European Journal of Business, Economics & Management

Vol. 1 No. 3 2025

Article Open Access

Exploration of Data-Driven Capital Market Investment Decision Support Model

Chenyang An [1,*]

[1] Economics and Finance, Queen Mary University of London, London, E1 4NS, United Kindom
* Correspondence: Chenyang An, Economics and Finance, Queen Mary University of London, London, E1 4NS, United Kindom

Received: 27 June 2025
Revised: 04 July 2025
Accepted: 16 July 2025
Published: 22 July 2025

Abstract: With the increasing complexity of the capital market and the explosive growth of available information, traditional investment decision-making methods have become increasingly insufficient in meeting the complex and dynamic needs of modern investors. As a new research direction, data-driven investment decision support model has gradually attracted the attention of the academic and practical circles. This paper discusses the influence of market efficiency theory, behavioral finance and portfolio theory on investment decision, and analyzes the application of data-driven model in individual stock investment, portfolio optimization, market trend prediction and risk management. A data-driven capital market investment decision support model is further designed, which covers the key links of data acquisition and preprocessing, model selection and construction, model training and optimization, etc., providing theoretical and practical basis for improving the scientificity and effectiveness of investment decision.

Keywords: data-driven; capital markets; investment decision

1. Introduction

At present, in the modern capital market, the complexity and diversity of investment decisions make the traditional investment methods face great challenges. With the inflow and rapid increase of massive information and the development of analytical technology, a data-driven investment decision support model has gradually become the focus of research and practice [1]. These models can integrate market information, economic data, and investor behavior, and provide investors with more scientific decision-making basis through data analysis. Using a data-driven approach, capital markets investment decisions can be fully supported, covering key areas such as individual stock investment, portfolio optimization, and market trend forecasting and risk management.

2. Capital Market Theory

2.1. Market Efficiency Theory

In the 1960s, Eugene Fama developed the theory of market efficiency, which argued that in an efficient capital market, all available information is immediately reflected in asset prices, making it impossible for any investor to analyze publicly available information to achieve excess returns. According to the theory, prices in the capital market are rational judgments made by investors based on all currently known information, so market prices adjust rapidly to reflect the impact of changes in information. EMH can be divided into three forms: weak, semi-strong, and strong forms, which correspond to the degree to which asset prices reflect historical data, publicly available information, and all information including insider knowledge, respectively. The importance of the theory lies in the fact that it challenges the possibility of predicting market movements through technical analysis or fundamental analysis, emphasizing the central role of information in the formation of market prices [2].

2.2. Behavioral Finance

Behavioral finance is an extension of traditional financial theory that aims to explain how investors are affected by psychological biases and irrational factors in their decision-making process. Unlike traditional finance, which assumes that investors are rational decision makers, behavioral finance argues that investors are often disturbed by cognitive biases, mood swings, and psychosocial factors that cause market prices to deviate from their intrinsic value. Common behavioral biases include overconfidence, loss aversion, anchoring effect and group behavior. For example, loss aversion refers to the fact that investors react more strongly to a loss than to an equivalent gain, which may lead them to hold on to a losing investment excessively and be unwilling to cut their losses in time. The anchoring effect is that investors will over-rely on certain initial information when making decisions, affecting subsequent judgments. By revealing these irrational behaviors, behavioral finance tries to explain many market phenomena that cannot be explained by traditional finance, such as asset bubbles, stock market fluctuations, and collective investor behavior. By revealing these irrational behaviors, behavioral finance tries to explain many market phenomena that cannot be explained by traditional finance, such as asset bubbles, stock market fluctuations and collective behavior of investors. Its core idea is that markets are not always perfectly rational and that investor psychology plays an important role in decision-making and market dynamics.

2.3. Portfolio Theory

Portfolio theory, proposed by Harry Markowitz in 1952, focuses on optimizing the balance between investment return and risk through rational asset allocation. The core idea is that by combining different assets together, investors can maximize expected returns with a certain level of risk. Markowitz's theory emphasizes that diversification reduces the risk of a single asset, thereby improving the stability and returns of the overall portfolio. In portfolio theory, the most important concept is diversification, which aims to reduce risk by spreading investments across different assets. Markowitz points out that different assets in a portfolio may exhibit different risk and return characteristics, and by combining these assets properly, you can reduce the volatility of the overall portfolio. The efficient frontier curve describes the form of the optimal investment portfolio and provides guidance for investors in choosing an appropriate combination of assets. Portfolio theory is based on the assumption that markets are rational and that investors make decisions based on the trade-off between return and risk when choosing a portfolio. Although markets are not always rational in practical applications, portfolio theory still provides a powerful framework for asset allocation and risk management, becoming one of the cornerstones of modern investment theory [3].

3. Application of Investment Decision Support Model in Capital Market

3.1. Individual Stock Investment Decision

Individual stock investment decisions play a crucial role in the capital market, investors rely on market data, company fundamentals, and technical indicators to guide their decisions. Therefore, the data-driven decision support model can provide investors with a strong decision basis by analyzing historical data, corporate financial status, and market trends. By leveraging data analysis models, investors can more effectively assess the investment potential of individual stocks. Common investment decision models for individual stocks include models based on financial ratios, technical analysis models and machine learning models. Through these models, investors are able to identify potential quality stocks, reduce investment risk and increase returns. Table 1 below shows several core indicators commonly used in individual stock investment decisions and their descriptions and calculation methods.

Table 1. Core financial and technical indicators in investment decisions of individual stocks.

Index	Description	Calculation method
Price/earnings ratio	A measure of a company's profitability, a low P/E ratio can mean a stock is undervalued.	P/E ratio = stock price/earnings per share
Price-to-book ratio	A ratio that reflects the value of a company's assets and is often used to assess the market pricing of a company's stock.	Price-to-book ratio = stock price/net asset value per share
Return on equity	The higher the ROE, the stronger the profitability of the company.	ROE = Net profit/shareholders' equity
5-day moving average	A technical indicator that reflects the short-term trend of a stock and is often used to determine a short-term buying opportunity.	Calculate the closing of the last 5 days

By combining price-to-earnings ratios, return on equity (ROE) and 5-day moving averages, investors are able to identify individual stocks with good growth potential and stable returns. At the same time, the use of machine learning algorithms (such as random forest, support vector machine, etc.) for forecasting, investors can also obtain more accurate forecast of individual stock trends and further optimize investment decisions.

3.2. Portfolio Optimization

Portfolio optimization is the key to balancing risk and return in the capital market. Its core goal is to maximize the expected return of the portfolio or minimize the risk through reasonable allocation of asset weights. In the data-driven investment decision support model, portfolio optimization is usually achieved by means of mean-variance optimization model. Let's say the portfolio contains n assets, Asset i has an expected return of $E(R_i)$, Assets i the risk (variance)is σ_i^2 ,the covariance between assets is σ_{ij}. The expected return $E(R_p)$ and risk σ_p of a portfolio can be expressed as:

$$E(Rp) = \sum_{i=1}^{n} w_i\, E(R_i) \quad (1)$$

$$\sigma\frac{2}{p} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i\, w_j \sigma_{ij} \quad (2)$$

Among them, w_i is assets i the weight of, $E(R_i)$ is the expected return on the asset, σ_{ij} As asset i and asset j between the covariance. In order to achieve optimization, the goal is usually to maximize the sharpe ratio of the portfolio:

$$S = \frac{E(Rp) - R_f}{\sigma p} \quad (3)$$

Among them,R_f is the risk-free rate, the Sharpe ratio reflects the excess return per unit of risk. By weighting the portfolio w_i Optimization of, It is possible to find the combination that achieves the greatest return for a given risk. Typically, this optimization problem can be addressed using numerical techniques such as quadratic programming or heuristic algorithms like genetic algorithms. In addition, modern data-driven investment decision models also combine machine learning and artificial intelligence techniques to estimate the expected return and covariance matrix of assets through big data analysis, thereby further improving the optimization effect [4].

3.3. Market Trend Forecasting and Risk Management

In the capital market, market trend prediction and risk management are key components of the investment decision support model. With data-driven technology, historical market data, technical analysis metrics and machine learning models can be combined to predict future market movements and monitor and manage risks in real time. Time series analysis is one of the commonly used methods in market trend forecasting, which predicts future market trends by modeling historical market data (such as stock prices, trading volumes, etc.). The Autoregressive Integrated Moving Average (ARIMA) model and the Seasonal ARIMA (SARIMA) model are classical tools in time series analysis. In addition, machine learning-based models such as support vector machines, random forests, and deep learning models are able to handle more complex non-linear data relationships to help predict market movements more accurately. In risk management, the value risk model is widely used to help investors assess and control the level of risk by quantifying the maximum loss a portfolio may suffer. Monte Carlo simulation is a common risk assessment tool that evaluates the impact of different market fluctuations on a portfolio by simulating a variety of market scenarios. Table 2 summarizes several common methods for market trend forecasting and risk management, along with their application areas and key features.

Table 2. Comparison of market trend forecasting and risk management methods.

Methods/Tools	Application field	peculiarity
Time series analysis	Market trend forecasting	It is suitable for data with obvious time series characteristics, and the model has strong explanatory ability
Support vector machine	Market trend prediction	It is suitable for processing complex nonlinear data and has good generalization ability
Value risk	Risk management	Quantitative risk loss, a measure of risk applicable to a portfolio
Monte Carlo simulation	Risk assessment and management	Can simulate a variety of market scenarios to help with situational analysis
Sentiment analysis	Market sentiment forecasting and trend forecasting	It can quickly reflect market sentiment fluctuations and is suitable for short-term forecasting

Through the combination of the above methods, it can provide investors with more comprehensive and accurate market trend prediction and risk management support, and optimize the investment decision-making process.

4. Design of Data-Driven Capital Market Investment Decision Support Model

4.1. Data Acquisition and Preprocessing

In the process of designing a data-driven capital market investment decision support model, data acquisition and preprocessing are the basic steps. First, the goal of data collection is to gather historical data relevant to investment decisions, mainly from financial market data, macroeconomic data, corporate financial data, and market sentiment data. For example, financial market data usually includes the price and trading volume of financial products such as stocks, bonds and futures, etc. Macroeconomic data includes GDP, inflation rate, interest rate, and exchange rate, while corporate financial data, derived from annual and quarterly reports, mainly includes income, profit, and liabilities. Market sentiment data captures investor emotions through unstructured sources such as news articles and social media posts. These data provide multi-dimensional support for the subsequent decision model. However, the collected data often has missing values, outliers or inconsistent formats, so it needs to be cleaned and preprocessed [5].

The process of data cleaning includes missing value processing, outlier detection, data standardization and feature engineering. In terms of missing value processing, methods such as mean interpolation and nearest neighbor interpolation can be used to fill the missing data, or directly delete the samples with too many missing values. The detection of outliers can use statistical methods such as Z-score or IQR to identify abnormal fluctuations in the data, and outliers can be corrected or deleted. Since the scales of different data sources may vary significantly, it is necessary to standardize or normalize the data. The commonly used standardization methods include Z-score standardization and Min-Max normalization. Feature engineering generates new features by processing original data, such as calculating technical indicators like moving averages and rates of return from time series data, or extracting emotion scores through sentiment analysis of text data. These steps ensure the quality of the data and provide a solid foundation for subsequent model training and prediction. (See Figure 1).

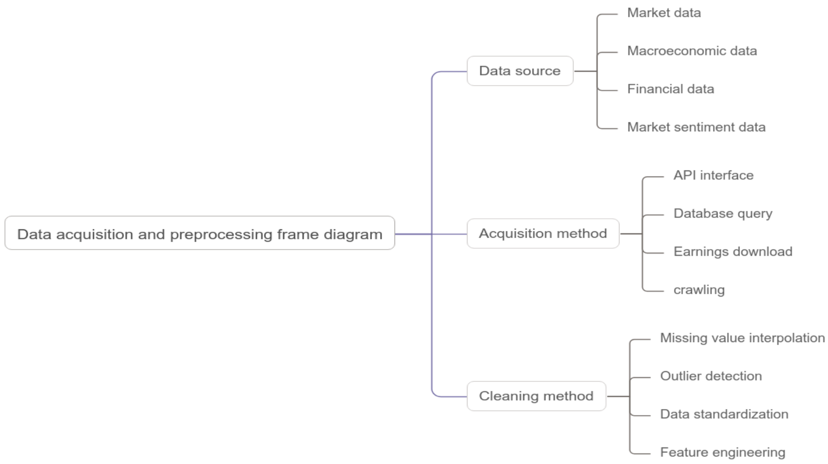

Figure 1. Framework of data acquisition and preprocessing.

Through the above collection and pre-processing steps, the quality and consistency of data can be ensured, laying a solid foundation for the subsequent model construction, and thus improving the accuracy and effectiveness of the capital market investment decision support model.

4.2. Model Selection and Construction

When constructing the capital market investment decision support model, it is essential to choose an appropriate mathematical model. By analyzing different data characteristics and market behavior, this study selects a combination of machine learning algorithms and statistical models to maximize prediction accuracy and model generalization ability. Firstly, according to the characteristics of time series data of capital market, we choose the method of combining ARIMA (autoregressive integral moving average model) and LSTM (long short-term memory network). The ARIMA model has an advantage in handling short-term market fluctuations and seasonal changes, while the LSTM network is effective in capturing long-term market trends and non-linear relationships. The combined model can balance prediction accuracy and interpretability.

$$Y_T = \mu + \sum_{i=1}^{p} \varphi_i \, Y_{t-i} + \sum_{j=1}^{q} \theta_j \in_{t-j} + \in_t \tag{4}$$

Among them, Y_t is the observed value of the time series, $\emptyset_i$ and θ_j Is autoregressive coefficient and moving average coefficient, ϵ_t Is the white noise error term,p and q are the order of autoregressive and moving average terms respectively. The LSTM model uses the following activation functions and error backpropagation mechanisms to optimize the parameters and capture the long-term non-linear trend of the market:

$$h_t = \tanh(W_h x_t + U_h h_{t-1} + b_h) \tag{5}$$

$$y_t = W_y h_t + b_y \tag{6}$$

Among them, x_t is the input data, h_t is hidden state,W_h、U_h、$b_h$is the weight and bias of LSTM, y_t is the predicted value.By combining the advantages of ARIMA and

LSTM models, the investment decision support model has high accuracy in time series prediction and trend analysis, and provides a reliable decision basis for capital market investors.

4.3. Model Training and Optimization

Through rigorous training, the model's predictive accuracy and decision-making performance can be enhanced. First, the training data set includes historical market data, company financial indicators, macroeconomic data, etc. Feature engineering is then applied to this data to eliminate noise and extract meaningful input features. Next, the appropriate model is selected for training, such as support vector machines, random forests, or deep neural networks. According to different models, different training algorithms and loss functions are selected to optimize the model parameters. In order to improve the generalization ability of the model, the cross-validation method is used to evaluate the performance of the model and avoid overfitting. Based on the results of cross-validation, the model is optimized to select the best learning rate, regularization coefficient and other parameters. In the optimization process, optimization algorithms such as gradient descent are used to adjust the weight and bias and minimize the loss function [6]. In order to accelerate the training process, parallel computing and distributed computing techniques are used to process large-scale financial data sets. Table 3 shows the adjustment of various hyperparameters during model training.

Table 3. Hyperparameter adjustment of the model.

Parameter name	Initial value	Optimal value	Instructions
Learning rate	0.01	0.001	Adjust the update step of model parameters
Regularization coefficient	0.1	0.05	Control the complexity of the model to prevent overfitting
Number of hidden layers	3	4	The number of intermediate layers in the neural network
Number of hidden layer nodes	128	256	The number of nodes in each layer affects the expressiveness of the model
Batch size	32	64	The number of data samples used in each iteration

Through the optimization of the above parameters, the model can better adapt to different market environments, improve the accuracy of prediction and the ability of decision support.

5. Conclusion

With the continuous development of capital markets, data-driven investment decision support models play a crucial role in optimizing investment strategies, improving market prediction accuracy and enhancing risk management capabilities. By combining market efficiency theory, behavioral finance theory and portfolio theory, the DSS model designed has strong practical significance in individual stock investment, portfolio optimization and market trend prediction. With ongoing advancements in data technology and continuous model refinement, the model will provide more accurate decision support for investors in the future, and further improve the efficiency and stability of the capital market.

References

1. L. Yun, "Analyzing Credit Risk Management in the Digital Age: Challenges and Solutions", *Econ. Manag. Innov.*, vol. 2, no. 2, pp. 81–92, Apr. 2025, doi: 10.71222/ps8sw070.

2. H. Javanmard, A. Khodamipour, and O. Pourheidari, "The effectiveness of investors' decision-making from legal requirements in the capital market," *Financ. Manag. Perspect.*, vol. 14, no. 45, pp. 9–36, 2024.
3. M. Eklund, A. M. Omidvar, J. van der Horst, T. Smith, C. Chen, S. Ramesh, et al., "Community microgrids: A decision framework integrating social capital with business models for improved resiliency," *Appl. Energy*, vol. 367, Art. no. 123458, 2024, doi: 10.1016/j.apenergy.2024.123458.
4. L. Chen, Y. Wang, Q. Liu, H. Zhang, F. Sun, R. Li, *et al.*, "Uncertainty analysis and optimization modeling with application to supply chain management: A systematic review," *Mathematics*, vol. 11, no. 11, Art. no. 2530, 2023, doi: 10.3390/math11112530.
5. P. Q. Khang, V. D. Thuan, T. V. Hoang, N. H. Binh, M. T. Thanh, and Q. H. Nam, "Machine learning for liquidity prediction on Vietnamese stock market," *Procedia Comput. Sci.*, vol. 192, pp. 3590–3597, 2021, doi: 10.1016/j.procs.2021.09.132.
6. J. Ren, "Research on financial investment decision based on artificial intelligence algorithm," *IEEE Sens. J.*, vol. 21, no. 22, pp. 25190–25197, 2021, doi: 10.1109/JSEN.2021.3104038.