International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 7864-7887

https://doi.org/10.70517/ijhsa464677

Collaborative filtering-driven financial data analysis - dual-module model for investment decision support

Chengcheng Zhu[1,*]
[1] Accounting College, Zhengzhou Vocational College of Finance and Taxation, Zhengzhou, Henan, 450000, China
Corresponding authors: (e-mail: Zcc0720050044@126.com).

Abstract Financial data analysis plays a crucial role in investment decision-making, but investors must also be wary of the issues that may arise. To address issues such as information lag, financial data manipulation, and distorted financial indicators. This paper employs hierarchical clustering and principal component analysis methods to conduct a diagnostic study of the financial condition of HK Company. The study primarily involves selecting financial condition diagnostic indicators, performing dimensionality reduction on the diagnostic indicators, and extracting the principal components of the diagnostic indicators. Subsequently, a comprehensive evaluation function is constructed based on the contribution rates of each principal component. This evaluation function enables the determination of the company's current financial condition, providing reliable data support for investment decisions. A collaborative filtering algorithm based on weighted triads is proposed as an investment decision-support model to provide investment decision schemes for investors. Experimental analysis indicates that the proposed model outperforms the benchmark method in estimating user preferences with greater accuracy. It also addresses the data sparsity issue where most results are zero when calculating the similarity between investment products using traditional collaborative filtering recommendation algorithms.

Index Terms weighted triads, collaborative filtering algorithm, hierarchical clustering, principal component analysis, investment decision-making

I. Introduction

In today's complex and ever-changing economic environment, the accuracy of corporate investment decisions directly impacts their market competitiveness and long-term development [1]. Financial analysis, as a crucial tool for evaluating the economic benefits of investment projects and predicting investment risks, plays an irreplaceable role in optimizing corporate investment decisions [2]. In the digital age, businesses possess vast amounts of data; however, traditional financial analysis methods are constrained by data scale and analytical capabilities, making it difficult to fully unlock the value of this data and convert it into actionable business insights [3]-[5]. Additionally, traditional investment decision-making methods often overlook the comprehensiveness and depth of financial analysis, leading to blind spots and risks in investment decisions [6], [7]. Therefore, by analyzing investment decision-making models based on financial analysis, their specific applications, and optimization, we aim to provide companies with more scientific and reasonable investment decision-making criteria [8], [9].

The emergence of big data technology has brought revolutionary changes to corporate financial decision-making, offering more comprehensive, accurate, and timely decision support [10]. Data-driven analytical paradigms enable enterprises to break through the limitations of traditional single data sources, achieving multidimensional integration of financial, market, and operational data [11], [12]. The application of machine learning and deep learning technologies in predictive analytics enhances the accuracy of financial indicator predictions, providing reliable data support for investment decisions [13]-[15]. Based on this, big data-driven financial analysis, as an important trend in the future development of financial analysis, will provide strong data support for corporate decision-making and drive the transformation of enterprises into data-driven organizations [16], [17]. Enterprises need to actively embrace big data technology, cultivate data analysis talent, and establish data security systems to remain competitive in the future [18], [19].

The deep integration of big data analysis technology with financial decision-making has enabled a shift from static analysis to dynamic monitoring, and from experience-based judgment to data-driven decision-making. On one hand, cloud computing and distributed processing technologies have enabled enterprises to build efficient data processing platforms, making real-time financial analysis possible. Literature [20] developed a mathematical model for identifying key variables in cloud capacity decision-making, which can assess investment opportunities in cloud capabilities for enterprise users under probabilistic conditions, thereby providing valuable insights for users' investment decisions. Literature [21] constructed a cloud accounting IT audit model to analyze financial status indicator data for enterprises. This model can accurately identify and judge abnormal situations in data on cloud accounting platforms, providing support for enterprises' sustainable development investment decisions. Literature [22] introduced cloud accounting information systems into the hotel industry, exploring the utility of cloud-based accounting functions in enhancing the competitive advantage of the hotel industry and promoting its long-term economic growth.

On the other hand, the deepening application of artificial intelligence technology in financial analysis has enabled enterprises to establish intelligent financial early warning mechanisms that can identify risks in a timely manner. Literature [23] proposes an enterprise financial risk prediction model based on a BP neural network. By focusing on changes in enterprise financial performance, it can provide enterprises with relatively accurate financial crisis predictions and has played an important role in the practice of enterprise financial early warning. Literature [24] combines the LSTM neural network model with the attention mechanism to predict systemic risks in China's financial markets. It incorporates online public opinion indicators into the early warning model, demonstrating high generalization performance and prediction accuracy. Literature [25] compares the predictive effectiveness of different early warning systems for financial crises, finding that early warning indicators based on price volatility feedback rates are more effective than logistic regression and contrast models in improving crisis prediction accuracy. Literature [26] builds upon enterprise financial data mining using a BP neural network, introducing mobile edge computing services and a geographic point of interest information optimization model, significantly improving the accuracy of enterprises' predictions of their own financial health and crises, thereby assisting enterprises in improving their financial management.

It is evident that the application of technologies such as artificial intelligence and cloud computing will further drive the intelligent development of financial analysis, making it more efficient, intelligent, and secure, and providing stronger support for corporate decision-making to help enterprises achieve more efficient and sustainable development. Of course, this cannot be achieved without the support of intelligent algorithms.

This paper uses HK Company's financial and non-financial data from 2019 to 2024 as a basis to establish a financial indicator system that combines financial and non-financial indicators. It uses hierarchical clustering methods to preprocess the data and preliminarily screen the required analysis variables. It uses principal component extraction methods to further reduce the dimensionality of the indicator data after hierarchical clustering and calculates the comprehensive financial status scores for each year. To address the issue of low diversity in traditional collaborative filtering algorithms, a collaborative filtering algorithm based on weighted triads is proposed. By introducing label information in the context of sparse data and limited additional information, both user interests and investment product attributes can be reflected simultaneously. A triad graph is constructed using the triadic relationships between users, investment products, and labels to calculate user preference scores and integrate product similarity. Finally, resource reallocation is performed on the weighted tripartite graph using the heat conduction method to uncover more similarity relationships, and collaborative filtering frameworks are employed for investment decision recommendations.

II. Principal component analysis method

Principal component analysis is a mathematical method that reduces the dimensionality of a large amount of data that would otherwise need to be compared, using a few principal components that represent most of the data information to represent the data. Not only are the principal components not highly correlated with each other, but they also cover most of the information represented by the previous indicators. Using SPSS 20.0 software to process the financial performance evaluation indicators selected for this paper, several principal components that represent the majority of the financial performance evaluation indicators are obtained after data processing. The specific values of each principal component can be calculated using the factor score matrix in the SPSS analysis results. Each principal component in the results represents one of the “four capabilities,” with only weak correlations among them, which is beneficial for further financial analysis [27], [28].

The following explains how principal component analysis is applied to actual financial performance evaluation. Below are the specific steps used in this paper to evaluate the financial performance of HK Company using principal component analysis:

The first step is to select financial performance evaluation indicators based on indicator selection principles, and then construct a raw data matrix based on the selected indicators. The raw evaluation matrix X used in this paper represents m financial indicators horizontally and n sample companies vertically. The specific matrix is as follows:

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \tag{1}$$

The second step involves standardizing the raw data. First, due to the large number of sample companies and significant differences in data, if standardization is not performed, the final principal component scores will also vary greatly, which does not reflect the actual financial disparities between companies and is inconsistent with objective reality. Second, the units of the selected financial performance evaluation indicators are not comparable. During data processing, it is necessary to reduce the differences in analysis results caused by different units. The standardized processing formula used in the paper is:

$$x'_{ij} = \frac{x_{ij-x_j}}{\sqrt{Var(x_j)}} \tag{2}$$

In the paper, SPSS 20.0 software was used to standardize the data. In the equation, $\overline{x}_j$ represents the standard deviation of the variable. In the equation, $\sqrt{Var\left(x_j\right)}$ represents the standard deviation of the variable.

The third step involves combining the results of the suitability test in the SPSS software to determine whether the selected original financial performance evaluation indicators can be subjected to principal component analysis. The most important criteria for this determination are whether the KMO value and Bartlett's sphericity test value in the suitability test results meet the minimum requirements. If the requirements are not met, it indicates that the principal components cannot cover most of the original indicator information, failing to meet the requirements for analysis in the paper. The specific threshold values are that the KMO value must exceed 0.6, and the Bartlett sphericity test result must be less than 0.05.

The fourth step is to extract the principal components. When determining the number of principal components, consider the contribution rate of the jth eigenvalue and the contribution rate of the first m eigenvalues. e_j is the contribution rate of a specific eigenvalue, and E_m is the sum of the contribution rates of the first m eigenvalues. To determine the number of principal components, identify how many eigenvalues satisfy the conditions of having a value greater than 1 and a contribution rate sum greater than 90%.

$$e_j = \frac{\lambda_j}{\sum_{i=1}^{p} \lambda_j} \tag{3}$$

$$E_m = \sum_{j=1}^{m} e_j \tag{4}$$

The fifth step is to interpret the extracted principal components. Determining the loadings of the principal components in each original indicator is done in order to objectively interpret the principal components. The relationship formula for principal component loadings is:

$$I_{ij} = \sqrt{\lambda} u_{ij} \left(i = 1, 2, \cdots, m; j = 1, 2, \cdots, p\right) \tag{5}$$

Each principal component contains the original data, but the correlation between each principal component and the original indicators is different. However, the loadings between the principal components and each original indicator are different. Based on the factor loading table, the loadings between each principal component and each original indicator can be clearly seen. Loadings are represented by numerical values, which can be either positive or negative. In determining the principal components, the paper examines the absolute values of the loadings, selects the principal component with the largest absolute loading value among the original indicators, and names it accordingly. This is because the absolute value of the loading represents the extent to which the principal component encompasses the information of the indicators. The closer the absolute value of the loading is to 1, the more information from the original indicators it can explain.

Finally, determine the formula for each principal component. Following the above steps, each principal component is named, and then the linear formula for each principal component is determined. The formula is determined by referring to the component matrix table, where each column in the matrix table represents the coefficient of each financial indicator, and each column represents a principal component. The formula is expressed as follows:

$$F_j = b_{j1}X_1 + b_{j2}X_2 + \cdots + b_{jp}X_p \text{ , } j = 1,2,\cdots,m \tag{6}$$

Among them, F_j represents the jth principal component, and there are a total of m principal components. X_i represents the ith financial indicator, and there are a total of p financial indicators. Since the paper focuses on the important indicators that constitute each principal component, each principal component is composed of the indicator factors that have the greatest impact on it, so the indicator factors included in each principal component are different.

III. Financial condition diagnosis research

III. A. Construction of a Financial Condition Diagnosis Indicator System

Taking all of the above factors into consideration and combining them with HK Company's financial statement data, this paper categorizes the company's data into six major categories based on profitability, solvency, growth potential, operational efficiency, cash generation capacity, and non-financial indicators. The financial condition diagnostic indicators used in this paper are summarized in Table 1.

Table 1: Diagnostic index system for financial situation

Index	Code	Index
Profitability	X1	Return on equity (weighted) (%)
	X2	Net asset yield (amortized) (%)
	X3	Net asset profit rate (%)
	X4	Sales net profit rate (%)
	X5	Fee adoption rate (%)
	X6	Cost profit per cent (%)
Solvency	X7	Mobility ratio
	X8	Speed ratio
	X9	Overspeed ratio
	X10	Equity ratio
	X11	Cash current liability ratio
Growth ability	X12	Revenue growth (%)
	X13	Rate of operating profit (%)
	X14	Net profit growth rate (%)
	X15	Net equity growth rate (%)
	X16	Total asset growth rate (%)
Operational capacity	X17	Inventory turnover (times)
	X18	Receivable turnover (times)
	X19	Turnover of current assets (times)
	X20	Fixed asset turnover (times)
	X21	Equity turnover (times)
	X22	Total asset turnover (times)
Ability to acquire	X23	Sales cash ratio (%)
	X24	Cash recovery (%)
	X25	Total asset cash recovery (%)
Non-financial indicators	X26	z-index
	X27	Independent proportion (%)

III. B. Preprocessing of financial condition diagnostic indicators

III. B. 1) Preliminary data processing

This paper uses data from HK Company for the six-year period from 2019 to 2024 as a basis, employing Excel spreadsheet functions to calculate the required financial ratios. These ratios are compiled into an indicator data table to reflect the company's profitability, debt-repayment capacity, growth potential, operational efficiency, and cash-generating ability. The financial condition diagnostic indicators for each year from 2019 to 2024 are shown in Table 2.

Table 2: HK's indicator data of 2019-2024

Index	Code	2019	2020	2021	2022	2023	2024
Profitability	X1	-0.0261	-0.0575	0.2934	0.0513	0.0564	0.0591
	X2	-0.0111	-0.0662	0.2863	0.0531	0.0750	0.0921
	X3	-0.0016	-0.0328	0.1781	0.2546	0.3643	1.6221
	X4	-0.2363	0.0218	0.0739	0.0585	0.0744	-0.9683
	X5	4.1782	4.4021	8.0916	7.8177	9.4845	14.8501
	X6	-0.2315	-0.9599	4.9504	6.7605	10.9745	29.7118
Solvency	X7	2.0133	1.9699	1.5116	1.9915	1.5336	1.4470
	X8	0.9287	1.0844	0.7139	0.9511	0.7847	0.0969
	X9	0.9935	1.2214	1.0822	1.0715	0.4930	1.1308
	X10	0.4791	0.6033	0.7121	0.7668	0.8453	0.8594
	X11	0.0087	-0.0187	-0.0201	-0.0146	0.0048	0.0292
Growth ability	X12	2.7922	38.9701	27.8960	12.8874	-3.5099	95.4025
	X13	-99.2614	10.1137	-3.4643	1.0597	0.4207	3.7184
	X14	0.0889	0.1014	0.0837	0.0978	0.0986	0.0797
	X15	-0.0464	-0.2500	0.1193	1.0170	1.3163	4.4900
	X16	0.1800	7.1051	6.9847	3.9895	5.1126	6.8126
Operational capacity	X17	2.9814	2.9571	2.9448	2.8656	2.5065	2.9962
	X18	2.9347	2.9872	2.7278	2.4011	2.5424	2.2958
	X19	0.0363	0.0521	0.0719	0.0828	0.0754	0.1359
	X20	0.1477	0.2065	0.2602	0.2878	0.2585	0.5302
	X21	0.0351	0.0862	0.0604	0.0859	0.0878	0.1679
	X22	0.0250	0.0504	0.0505	0.0559	0.0429	0.0730
Ability to acquire	X23	0.2029	0.1939	0.1666	0.1926	0.2123	0.1969
	X24	-48.2357	1.0970	-241.4955	-55.9379	64.4633	53.9458
	X25	0.3319	-0.0069	-0.5787	-0.1835	0.3057	1.1224
Non-financial indicators	X26	1.55	1.55	1.55	1.55	1.55	1.55
	X27	33.17	40.55	31.26	46.70	31.29	31.47

III. B. 2) Data dimension reduction processing

In this paper, 27 financial indicators from HK Company were selected for diagnosis, and a hierarchical clustering method was used to reduce the dimensionality of these indicators. All financial and non-financial data indicators from 2019 to 2024 were imported from Excel spreadsheets into the SPSS system, and variables were set according to the system's requirements. Subsequently, data mining analysis was conducted on these data. The final filtered variable system is shown in Table 3. After dimensionality reduction of the indicator data, 11 redundant indicators were removed. The remaining 16 indicators fully represent the information of the original indicator system, so these 16 diagnostic indicators were selected for subsequent diagnostic research.

Table 3: Index system after pretreatment

Index	Code	Index
Profitability	X2	Net asset yield (amortized) (%)
	X4	Sales net profit rate (%)
	X5	Fee adoption rate (%)
	X6	Cost profit per cent (%)
Solvency	X7	Mobility ratio
	X11	Cash current liability ratio
Growth ability	X12	Revenue growth (%)
	X13	Rate of operating profit (%)
	X16	Total asset growth rate (%)
Operational capacity	X18	Receivable turnover (times)
	X21	Equity turnover (times)
	X22	Total asset turnover (times)
Ability to acquire	X23	Sales cash ratio (%)
	X24	Cash recovery (%)
	X25	Total asset cash recovery (%)
Non-financial indicators	X27	Independent proportion (%)

III. C. Extraction of principal components of financial condition diagnostic indicators

Dimension reduction of the indicators was performed using hierarchical clustering, and 16 financial indicators were ultimately selected. However, it was difficult to determine whether there were any financial indicators that could be further merged. Therefore, this paper used principal component analysis to first extract variable factors, further reduce dimensions, and perform data analysis to derive a comprehensive evaluation function.

III. C. 1) Extraction of principal components

Principal component analysis (PCA) is not a simple dimension reduction technique, but rather a method of integrating the information contained in the data to generate new indicators that represent the original data information. It is another approach to data dimension reduction. The extracted principal components are comprehensive representations of the original financial diagnostic indicators, and each principal component is independent of the others, meaning that the corresponding vectors are orthogonal to one another. Additionally, the final principal components must satisfy the condition that their eigenvalues are greater than 0.5.

The results of principal component extraction are shown in Table 4. The four principal components extracted contain more than 90% of the information from the original financial diagnostic indicators. To illustrate this more clearly, a scatter plot of the eigenvalues extracted from the principal components is shown in Figure 1. The first four principal components cover most of the original data information without causing any loss of data information.

Table 4: Schematic diagram of the total variance decomposition table

Constituent	Initial eigenvalue			Extract the sum of squares and load		
	Total	Variance/%	Cumulation/%	Total	Variance/%	Cumulation/%
1	8.224	0.514	0.514	8.224	0.514	0.514
2	4.528	0.283	0.797	4.528	0.283	0.797
3	1.375	0.086	0.883	1.375	0.086	0.883
4	0.624	0.039	0.922	0.624	0.039	0.922
5	0.381	0.024	0.946			
6	0.21	0.013	0.959			
7	0.143	0.009	0.968			
8	0.099	0.006	0.974			
9	0.088	0.006	0.98			
10	0.078	0.005	0.985			
11	0.061	0.004	0.989			
12	0.059	0.004	0.993			
13	0.052	0.003	0.996			
14	0.039	0.002	0.998			
15	0.021	0.001	0.999			
16	0.018	0.001	1.000			

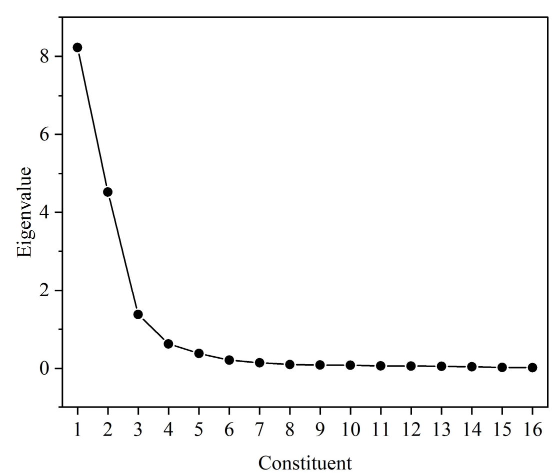


Figure 1: Characteristic value gravel diagram

To clearly illustrate the relationship between the principal component factors and the original variables, the extreme variance method was used to rotate the component loading matrix, and the rotated component loading matrix is shown in Table 5.

Principal Component 1 (F1) is primarily influenced by five financial condition diagnostic indicators: return on equity (diluted), net profit margin, cost-to-revenue ratio, equity turnover rate, and total asset turnover rate. F1 is named the operational capability factor.

Principal component 2 (F2) is primarily explained by six financial indicators: current ratio, cash-to-current liabilities ratio, accounts receivable turnover rate, sales cash ratio, operating income cash recovery rate, and total assets cash recovery rate. F2 is named the cash generation capability factor.

Principal component 3 (F3) is primarily explained by three financial indicators: operating profit growth rate, total assets growth rate, and independent director ratio. F3 is named the growth capability factor.

Principal component 4 (F4) is primarily explained by two financial condition diagnostic indicators: management expense ratio and operating revenue growth rate. F4 is named the profitability factor.

Table 5: Rotational load matrix

Code	Constituent			
	1	2	3	4
X2	**0.920**	-0.236	-0.007	0.322
X4	**0.976**	-0.160	0.021	0.104
X5	-0.967	0.032	0.124	**0.207**
X6	**0.948**	0.192	0.015	0.173
X7	-0.004	**0.985**	-0.01	-0.096
X11	0.479	**0.729**	-0.345	0.308
X12	0.595	0.047	0.203	**0.771**
X13	0.382	-0.147	**0.888**	-0.052
X16	0.395	-0.265	**0.729**	0.336
X18	0.009	**0.966**	0.054	0.187
X21	**0.923**	-0.118	0.183	-0.258
X22	**0.925**	-0.123	0.218	0.293
X23	0.059	**0.981**	-0.125	-0.043
X24	0.238	**0.945**	0.082	-0.006
X25	0.568	**0.711**	-0.316	0.347
X27	0.481	0.302	**0.665**	-0.013

III. C. 2) Formation of the comprehensive evaluation function

The impact factor scores obtained in the previous step are used as the coefficients for each principal component. The values of each principal component are calculated based on the contribution rates of the respective impact indicator variables and the values of the indicator variables themselves. Specifically, the factor score coefficient matrix obtained during the principal component extraction process is shown in Table 6.

Table 6: Factor score coefficient matrix

Code	Constituent			
	1	2	3	4
X2	0.122	-0.003	-0.062	0.103
X4	0.187	-0.008	0.015	0.197
X5	-0.289	0.019	0.044	0.535
X6	0.162	-0.002	0.049	-0.057
X7	-0.012	0.236	-0.081	-0.183
X11	0.005	0.086	-0.180	0.235
X12	-0.104	-0.049	0.031	0.73
X13	0.058	0.027	0.418	-0.215
X16	-0.033	-0.036	0.296	0.275
X18	-0.086	0.227	0.081	0.145
X21	0.138	-0.017	0.054	-0.007
X22	0.103	-0.028	0.06	0.086
X23	-0.001	0.194	-0.009	-0.16
X24	0.03	0.218	-0.106	-0.156
X25	0.007	0.088	-0.136	0.228
X27	-0.144	0.132	0.386	0.075

According to Table 6, the specific representation of each principal component factor can be obtained, where the coefficients of each original diagnostic indicator are the contribution rates of their corresponding principal components. The specific expressions are as follows:

F1=0.122*X2+0.187*X4-0.289*X5+0.162*X6-0.012*X7+0.005*X11-0.104*X12+0.058*X13-0.033*X16-0.086*X18+0.138*X21+0.103*X22-0.001*X23+0.03*X24+0.007*X25-0.144*X27

F2=-0.003*X2-0.008*X4+0.019*X5-0.002*X6+0.236*X7+0.086*X11-0.049*X12+0.027*X13-0.036*X16+0.227*X18-0.017*X21-0.028*X22+0.194*X23+0.218*X24+0.088*X25+0.132*X27

F3=-0.062*X2+0.015*X4+0.044*X5+0.049*X6-0.081*X7-0.18*X11+0.031*X12+0.418*X13+0.296*X16+0.081*X18+0.054*X21+0.06*X22-0.009*X23-0.106*X24-0.136*X25+0.386*X27

F4=0.103*X2+0.197*X4+0.535*X5-0.057*X6-0.183*X7+0.235*X11+0.73*X12-0.215*X13+0.275*X16+0.145*X18-0.007*X21+0.086*X22-0.16*X23-0.156*X24+0.228*X25+0.075*X27

Based on the composition expressions of the four principal components and the coefficients of each principal component calculated using the component loading matrix, a comprehensive evaluation function capable of assessing a company's financial condition is derived, namely Y = 0.143*F1 + 1.084*F2 + 0.86*F3 + 1.831*F4. By applying the relevant indicator data of HK Company from 2019 to 2024 to this formula, we can calculate the values of each principal component factor for the past six years and the comprehensive evaluation function score for the company's overall financial condition, as shown in Table 7.

This allows for a clear understanding of the company's operational performance in each year and identifies which years experienced financial issues. To determine which specific indicators caused problems and impacted HK Company's overall financial condition, analysis can be conducted based on the key influencing factors identified during the principal component extraction process. Principal components are a comprehensive representation of specific indicators. After identifying which principal component is the primary influencing factor, it is necessary to conduct a detailed analysis of the main constituent indicators of that principal component, grounding the indicators in the company's actual operational activities, thereby addressing the issues in practice. This enables a systematic analysis of the company's operational strategy based on actual conditions and targeted solutions to the financial challenges faced by HK Company.

Table 7: HK company's financial comprehensive score for the past six years

	2019	2020	2021	2022	2023	2024
F1	-2.7022	2.4294	-15.2053	-2.4884	-7.3197	-4.8529
F2	6.1721	5.0997	4.7299	5.4027	5.5750	-11.5886
F3	34.0790	32.0589	31.3065	31.9189	19.5596	12.5473
F4	42.0201	39.6858	23.4185	35.4346	7.4485	12.7445
Y	100.2380	93.8010	-4.5558	87.9875	35.3514	17.5413

IV. Investment decision support model

Innovation in investment decision support models has become key to companies seeking efficient returns on investment. With the growth of data volumes and advances in analytical technology, traditional investment models are unable to fully capture the complexity and dynamism of the market. Therefore, this paper proposes a collaborative filtering algorithm based on weighted triads to deeply mine hidden patterns and correlations in data, thereby providing more accurate guidance for investment decisions.

IV. A. Construction of the Three-Part Model

The item-user+ag tripartite graph model treats users U, investment products I, and labels T as abstract nodes and constructs a tripartite graph using the relationships between the nodes. $U=\{u_1,u_2,\cdots,u_m\}$ represents the set of m users, $I=\{i_1,i_2,\cdots,i_n\}$ represents the set of n investment products, $T=\{t_1,t_2,\cdots,t_r\}$ denotes the set of r tags. The relationships among the three sets are represented by the tripartite graph $IUT=(U,I,T,E)$ is used to represent the relationship between the three sets, where $U\cap I\cap T=\varnothing$, and E represents the set of connections between users and investment products and between users and tags. The edges between users and investment products represent users' selection of investment products, and the edges between users and tags represent users' use of tags.

The tripartite graph consists of two subgraphs (the bipartite graphs user-item and user-tag), which are represented as follows:

$$A=(a_{ui})_{m\times n}=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1i} \\ a_{21} & a_{22} & \cdots & a_{2i} \\ \vdots & \vdots & & \vdots \\ a_{u1} & a_{u2} & \cdots & a_{ui} \end{pmatrix} \tag{7}$$

$$A'=(a'_{ut})_{m\times r}=\begin{pmatrix} a'_{11} & a'_{12} & \cdots & a'_{1t} \\ a'_{21} & a'_{22} & \cdots & a'_{2t} \\ \vdots & \vdots & & \vdots \\ a'_{u1} & a'_{u2} & \cdots & a'_{ut} \end{pmatrix} \tag{8}$$

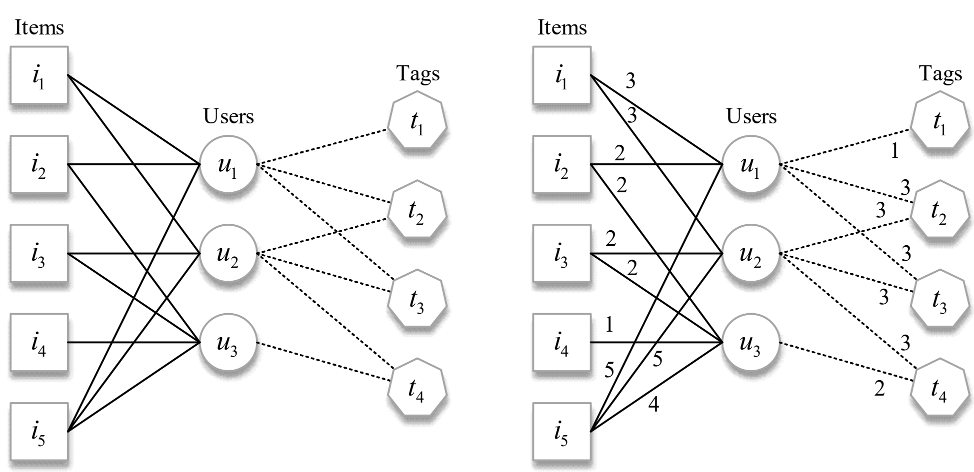
If user u chooses investment product i, $a_{ui}=1$ otherwise $a_{ui}=0$ Similarly, if user u uses the tag t, $a^{'}_{ut}=1$, otherwise $a^{'}_{ut}=0$ unweighted item-user-tag tripartite graph model is shown in Figure 2(a).

In the tripartite graph, the weights of the edges between users, investment products, and tags typically represent the degree to which users like investment products and tags. In this paper, the degree to which users like investment products and tags is defined as user preference, including user preference for investment products (UPI) and user preference for tags (UPT). Then, the user preference values are used as the weights of the corresponding edges in the item-user-tag tripartite graph to construct a weighted item-user-tag tripartite graph model. The definitions of user preference for investment products and user preference for tags are as follows.

Definition 1: In the tripartite graph IUT, if there is an edge between user u and investment product i, the weight of the edge between user u and investment product i represents user u's preference for investment product i, defined as user u's preference for investment product i, denoted as $upi(i,u)$.

Definition 2: In the tripartite graph IUT, if there is a connected edge between user u and tag t, then the weight of the connected edge between user u and tag t represents user u's preference for tag t, defined as user u's preference for tag t, denoted as $upt(t,u)$.

Based on the above definitions, the weighted item-user-tag tripartite graph model is shown in Figure 2(b). In the weighted user-item and user-tag bipartite graphs A and $A^{'}$, if user u chooses to invest in product i, then $a_{ui}=upi(i,u)$, otherwise $a_{ui}=0$. Similarly, if user u uses tag t, $a^{'}_{ut}=upt(t,u)$, otherwise $a^{'}_{ut}=0$.



(a) Unweighted tripartite graph

(b) weighted tripartite graphs

Figure 2: Tripartite network model

In the weighted user-item bipartite graph, denote the sum of the edge weights of all edges connected to investment product i as:

$$k(i)=\sum_{u=1}^{m} a_{ui}=\sum_{u=1}^{m} upi(i,u) \tag{9}$$

Indicates the degree of investment product i. Denotes the sum of the edge weights of all edges connected to user u:

$$k(u)=\sum_{i=1}^{n}a_{ui}=\sum_{i=1}^{n}upi(i,u) \tag{10}$$

Indicates the degree of user u. The transpose matrix of A is $A^T=(a_i u)_{n\times m}$. Based on this, construct a diagonal matrix related to user degree:

$$D_U=diag(d_1^U,\cdots,d_u^U,\cdots,d_m^U) \tag{11}$$

$$d_u^v=\begin{cases}\frac{1}{k(u)}, & k(u)\neq 0\\ 0, & k(u)=0\end{cases} \tag{12}$$

Construct a diagonal matrix related to the degree of investment products:

$$D_I=diag(d_1^{'},\cdots,d_i^{'},\cdots,d_n^{'}) \tag{13}$$

$$d_i^I=\begin{cases}\frac{1}{k(i)}, & k(i)\neq 0\\ 0, & k\,|\,(i)=0\end{cases} \tag{14}$$

In the weighted user-tag bipartite graph, denote the sum of the edge weights of all edges connected to tag i as:

$$k(t)=\sum_{u=1}^{m}a_{ut}^{'}=\sum_{u=1}^{m}upt(t,u) \tag{15}$$

Denotes the degree of label t. Let denote the sum of the edge weights of all edges connected to user u:

$$k'(u)=\sum_{t=1}^{r}a'_{ut}=\sum_{t=1}^{r}upt(t,u) \tag{16}$$

The transpose matrix $A^{'T}=(a_{tu}^{'})_{r\times m}$ of the degree $A^{'}$ of user u constructs a diagonal matrix related to the user degree:

$$D_U^{'}=diag(d_1^{'U},\cdots,d_u^{'U},\cdots,d_m^{'U}) \tag{17}$$

$$d_u^{'U}=\begin{cases}\frac{1}{k'(u)}, & k'(u)\neq 0\\ 0, & k'(u)=0\end{cases} \tag{18}$$

Construct a diagonal matrix related to tag degrees:

$$D_T=diag(d_1^T,\cdots,d_t^T,\cdots,d_r^T) \tag{19}$$

$$d_t^T=\begin{cases}\frac{1}{k(t)}, & k(t)\neq 0\\ 0, & k(t)=0\end{cases} \tag{20}$$

IV. B. User Preference Calculation

Consider the three-part graph IUT as two bipartite graphs UI and UT. Using projection techniques, map the binary relationships in the two bipartite graphs to the single-mode networks U^I and U^T associated with two users. Calculate the users' preferences for investment products and tags, which serve as the weights for the corresponding edges between users and investment products, and between users and tags in the three-part graph. The matrix $Q=PP^T$ in Figure U' represents the relationships between each pair of users who have selected the same investment product, while the matrix $Q^{'}=P^{'}P^{'T}$ in Figure U^T represents the relationships between each pair of users who have used the same tag. Matrices P^r and $P^{'T}$ are the transposed matrices of matrices P and P', respectively. The process of converting the tripartite graph into bipartite graphs and single-mode networks is illustrated in Figure 3.

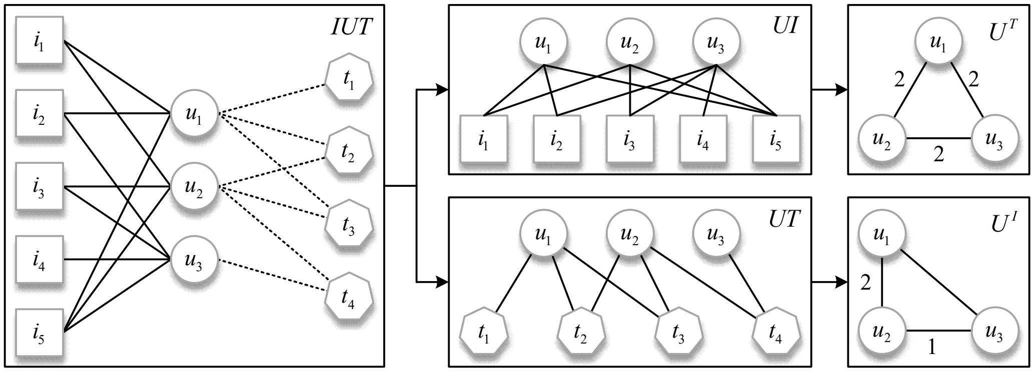


Figure 3: Transformation process

The similarity of users' preferences for investment products indicates that if two users choose certain identical investment products, their preferences for investment products are similar. The element x_{uv} in the similarity matrix X of users' preferences for investment products represents the similarity of preferences for investment products between users u and v:

$$x_w = \frac{q_w}{q_w + q_w - q_w}, q_w \in Q \tag{21}$$

Based on the similarity of users' preferences for investment products as described above, we can obtain users' preferences for investment products:

$$Y = P^T X \tag{22}$$

The element γ_{iu} in Y represents user u's preference for investment product i, $upi(i,u)$.

The element $x_{uv}^{'}$ in the user preference similarity matrix X represents the preference similarity between users u and v for a label:

$$x_{uv}^{'} = \frac{q_{uv}^{'}}{q_{uv}^{'} + q_{uv}^{'} - q_{uv}^{'}}, q_{uv}^{'} \in Q^{'} \tag{23}$$

Based on the above user preference similarity for tags, we can obtain the user's preference for tags:

$$Y^{'} = P^{'T} X^{'} \tag{24}$$

The element y_a in Y' is user u's preference degree $upt(t,u)$ for label t.

IV. C. Similarity calculation

First, consider a weighted user-item bipartite graph. Assume that the target user u is allocated “1” unit of resources, while other users have 0 resources, resulting in an initial m-dimensional resource vector f_1. After heat conduction diffusion, the final resource vector for all users is obtained as $f_1^{'} = Wf_1$, where W is the state transition matrix for the heat conduction process. In the first step, users allocate resources to each investment product based on the ratio of the edge weights between users and investment products to the sum of the edge weights for each investment product. After diffusion, the resource vector for investment products is obtained as:

$$g_1 = D_I A^T f_1 \tag{25}$$

The second step is to distribute resources to users according to the ratio of the edge weight between investment products and users and the sum of each user's edge weight, obtaining the final resource vector for all users:

$$f_1^{'} = D_U A g_1 = D_U A D_I A^{\text{û}} f_1 \tag{26}$$

From this, we obtain the state transition matrix $W = D_U A D_r A^T$. The element w_u in the v th row and u th column of W represents the resources obtained by user v from user u :

$$\begin{aligned} w_m &= (D_v A D_I A^T)_{vu} \\ &= \sum_{i \in I} \left[(D_U A)_n \cdot (D_I A^T)_{iu} \right] \\ &= \sum_{i \in I} \left[\frac{a_{si}}{k(v)} \cdot \frac{a_{iu}}{k(i)} \right] \\ &= \frac{1}{k(v)} \sum_{i \in I} \frac{a_{ni} a_{iu}}{k(i)} \end{aligned} \tag{27}$$

Define the resource w_{uv} obtained by user v from user u as the similarity between target user u and user v in the weighted user-item bipartite graph:

$$sim(v,u) = w_{su} = \frac{1}{k(v)} \sum_{i \in I} \frac{a_{si} a_{ui}}{k(i)} \tag{28}$$

$$a_{ui} = \begin{cases} upi(i,u), & \text{Existence } E_{ui} \\ 0, & \text{other} \end{cases} \tag{29}$$

Among them, $upi(i,u)$ is user u 's preference for investment product i :

$$k(v) = \sum_{i=1}^{n} upi(i,v) \tag{30}$$

Represents the degree of user v in the user-item weighted bipartite graph:

$$k(i) = \sum_{u=1}^{m} upi(i,u) \tag{31}$$

Represents the degree of investment product i in the user-item weighted bipartite graph. E_{ui} represents the edge between user u and investment product i in the user-item weighted bipartite graph.

After heat conduction diffusion, the final resource vector for all users is obtained: $f_2^{'} = W f_2$ In the first step, users allocate resources to each label according to the ratio of the edge weight between the user and the label to the sum of the edge weights of each label. After diffusion, the label resource vector is obtained:

$$g_2 = D_T A^{'T} f_2 \tag{32}$$

The second step label returns resources to users in the same way according to the ratio of the edge weight between the label and the user to the sum of each user's edge weight, obtaining the final resource vector for all users:

$$f_2^{'} = D_U A^{'} g_2 = D_U A^{'} D_T A^{'} f_2 \tag{33}$$

The state transition matrix can be obtained as follows:

$$W^{'} = D_v^{'} A D_r A^{'T} \tag{34}$$

The element $w_{vu}^{'}$ in row v and column u of $W^{'}$ represents the resources obtained by user v from user u :

$$\begin{aligned} w_{m} &= (D_{v} A D_{T} A^{,T})_{su} \\ &= \sum_{t \in T} \left[(D_{U} A)_{u} \cdot (D_{T} A^{,T})_{u} \right] \\ &= \sum_{t \in T} \left[\frac{a_{ut}^{'}}{k^{'}(v)} \cdot \frac{a_{tu}^{'}}{k^{'}(t)} \right] \\ &= \frac{1}{k^{'}(v)} \sum_{t \in T} \frac{a_{st}^{'} a_{tu}^{'}}{k(t)} \end{aligned} \tag{35}$$

Define $w_{vu}^{'}$ as the similarity between target users u and v in the weighted usertag bipartite graph:

$$sim^{'}(v,u) = w_{ru}^{'} = \frac{1}{k^{'}(v)} \sum_{t \in T} \frac{a_{st}^{\prime'} a_{st}^{'}}{k(t)} \tag{36}$$

$$a_{ut}^{'} = \begin{cases} upt(t,u), & \text{Existence } E_{ut} \\ 0, & \text{other} \end{cases} \tag{37}$$

Among them, $upt(t,u)$ is user u 's preference for label t.

$$k(t) = \sum_{u=1}^{m} upt(t,u) \tag{38}$$

Indicates the degree of tag t in the bipartite graph usertag.

$$k^{'}(v) = \sum_{t=1}^{r} upt(t,v) \tag{39}$$

The degree of user v in the user-tag bipartite graph is denoted by E_{ut}, which represents the edge between user u and tag t in the user-tag weighted bipartite graph.

The target user is u_1, and the process of solving user similarity on the weighted user-item bipartite graph is shown in Figure 4. The diffusion process on the weighted user-tag bipartite graph is similar.

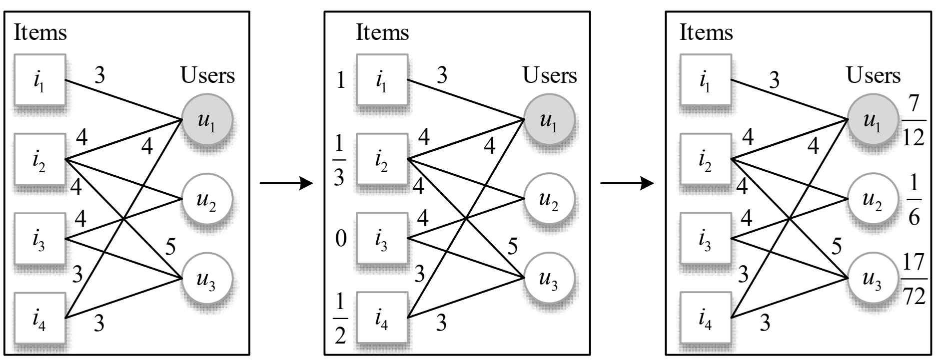


Figure 4: Resource diffusion on a weighted user-item bipartite network

Finally, we introduce the parameter λ to integrate the user similarity $sim(v,u)$ and $sim'(v,u)$ between investment products and tags, obtaining the final user similarity:

$$similarity(v,u) = \lambda sim(v,u) + (1-\lambda) sim(v,u) \tag{40}$$

Among them, $\lambda \in [0,1]$ is an adjustable parameter. When $\lambda = 0$, the similarity becomes the user similarity on the separate weighted user-tag bipartite graph; when $\lambda = 1$, the similarity becomes the user similarity on the separate weighted user-item bipartite graph.

IV. D. Prediction Scores and Recommendations

Based on a user-based collaborative filtering framework, predict scores for investment products that target users have not selected. Given a target user u and an unselected investment product i, the predicted score for investment product i for user u is:

$$pred_{u,i} = \overline{r}_u + \frac{\sum_{v \in N_u(S)} similarity(u,v)(r_{v,i} - \overline{r}_v)}{\sum_{v \in N_u(S)} |similarity(u,v)|} \tag{41}$$

Among them, v represents users who have already rated investment product i, $r_{v,i}$ represents user v's rating of investment product i, and $\overline{r}_u, \overline{r}_v$ represent the average ratings of all ratings by users u and v, respectively, and $N_u(s)$ denotes the s nearest neighbors of user u.

The predicted ratings of investment products not selected by target user u are then sorted in descending order to generate a recommendation list, with the top L investment products being recommended to user u.

V. Experimental analysis of investment decision support models

To evaluate the effectiveness of the investment decision support model's recommendations, this section conducts experiments based on user investment opinion data collected from Stocktwits. The analysis of the experimental results is divided into two parts. First, the effectiveness of the personalized preference modeling method is evaluated. Second, the experimental results of the hybrid recommendation algorithm based on high quality and personalization are analyzed in terms of personalized recommendations and profitability.

V. A. Evaluation Indicators

This paper uses all data in the historical time window to estimate the preferences of users who appear on a certain day d_t in the test set for all stocks, and recommends the top K stocks based on the preference values. Therefore, in the experiment, this paper uses the average Precision@K and Recall@K for each day and each user in the test set as evaluation metrics, which are defined as follows:

$$\Pr ecision@K = \frac{1}{|T|} \sum_{d_t \in T} \left(\frac{1}{|U^{d_t}|} \sum_{u \in U^{d_t}} \frac{|R_u^{d_t} \cap L_u^{d_t}|}{K} \right) \tag{42}$$

$$\mathrm{Re}\,call@K = \frac{1}{|T|} \sum_{d_t \in T} \left(\frac{1}{|U^{d_t}|} \sum_{u \in U^{d_t}} \frac{|R_u^{d_t} \cap L_u^{d_t}|}{|L_u^{d_t}|} \right) \tag{43}$$

where T denotes the date in the test set, U^{d_t} denotes the set of users appearing on day d_t, $R_u^{d_t}$ denotes the set of stocks recommended by the model to user u, and $L_u^{d_t}$ denotes the set of stocks that user u actually prefers, i.e., the set of stocks appearing in user u's comments.

In experiments involving high-quality and personalized mixed stock recommendations, to evaluate the profitability of the recommended results, the average daily returns of the portfolio in actual trading are used. Specifically, for the S^r and A^r generated on trading day d, this paper divides the return of $r(S^r, \mathrm{A}^r)$ by the holding period length to obtain the daily investment return:

$$r_d(S^r, \mathrm{A}^r) = \frac{r(S^r, \mathrm{A}^r)}{|\Delta|} \tag{44}$$

where $|\Delta|$ represents the length of the holding period Δ (i.e., the number of days). Then, the average of the daily average returns for all test days is expressed as the daily average return:

$$Daily\_\mathrm{Re}turn = \frac{1}{\mathrm{N}^d} \sum_{d=1}^{\mathrm{N}^d} r_d(S^r, \mathrm{A}^r) \tag{45}$$

V. B. Experimental Analysis and Comparison of Personalized Preference Modeling

To validate the personalized preference modeling method for stock trend perception proposed in this paper, this section conducts a personalized recommendation experiment based on user data from the Stocktwits platform in 2023, using Precision@K and Recall@K as evaluation metrics. Tables 8 and 9 present the experimental results of SM-UCF and the other three comparison methods under different stock trend time window lengths t', stock recommendation quantities K, and historical data time window lengths Δt. The experimental results lead to the following conclusions: (1) User-based collaborative filtering algorithms outperform item-based collaborative filtering overall, indicating that personalized preference modeling based on similar users is more accurate in online investment communities. (2) The method proposed in this paper (SM-UCF) outperforms the baseline method, suggesting that the nearest neighbor user mining method for stock trend perception can more effectively estimate user preferences. (3) As the historical data time window length Δt increases, the recommendation performance of all methods deteriorates. On one hand, this suggests that recent data can more accurately reflect users' personalized preferences. On the other hand, it may be due to differing coverage rates of users on the test day across historical data of varying lengths, leading to biased results. (4) For different stock price trend time window lengths t', both excessively long or short durations degrade SM-UCF's recommendation performance. Therefore, selecting an appropriate parameter t' is crucial in practical applications.

Table 8: Different parameters of Precision@K (%)

Time-window of History Δ=7									
	Our Approach						Traditional Approaches		
	t′=1	t′=14	t′=30	t′=90	t′=180	t′=365	UCF	ICF	POP
K=1	33.388	34.372	34.502	33.824	33.105	32.701	28.274	25.361	3.612
K=4	13.354	13.774	13.701	13.486	13.129	13.077	11.407	10.875	2.666
K=8	8.282	8.451	8.495	8.361	8.143	8.132	7.026	6.822	2.103
K=15	4.314	4.403	4.398	4.395	4.267	4.182	3.617	3.489	1.595
K=25	2.299	2.329	2.312	2.299	2.212	2.216	1.886	1.774	1.12
K=50	1.003	1.01	1.036	0.993	0.944	0.968	0.835	0.77	0.668
Time-window of History Δ=14									
	Our Approach						Traditional Approaches		
	t′=1	t′=14	t′=30	t′=90	t′=180	t′=365	UCF	ICF	POP
K=1	30.066	31.448	31.382	30.885	30.801	30.12	25.965	23.347	3.564
K=4	12.805	13.391	13.329	13.112	13.088	12.843	11.259	10.79	2.554
K=8	8.118	8.366	8.395	8.244	8.196	8.093	7.119	6.963	2.132
K=15	4.295	4.423	4.422	4.373	4.315	4.259	3.723	3.598	1.583
K=25	2.3	2.367	2.355	2.304	2.25	2.268	1.939	1.865	1.086
K=50	1.006	1.047	1.045	1.012	0.988	0.996	0.861	0.785	0.664
Time-window of History Δ=30									
	Our Approach						Traditional Approaches		
	t′=1	t′=14	t′=30	t′=90	t′=180	t′=365	UCF	ICF	POP
K=1	26.461	27.669	28.104	27.827	27.29	27.067	23.844	21.528	3.501
K=4	12.074	12.476	12.667	12.517	12.21	12.285	10.999	10.622	2.506
K=8	7.804	8.025	8.133	8.019	7.881	7.909	7.116	6.934	2.084
K=15	4.208	4.338	4.343	4.275	4.217	4.255	3.714	3.664	1.505
K=25	2.266	2.323	2.327	2.29	2.294	2.217	1.986	1.882	1.113
K=50	1.039	1.054	1.056	1.042	1.068	1.024	0.86	0.814	0.652

Table 9: Different parameters of Recall@K (%)

Time-window of History Δ=7									
	Our Approach						Traditional Approaches		
	t′=1	t′=14	t′=30	t′=90	t′=180	t′=365	UCF	ICF	POP
K=1	9.629	9.951	9.962	9.737	9.521	9.4	8.113	7.327	1.136
K=4	11.575	11.833	11.861	11.643	11.372	11.319	9.884	9.384	2.346
K=8	12	12.27	12.264	12.078	11.776	11.672	10.176	9.855	3.108
K=15	12.514	12.785	12.793	12.574	12.262	12.165	10.461	10.116	4.644
K=25	13.197	13.449	13.469	13.188	12.869	12.79	10.834	10.4	6.491
K=50	14.456	14.71	14.768	14.492	14.074	13.996	11.737	11.086	9.733
Time-window of History Δ=14									
	Our Approach						Traditional Approaches		
	t′=1	t′=14	t′=30	t′=90	t′=180	t′=365	UCF	ICF	POP
K=1	8.651	8.961	9.041	8.913	8.813	8.622	7.518	6.679	1.11

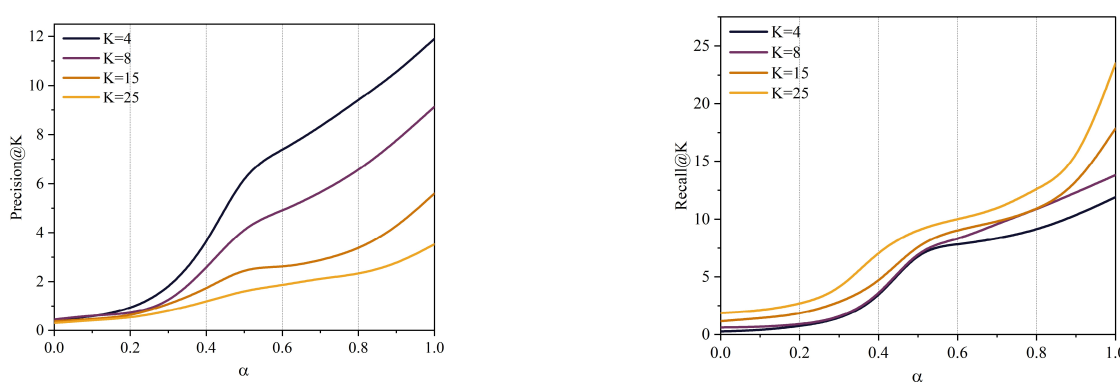
K=4	11.115	11.562	11.543	11.364	11.246	11.048	9.759	9.349	2.294
K=8	11.692	12.09	12.124	11.904	11.806	11.609	10.282	10.01	3.076
K=15	12.482	12.788	12.827	12.588	12.447	12.279	10.688	10.417	4.469
K=25	13.259	13.66	13.628	13.397	13.198	12.992	11.088	10.81	6.318
K=50	14.718	15.166	15.058	14.816	14.454	14.287	11.936	11.413	9.59
Time-window of History Δ=30									
	Our Approach						Traditional Approaches		
	t′=1	t′=14	t′=30	t′=90	t′=180	t′=365	UCF	ICF	POP
K=1	7.609	7.926	8.059	8.032	7.918	7.706	6.869	6.181	1.092
K=4	10.407	10.786	10.993	10.848	10.612	10.543	9.622	9.222	2.201
K=8	11.212	11.609	11.713	11.584	11.381	11.37	10.287	10.031	2.998
K=15	12.219	12.548	12.628	12.447	12.244	12.194	10.792	10.698	4.346
K=25	13.222	13.606	13.642	13.434	13.312	13.067	11.271	11.129	6.203
K=50	14.966	15.366	15.458	15.055	15.208	14.57	12.202	11.761	9.315

V. C. Analysis of recommended experimental results

The primary objective of this section's experiment is to explore the impact on recommendation effectiveness when considering both profit potential and personalized preferences simultaneously. Let the proportion factor α be {0, 0.1, ..., 0.9, 1.0}, and the number of recommendations K be {4, 8, 15, 25}, to examine the effects on recommendation results in terms of both revenue and personalization.

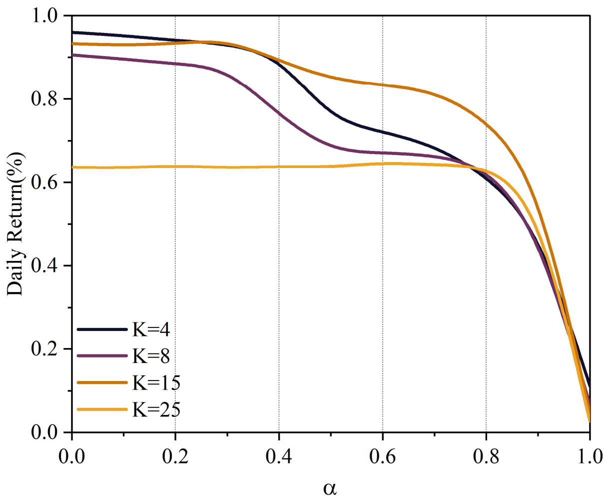
First, we examine the impact of the proportional factor α on personalized recommendation effectiveness, using Precision@K and Recall@K as evaluation metrics. The experimental results are shown in Figure 5. From the experimental results, we can observe the following: (1) As the proportional factor α increases, personalized recommendation effectiveness gradually improves, indicating that adjusting the proportional factor can incorporate personalized factors into recommendation effectiveness. (2) The change in recommendation effectiveness with α is not linear. It can be seen that the rate of improvement in personalized recommendation effectiveness is highest around α = 0.5, after which it decreases and then continues to increase. (3) As the number of recommendations K increases, Precision@K decreases overall, while Recall@K increases overall. This indicates that the hybrid recommendation algorithm has a ranking function, with stocks ranked higher in the list better aligning with users' personalized preferences.

To examine the impact of the proportional factor α on high-quality recommendation effectiveness, holding periods of one day, two weeks, and one month were set, with the evaluation metric being daily average returns. After averaging the results for each trading day and each user, the experimental results are shown in Figure 6. From the experimental results, it can be concluded that: (1) As α increases, the “high-quality” (profitability) nature of recommended stocks gradually weakens, and daily average returns gradually decrease, indicating that α can regulate the relative influence of high-quality and personalized factors. (2) The return performance does not change linearly with α, but decreases at an increasingly faster rate overall. (3) As the number of recommendations K increases, the average daily return decreases, especially when the holding period is one week, indicating that the hybrid recommendation algorithm has a “high-quality” sorting effect, with stocks ranked higher having greater return potential. (4) The return performance of recommended stocks deteriorates as the holding period increases. In practical applications, shorter holding periods can be chosen, but they are more susceptible to transaction costs.

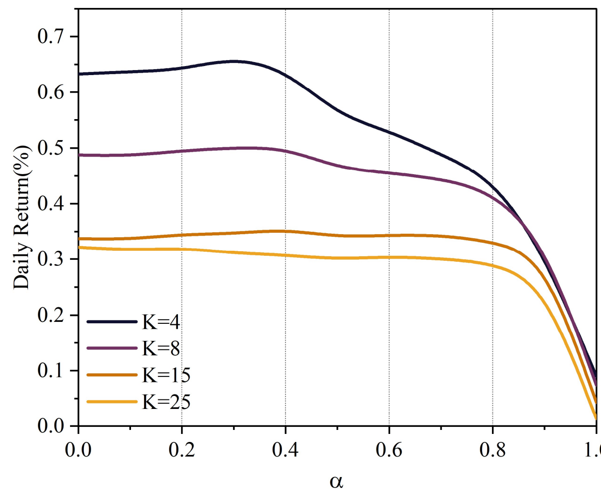


(a) Precision@K (b) Recall@K

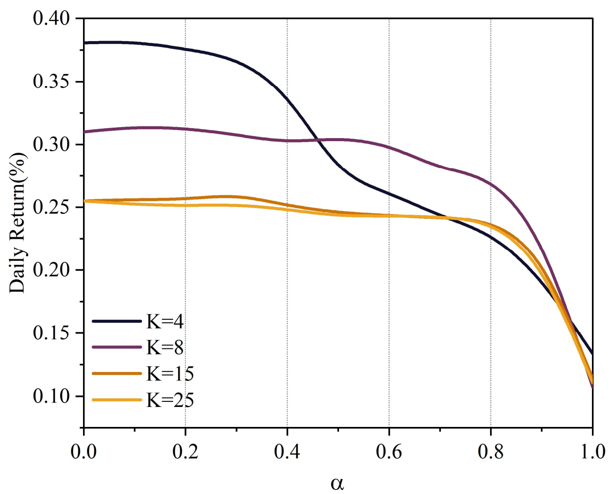
Figure 5: Precision@K and Recall@K under different α



(a) Daily Trading



(b) Weekly Trading



(c) Monthly Trading

Figure 6: Is different from α and the daily income of holding

To more clearly compare how the proportion factor α affects personalized and high-quality recommendation performance, Figure 7 shows the recommendation performance under different recommendation quantities K when the holding period is one week. The experimental results show that: (1) As α increases, the proportion of personalized factors increases, so personalized recommendation performance gradually improves, but revenue performance deteriorates. (2) The changes in personalized and high-quality recommendation effectiveness are asymmetric. Taking K=4 as an example, when α is less than 0.4, it is possible to improve personalized recommendation effectiveness without compromising high-quality recommendation performance (daily average revenue). Similar results are observed when K takes other values. (3) As K increases, the "inflection point" where α significantly impacts high-quality recommendations shifts to the right. At this point, α can be set to a larger value to maximize the consideration of personalized factors without compromising revenue performance, thereby enhancing user experience.

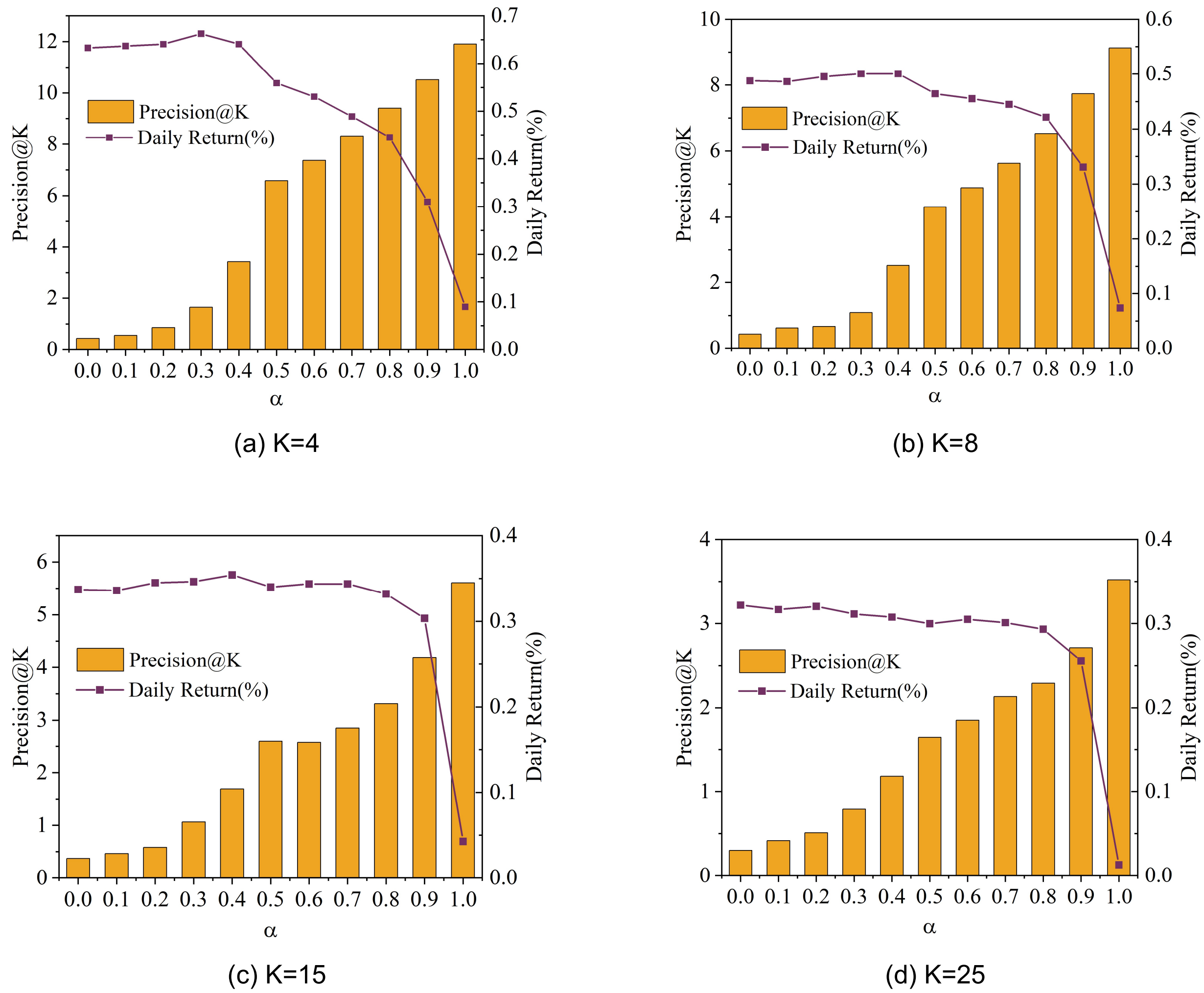


Figure 7: Different α and recommended number of daily income and precision@k

VI. Application Analysis

VI. A. Preparation of experimental data

Using the well-known investment platform X as the research context, this study simulates its personalized recommendation environment to explore the rationality of introducing personalized recommendations and the applicability and effectiveness of improved algorithms. The initial data required for this paper consists of four parts: first, the investment product set, including attribute detail data; second, the standardized table of investment product attributes; third, investor purchase records; and fourth, the importance matrix for attribute relationships. Since this experiment is solely aimed at testing the feasibility and effectiveness of the improved algorithm, only 50 investment products are selected as the candidate product set, with four key feature attributes set as the attribute set for each product, and 150 purchase records generated for 40 investors.

(1) Investment product set: To ensure the rationality of the product attribute combinations, this paper obtained 50 real investment product data points from the X platform, where the i-th product is represented. The investment product data is shown in Table 10.

Table 10: 50 online loan product tables

Product (Bi)	Annual interest rate /%	Time limit /Month	Total item / (thousand yuan)	Repayment mode
B1	3.37	12	170	Monthly interest repayment and principal repayment upon maturity
B2	2.90	3	210	Daily interest calculation
B3	8.05	24	3030	Monthly interest repayment and principal repayment upon maturity
B4	9.61	12	2520	Monthly interest repayment and principal repayment upon maturity
B5	5.22	3	2840	Monthly interest repayment and principal repayment upon maturity
B6	5.86	12	2010	Monthly interest repayment and principal repayment upon maturity
B7	6.55	6	1980	Equivalent interest
B8	8.02	12	470	Monthly interest repayment and principal repayment upon maturity
B9	6.95	3	150	Equivalent interest
B10	5.41	3	610	Monthly interest repayment and principal repayment upon maturity
B11	7.43	3	1180	Pay off the principal and interest in a lump sum upon maturity
B12	6.26	12	1330	Monthly interest repayment and principal repayment upon maturity
B13	7.29	6	2250	Equivalent interest
B14	4.49	24	1570	Monthly interest repayment and principal repayment upon maturity
B15	9.94	36	2440	Equivalent interest
B16	9.74	3	140	Monthly interest repayment and principal repayment upon maturity
B17	3.87	24	540	Equivalent interest
B18	8.18	36	2560	Pay off the principal and interest in a lump sum upon maturity
B19	6.24	24	2040	Monthly interest repayment and principal repayment upon maturity
B20	8.00	3	1680	Equivalent interest
B21	8.06	12	840	Pay off the principal and interest in a lump sum upon maturity
B22	3.06	6	430	Monthly interest repayment and principal repayment upon maturity
B23	5.30	24	1590	Monthly interest repayment and principal repayment upon maturity
B24	9.85	12	1870	Equivalent interest
B25	5.82	36	170	Equivalent interest
B26	8.50	12	1630	Equal principal
B27	11.47	1	380	Monthly interest repayment and principal repayment upon maturity
B28	6.00	24	3100	Equal principal
B29	7.69	3	1040	Monthly interest repayment and principal repayment upon maturity
B30	9.29	6	2560	Pay off the principal and interest in a lump sum upon maturity
B31	10.80	1	2420	Monthly interest repayment and principal repayment upon maturity
B32	8.95	1	2700	Equal principal
B33	6.26	3	1590	Pay off the principal and interest in a lump sum upon maturity
B34	4.83	1	1840	Pay off the principal and interest in a lump sum upon maturity
B35	4.95	6	550	Pay off the principal and interest in a lump sum upon maturity

B36	6.64	3	870	Equivalent interest
B37	2.50	24	630	Daily interest calculation
B38	8.06	24	1750	Pay off the principal and interest in a lump sum upon maturity
B39	5.08	12	630	Monthly interest repayment and principal repayment upon maturity
B40	4.75	3	500	Equal principal
B41	7.37	24	1080	Equal principal
B42	8.50	12	1980	Monthly interest repayment and principal repayment upon maturity
B43	5.63	3	70	Pay off the principal and interest in a lump sum upon maturity
B44	6.82	6	1140	Equal principal
B45	9.12	6	2200	Monthly interest repayment and principal repayment upon maturity
B46	10.00	6	2300	Equal principal
B47	7.77	6	690	Monthly interest repayment and principal repayment upon maturity
B48	5.03	1	260	Equal principal
B49	6.86	12	960	Pay off the principal and interest in a lump sum upon maturity
B50	5.79	3	410	Equivalent interest

(2) Standardized table of investment product attributes: Investment products have complex characteristics. In the experiment, Qi(i=1, 2, 3, 4) represents the annual interest rate, term, total project amount, and repayment method of the investment product, respectively. These four attributes are used as the basis for personalized product recommendations, denoted as q_{ij_i} denotes the j_i th attribute value of the ith attribute (where (i = (1, 2, 3, 4)), j_1 = 1, 2, ..., 9, j_2 =1, 2, ..., 6, j_3 =1, 2, ..., 5, j_4 =1, 2, ..., 5), and Table 11 is the product attribute standardization table.

Table 11: Standardization of Online Loan Product Attributes

Attribute value	Q1 Annual interest rate	Q2 Time limit	Q3 Total item	Q4 Repayment mode
1	3%	1 Month	0-300 thousand yuan	Daily interest calculation
2	4%	3 Month	300-500 thousand yuan	Monthly interest repayment and principal repayment upon maturity
3	5%	6 Month	500-1000 thousand yuan	Equivalent interest
4	6%	12 Month	1000-2000 thousand yuan	Equal principal
5	7%	24 Month	2000-3000 thousand yuan	Pay off the principal and interest in a lump sum upon maturity
6	8%	36 Month		
7	9%			
8	10%			
9	11%			

(3) Investor purchase attribute record table: Assuming that the platform has a total of 40 investor records, Ui (i=1, 2, ..., 40) represents the i-th investor. Using the RAND function in Excel, 150 purchase records for 40 investors are randomly generated. Referring to the product attribute standardization table, each product purchase record can be converted into a product attribute record $R_i = \{q_{1j_1}, q_{2j_2}, q_{3j_3}, q_{4j_4}\}$ In this experiment, Investor 1 (U1) is selected as the target user for recommendations, and all subsequent algorithmic processes are based on U1's historical investment decisions.

(4) Attribute Importance Matrix: In real-world scenarios, the platform website conducts surveys before users invest to assess individual investors' understanding and familiarity with financial knowledge, as well as their investment behavior habits and preferences. The experiment selected 40 investment platform users to participate in a Saaty nine-level gradient questionnaire survey, and randomly selected one as the experimental basis for U1's personalized recommendations. The pairwise comparison matrix for attribute importance is shown in Table 12.

Table 12: Pairwise comparison Judgment Matrix of the importance between attributes

Attribute	Annual interest rate	Time limit	Total item	Repayment mode
Annual interest rate	1	0.33	4	6
Time limit	3	1	5	4
Total item	0.25	0.2	1	3
Repayment mode	0.17	0.25	0.33	1

VI. B. Personalized Recommendation Process

VI. B. 1) Calculation of investor preference

This paper selects investor U1 as the target user and uses the method described in Section 4.2 to calculate U1's preference for each attribute of the product. U1's preference for the attribute values of all investment products is shown in Table 13. Taking U1's preference for the attributes of product B1 as an example, the attribute values of product B1 are {q11, q21, q31, q42]. First, calculate U1's preference for the attribute q11 of Q1. Let K = 3, i.e., find the three attribute values with the highest similarity to q11 in the similarity matrix, excluding the current attribute itself. $w(i,k) = w(q11,3) = \{q14, q16, q17\}$, then U1's preference for attribute q11 of Q1 is 0.32.

Table 13: U1——investment product properties value interest

Upt	Q1	Q2	Q3	Q4	Upt	Q1	Q2	Q3	Q4
B1	0.32	0.58	0.78	0.61	B26	0.79	1.1	0.54	0.38
B2	0.34	0.6	0.83	0.39	B27	0.31	0.24	0.6	0.61
B3	0.37	0.61	0.95	0.51	B28	0.57	0.57	0.98	0.41
B4	0.41	0.85	0.95	0.55	B29	0.89	1.03	0.57	0.56
B5	0.4	0.36	0.99	0.47	B30	0.48	0.33	0.97	1.29
B6	0.5	0.28	0.55	0.54	B31	0.5	1.23	0.93	0.58
B7	0.46	0.52	0.57	0.62	B32	0.55	0.6	0.95	0.47
B8	0.78	0.39	0.61	0.55	B33	0.45	0.89	0.56	1.2
B9	0.42	0.94	0.82	0.47	B34	0.55	0.81	0.53	1.2
B10	0.59	0.6	0.89	0.52	B35	0.45	0.85	0.91	1.24
B11	0.78	1.11	0.49	1.16	B36	0.57	0.79	0.86	0.54
B12	0.52	1.16	0.62	0.47	B37	0.63	0.38	0.7	0.44
B13	0.54	0.6	1.02	0.54	B38	0.42	0.62	0.49	1.14
B14	0.62	0.54	0.6	0.55	B39	0.58	0.52	0.93	0.49
B15	0.49	1.12	0.95	0.56	B40	0.48	0.82	0.9	0.43
B16	0.44	0.7	0.82	0.53	B41	0.46	0.6	1.02	0.4
B17	0.45	0.52	0.99	0.53	B42	0.46	0.59	0.54	0.44
B18	0.52	0.5	0.95	1.21	B43	0.45	0.84	0.85	1.19
B19	0.44	0.81	0.49	0.41	B44	0.53	1.17	0.58	0.47
B20	0.51	0.64	0.57	0.41	B45	0.44	0.54	0.9	0.55
B21	0.95	0.63	0.98	1.18	B46	0.5	0.26	1.06	0.45
B22	0.38	0.53	0.61	0.54	B47	0.55	0.64	0.98	0.5
B23	0.55	0.57	0.52	0.45	B48	0.56	0.4	0.81	0.53
B24	0.52	0.61	0.55	0.39	B49	0.88	0.97	1.01	1.05
B25	0.48	0.88	0.82	0.43	B50	0.7	0.7	0.74	0.54

VI. B. 2) Measurement of Investment Preferences

Based on investors' preference weights for different attributes and their preference degrees for each attribute value of a product, the algorithm U1 is then used to measure the overall preference degree for the product. Taking investment product B1 as an example, the calculated preference degree of U1 for product B1 is 0.53. By implementing the above algorithm steps using the Python programming language, the preference degrees of investor U1 for 50 investment products can be obtained, as shown in Table 14.

Table 14: U1's interest meter for 50 investment products

Bi	Upi	Bi	Upi	Bi	Upi	Bi	Upi	Bi	Upi
B1	0.53	B11	0.9	B21	0.84	B31	0.87	B41	0.67
B2	0.55	B12	0.84	B22	0.57	B32	0.66	B42	0.53
B3	0.62	B13	0.59	B23	0.52	B33	0.71	B43	0.76
B4	0.73	B14	0.6	B24	0.62	B34	0.69	B44	0.87
B5	0.45	B15	0.88	B25	0.63	B35	0.75	B45	0.55
B6	0.47	B16	0.6	B26	0.92	B36	0.73	B46	0.5
B7	0.58	B17	0.61	B27	0.3	B37	0.5	B47	0.54
B8	0.58	B18	0.68	B28	0.52	B38	0.5	B48	0.48
B9	0.66	B19	0.64	B29	0.99	B39	0.63	B49	1.01
B10	0.56	B20	0.56	B30	0.35	B40	0.59	B50	0.6

VI. B. 3) Generating Recommended Results

After model calculation, the preference degree of investor U1 for 50 investment products is obtained. The TOP-N investment products are selected according to their preference degree to form a recommendation list displayed to investors. In this example, N=10, generating the personalized recommendation list of investment products for investor U1 as shown in Table 15.

Table 15: investor investment product personalized recommendation

Ui	TOP-N	Bi	Upi
U40	1	B26	1.15
	2	B28	0.92
	3	B34	0.96
	4	B31	0.92
	5	B15	0.81
	6	B23	0.83
	7	B44	0.78
	8	B42	0.88
	9	B32	0.76
	10	B16	0.71

VI. C. Analysis of Recommended Effects

In this example, if we directly calculate investors' interest preferences for products using the traditional collaborative filtering recommendation algorithm—that is, by using investor-investment product purchase records to calculate and generate a similarity matrix between products, as shown in Figure 8 below—the similarity between most investment products is 0, and the data sparsity problem is severe. The improved algorithm decomposes products into individual attributes, uses the investor-attribute transaction table to obtain a similarity matrix between the values of each attribute, and calculates investors' preferences for each attribute value of the product's individual attributes. This effectively avoids the data sparsity issue and achieves better recommendation results.

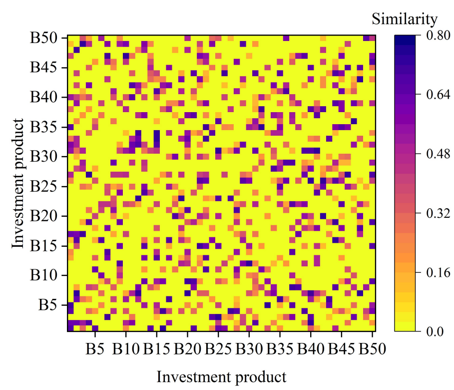


Figure 8: 50 net loan product similarity matrix

VII. Conclusion

For investors, financial data analysis is the primary reference for investment decisions. To avoid the adverse effects that risks inherent in financial data analysis may have on investment decisions. This paper selects 16 indicators representing a company's financial condition based on the principles for selecting financial diagnostic indicators and preprocessing of indicators. Using principal component analysis, four main factors affecting a company's financial condition are identified, and a comprehensive financial score is calculated to comprehensively assess the financial condition of each year. Based on this, an investment decision-making support model is constructed—a collaborative filtering algorithm based on weighted triads.

In the experimental section, the impact of different proportion factors α on personalized and high-quality recommendation effectiveness was investigated. The experimental results show that as α increases, the proportion of personalized factors increases, thereby gradually improving personalized recommendation effectiveness but reducing revenue performance. Taking K=4 as an example, when α is less than 0.4, it is possible to improve personalized recommendation effectiveness without compromising high-quality recommendation performance. This method was also validated as capable of improving personalized recommendation effectiveness without significantly compromising the performance of recommended stock returns.

Finally, the personalized recommendation process for investment products on the platform was simulated. One investor was selected from a group of 40, their preferences for 50 investment products were calculated, and investment recommendation results were generated for them. This further validated the success and effectiveness of the algorithm improvements.

References

[1] Ivanenko, T., Hrushko, V., & Frantsuz, A. (2018). Optimal investment decision making on the model of production enterprise with limited resources. Investment Management and Financial Innovations, (15, Iss. 4), 61-68.

[2] Mulyk, T., Tomchuk, O., & Fedoryshyna, L. (2020). Analysis of financial stability of the enterprise. The scientific heritage, (43-3 (43)), 46-55.

[3] Levytska, S., Akimova, L., Zaiachkivska, O., Karpa, M., & Gupta, S. K. (2020). Modern analytical instruments for controlling the enterprise financial performance. Financial and credit activity problems of theory and practice, 2(33), 314-323.

[4] Valaskova, K., Kliestik, T., & Kovacova, M. (2018). Management of financial risks in Slovak enterprises using regression analysis. Oeconomia copernicana, 9(1), 105-121.

[5] Bartolacci, F., Caputo, A., & Soverchia, M. (2020). Sustainability and financial performance of small and medium sized enterprises: A bibliometric and systematic literature review. Business Strategy and the Environment, 29(3), 1297-1309.

[6] Ashraf, S., GS Félix, E., & Serrasqueiro, Z. (2019). Do traditional financial distress prediction models predict the early warning signs of financial distress?. Journal of Risk and Financial Management, 12(2), 55.

[7] Hamza, N., & Arif, I. (2019). Impact of financial literacy on investment decisions: The mediating effect of big-five personality traits model. Market Forces, 14(1), 43-60.

[8] Olayinka, A. A. (2022). Financial statement analysis as a tool for investment decisions and assessment of companies' performance. International Journal of Financial, Accounting, and Management, 4(1), 49-66.

[9] Chandramowleeswaran, G., Alzubaidi, L. H., Liz, A. S., Khare, N., Khurramov, A., & Baswaraju, S. (2023, August). Design of financing strategy model of financial management based on data mining technology. In 2023 Second International Conference On Smart Technologies For Smart Nation (SmartTechCon) (pp. 1179-1183). IEEE.

[10] Kumar, S., Sharma, D., Rao, S., Lim, W. M., & Mangla, S. K. (2025). Past, present, and future of sustainable finance: insights from big data analytics through machine learning of scholarly research. Annals of Operations Research, 345(2), 1061-1104.

[11] Raguseo, E. (2018). Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. International Journal of Information Management, 38(1), 187-195.

[12] Sharairi, J. A., Saatchi, S. G., Rahahle, M. Y., Maabreh, H. M. A., Sarram, M., Anagreh, S., ... & Al-Hawary, S. I. S. (2024). How big data governance meets financial decision-making: evidence from banking sector in emerging economies. In Artificial Intelligence and Economic Sustainability in the Era of Industrial Revolution 5.0 (pp. 1295-1311). Cham: Springer Nature Switzerland.

[13] Tong, D., & Tian, G. (2023). Intelligent financial decision support system based on big data. Journal of Intelligent Systems, 32(1), 20220320.

[14] Müller, O., Fay, M., & Vom Brocke, J. (2018). The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. Journal of management information systems, 35(2), 488-509.

[15] Ghasemaghaei, M., & Calic, G. (2020). Assessing the impact of big data on firm innovation performance: Big data is not always better data. Journal of business research, 108, 147-162.

[16] Richins, G., Stapleton, A., Stratopoulos, T. C., & Wong, C. (2017). Big data analytics: opportunity or threat for the accounting profession?. Journal of information systems, 31(3), 63-79.

[17] Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big data and analytics in the modern audit engagement: Research needs. Auditing: A Journal of Practice & Theory, 36(4), 1-27.

[18] Wang, T., & Song, T. (2024). Application of Financial Big Data Analysis Method Based on Collaborative Filtering Algorithm in Supply Chain Enterprises. International Journal of Cooperative Information Systems, 33(04), 2350022.

[19] Pereira, N., & Varma, S. L. (2018). Financial planning recommendation system using content-based collaborative and demographic filtering. In Smart Innovations in Communication and Computational Sciences: Proceedings of ICSICCS 2017, Volume 1 (pp. 141-151). Singapore: Springer Singapore.

[20] Lee, I. (2019). An optimization approach to capacity evaluation and investment decision of hybrid cloud: a corporate customer's perspective. Journal of Cloud Computing, 8, 1-13.

[21] Sun, L. (2024). Management research of big data technology in financial decision-making of enterprise cloud accounting. Journal of Information & Knowledge Management, 23(01), 2350067.

[22] Syah, D. H., Muda, I., Lumbanraja, P., & Kholis, A. (2023). The role of cloud computing on accounting information system quality: a study in hotel industry. TEM Journal, 12(3), 1890.

[23] Sun, X., & Lei, Y. (2021). Research on financial early warning of mining listed companies based on BP neural network model. Resources Policy, 73, 102223.

[24] Ouyang, Z. S., Yang, X. T., & Lai, Y. (2021). Systemic financial risk early warning of financial market in China using Attention-LSTM model. The North American Journal of Economics and Finance, 56, 101383.

[25] Allaj, E., & Sanfelici, S. (2023). Early Warning Systems for identifying financial instability. International Journal of Forecasting, 39(4), 1777-1803.

[26] Zeng, H. (2022). Influences of mobile edge computing-based service preloading on the early-warning of financial risks. The Journal of Supercomputing, 78(9), 11621-11639.

[27] Jinjie Teng. (2025). Financial Data Reduction and Information Retention Strategy based on Principal Component Analysis (PCA) Algorithm. Procedia Computer Science,262,218-226.

[28] Jung Kwangmin,Kim Donggyu & Yu Seunghyeon. (2022). Next generation models for portfolio risk management: An approach using financial big data. Journal of Risk and Insurance,89(3),765-787.