’ Technology Transfer Capabilities: The Case of a Power Utility in the Northwest USA,” Jan. 2000. doi: 10.15760/etd.6871.
[195] K. Phan, “Innovation Measurement: A Decision Framework to Determine Innovativeness of a Company,” Jan. 2000. doi: 10.15760/etd.1017.
[196] D. F. Kocaoglu, “A participative approach to program evaluation,” *IEEE Trans. Eng. Manag.*, vol. EM-30, no. 3, pp. 112–118, Aug. 1983, doi: 10.1109/TEM.1983.6448602.
[197] E. Garces, K. van Blommestein, J. Anthony, J. Hillegas-Elting, T. Daim, and B.-S. Yoon, “Technology domain analysis: A case of energy-efficient advanced commercial refrigeration technologies,” *Sustain. Prod. Consum.*, vol. 12, pp. 221–233, Oct. 2017, doi: 10.1016/j.spc.2017.08.002.
[198] “VOSviewer - Visualizing scientific landscapes,” VOSviewer. Accessed: Jun. 15, 2021. [Online]. Available: https://www.vosviewer.com//
[199] T. L. Saaty, “Fundamentals of the analytic network process — Dependence and feedback in decision-making with a single network,” *J. Syst. Sci. Syst. Eng.*, vol. 13, no. 2, pp. 129–157, Apr. 2004, doi: 10.1007/s11518-006-0158-y.
[200] T. L. Saaty, “Decision making with the analytic hierarchy process,” *Int. J. Serv. Sci.*, vol. 1, no. 1, pp. 83–98, Jan. 2008, doi: 10.1504/IJSSci.2008.01759.
[201] T. G. Martin *et al.*, “Eliciting Expert Knowledge in Conservation Science,” *Conserv. Biol.*, vol. 26, no. 1, pp. 29–38, 2012, doi: 10.1111/j.1523-1739.2011.01806.x.
[202] M. C. Runge, S. J. Converse, and J. E. Lyons, “Which uncertainty? Using expert elicitation and expected value of information to design an adaptive program,” *Biol. Conserv.*, vol. 144, no. 4, pp. 1214–1223, Apr. 2011, doi: 10.1016/j.biocon.2010.12.020.
[203] S. Beecham, T. Hall, C. Britton, M. Cottee, and A. Rainer, “Using an expert panel to validate a requirements process improvement model,” *J. Syst. Softw.*, vol. 76, no. 3, pp. 251–275, Jun. 2005, doi: 10.1016/j.jss.2004.06.004.
[204] C. W. Cai, M. K. Linnenluecke, M. Marrone, and A. K. Singh, “Machine Learning and Expert Judgement: Analyzing Emerging Topics in Accounting and Finance Research in the Asia–Pacific,” *Abacus*, vol. 55, no. 4, pp. 709–733, 2019, doi: 10.1111/abac.12179.

[205] S. Lin and M. Lu, "Characterizing disagreement and inconsistency in experts' judgments in the analytic hierarchy process," *Manag. Decis.*, vol. 50, no. 7, pp. 1252–1265, Jan. 2012, doi: 10.1108/00251741211246996.
[206] D. L. Olson, "Comparison of weights in TOPSIS models," *Math. Comput. Model.*, vol. 40, no. 7, pp. 721–727, Oct. 2004, doi: 10.1016/j.mcm.2004.10.003.
[207] J. Landeta, "Current validity of the Delphi method in social sciences," *Technol. Forecast. Soc. Change*, vol. 73, no. 5, pp. 467–482, Jun. 2006, doi: 10.1016/j.techfore.2005.09.002.
[208] 1776 Main Street Santa Monica and California 90401-3208, "Delphi Method." Accessed: May 16, 2021. [Online]. Available: https://www.rand.org/topics/delphi-method.html
[209] G. J. Skulmoski, F. T. Hartman, and J. Krahn, "The Delphi Method for Graduate Research," *J. Inf. Technol. Educ. Res.*, vol. 6, no. 1, pp. 1–21, Jan. 2007.
[210] "The Delphi Method for Graduate Research - Learning & Technology Library (LearnTechLib)." Accessed: May 16, 2021. [Online]. Available: https://www.learntechlib.org/p/111405/
[211] T. Krueger, T. Page, K. Hubacek, L. Smith, and K. Hiscock, "The role of expert opinion in environmental modelling," *Environ. Model. Softw.*, vol. 36, pp. 4–18, Oct. 2012, doi: 10.1016/j.envsoft.2012.01.011.
[212] R. Smith, "Peer Review: A Flawed Process at the Heart of Science and Journals," *J. R. Soc. Med.*, vol. 99, no. 4, pp. 178–182, Apr. 2006, doi: 10.1177/014107680609900414.
[213] P. M. Kuhnert, T. G. Martin, and S. P. Griffiths, "A guide to eliciting and using expert knowledge in Bayesian ecological models," *Ecol. Lett.*, vol. 13, no. 7, pp. 900–914, 2010, doi: 10.1111/j.1461-0248.2010.01477.x.
[214] B. Mellers, "Review of Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice," *Polit. Psychol.*, vol. 20, no. 2, pp. 416–419, 1999.
[215] K. R. Hammond, *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. Oxford University Press, 1996.
[216] A. Saltelli, "Sensitivity Analysis for Importance Assessment," *Risk Anal.*, vol. 22, no. 3, pp. 579–590, 2002, doi: https://doi.org/10.1111/0272-4332.00040.
[217] E. Borgonovo and E. Plischke, "Sensitivity analysis: A review of recent advances," *Eur. J. Oper. Res.*, vol. 248, no. 3, pp. 869–887, Feb. 2016, doi: 10.1016/j.ejor.2015.06.032.
[218] J. Butler, J. Jia, and J. Dyer, "Simulation techniques for the sensitivity analysis of multi-criteria decision models," *Eur. J. Oper. Res.*, vol. 103, no. 3, pp. 531–546, Dec. 1997, doi: 10.1016/S0377-2217(96)00307-4.
[219] H. Chen and D. F. Kocaoglu, "A sensitivity analysis algorithm for hierarchical decision models," *Eur. J. Oper. Res.*, vol. 185, no. 1, pp. 266–288, Feb. 2008, doi: 10.1016/j.ejor.2006.12.029.
[220] M. Behzadian, S. Khanmohammadi Otaghsara, M. Yazdani, and J. Ignatius, "A state-of the-art survey of TOPSIS applications," *Expert Syst. Appl.*, vol. 39, no. 17, pp. 13051–13069, Dec. 2012, doi: 10.1016/j.eswa.2012.05.056.

[221] H. Deng, C.-H. Yeh, and R. J. Willis, "Inter-company comparison using modified TOPSIS with objective weights," *Comput. Oper. Res.*, vol. 27, no. 10, pp. 963–973, Sep. 2000, doi: 10.1016/S0305-0548(99)00069-6.

[222] Y.-J. Lai, T.-Y. Liu, and C.-L. Hwang, "TOPSIS for MODM," *Eur. J. Oper. Res.*, vol. 76, no. 3, pp. 486–500, Aug. 1994, doi: 10.1016/0377-2217(94)90282-8.

[223] H.-S. Shih, H.-J. Shyur, and E. S. Lee, "An extension of TOPSIS for group decision making," *Math. Comput. Model.*, vol. 45, no. 7, pp. 801–813, Apr. 2007, doi: 10.1016/j.mcm.2006.03.023.

[224] A. Albadvi, S. K. Chaharsooghi, and A. Esfahanipour, "Decision making in stock trading: An application of PROMETHEE," *Eur. J. Oper. Res.*, vol. 177, no. 2, pp. 673–683, Mar. 2007, doi: 10.1016/j.ejor.2005.11.022.

[225] J.-P. Brans and Y. De Smet, "PROMETHEE Methods," in *Multiple Criteria Decision Analysis: State of the Art Surveys*, S. Greco, M. Ehrgott, and J. R. Figueira, Eds., in International Series in Operations Research & Management Science. , New York, NY: Springer, 2016, pp. 187–219. doi: 10.1007/978-1-4939-3094-4_6.

[226] J. P. Brans, Ph. Vincke, and B. Mareschal, "How to select and how to rank projects: The Promethee method," *Eur. J. Oper. Res.*, vol. 24, no. 2, pp. 228–238, Feb. 1986, doi: 10.1016/0377-2217(86)90044-5.

[227] C. Macharis, J. Springael, K. De Brucker, and A. Verbeke, "PROMETHEE and AHP: The design of operational synergies in multicriteria analysis.: Strengthening PROMETHEE with ideas of AHP," *Eur. J. Oper. Res.*, vol. 153, no. 2, pp. 307–317, Mar. 2004, doi: 10.1016/S0377-2217(03)00153-X.

[228] K. M. A.-S. Al-Harbi, "Application of the AHP in project management," *Int. J. Proj. Manag.*, vol. 19, no. 1, pp. 19–27, Jan. 2001, doi: 10.1016/S0263-7863(99)00038-1.

[229] D.-Y. Chang, "Applications of the extent analysis method on fuzzy AHP," *Eur. J. Oper. Res.*, vol. 95, no. 3, pp. 649–655, Dec. 1996, doi: 10.1016/0377-2217(95)00300-2.

[230] C. Kahraman, U. Cebeci, and Z. Ulukan, "Multi-criteria supplier selection using fuzzy AHP," *Logist. Inf. Manag.*, vol. 16, no. 6, pp. 382–394, Jan. 2003, doi: 10.1108/09576050310503367.

[231] R. Handfield, S. V. Walton, R. Sroufe, and S. A. Melnyk, "Applying environmental criteria to supplier assessment: A study in the application of the Analytical Hierarchy Process," *Eur. J. Oper. Res.*, vol. 141, no. 1, pp. 70–87, Aug. 2002, doi: 10.1016/S0377-2217(01)00261-2.

[232] S. Sipahi and M. Timor, "The analytic hierarchy process and analytic network process: an overview of applications," *Manag. Decis.*, vol. 48, no. 5, pp. 775–808, Jan. 2010, doi: 10.1108/00251741011043920.

[233] J. R. Figueira, V. Mousseau, and B. Roy, "ELECTRE Methods," in *Multiple Criteria Decision Analysis: State of the Art Surveys*, S. Greco, M. Ehrgott, and J. R. Figueira, Eds., in International Series in Operations Research & Management Science. , New York, NY: Springer, 2016, pp. 155–185. doi: 10.1007/978-1-4939-3094-4_5.

[234] H. Li and J. Sun, "Hybridizing principles of the Electre method with case-based reasoning for data mining: Electre-CBR-I and Electre-CBR-II," *Eur. J. Oper. Res.*, vol. 197, no. 1, pp. 214–224, 2009, doi: 10.1016/j.ejor.2008.05.024.
[235] J. S. Dyer, "Maut — Multiattribute Utility Theory," in *Multiple Criteria Decision Analysis: State of the Art Surveys*, J. Figueira, S. Greco, and M. Ehrogott, Eds., in International Series in Operations Research & Management Science. , New York, NY: Springer, 2005, pp. 265–292. doi: 10.1007/0-387-23081-5_7.
[236] J. S. Dyer, P. C. Fishburn, R. E. Steuer, J. Wallenius, and S. Zionts, "Multiple Criteria Decision Making, Multiattribute Utility Theory: The Next Ten Years," *Manag. Sci.*, vol. 38, no. 5, pp. 645–654, May 1992, doi: 10.1287/mnsc.38.5.645.
[237] S. J. T. Jansen, "The Multi-attribute Utility Method," in *The Measurement and Analysis of Housing Preference and Choice*, S. J. T. Jansen, H. C. C. H. Coolen, and R. W. Goetgeluk, Eds., Dordrecht: Springer Netherlands, 2011, pp. 101–125. doi: 10.1007/978-90-481-8894-9_5.
[238] G. W. Torrance, M. H. Boyle, and S. P. Horwood, "Application of Multi-Attribute Utility Theory to Measure Social Preferences for Health States," *Oper. Res.*, vol. 30, no. 6, pp. 1043–1069, Dec. 1982, doi: 10.1287/opre.30.6.1043.
[239] D. Von Winterfeldt and G. W. Fischer, "Multi-Attribute Utility Theory: Models and Assessment Procedures," in *Utility, Probability, and Human Decision Making: Selected Proceedings of an Interdisciplinary Research Conference, Rome, 3–6 September, 1973*, D. Wendt and C. Vlek, Eds., in Theory and Decision Library. , Dordrecht: Springer Netherlands, 1975, pp. 47–85. doi: 10.1007/978-94-010-1834-0_3.
[240] J. Wallenius, J. S. Dyer, P. C. Fishburn, R. E. Steuer, S. Zionts, and K. Deb, "Multiple Criteria Decision Making, Multiattribute Utility Theory: Recent Accomplishments and What Lies Ahead," *Manag. Sci.*, vol. 54, no. 7, pp. 1336–1349, Jul. 2008, doi: 10.1287/mnsc.1070.0838.
[241] C. Okoli and S. D. Pawlowski, "The Delphi method as a research tool: an example, design considerations and applications," *Inf. Manage.*, vol. 42, no. 1, pp. 15–29, Dec. 2004, doi: 10.1016/j.im.2003.11.002.
[242] J. B. Heaton, N. Polson, and J. Witte, "Deep Learning for Finance: Deep Portfolios," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2838013, Sep. 2016. Accessed: Jan. 20, 2020. [Online]. Available: https://papers.ssrn.com/abstract=2838013
[243] "The Book of Alternative Data: A Guide for Investors, Traders and Risk Managers | Wiley," Wiley.com. Accessed: Feb. 16, 2021. [Online]. Available: https://www.wiley.com/en-us/The+Book+of+Alternative+Data%3A+A+Guide+for+Investors%2C+Traders+and+Risk+Managers-p-9781119601807
[244] M. S. Gal and D. L. Rubinfeld, "Data standardization," *N. Y. Univ. Law Rev.*, vol. 94, no. 4, pp. 737–770, 2019, doi: 10.2139/ssrn.3326377.
[245] M. C. Thrun, "Knowledge discovery in quarterly financial data of stocks based on the prime standard using a hybrid of a swarm with SOM," in *ESANN 2019 - Proceedings, 27th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, ESANN (i6doc.com), 2019,

pp. 397–402. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85071284432&partnerID=40&md5=90e9b903406e179c9b76951f09e4fbcc

[246] "Phys. Rev. E 85, 056108 (2012) - Quantifying reflexivity in financial markets: Toward a prediction of flash crashes." Accessed: Feb. 16, 2021. [Online]. Available: https://journals.aps.org/pre/abstract/10.1103/PhysRevE.85.056108

[247] E. F. Fama, "Efficient Capital Markets: II," *J. Finance*, vol. 46, no. 5, pp. 1575–1617, 1991, doi: 10.1111/j.1540-6261.1991.tb04636.x.

[248] A. W. Lo, "Efficient Markets Hypothesis," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 991509, Jun. 2007. Accessed: Apr. 15, 2019. [Online]. Available: https://papers.ssrn.com/abstract=991509

[249] L. H. Pedersen, *Efficiently Inefficient: How Smart Money Invests and Market Prices Are Determined*. Princeton University Press, 2015.

[250] A. Bogatu, N. W. Paton, A. A. A. Fernandes, and M. Koehler, "Towards Automatic Data Format Transformations: Data Wrangling at Scale," *Comput. J.*, vol. 62, no. 7, pp. 1044–1060, Jul. 2019, doi: 10.1093/comjnl/bxy118.

[251] F. Endel and H. Piringer, "Data Wrangling: Making data useful again," *IFAC-Pap.*, vol. 48, no. 1, pp. 111–112, Jan. 2015, doi: 10.1016/j.ifacol.2015.05.197.

[252] S. Jiang and J. Kahn, "Data wrangling practices and collaborative interactions with aggregated data," *Int. J. Comput.-Support. Collab. Learn.*, vol. 15, no. 3, pp. 257–281, Sep. 2020, doi: 10.1007/s11412-020-09327-1.

[253] S. Kandel *et al.*, "Research directions in data wrangling: Visualizations and transformations for usable and credible data," *Inf. Vis.*, vol. 10, no. 4, pp. 271–288, Oct. 2011, doi: 10.1177/1473871611415994.

[254] Z. Milosevic, W. Chen, A. Berry, and F. A. Rabhi, *Real-Time Analytics*. Elsevier Inc., 2016. doi: 10.1016/B978-0-12-805394-2.00002-7.

[255] D. H. Bailey and M. L. de Prado, "The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting, and Non-Normality," *J. Portf. Manag.*, vol. 40, no. 5, pp. 94–107, Sep. 2014, doi: 10.3905/jpm.2014.40.5.094.

[256] N. Ståhl, G. Mathiason, G. Falkman, and A. Karlsson, "Using recurrent neural networks with attention for detecting problematic slab shapes in steel rolling," *Appl. Math. Model.*, vol. 70, pp. 365–377, 2019, doi: 10.1016/j.apm.2019.01.027.

[257] G. Luo, "A review of automatic selection methods for machine learning algorithms and hyper-parameter values," *Netw. Model. Anal. Health Inform. Bioinforma.*, vol. 5, no. 1, p. 18, May 2016, doi: 10.1007/s13721-016-0125-6.

[258] P. Neary, "Automatic Hyperparameter Tuning in Deep Convolutional Neural Networks Using Asynchronous Reinforcement Learning," in *2018 IEEE International Conference on Cognitive Computing (ICCC)*, Jul. 2018, pp. 73–77. doi: 10.1109/ICCC.2018.00017.

[259] P. Schratz, J. Muenchow, E. Iturritxa, J. Richter, and A. Brenning, "Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data," *Ecol. Model.*, vol. 406, pp. 109–120, Aug. 2019, doi: 10.1016/j.ecolmodel.2019.06.002.

[260] T. Hu and A. Tripathi, "The performance evaluation of textual analysis tools in financial markets," in *25th Annual Workshop on Information Technologies and*

Systems, WITS 2015, University of Texas at Dallas, 2015. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85006988312&partnerID=40&md5=51f2c3730142741b90781c8dc75d62b7

[261] J. Ding, V. Tarokh, and Y. Yang, “Model Selection Techniques: An Overview,” *IEEE Signal Process. Mag.*, vol. 35, no. 6, pp. 16–34, 2018, doi: 10.1109/MSP.2018.2867638.

[262] W. Zhang and C. Zhou, “Deep Learning Algorithm to solve Portfolio Management with Proportional Transaction Cost,” in *CIFEr 2019 - IEEE Conference on Computational Intelligence for Financial Engineering and Economics*, Institute of Electrical and Electronics Engineers Inc., 2019. doi: 10.1109/CIFEr.2019.8759056.

[263] A. J. Bean and A. C. Singer, “Factor graph switching portfolios under transaction costs,” in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Prague, 2011, pp. 5748–5751. doi: 10.1109/ICASSP.2011.5947666.

[264] H. I. Bülbül and Ö. Ünsal, “Comparison of classification techniques used in machine learning as applied on vocational guidance data,” in *Proceedings - 10th International Conference on Machine Learning and Applications, ICMLA 2011*, Honolulu, HI, 2011, pp. 298–301. doi: 10.1109/ICMLA.2011.49.

[265] “Performance Attribution of US Institutional Investors - Binay - 2005 - Financial Management - Wiley Online Library.” Accessed: Feb. 16, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1755-053X.2005.tb00102.x

[266] J. Simonian, C. Wu, D. Itano, and V. Narayanam, “A Machine Learning Approach to Risk Factors: A Case Study Using the Fama–French–Carhart Model,” *J. Financ. Data Sci.*, vol. 1, no. 1, pp. 32–44, Jan. 2019, doi: 10.3905/jfds.2019.1.032.

[267] I. Iskin, “An Assessment Model for Energy Efficiency Program Planning in Electric Utilities: Case of the Pacific of Northwest U.S.A.,” pp. 6–8, 2014, doi: 10.15760/etd.1850.

[268] D. Blitz, T. Hoogteijling, H. Lohre, and P. Messow, “How Can Machine Learning Advance Quantitative Asset Management?” Rochester, NY, Jan. 09, 2023. doi: 10.2139/ssrn.4321398.

[269] P. M. Mirete-Ferrer, A. Garcia-Garcia, J. S. Baixauli-Soler, and M. A. Prats, “A Review on Machine Learning for Asset Management,” *Risks*, vol. 10, no. 4, Art. no. 4, Apr. 2022, doi: 10.3390/risks10040084.

[270] R. N. Shukla, V. Vyas, A. Roy, and A. Chaturvedi, “Application of Data Science and Artificial Intelligence in Financial decisions: A survey,” in *2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Dec. 2022, pp. 1–6. doi: 10.1109/IATMSI56455.2022.10119241.

[271] D. Avramov, S. Cheng, and L. Metzker, “Machine Learning versus Economic Restrictions: Evidence from Stock Return Predictability,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3450322, Aug. 2020. doi: 10.2139/ssrn.3450322.

[272] Y. Tang *et al.*, “A survey on machine learning models for financial time series forecasting,” *Neurocomputing*, vol. 512, pp. 363–380, Nov. 2022, doi: 10.1016/j.neucom.2022.09.003.
[273] K. Kaczmarczyk and K. Miałkowska, “Backtesting comparison of machine learning algorithms with different random seed,” *Procedia Comput. Sci.*, vol. 207, pp. 1901–1910, Jan. 2022, doi: 10.1016/j.procs.2022.09.248.
[274] K. Olorunnimbe and H. Viktor, “Deep learning in the stock market—a systematic survey of practice, backtesting, and applications,” *Artif. Intell. Rev.*, vol. 56, no. 3, pp. 2057–2109, Mar. 2023, doi: 10.1007/s10462-022-10226-0.
[275] N. Nazareth and Y. V. Ramana Reddy, “Financial applications of machine learning: A literature review,” *Expert Syst. Appl.*, vol. 219, p. 119640, Jun. 2023, doi: 10.1016/j.eswa.2023.119640.
[276] H. Ko and J. Lee, “Can Chatgpt Improve Investment Decision? From a Portfolio Management Perspective.” Rochester, NY, Mar. 16, 2023. doi: 10.2139/ssrn.4390529.
[277] M. M. Dowling and B. M. Lucey, “ChatGPT for (Finance) Research: The Bananarama Conjecture.” Rochester, NY, Jan. 11, 2023. doi: 10.2139/ssrn.4322651.
[278] H. Umer and M. S. Khan, “ChatGPT in Finance: Addressing Ethical Challenges.” Rochester, NY, Apr. 29, 2023. doi: 10.2139/ssrn.4433171.
[279] F. Lu, L. Huang, and S. Li, “ChatGPT, Generative AI, and Investment Advisory.” Rochester, NY, Jul. 24, 2023. doi: 10.2139/ssrn.4519182.
[280] A. Zaremba and E. Demir, “ChatGPT: Unlocking the Future of NLP in Finance.” Rochester, NY, Jan. 13, 2023. doi: 10.2139/ssrn.4323643.
[281] I. Aldridge, “The AI Revolution: From Linear Regression to ChatGPT and Beyond and How It All Connects to Finance,” *J. Portf. Manag.*, Jul. 2023, doi: 10.3905/jpm.2023.1.519.
[282] Z. Feng, G. Hu, and B. Li, “Unleashing the Power of ChatGPT in Finance Research: Opportunities and Challenges.” Rochester, NY, Sep. 11, 2023. doi: 10.2139/ssrn.4424979.

APPENDICES

Appendix A: Letter of Invitation to Experts

Dear Expert X,

I am Farshad Saadatmand, a Ph.D. student in the Engineering and Technology Management Department at Portland State University.

I am researching the challenges of machine learning/data science (ML/DS) based investment research in developing quantitative investment strategies.

The core of my research is developing a model that investment companies can use to assess the readiness of developing strategies to be implemented in practice. To achieve this goal, subject-matter experts should validate and quantify the model.

It is a privilege to have your contribution as an expert by providing your invaluable inputs and insights to this research.

How to participate:

There are three main steps to join in this research:

1. Survey - to validate the model factors and perspectives (10 mins) - (https://portlandstate.qualtrics.com/jfe/form/SV_9MEhDo6dc3bPSF8)

2. HDM model - to do a pairwise comparison and ranking the factors (10 mins) - (http://research1.etm.pdx.edu/hdm2/expert.aspx?id=89fd984f28d37d88/ffda87ad54c43c75)

3. Survey - to quantify Desirability Curves. (10 mins) - (https://portlandstate.qualtrics.com/jfe/form/SV_5sX6GAyrbBSsOJU)

I look forward to hearing from you, and your participation and precious time are greatly appreciated in advance!

Farshad Saadatmand

Ph.D. Student

Department of Engineering and Technology Management, Portland State University

Member | CFA Society of Portland

moham29@pdx.edu

www.linkedin.com/in/farshad-saadatmand

Appendix B: Letter of Model Validation to Experts
Dear Expert X,

Thanks for accepting my invitation to participate as a subject matter expert in research titled “Strategy Readiness Assessment: A Hierarchical Decision Model to Evaluate Strategy’s Readiness of quantitative machine learning/data science (ML/DS) driven investment strategies.”

The core of my research is developing a model that investment companies can use to assess the readiness of developing strategies to be implemented in practice. To achieve this goal, subject-matter experts should validate and quantify the model.

Model perspectives and criteria have been identified through an extensive literature review. To validate them, please click on the link below to get access to the validation survey:

(https://portlandstate.qualtrics.com/jfe/form/SV_9MEhDo6dc3bPSF8)

Please follow the instructions as provided in the survey and give your responses to validate the model perspectives and criteria. The following phases will be sent out to you via email later. Thanks for your precious time, and I greatly appreciate your invaluable insights.

Sincerely,

Farshad Saadatmand

Ph.D. Student
Department of Engineering and Technology Management, Portland State University

Member | CFA Society of Portland

moham29@pdx.edu

www.linkedin.com/in/farshad-saadatmand

Appendix C: Letter of Model and Desirability Curve Quantifications
Dear Expert X,

Thanks for accepting my invitation to participate as a subject matter expert in research titled "Strategy Readiness Assessment: A Hierarchical Decision Model to Evaluate Strategy's Readiness of quantitative machine learning/data science (ML/DS) driven investment strategies."

The core of my research is developing a model that investment companies can use to assess the readiness of developing strategies to be implemented in practice. To achieve this goal, subject-matter experts should validate and quantify the model.

You will be asked to respond to model and desirability curve quantifications in this research phase. Please note that all your professional information and model responses will remain strictly confidential, and the researcher will only report the results at the aggregated level. Please let me know if you have any questions about quantifications.

To conduct the model and desirability curve quantifications, please click the following links and follow the instructions.

HDM model - to do a pairwise comparison and ranking of the factors (http://research1.etm.pdx.edu/hdm2/expert.aspx?id=89fd984f28d37d88/5142b057e20c44a3)

Survey - to quantify Desirability Curves (https://portlandstate.qualtrics.com/jfe/form/SV_5sX6GAyrbBSsOJU)

Thanks for your precious time, and I greatly appreciate your invaluable insights.

Sincerely,

Farshad Saadatmand

Ph.D. Student
Department of Engineering and Technology Management, Portland State University

Member | CFA Society of Portland

moham29@pdx.edu

www.linkedin.com/in/farshad-saadatmand

Appendix D. Supplemental File Information

The following supplemental files accompany this dissertation.

Tables of Improvement Simulation Cases

This supplemental file includes Tables 51 and 52 that demonstrate improvement simulations for case study 1 and case study 2 respectively.

File type: xlsx

File name: Tables of Improvement Simulation Cases

File size: 160 KB

Required software: Microsoft Excel

Model Validation Survey

This supplemental file includes all the survey questions, descriptions, and tables needed to validate all of the model’s perspectives and criteria.

File type: xlsx

File name: Model Validation Survey

File size: 3,905 KB

Required software: Microsoft Excel

Desirability Curve Quantification Survey

This supplemental file includes all the survey questions, descriptions, and tables to capture each decision criterion's desirability level.

File type: xlsx

File name: Desirability Curve Quantification Survey

File size: 3,520 KB

Required software: Microsoft Excel