Research on the Application of Machine Learning in the Asset Management Industry: A Literature Reviewof Risk Assessment and Investment Decision-Making

Zhengtao Xu
Civil Engineering (Leeds Joint School), Southwest Jiaotong University, Chengdu, Sichuan, China

Abstract: **Against the backdrop where the digital economy accounts for 42.8% of GDP, coupled with policy support and growing technology investment, the asset management industry is transitioning from experience-driven to technology-driven. This paper focuses on the core scenarios of risk assessment and investment decision-making, sorts out three core machine learning algorithms (supervised learning, unsupervised learning, and reinforcement learning), and analyzes their application value in credit risk assessment, market risk assessment, intelligent portfolio construction, and market trend prediction. Combined with empirical cases, it examines the challenges at the data, technology, and industry levels and puts forward corresponding suggestions. The research shows that machine learning, with its capabilities of multi-source data fusion, nonlinear relationship capture, and dynamic optimization, significantly improves the accuracy and efficiency of asset management business, providing core impetus for the high-quality development of the industry.**

Keywords: **Machine Learning; Asset Management; Risk Assessment; Investment Decision-Making; Financial Technology; Data Fusion; Algorithm Optimization**

1. Introduction

1.1 Research Background and Significance

Amid the digital wave where the digital economy accounts for 42.8% of GDP, machine learning and big data technologies are reshaping the operational logic of the financial industry. [1] As a core component of the financial sector, the traditional experience-driven and linear analysis models of the asset management industry are no longer able to cope with massive transaction data, complex market risk spillover, and the new demand of investors for "personalized allocation under risk control". The policy support from China's *New Generation Artificial Intelligence Development Plan* and the "Artificial Intelligence +" initiative provides a solid guarantee for the application of machine learning in the asset management field.[2] Its powerful capabilities in data processing and non-linear relationship capture have become the core driving force for the industry to break through development bottlenecks.

At the theoretical level, this paper integrates the technical principles of machine learning with the business logic of asset management, constructs a "technology-scenario-value" framework, expands the boundaries of traditional financial theories, and fills the gap in interdisciplinary research. At the practical level, it provides references for asset management institutions in technology selection and process restructuring, and also offers a basis for regulators to formulate technical application standards, contributing to the standardized and sustainable innovation of the industry.

1.2 Research Content and Framework

Focusing on the core applications of machine learning in the asset management field, this paper takes risk assessment and investment decision-making as the entry points, and constructs a research framework of "technical foundation-specific application-case verification-challenge and prospect". Specifically, it sorts out the core algorithms and development status of machine learning; analyzes its practical applications in various risk assessment and key investment decision-making scenarios; verifies the technical effects with empirical cases; examines the multiple challenges faced in applications and proposes optimization paths, so as to provide references for the development of the industry.

2. Current Status and Core Algorithms

2.1 Development Status of Machine Learning in the Asset Management Industry

In the asset management field, machine learning has formed a differentiated scenario-adaptation pattern: as the most mature method, supervised learning has achieved large-scale implementation in core links such as risk control, return prediction, and portfolio optimization. Mainstream institutions adopt GBDT, XGBoost, and deep learning models to process structured financial data and unstructured alternative data, with feature engineering supporting their coverage of key scenarios including credit rating and default prediction; unsupervised learning focuses on market structure analysis and anomaly detection—clustering algorithms such as K-Means assist in investor behavior and market state identification, while dimensionality reduction technologies like PCA and t-SNE provide support for factor analysis and risk attribution, with anomaly detection systems becoming a core risk control configuration. However, the insufficient output stability of such models, characterized by significant differences in clustering results across different time windows, limits their direct application in core investment decision-making; reinforcement learning has gradually moved from theoretical exploration to practice, demonstrating unique potential in scenarios such as dynamic decision-making, portfolio optimization, and algorithmic trading. Leading hedge funds, quantitative institutions, and intelligent investment advisory platforms have begun to apply Deep Reinforcement Learning (DRL) in businesses like asset allocation and risk hedging, with the core logic of optimizing long-term returns through continuous interaction between the agent and the market environment.

2.2 Core Machine Learning Algorithms

Machine learning-based risk management models in asset management are built on multi-dimensional big data analysis, greatly expanding the coverage of risk factors and forming the core support for "big data risk control".[3] Three core algorithms perform distinct functions and adapt to different scenarios in risk assessment and investment decision-making:

Supervised learning algorithms are trained on labeled data and suitable for precise prediction and classification. Traditional algorithms such as logistic regression and decision trees are widely used in credit risk assessment.[4] Deep learning algorithms offer prominent advantages: CNN combined with Bert-Transformer achieves an AUC value of 0.6130 in farmer default discrimination.[5] DNN fuses multi-source data, improving default recognition accuracy by approximately 30% compared with logistic regression; GNN series algorithms significantly reduce MSE and MAE relative to XGBoost models in systemic risk early warning.[6]

Unsupervised learning algorithms do not require labeled data and focus on mining intrinsic data patterns. Clustering algorithms such as K-Means support personalized asset allocation, while dimensionality reduction techniques like PCA simplify portfolio construction. In scenarios such as bank branch efficiency prediction, supplementary predictive dimensions can be derived by calculating the Euclidean distance matrix of feature data.[7]

Deep Reinforcement Learning (DRL) models portfolio management as a Markov Decision Process (MDP), serving as an important tool for financial market modeling and decision-making.[8] It enables dynamic adjustment of allocation ratios to balance returns and risks in asset allocation, and optimizes order execution strategies to reduce costs and improve efficiency in algorithmic trading.

2.3 Comparative Table of Core Algorithm Scenario Adaptability

Table 1. Comparative Table of Core Algorithm Scenario Adaptability

Application Scenario	Mainstream Algorithms	Advantages	Limitations	Applicable Institution Types	Key Performance Indicators Example
Credit Risk Assessment	XGBoost	Efficient processing of structured data;fast training speed	Weak adaptability to unstructured data	Commercial banks,consumer finance companies	30% improvement in default recognition accuracy
Credit Risk Assessment	CNN+Bert	Strong text data processing capability;effective capture of semantic information	Large demand for training data (needing over 10,000 samples)	Large asset management institutions,leading banks	AUC value of 0.6130 (farmer default discrimination)
Systemic	GNN	Captures network	Weak model	Regulatory	32.84% reduction in

Risk Early Warning	(GraphSAGE)	transmission paths;fuses multi-source indicators	interpretability;high computing power requirements	authorities,leading securities firms	MSE.18.94% reduction in MAE
Asset Allocation	DRL	Dynamically optimizes allocation ratio;adapts to market volatility	Sensitive to parameters;long training cycle	Quantitative institutions,hedge funds	4.2% improvement in annualized return;3.8% reduction in maximum drawdown
Market State Identification	K-Means	Applicable without labeled data;efficient computation	Poor result stability;dependent on initial cluster numbers	All types of asset management institutions (for auxiliary decision making)	Over 75% accuracy in market state identification

3. Application of Machine Learning in Risk Assessment of the Asset Management Industry

3.1 Credit Risk Assessment: Multi-Source Data Fusion and Model Optimization

Traditional credit risk assessment relies on single structured data, with lag and one-sidedness. Machine learning achieves a significant improvement in assessment accuracy through multi-source information integration and sample optimization. In corporate credit risk assessment, interpretable models incorporate non-linear indicators such as group decentralization level and dividend ratio, increasing the prediction accuracy by 12.7% compared with traditional models.[9] For groups with sparse data such as farmers, the MHLS-ALR-LightGBM-SHAP model achieves a default recall rate of 85.73% through sample optimization and decision interpretation.[10] The CNN+Bert model processes text data from credit investigation reports, achieving an AUC value of 0.6130 in farmer default discrimination.[5]

At the commercial bank level, the CLXT fusion model integrates 33 risk indicators, controlling the prediction errors of stable and non-stable risk indices within 10.41% and 1.53% respectively.[11] The K-means-SMOTE oversampling method solves the sample imbalance problem, and the Stacking model constructed based on it achieves a test set AUC of 0.7568, with default recognition accuracy improved by about 30% compared with traditional logistic regression.[12]

3.2 Market Risk Assessment: Dynamic Prediction and Scenario Simulation Upgrade

The core of market risk assessment is to capture volatility laws. Machine learning achieves precise prediction through time-series analysis and multi-factor integration. In the field of green finance, dual machine learning models identify the non-linear correlation between green bond risks and policies as well as industry prosperity, enabling accurate assessment of the risk premium level of issuers.[13,14] Digital green finance platforms provide an innovative path for environmental risk assessment in the asset management field. Based on panel data of 287 cities from 2010 to 2023, Lü Wanqing & Xiao Zhaofu empirically show through dual machine learning models that such platforms can promote the quality and quantity improvement of green credit, reduce the risk-taking of commercial banks, and significantly enhance the synergistic efficiency of pollution reduction and carbon emission reduction.[15] At the macro level, the financial risk index of commercial banks is strongly correlated with the economic cycle, and there is a risk superposition effect when multiple indicators are triggered. LSTM combined with natural language processing parses central bank policy texts, enabling prediction of interest rate fluctuations 1-3 months in advance.[11]

Unsupervised learning plays an important supporting role: dimensionality reduction technologies such as PCA extract core risk factors such as macroeconomics and liquidity from massive data, simplifying model complexity; clustering algorithms such as K-Means can identify market states such as bull markets, bear markets, and volatile markets, providing adaptive schemes for risk assessment in different environments.

3.3 Systemic Risk Early Warning: Network Modeling and Multi-Source Monitoring

The core of systemic risk early warning is to capture the transmission paths of risks across institutions and markets. With its capabilities in multi-source data integration and pattern recognition, machine learning has become a key technical support in this field.

Research by Yang Zihui et al. confirms that risk preference is a forward-looking indicator of China's financial risks. The machine learning early warning model constructed based on bank risk preference integrates multi-source data such as bank balance sheets, market transactions, and macroeconomics, focusing on monitoring key indicators such as changes in risk preference, leverage level, and related-party transactions. It can warn of systemic risk hazards 3-6 months in advance, with the out-of-sample prediction goodness-of-fit reaching up to 53.97%.[16]

Graph Neural Networks (GNN) have further innovated early warning technologies. Based on 99,185 valid comment data from East Money Stock Bar over 214 days, Tu Yan & Jin Guanhao constructed an inter-bank social network with banks as nodes and the number of common investors as edge weights, embedding multi-dimensional indicators such as macroeconomics, bank systems, micro-operations, and financial sentiment. [6] They proposed a GNN early warning model. Empirical results show that after adding financial sentiment indicators, the GraphSAGE model achieves optimal performance when the sliding window is 2, with MSE and MAE reduced by 32.84% and 32.84% respectively compared with traditional models such as XGBoost.

In addition, the combination of knowledge graphs and machine learning can clearly depict the correlation between financial institutions such as inter-bank lending and cross-shareholding, accurately identifying key nodes of risk transmission, and providing targeted basis for systemic risk prevention and control.

4. Application of Machine Learning in Investment Decision-Making of the Asset Management Industry

4.1 Intelligent Portfolio Construction: Dynamic Optimization and Personalized Matching

Machine learning has promoted the transformation of portfolios from static allocation to dynamic optimization, realizing the precise personalized matching of "risk-return".

In macro asset allocation, Cheng Shi & Xu Jie combined the Cobb-Douglas production function with random forest and LSTM models to calculate Hong Kong's potential growth rate, confirming that the performance of the financial market is a key influencing factor, providing support for the selection of major asset classes. [17] The dynamic allocation model based on deep reinforcement learning of leading quantitative institutions achieved a 4.2% improvement in annualized return and a 3.8% reduction in maximum drawdown compared with the traditional mean-variance model in the 2017-2023 market.

In terms of personalized allocation, unsupervised learning clustering algorithms can accurately divide customer groups and generate differentiated schemes based on risk preferences and investment goals. China Merchants Bank's "Mojie Zhitou" integrates the Markowitz model with machine learning to dynamically adjust the stock-bond ratio. In the early stage of the 2020 epidemic, it reduced the weight of equity assets to 40%, and the user holding period was 30% longer than that of traditional fund customers. In segmented scenarios, the postal savings bank branch efficiency prediction system constructed by Wang Zifeng et al. based on the XGBoost model integrates 5 core indicators, with a prediction error MAPE of only 3.9%, providing quantitative basis for offline channel investment.[7]

4.2 Market Trend Prediction and Timing: Pattern Recognition and Dynamic Decision-Making

Market trend prediction and timing selection are the core of investment decision-making. With its capabilities in pattern recognition and dynamic optimization, machine learning has significantly improved the accuracy and efficiency of decision-making.

In terms of trend prediction, deep learning models such as CNN and LSTM can mine non-linear patterns and long-term dependency relationships in transaction data and macro indicators.[18] Natural language processing technology can parse unstructured data such as news sentiment and policy texts, extract market sentiment and policy orientation information, and timely perceive trend changes.

4.3 Industry and Individual Stock Selection: Factor Mining and Value Judgment

Through multi-dimensional factor mining and precise valuation, machine learning provides a scientific basis for industry allocation and individual stock selection, breaking through the limitations of traditional fundamental analysis.

In terms of industry selection, machine learning integrates multi-dimensional data such as industry prosperity and policy support to identify industries with growth potential. Based on the causal inference of dual machine learning, Wu Yongxia et al. found the heterogeneous impact of big data applications on corporate financialization, providing references for industry risk assessment and allocation; through dual machine learning models, Shang Ying et al. confirmed that financial technology has a more significant improvement effect on the efficiency of listed banks and banks in eastern regions, pointing out the segmented direction for financial sector investment. [19,20]

In terms of individual stock selection, the model processes multi-source data such as corporate finance and public opinion to mine profit-driven factors and undervalued opportunities. Through text analysis, Zhu Kang & Tang Yong constructed a data factor utilization indicator, confirming its significant positive correlation with the intensity of corporate digital technology innovation; through feature engineering and model optimization, Lu Nachuan effectively identified the key factors of corporate default risk, providing support for individual stock risk assessment. [21,12]

5. Challenges Faced by the Application of Machine Learning in the Asset Management Industry

5.1 Data-Level Challenges

Data quality and compliance are the core constraints on the application of machine learning models. In terms of data quality, structured data has problems such as inconsistent calibers and many missing values, while unstructured data has low standardization and high noise interference. Imperfect data governance can easily lead to the dilemma of "garbage in, garbage out" in model training; the high cost of alternative data procurement and processing forms a threshold for small and medium-sized asset management institutions. [10]

The pressure of data security and privacy protection is prominent. Asset management institutions involve a large amount of sensitive customer information and transaction data, and compliance requirements are increasingly strict. Balancing data sharing and security protection has become an industry problem.[22] In addition, differences in data compliance rules across regions in cross-border businesses, imbalanced data timeliness, and noise interference may lead to model learning biases, affecting prediction accuracy.

5.2 Technical-Level Challenges

Insufficient model interpretability is a core bottleneck. Complex algorithms such as deep learning and integrated learning are mostly "black-box" models, whose decision logic is difficult to explain, affecting investor trust and triggering regulatory compliance risks. [11] Although there are interpretable tools such as SHAP and LIME, it is still difficult to fully restore the decision-making process in complex scenarios.

The generalization ability and stability of models need to be improved. The volatility and time-variability of the financial market make models trained on historical data prone to "overfitting".[12] And predictions are likely to fail in extreme market environments. Macroeconomic prediction models are sensitive to scenario assumptions.[17] At the same time, there are large differences in model adaptability across different asset management scenarios, and general-purpose models are difficult to meet segmented needs.[14]

5.3 Industry-Level Challenges

Industry standards and regulatory norms are not yet sound. The existing framework is difficult to adapt to asset management businesses driven by machine learning, and conflicts in regulatory rules in cross-border businesses increase compliance costs.[9]

The shortage of interdisciplinary talents has become a key constraint. The market has an urgent demand for interdisciplinary talents with "technology + business" capabilities, which is particularly prominent in small and medium-sized institutions. The insufficient technical literacy of some practitioners leads to deviations in model applications.

Traditional business processes and organizational structures are not adapted to intelligent decision-making. Cumbersome approval processes and lack of cross-departmental collaboration affect the timeliness of intelligent decision-making and the efficiency of technology-business integration. There are significant differences in model adaptability among institutions of different sizes, and the lack of industry standards leads to incomparable model outputs, which may trigger risks such as algorithmic discrimination.[7]

6. Conclusions and Prospects

6.1 Research Conclusions

With its capabilities in data fusion, model innovation, and scenario adaptability, machine learning has achieved key breakthroughs in risk assessment and investment decision-making in the asset management field, becoming the core driving force for the digital transformation of the industry.

At the risk assessment level, machine learning has broken the reliance on traditional single data, covering various risk subjects and risk types, and significantly improved assessment accuracy and response speed. Supervised learning and integrated learning solve the problems of data imbalance and non-linearity in credit risk assessment; unsupervised learning realizes risk factor extraction and market state identification; graph neural networks accurately capture systemic risk transmission paths; dual machine learning models provide causal inference support for green finance risk prevention and control.

At the investment decision-making level, quantitative strategy generation, dynamic asset allocation, and human-machine collaboration models have promoted the transformation of asset management businesses from experience-driven to data-driven. Deep reinforcement learning optimizes dynamic asset allocation; multi-dimensional data integration improves the accuracy of market trend prediction; complex algorithms provide a scientific basis for industry and individual stock selection; human-machine collaboration models balance decision-making efficiency and robustness.

However, the application of machine learning still faces multiple challenges: at the data level, there are problems such as uneven quality, prominent compliance risks, and high acquisition costs; at the technical level, it is limited by bottlenecks such as insufficient model interpretability, limited generalization ability, and to-be-improved stability; at the industry level, there are obstacles such as lack of standards and norms, shortage of interdisciplinary talents, and inadequate adaptation of business processes and organizational structures. These problems need to be solved through technological innovation, institutional improvement, and process restructuring.

6.2 Future Prospects

At the technical level, the interpretability and generalization ability of models will continue to improve. Technologies such as SHAP and LIME will be deeply integrated with complex models; transfer learning and federated learning will be widely applied to solve the problems of model adaptability and data silos; at the same time, the integration of machine learning with blockchain, knowledge graphs and other technologies will be further deepened.

In terms of application scenarios, it will expand towards full-process and refined development. Real-time risk control systems and personalized intelligent investment advisory will become mainstream; scenario-based applications will extend from traditional risk assessment and portfolio optimization to multiple fields such as intelligent risk control and green investment.

At the industry level, standards and norms as well as the ecological system will be gradually improved. Regulators will clarify technical application standards; asset management institutions will restructure business processes and organizational structures; universities and enterprises will collaborate to cultivate interdisciplinary talents; data sharing mechanisms under compliance premises will be gradually established.

References

[1] CAICT. *Research Report on China's Digital Economy Development 2024*[R]. 2024.

[2] The State Council. New Generation Artificial Intelligence Development Plan[Z]. 2017.

[3] Liu Chunhang. Artificial Intelligence, Big Data and Financial Risk Management[J]. Financial Regulation Research, 2025,(05):1-11.

[4] Xu Yuncheng, Yang Yongqing, Wu Hengliang, et al. Research on Customer Asset Allocation Decision Preference Prediction Based on Machine Learning[J]. Financial Technology Research, 2021,(03):45-56.

[5] Dong Bingjie. Bank Farmer Microcredit Investment Decision Model Based on Default Probability[D]. Dalian University of Technology, 2024.

[6] Tu Yan, Jin Guanhao. Research on Banking Systemic Risk Early Warning Integrating Financial Sentiment Analysis [J]. Journal of the Central University of Finance and Economics, 2025(11):56-73.
[7] Wang Zifeng, Chen Juanjuan, Wang Tengrui. Predicting the Efficiency of Newly Established Branches of Postal Savings Bank in the Context of Digital Finance: An Explainable Machine Learning Approach [J]. Postal Research, 2025(6):15-23.
[8] Xu Bo, He Yijun, Wen Jiancheng, et al. A Review of Deep Reinforcement Learning Applications in Quantitative Trading for Financial Markets [J]. Journal of Intelligent Science and Technology, 2024, 6(04): 416-428.
[9] Liu Bin, Xu Shun, Li Haoran, et al. Do Group Characteristics Aid in Predicting Corporate Debt Crises? Empirical Evidence Based on Explainable Machine Learning Models [J]. China Industrial Economics, 2025(7):174-192.
[10] Wang Shouyang, Zhang Yi, Zhang Tingting, et al. Research on Intelligent Assessment of Farmer Credit Risk Based on Sample Optimisation and Explainable Machine Learning [J]. Agricultural Economic Issues, 2025(11): 62-84.
[11] Cao Zeren, Li Di. Research on Financial Risk Prediction and Early Warning Mechanisms for Chinese Commercial Banks Based on Big Data [J]. Shanghai Economy, 2025(6):93-112.
[12] Lu Nachuan. Research on Credit Default Risk Assessment Based on K-means-SMOTE and Improved Stacking Methods [D]. Yangzhou University, 2025.
[13] Song Gaoyan, Deng Xin, Zong Jiafeng. The Impact of Green Bonds on Corporate Green Total Factor Productivity: Empirical Evidence Based on Multi-Period Double Difference and Dual Machine Learning [J]. Exploration of Financial Theory, 2025(6):42-53.
[14] Wu Guowei, Shang Jin. The Cost Advantage of Green Bond Issuance and Its Determinants: Evidence from Machine Learning [J]. Bonds, 2025(10):28-34.
[15] Lü Wanqing, Xiao Zhaofu. Synergistic Effects of Digital Green Finance Platforms on Pollution Reduction and Carbon Mitigation: Evidence from Dual Machine Learning Causal Inference [J]. Journal of Guizhou University of Finance and Economics, 2025
[16] Yang Zhihui, Li Dongcheng, Chen Yutian. Can Risk Preference Serve as a Forward-Looking Indicator of China's Financial Risk? — New Evidence from Cutting-Edge Machine Learning Methods [J]. Management World, 2025, (10): 21-37.
[17] Cheng Shi, Xu Jie. Estimation of Hong Kong's Potential Growth Rate: A Machine Learning Approach[J]. Shanghai Finance, 2025(11):14-23.
[18] Liu Lingjun, Chen Xiaoliang, Chen Yanbin. Re-evaluating Factors Influencing Monetary Policy Effectiveness: An Approach Based on Explainable Machine Learning Methods [J]. Journal of Quantitative Economics and Technical Economics, 2026, (01): 27-52.
[19] Wu Yongxia. The Impact of Big Data Applications on the Financialisation of Real Economy Enterprises: A Causal Inference Approach Based on Dual Machine Learning [J]. Accounting Research, 2024, (02): 56-68.
[20] Shang Ying, Wang Zhucun, Yin Chengyuan. Research on Fintech Empowering Bank Efficiency in the Context of New Quality Productive Forces: Causal Inference Based on Dual Machine Learning[J]. Statistics Theory and Practice, 2025(10):4-11.
[21] Zhu Kang, Tang Yong. Data Element Utilisation and Corporate Financial Asset Allocation: Evidence from Machine Learning and Text Analysis [J]. Accounting Research, 2025(6):1-19.
[22] Ouyang Ruihui, Li Wenjie. Research on Commercial Bank Data Asset Valuation Based on the Shapley Value Model [J]. Modern Financial Research, 2025(8):32-46.