J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 6759--6777

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Collaborative Filtering Algorithm Based Financial Data Analysis and Investment Decision Support Modeling

Chengcheng Zhu[1,✉]

[1] *Accounting College, Zhengzhou Vocational College of Finance and Taxation, Zhengzhou, Henan, 450000, China*

ABSTRACT

The article solves problems such as personalized investment, and then achieves the expected effect of investment decision-making. The article firstly designs an investment decision support model based on collaborative filtering, elaborates the implementation path to realize investment decision support from the perspective of machine learning, and then combines the user image technology to design the user image labeling system and model construction. Finally, the effectiveness and rationality of the proposed method in this paper are verified through experiments. Experiments on a corporate investment decision support task on a company's dataset reveal that the method proposed in this paper has good performance on all metrics, with the highest value of 0.6985 on AUC.This gives an indication of the effectiveness of the financial data analysis and investment decision support model proposed in this paper.

Keywords: incremental learning, user profiling, financial data analysis, investment decision support, collaborative filtering

1. Introduction

In this era of information explosion, a large amount of data is generated, stored and processed every day, which covers a wide range of fields from consumer behavior, market trends to corporate financial status [1-2]. Particularly in the financial domain, with the advancement of technology, traditional accounting and financial analysis methods can no longer meet the needs of modern businesses and investors, who require more accurate, timely and in-depth data analysis to support their decision-making process [3-4]. This reliance on data has not only changed the methods and tools of financial analysis, but also had a profound impact on the process and results of investment decisions [5].

In this context, big data technology has emerged to provide new possibilities for financial analysis and investment decision making. Big data provides a new way of capturing, analyzing, and interpreting data, enabling decision makers to gain deeper and broader insights from it [6-7]. This technological

✉ Corresponding author.
E-mail address: Zcc0720050044@126.com (C. Zhu).
Received 25 March 2024; Revised 05 June 2024; Accepted 15 December 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-375

advancement brings opportunities as well as new challenges to investment decision making. How to effectively utilize these data in the era of big data in order to improve the quality and efficiency of investment decisions has become a key issue for enterprises and financial institutions [8-9]. Financial big data can be defined as large, diverse, and rapidly generated data sets involving financial activities, the size and complexity of which exceed the processing capabilities of traditional data processing software [10-11]. These data are generated not only from internal financial systems, such as accounting books, financial statements, and budgeting systems, but also from external sources, such as stock markets, social media, news reports, and other information related to the financial condition and performance of the organization. This integration and analysis of data provides decision makers with a comprehensive view of the financial situation and future business trends of an organization, enabling them to better understand and predict the financial situation and future business trends [12-14].

The main characteristics of financial big data include huge volume of data, high speed data generation and processing, diversity and authenticity. Financial big data involves a huge amount of financial and fiscal information, including corporate financial statements, transaction data, customer behavior data, and so on. The huge amount of data can help decision makers to understand market trends and consumer behavior in a more comprehensive and in-depth way so that they can make more informed decisions [15-17]. Financial big data are generated at a very fast speed, especially in areas such as Internet finance and electronic payments. At the same time, the processing speed of financial big data is also very important, which is because timely access to and analysis of data is crucial for business decisions and operations [18-19]. Financial big data covers a wide range of data sources and types, including data from the market, social media, external data providers, and other sources, in addition to financial data within the organization [20-21]. This diversity can provide enterprises with more comprehensive business insights and help them more accurately grasp market demands and opportunities. Finally, financial big data has authenticity, which means it is generated based on real transactions and economic activities [22-23]. Compared with traditional questionnaires or manual statistics, financial big data has higher authenticity and can more accurately reflect market conditions and enterprise performance [24].

With the constant changes in the global economic environment and the increasingly fierce market competition, enterprises are facing unprecedented challenges and opportunities in the business decision-making process [25]. As an important tool for modern enterprise management, financial data analysis plays an increasingly important role in the process of making business decisions. However, in the process of financial data analysis, enterprises encounter many challenges [26-27]. Therefore, it is necessary to continuously optimize and improve the process of enterprise financial data analysis, so that through effective financial data analysis, enterprises can make correct strategic decisions, operational decisions, and risk management decisions, thus enhancing their competitive advantages and sustainable development capabilities [28].

Literature [29] describes the operation mechanism and underlying logic of artificial intelligence technology in facilitating enterprise financing decision-making, which focuses on the active role of natural language processing (NLP) and robotic process automation (RPA) technology in exploring the value of information data and improving the efficiency of financial and monetary data analysis. Literature [30] conceptualized a management accounting analysis framework based on the logic underlying the balanced scorecard theory in the context of business intelligence, which provides strong support for corporate financial performance evaluation and business decision-making. Literature [31] constructs four educational curriculum design components with reference to the comprehensive model of curriculum design and applies them to the teaching of accounting courses, and points out that information technology such as blockchain technology, big data analysis technology will become an important driving force for reform and innovation in the field of accounting in the future. Literature [32] builds an integrated model of accounting, auditing and financial data analysis applicable to foreign economic management, which effectively supports foreign economic cooperation and investment and trade. Literature [33] proposed the introduction of artificial intelligence technology in enterprise financial auditing work, namely (genetic algorithm optimized deep learning neural network model) to improve the accuracy of financial audit classification, and data conversion efficiency. Literature [34] constructed a set of cloud accounting IT auditing model based on big data technology and in simulated numerical tests, it was proved that the cloud accounting auditing model can provide accurate and effective data analysis results for enterprise financial analysis and financial decision-making.

The article firstly carries out theoretical research on collaborative filtering algorithm, incremental learning and asset allocation, and proposes a recommendation algorithm based on incremental over-arching learning machine on the basis of relevant theoretical analysis. That is, new neurons are added according to investor characteristics and asset allocation until the whole process is stable. Subsequently, the design of the user portrait labeling system and the construction of the model are carried out to realize the accurate push of intelligent investment decisions. The initial investor data collection is formed by collecting relevant information about investors, and then the basic data is classified by combing and summarizing the big data cleaning and data analysis. Secondly, the investor's behavior model is constructed and the investor is labeled. Finally form a personal planning report based on the algorithm proposed in this paper. And the experiments on real company datasets with adequate enterprise investment decision support tasks are carried out, and the automated and interpretable investment recommendations given by the method of this paper are demonstrated through case studies to validate the effectiveness and reasonableness of the method proposed in this paper.

2. Design and modeling of investor profile labeling system

2.1. Investor user profile construction process

Currently, the process of creating user profiles is considered as collecting user profile data, exploring user information, sharing tags and enriching user profiles. The building process involves extracting basic characteristics, needs, preferences and other functional information about the user. The process of building a user profile is divided into three stages: data collection, data mining and filtering, and label extraction and reorganization [35]. Based on the process of user profiling, the process of investor profiling is subdivided by combining the investment product market and relevant information about investors, as well as data types and service requirements related to investment products, and the process of investor profiling is refined, and the process of constructing user profiles is shown in Figure 1.

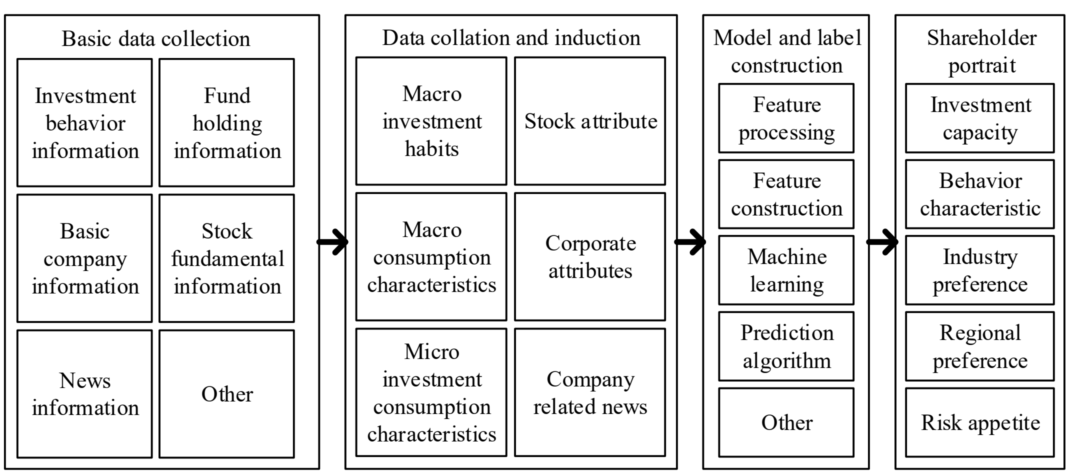

Fig. 1. Flowchart for creating a user portrait

2.2. User Profile Construction for Investors

2.2.1. Collection of portrait indicators. Investment users need to consider many factors in the investment process, not only real-time investment product price changes and trends over time, but also to consider the industry, region, concept, market sector and other investment product data, and make their investment decisions to adapt to the national policy and listed companies of good news and bad news, so the construction of the investment user knowledge base to include the user and the market of the two major categories, the five data aspects: the user's basic information, simulated user transaction data, user transaction line release data, product market data, product attribute data. Only by thoroughly processing various data can this paper accurately describe the behavioral changes of users in the investment process.

1) Basic user information:

(1) Personal basic information. This paper selects seven indices in the personal basic data that can reflect the value of customer potential.

(2) User's existing assets. This paper selects two indexes, account opening length and customer assets, to measure the user's existing assets. The account opening time and customer asset accumulation are in fact related to the user's behavioral characteristics.

2) User trading behavior:

(1) User activity. User activity is related to its transaction frequency, this paper uses the number of new accounts, transaction amount, the frequency of changes in holdings measured. According to the user activity can be divided into long term customers, medium term customers and short term customers or it is divided into high turnover customers, turnover customers and low turnover customers.

(2) User preference characteristics. User preferences are portrayed through the user's choice category preferences, choice method preferences, preferences and acceptable risk preferences for investment products. The percentage description of investment category preferences contains the percentage of investment product type, fund type, and investment product type. Risk tolerance preference is usually obtained from the research questionnaire, but because of privacy issues customers usually can not fill in the real information, so the reference value is not great, so it can be based on the risk factor of its investment products itself, roughly portrayed its risk tolerance tendency.

3) User Social Information:

(1) Social Influence. On the other hand, the user's portrait can also be portrayed through the user's social behavior, in general, a person is committed to investing in his attention to the social platform or social circle for the user's investment preferences have an important impact, so it is also very important to portray the user through social behavior. Social influence indicates the user's status or authority on investment in his circle.

(2) Social Activity. For a person who has been interested in the investment field for a long time, investment is naturally his interest, so his actions on social platforms can reflect his interest level.

2.2.2. Labeling system construction. User profiling mainly consists of extracting, analyzing, classifying and labeling user information from data. The behavior of investment product investment users has strong domain characteristics, so when selecting and extracting functions, attention should be paid to the domain-specific characteristics of users. Therefore, it is necessary to first collect data on the behavior of investment product investment users and establish a comprehensive and representative identification system for investment product investment users.

In view of the specificity of investment product investment user behavior, investment behavior analysis should not only refer to user behavior analysis methods in other industries, but should also be specifically analyzed and designed according to the characteristics of the equity investment field. For ordinary Internet users, the user can be described by focusing on the key information and behavioral information obtained when interacting with the platform, but for investment product investment users and the market conditions related to the investment process. Therefore, the behavioral description for investment product investment users can be expressed as formula (1):

$$A_n = l_m \cup l_n \cup l_o \cup l_q \tag{1}$$

where A_n represents the collection of overall behavioral description of investment users, l_m represents the collection of basic attributes of investment users including personal profiles, l_n is the collection of user social behaviors, l_o is the collection of data on the expression of emotional tendencies of investment users, and l_q is the collection of transaction behaviors of investment users. Through the analysis of user behavior, combined with the characteristics of investment users, a total of five dimensions of user profile labels including basic investment ability labels, behavioral characteristics labels, industry preference labels, geographical preference labels and risk preference labels, the user profile indicator system is shown in Figure 2, these user labels represent the user's personal characteristics in a certain dimension, and these labels can be generated after a combination of matching with representative user descriptions, in terms of both group and individual. These user labels represent the user's personal characteristics in a certain dimension, and these labels can be combined and matched to generate representative user descriptions, which are interpretable at both the group and individual levels.

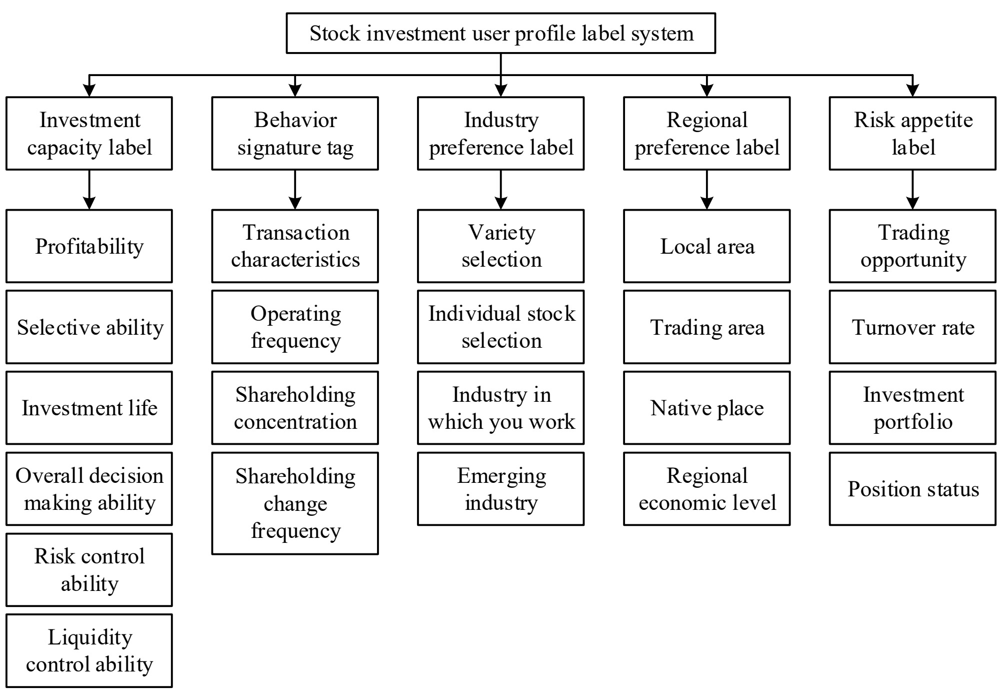

Fig. 2. User profile indicator system of stock investors

2.2.3. Classification of investor labels. The kernel of the user image is to “tag” the user, and the first step in building a user profile is to figure out what kind of tagging is required, and what kind of tagging is based on the industry needs and the actual state of the data. The method of “labeling” is determined by the different types and nature of the label. In general, according to the labeling "hit" method can be divided into three categories. (1) fact labeling: obtained directly from the database or after simplified calculations. (2) Model labeling: Labels constructed with machine learning and natural language processing techniques. (3) Advanced labels: obtained by modeling data based on factual labels or model labels.

3. Investment decision support model

3.1. Related work

3.1.1. Collaborative Filtering Algorithm (CF). Collaborative filtering (CF) can be said to be one of the most well-known classic algorithms in the recommendation field. In all kinds of recommendation scenarios, we often get massive information about users' feedback, ratings, comments, etc. on items, and the so-called collaborative filtering is a process to filter out the items that the target users may be interested in from this massive information [36]. In collaborative filtering recommendation algorithms, they can be categorized into memory-based collaborative filtering and model-based collaborative filtering. Among them, memory-based collaborative filtering can be further categorized into two common approaches: user-based collaborative filtering (User-CF) algorithm and item-based collaborative filtering (Item-CF) algorithm.

1) The User-CF algorithm makes recommendations from the user's point of view by comparing the behavioral similarities between users. First, the algorithm creates a user-item rating matrix, where each element represents the user's rating or behavior towards the item. Then, a set of users that are most similar to the target user is found by calculating the similarity between the users, such as cosine similarity or Pearson correlation coefficient. Finally, based on the rating data of this set of similar users, the target user's preference for the item that has not yet been rated is predicted and a recommendation is made. Then according to the algorithm we are a little more specific, naturally raises the question of how to measure the similarity between two users. For example, in real life how to accurately determine whether the interests of two people are similar. Suppose you like a certain type of clothes, it so happens that your roommate also likes this type of clothes, then obviously you will have more common topics, that is, you have common interests, and the similarity between you is much closer, and we can just use this to calculate.

For example, suppose there are user μ and user v, such that $N(\mu)$ represents the ensemble of items that user μ likes and $N(v)$ represents the ensemble of items that user v likes. Then $N(\mu) \cap N(v)$ represents the ensemble of items that both user μ and user v like at the same time, and $N(\mu) \cup N(v)$ represents the ensemble of items that both user μ and user v like, so the following formula can be used to calculate the similarity:

$$W_{\mu v} = \frac{|N(\mu) \cap N(v)|}{|N(\mu) \cup N(v)|} \tag{2}$$

The above formula is called Jaccard's formula, which is intuitively understood as the number of items in the set of items that both user μ and user v share a favorite divided by the sum of their favorites, and if it so happens that the items favored by μ and v are exactly the same with no difference, then μ and v have a similarity of 1.

It can also be calculated using the cosine similarity formula, which means that the denominator part becomes the product of the number of items liked by μ and the number of items liked by v and then squared, and no longer the intersection between them. The formula is as follows:

$$W_{\mu v} = \frac{|N(\mu) \cap N(v)|}{\sqrt{|N(\mu) N(v)|}} \tag{3}$$

2) Item-CF algorithm is a recommendation method based on item similarity. This algorithm makes recommendations from the perspective of items by comparing the similarity between items. First, the algorithm creates an item-user rating matrix, where each element represents the user's rating or behavior on the item. Then, it finds a set of items that are most similar to the target item by calculating the similarity between items, such as cosine similarity or Jaccard similarity. Finally, based on the target user's rating data for this set of similar items, the target user's preference for items that have not yet been rated is predicted and recommendations are made. Unlike those algorithms that directly use item attributes to calculate, this algorithm calculates the similarity between items by analyzing user behavior. This not only discovers the user's preference for certain items, but also can be based on these similarity degrees, thus recommending other items with similar characteristics to the user. In this way, recommendations can be obtained to the full satisfaction of the user. The formula for calculating the similarity of items:

$$W_{\mu v} = \frac{|N(\mu) \cap N(v)|}{N(\mu)} \tag{4}$$

In the above expression, the denominator $N(\mu)$ is the overall number of users who like the item, while the numerator $N(\mu) \cap N(v)$ is the number of users who like both items μ and v. So the above equation can be interpreted as how many of the users who like μ also like v, and if all the ones who like μ like v, then $W_{\mu v} = 1$. The above equation still looks reasonable, but if item v is very popular, and a super large number of people like it, then the numerator and the denominator in the above equation will be very close to each other, and at this point $W_{\mu v}$ will be very close to 1. That is, there is a high degree of similarity between any item and a popular item, and this would lead to the fact that the Item-CF algorithm will always recommend popular items, so this is not a good way to calculate.

So the following formula can be used:

$$W_{\mu v} = \frac{|N(\mu) \cap N(v)|}{\sqrt{|N(\mu)||N(v)|}} \tag{5}$$

The improvement of the above equation is that if $N(v)$ is very large, the denominator will be correspondingly large, which is equivalent to penalizing the weights of the items and mitigating the possibility that popular items will be similar to many other items.

After establishing the similarity matrix of items, like User-CF algorithm, we have to face the problem of how to select the most interesting items for users from many similar items. Therefore Item-CF algorithm calculates the degree of interest of user μ in item j by the following formula P:

$$P(\mu, j) = \sum_{i \in S(j,K) \cap N(\mu)} W_{ji} R_{\mu j} \tag{6}$$

$N(\mu)$ in the above equation is the set of items preferred by the user, and $S(j,K)$ is the set of K items with the highest similarity to item j, i.e., the similarity of items j and i. $R_{\mu i}$ is the user's interest in μ item i (for the implicit feedback dataset, if the user has a history of behavior towards item i, it can be ordered $R_{\mu i} = 1$). It is commonly understood that for each item i in the list of items μ preferred by the user, the closest K item to its similarity can be found according to the item similarity matrix, order $K_{ij}, j = 1,2,...K$, and then the item similarity W_{ji} is used to indicate the user's interest in the item j. Finally, the filtered items are sorted in reverse order according to the user's interest in them, and the whole list or the first K items of the list are recommended to the user, so the Item-CF algorithm is completed.

3.1.2. Incremental learning. Incremental learning, also known as continuous learning or lifelong learning, is one of the directions that seeks to solve this problem in the field of deep learning. Although current deep learning models have demonstrated very good performance in a wide range of tasks, standard multilayer perceptron architectures still face a serious problem of catastrophic forgetting, making it difficult to meet the needs of real environments as they dynamically change. In task-incremental learning, the model is confronted with a series of different tasks that differ in dataset, goal, or both. The goal of the model is to learn new tasks while maintaining knowledge of previous tasks and avoiding significant performance degradation on old tasks. In incremental data learning, the model focuses on changes in the data distribution while the categories and tasks remain the same. The model needs to adapt to such incremental changes in the data, e.g., the evolution of the data's quality, characteristics, or distribution over time. In category incremental learning, the goal of the model is to gradually expand the number of categories it can recognize [37]. This means that over time, the model needs to learn to recognize emerging categories without training from scratch. Category incremental learning can be further categorized into single-step incremental learning and multi-step incremental learning. Single-step incremental learning refers to adding a new category or set of categories only once in the learning task. Specifically, in a single-step incremental learning scenario, the model is first trained on a certain number of categories to form a baseline model. After that, a new set of categories is introduced and the model needs to incrementally learn these new categories without forgetting what it has already learned (the original categories). Multi-step incremental learning is even more complex, as it involves the process of adding new categories multiple times. In this scenario, the model needs to learn new categories at each step while maintaining the memory of all previous categories. This approach more closely resembles real-world scenarios in which new categories may appear continuously.

3.1.3. Classical asset allocation theory. Asset allocation is the rigorous execution of an investment strategy that attempts to balance the balance between risk and return by adjusting the percentage of each asset in a portfolio based on the investor's risk tolerance, objectives, and investment timeframe. A portfolio is a collection of investments held by investment firms, hedge funds, financial institutions or individuals. The term portfolio refers to any combination of financial risks such as investment products, bonds and cash. Portfolios may be held by individual investors and/or managed by financial professionals, hedge funds, banks and other financial institutions. It is a generally accepted principle that portfolios are designed according to the investor's risk tolerance, time frame and investment objectives. The monetary value of each asset may affect the risk/return ratio of the portfolio and is referred to as the asset allocation of the portfolio.

Asset allocation is shown in Figure 3 and there are several methods of calculating portfolio returns and performance. One traditional method is to use quarterly or monthly currency-weighted returns, but a true time-weighted approach is the method preferred by many investors in the financial markets. There are also several models that can be used to measure the performance attribution of portfolio returns as opposed to indexes or benchmarks, which are partially considered investment strategies. On freelance websites portfolios are considered work you have done in the past, or investment products you own in any company.

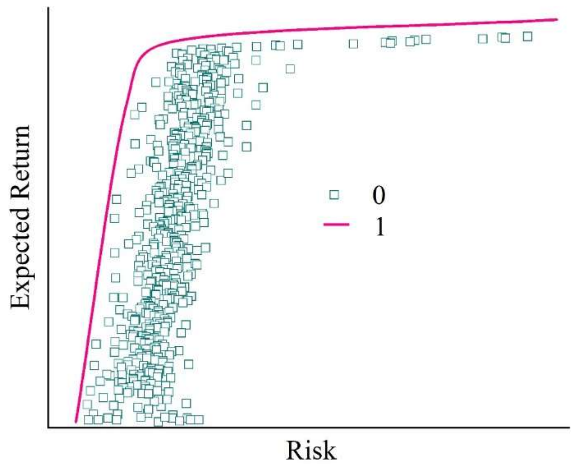

Fig. 3. Asset allocation

3.2. Algorithm implementation

3.2.1. Investment decision support model framework. The BD-IELM investment decision support model is shown in Figure 4.The BD-IELM investment decision support model consists of three parts: investor transaction prediction, financial service selection and recommendation part. Investor transaction prediction includes feature extraction and BD-IELM training. For friend recommendation, two features are extracted: content-based features and neighborhood-based features. The content-based feature detects the overall path structure in the financial time series. The neighborhood-based feature considers the number of mutual friends. For product recommendation, three features are extracted: the net worth impact feature, the popularity feature and the macroeconomic impact feature. The feature of net worth impact is extracted using power law distribution. The popularity feature is extracted using positional entropy. The popularity impacted features are extracted using a collaborative friend-based filtering approach [38].

In financial services selection, financial data can be represented as three graphs: user-user graph, product-product graph and user-product graph. In the user-user graph, each vertex represents a user and each undirected edge represents a friendship between two users. In a product-product graph, each vertex represents a product and each undirected edge represents the relevant distance between two products. In a user-product graph, there are two types of vertices: user vertices and product vertices. The directed edges starting from the user until the end of the product indicate that the user has purchased these products. In the recommendation part, BD-IELM divides the recommendation candidate space into two groups: positive instances and negative instances. A set of positive instances is the set of recommendation results.

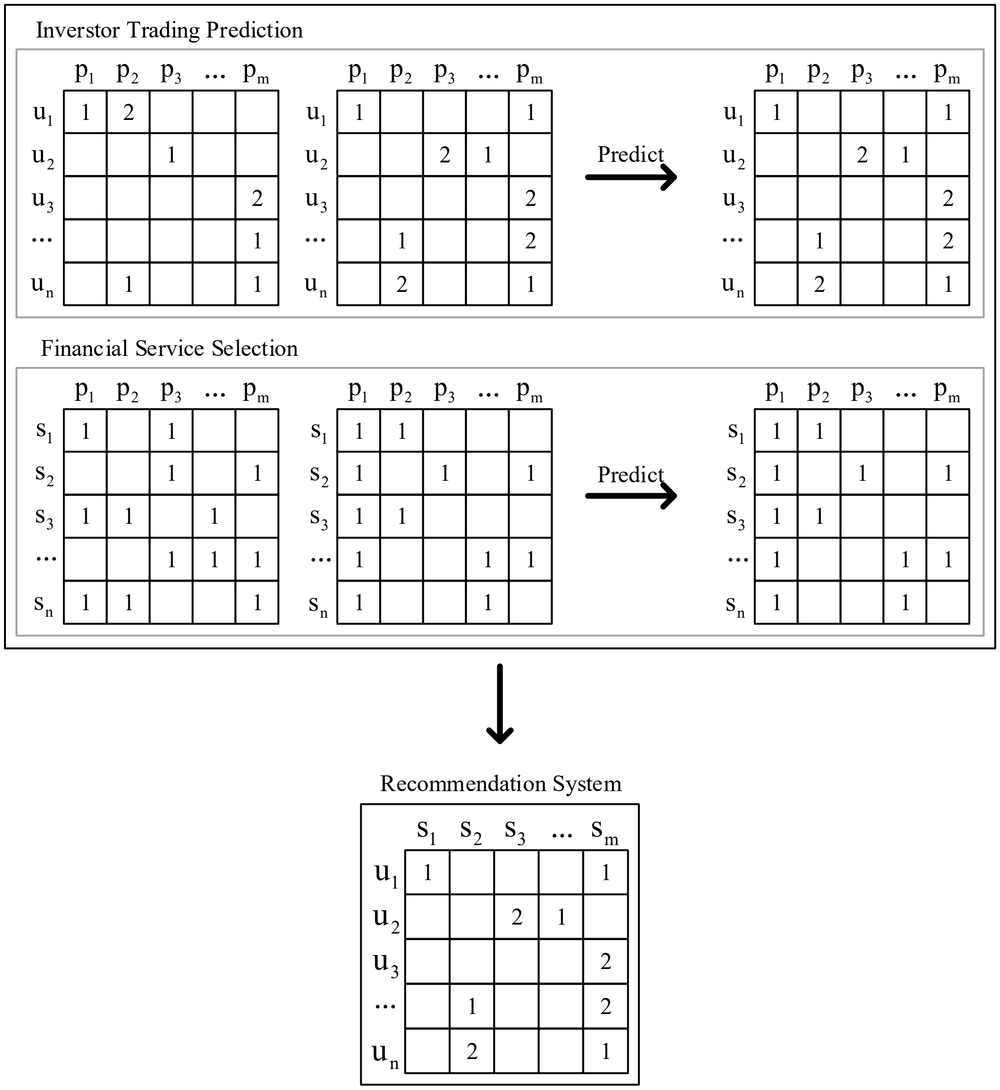

Fig. 4. Architecture diagram of BD-IELM recommendation system

3.2.2. Investor trading forecasts. Investor trading forecasts belong to the category of long-term forecasts, which require the estimation of multiple future values. Dynamic predictions of user investment behavior are particularly important for recommender systems. In general, appropriate program recommendations should not only consider the investor's current benefits and risks, but also the ways in which his or her future benefits may change. Based on the vector autoregressive model, the vector of investor potential factors $I_i' \in \mathbb{R}^K$ in time interval t can be used with the user in the previous $\tau(\tau < t)$ time interval:

$$I_i^t = C_i^1 I_i^{t-1} + C_i^2 I_i^{t-2} + \ldots + C_i^{\tau} I_i^{t-\tau} + \epsilon_i^t \tag{7}$$

As in the above equation, $\{C_i^j \in \mathbb{R}^{K\times K}\}_{j=1}^{\tau}$ is a coefficient matrix, $\epsilon_i^t \in \mathbb{R}^K$ is uncorrelated Gaussian noise with zero mean, and $D_i \in \mathbb{R}^{K\times K}$ is a time-invariant covariance matrix.

For user i given τ and T time intervals $\{I_i^t\}_{t=1}^T$ of the available series of investor interests, we use least squares to estimate the learned parameters $\{C_i^j\}_{j=1}^{\tau}$ and D_i in Eqs. We can then predict investor interests $\{I_i'\}_{t=T+1}$ and learned parameters based on the time series of observed investor interests. In our previous work, we proposed an efficient limit learning algorithm, called $\ell_{2,1}$ the limit learning machine based on paradigms and random Fourier mapping ($\ell_{2,1}$ RF-ELM), based on stochastic Fourier mapping and $\ell_{2,1}$ paradigms for regression, binary and multi-class classification. In this approach, RF-ELM is an explicit approximation of implicit Gaussian kernel ELM. Moreover, due to the integration of $\ell_{2,1}$ paradigms, $\ell_{2,1}$ RF-ELM can automatically prune irrelevant and redundant hidden neurons to form more discriminative and compact hidden layers. Therefore, we propose to use $\ell_{2,1}$ RF-ELM for financial time series forecasting and apply the algorithm to real-world datasets.

3.2.3. Choice of financial services. Asset allocation is a dynamic process, the investor's risk tolerance, investment funds are constantly changing, so for different investors, the risk of different asset allocation motivation means different, so the final choice is not the same portfolio.

4. Performance testing of the investment decision support model

4.1. Experimental setup

In order to evaluate the investment prediction results of the methods in this paper, this section slices the SkyEye dataset into a training set and a test set, and each method performs investment prediction on the same test set. Due to the complex dependencies between firms and the fact that one investment event may have a significant impact on another investment event at a later time, this paper slices the dataset especially according to the chronological order of the investment time. With the constructed training and test sets, this paper compares the prediction results of the framework proposed in this paper with the baseline approach using standard evaluation metrics used in recommender systems, including AUC, Recall, MAP, and NDCG. For the three metrics of Recall, MAP and NDCG, this paper evaluates the top k of the predicted recommendation results, where k =52052. Such a setup can better see the performance of the recommendation method proposed in this paper under different quantities.

4.2. Analysis of the effectiveness of investment decision support

This section will first validate the predictive performance comparison between the method proposed in this paper and each of the baseline methods on the investment decision support task. In order to validate the effectiveness of the framework proposed in this paper, this paper compares the SPGCN framework with some state-of-the-art methods on investment decision support tasks. These baseline methods fall into two main categories: those based on decision support and those based on graph characterization. Specifically, all the baselines used in this paper are shown below:

Popular: this is a simple and effective baseline method that predicts items based solely on their popularity in the training set.

NCF: This is a collaborative filtering recommender system in a neural network framework.

LightGCN: This is a graph-based investment decision support method.LightGCN uses a simplified GCN to achieve state-of-the-art results on many recommendation tasks.

GRU4Rec: This is the classic approach to sequence recommendation. Investments are treated as sequences based on investment date information and sequence prediction performance is evaluated.

Graphsage: this is a simple but widely used graph characterization method that can be applied to large-scale graph network data.

GAT: This is a method that introduces an attentional mechanism into graph neural network representations, through which the influence of each neighboring node on the current node is measured.

RGCN: This is the classical representation method for multi-relational graphs.

SGCN: Graph neural networks are combined with symbolic networks for graph representation to decouple and model the dependencies between cooperative and competitive relationships.

A comparison of the results on the company's investment decision support task is shown in Tables 1 and 2. From the experimental results:

1) Notice that the method proposed in this paper has the best or second best performance on all metrics. For example, the highest value on AUC is 0.6985, which shows the effectiveness of this paper's method on investment decision support tasks.

2) All decision support methods and uni-relational graph embedding methods (Graphsage, GAT) have the same inputs, i.e., with the same inputs, the decision support methods usually have better performance than purely graph-based characterization methods on investment decision support tasks. This is mainly due to the fact that investment graphs are dichotomous in nature, and more general graph characterization techniques are unable to capture specific patterns between investors and investees; on the contrary, investment decision support related methods, whether based on collaborative filtering or LightGCN are better adapted to investment decision support tasks.

3) RGCN and SGCN models outperform decision support methods and single-relationship graph characterization methods. However, RGCN's generic modeling of different relationships and SGCN's specific modeling of cooperation and competition are not accurate enough for investment decisions. In contrast, the method in this paper realizes the decomposition of propagation direction through an innovative network architecture, which further improves the performance of the investment decision support task.

Table 1. The effect of the company's investment recommendation task is compared

Model	AUC	Recall@5	Recall@10	Recall@20
Popular	0.5422	0.0449	0.1182	0.2214
NCF	0.6012	0.1704	0.2986	0.3941
LightGCN	0.6182	0.2026	0.3118	0.4169
GRU4Rec	0.5967	0.1834	0.2977	0.3992
Graphsage	0.5932	0.2185	0.3396	0.3748
GAT	0.59	0.1763	0.304	0.422
RGCN	0.6854	0.2956	0.4054	0.47
SGCN	0.5889	0.2274	0.3165	0.4534
Ours	0.6985	0.2872	0.415	0.4296

Table 2. The effect of the company's investment recommendation task is compared

Model	MAP@	MAP@	MAP@	NDCG@	NDCG@	NDCG@
Popular	0.0496	0.0423	0.042	0.0235	0.062	0.0897
NCF	0.1035	0.1193	0.1479	0.1104	0.16	0.2124
LightGCN	0.0757	0.121	0.1317	0.1587	0.2072	0.1877
GRU4Rec	0.0894	0.0368	0.1393	0.1312	0.1506	0.1875
Graphsage	0.0285	0.089	0.0774	0.1071	0.1708	0.1644
GAT	0.0581	0.1163	0.0935	0.1009	0.1251	0.2075
RGCN	0.1126	0.1643	0.1792	0.1753	0.2144	0.2457
SGCN	0.1144	0.1373	0.0909	0.1658	0.1738	0.2174
Ours	0.1626	0.1864	0.1452	0.2137	0.2177	0.274

5. Case studies

5.1. Individual investment decision support

By importing the historical customer portrait, combined with the financial data of customer Wang, analyzed by the algorithm, the customer asset allocation ratio is shown in Figure 5.

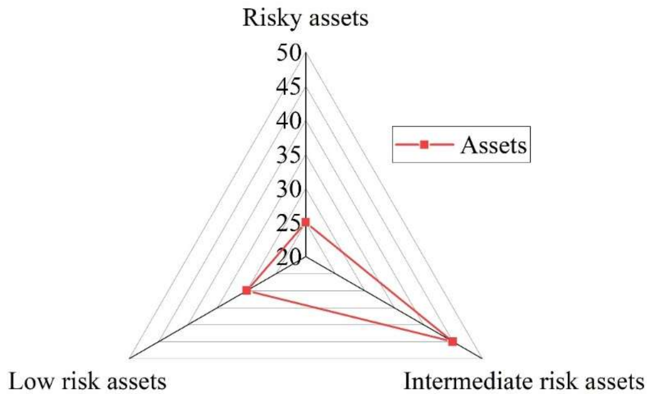

Fig. 5. The allocation of customer assets

According to the personal planning report of Xiao Wang automatically generated by the BD-IELM investment decision support model, the personal economy is shown in Table 3.

Table 3. Personal economy

Overall investment experience	The overall investment experience is defined as: experienced investors according to your response to the following questions	
	Which of the following highlights your investment experience?	Most of the investment in stocks, funds, foreign exchange, etc., is less invested in deposits and government bonds
	How many years of experience have you invested in stocks, funds, foreign exchange and financial derivatives?	5 years (excluding) to 8 years
Specific product knowledge and trading experience	Product type	Your product cognition and experience
	Open fund	Knowledge, no trading experience
	Bonds	Cognitive, more than 3 years of experience
	Structured investment products	Knowledge, no trading experience
	Investment link insurance	Knowledge, no trading experience
	Non-investment link class insurance	Cognitive, more than 3 years of experience

The financial situation is shown in table 4.

Table 4. Financial situation

(1) Basic financial situation			
Categories	Detail	Amount	Proportion
Revenue (Monthly income: ¥120,000)	Wage income	¥120,000	100%
Expenditure (Monthly expenditure: ¥45,000)	Loan/mortgage expenditure	¥25,000	55.56%
	Other expenses/non-living expenses	¥20,000	44.44%
Assets (Total assets: ¥22,900,000)	Residential property	¥16,000,000	69.87%
	Other current assets	¥2,500,000	10.92%
	Investment assets	¥2,200,000	9.61%
	Local pension	¥2,200,000	9.61%
Load (Total liability: ¥2,500,000)	Other private loans	¥2,500,000	100%
Disposable input (salary ratio)	Monthly disposable income	¥50,000	41.67%
(2) Net worth			
According to the information, your net worth is (positive)		¥18,000,000	
Emergency funds (six months of household spending)		¥235,000	
Amount of investment		¥18,550,000	
The number of people raised		0	
Emergency funds have been reserved		Yes	
(3) Risk level			
Your risk tolerance		4	

The goals and preferences are shown in Table 5.

Table 5. Goals and preferences

(1) Investment appreciation		
Detailed content	Initial information	After change
Your initial investment amount	¥600,000	¥600,000
Your balance investment amount	¥0	¥0
Your investment period	Five years	Five years
Your investment target value	¥450,000	¥450,000
Inflation rate	2.63 p.a.	2.63 p.a.
Investment value (premium return hypothesis)	¥1,043,265	¥1,043,265
Investment value (median return hypothesis)	¥720,948	¥720,948
Investment value (low return assumption)	¥385,512	¥385,512
Growth rate (median return hypothesis)	6.99 p.a.	6.99 p.a.
More than/below the investment target (median return hypothesis)	¥135,649	¥135,649
Additional investment amount per month	¥0	¥0
(2) Investment preference		
Investment limit	Three years or less	
Exchange rate risk	Yes	
Regional risk	Yes	
Liquidity preference	No preference	
preference	No preference	

The product selection is shown in Table 6.

Table 6. Product selection

(1) Discussed plan						
Product category	Product name	Product risk level*	currency	Amount of investment/amount of investment		
SID	The two years of structured investment products for the two years of the RMB	1	CNY	N/A		
SID	Transfer - one-year RMB structured investment products	2	CNY	N/A		
(2) You choose (other) products						
Product category	Product name	Product risk level*	currency	Amount of investment (initial amount)	Amount of investment (monthly investment)	Product screening condition
SID	The two years of structured investment products for the two years of the RMB	3	CNY	N/A	¥0	Risk level: high risk or below

5.2. *Formula investment decision support*

The investment decision support case for a company is shown in Figure 6. In the figure, the paper selects an investment firm from a realistic dataset and lists the top 5 candidate investment firms that this paper's methodology recommends for them. These two firms had made investments primarily in one industry, and the methodology of this paper was able to recommend multiple companies in other industries for each of them. As is already known from their past investment track record, the investment firms are currently focused on the real estate industry, while this paper's methodology recommends companies in the finance, IoT real estate, manufacturing, and construction industries. First of all, the methodology of this paper recommends companies directly associated with real estate as investment decision support. And the model also recommends finance as an alternative candidate industry, which may be implicit in other previous investment interests of the investing company. On the other hand, the methodology of this paper also recommends related companies that have complementary relationships with housing, such as those in the construction, manufacturing, and IoT industries. Moreover, it is possible to observe their relationships on the localization of the dependency network of the firms on the right hand side and verify that these complementary firms co-occur in the same dependency chain, while the alternative firms are perpendicular to the direction of the dependency chain. This case validates that the methodology of this paper can lead to interpretable investment recommendations in investment analysis.

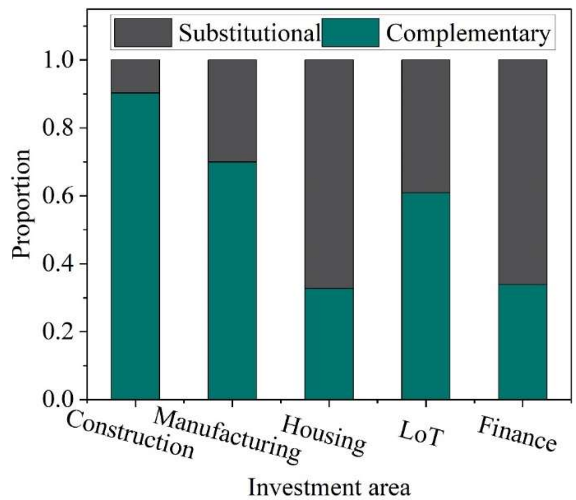

Fig. 6. A company's investment recommendation case

6. Conclusion

Accompanied by the rapid development of the financial market, the algorithmic model to obtain the most perfect asset allocation for investors, to provide personalized choice space for investors to buy investment products, and personalized recommendation technology to improve the learning efficiency and accuracy of the intelligent investment adviser is the top priority of financial data analysis and investment decision support. Therefore, the article proposes an investment decision support model based on collaborative filtering algorithm.

Through experiments on investment decision support tasks on real company datasets, the AUC value of this article's method is the highest 0.6985, which shows the effectiveness and validity of this article's method.

In the investment decision support case study of a company, it is found that the method of this paper can make good investment analysis, which leads to interpretable investment recommendations.

References

[1] Adesina, A. A., Iyelolu, T. V., & Paul, P. O. (2024). Leveraging predictive analytics for strategic decision-making: Enhancing business performance through data-driven insights. World Journal of Advanced Research and Reviews, 22(3), 1927-1934.

[2] Popovič, A., Hackney, R., Tassabehji, R., & Castelli, M. (2018). The impact of big data analytics on firms' high value business performance. Information Systems Frontiers, 20, 209-222.

[3] Hosen, M. S., Islam, R., Naeem, Z., Folorunso, E. O., Chu, T. S., Al Mamun, M. A., & Orunbon, N. O. (2024). Data-driven decision making: Advanced database systems for business intelligence. Nanotechnology Perceptions, 20(3), 687-704.

[4] Chintala, S., & Thiyagarajan, V. (2023). AI-Driven Business Intelligence: Unlocking the Future of Decision-Making. ESP International Journal of Advancements in ComputationalTechnology, 1, 73-84.

[5] Paramesha, M., Rane, N. L., & Rane, J. (2024). Big data analytics, artificial intelligence, machine learning, internet of things, and blockchain for enhanced business intelligence. Partners Universal Multidisciplinary Research Journal, 1(2), 110-133.

[6] Bakmaz, O., Bjelica, B., IVIĆ, L., Volf, D., & MAJSTOROVIĆ, A. (2017). THE SIGNIFICANCE OF THE AUDIT PROFESSION IN THE FINANCIAL ANALYSIS OF AGRICULTURAL ENTERPRISES OF THE REPUBLIC OF SERBIA. Annals of'Constantin Brancusi'University of Targu-Jiu. Economy Series/Analele Universităţii'Constantin Brâncuşi'din Târgu-Jiu Seria Economie, (3).

[7] Iyelolu, T. V., & Paul, P. O. (2024). Implementing machine learning models in business analytics: Challenges, solutions, and impact on decision-making. World Journal of Advanced Research and Reviews, 22(3), 1906-1916.

[8] Zaytsev, A., Dmitriev, N., Bunkovsky, D., & Faizullin, R. (2022). Audit of intellectual capital at an industrial enterprise: open data analysis digital-model. Int J Technol, 13, 1473-1483.

[9] Dwiastanti, A. (2017). Analysis of financial knowledge and financial attitude on locus of control and financial management behavior. MBR (Management and Business Review), 1(1), 1-8.

[10] Birt, J., Chalmers, K., Maloney, S., Brooks, A., Oliver, J., & Bond, D. (2020). Accounting: Business reporting for decision making. John Wiley & Sons.

[11] Arianti, B. F. (2018). The influence of financial literacy, financial behavior and income on investment decision. Economics and Accounting Journal, 1(1), 1-10.

[12] Badmus, O., Rajput, S. A., Arogundade, J. B., & Williams, M. (2024). AI-driven business analytics and decision making. World Journal of Advanced Research and Reviews, 24(1), 616-633.

[13] Hao, Y., & Qiu, F. (2022). Research on the Application of DM Technology with RF in Enterprise Financial Audit. Mobile Information Systems, 2022(1), 4051469.

[14] Black, K. (2023). Business statistics: for contemporary decision making. John Wiley & Sons.

[15] Guang, Z., Wang, Y., & Wang, Q. (2016, December). Enterprise financial audit modeling research based on data mining. In 2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS) (pp. 497-500). IEEE.

[16] Kimmel, P. D., Weygandt, J. J., & Kieso, D. E. (2020). Financial accounting: Tools for business decision making. John Wiley & Sons.

[17] Gauzelin, S., & Bentz, H. (2017). An examination of the impact of business intelligence systems on organizational decision making and performance: The case of France. Journal of Intelligence Studies in Business, 7(2).

[18] Albright, S. C., & Winston, W. L. (2020). Business analytics: Data analysis and decision making. Cengage Learning, Inc.

[19] Si, Y. (2022). Construction and application of enterprise internal audit data analysis model based on decision tree algorithm. Discrete Dynamics in Nature and Society, 2022(1), 4892046.

[20] Cubas-Díaz, M., & Martinez Sedano, M. A. (2018). Measures for sustainable investment decisions and business strategy–a triple bottom line approach. Business strategy and the environment, 27(1), 16-38.

[21] Yao, L. (2024, February). Application of Artificial Intelligence Technology in Enterprise Financial Audit. In 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS) (pp. 1-5). IEEE.

[22] Rashid, A., & Saeed, M. (2017). Firms' investment decisions–explaining the role of uncertainty. Journal of Economic Studies, 44(5), 833-860.

[23] Balachandran, B. M., & Prasad, S. (2017). Challenges and benefits of deploying big data analytics in the cloud for business intelligence. Procedia Computer Science, 112, 1112-1122.

[24] Xu, L., Gao, R., Xie, Y., & Du, P. (2019). To be or not to be? Big data business investment decision-making in the supply chain. Sustainability, 11(8), 2298.

[25] 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.

[26] Palepu, K. G., Healy, P. M., & Peek, E. (2016). Business analysis and valuation. Fourth Edition: Cengage learning.

[27] Uddin, M. M., Ullah, R., & Moniruzzaman, M. (2024). Data Visualization in Annual Reports–Impacting Investment Decisions. International Journal for Multidisciplinary Research, 6(5).

[28] Maaitah, T. (2023). The role of business intelligence tools in the decision making process and performance. Journal of intelligence studies in business, 13(1), 43-52.

[29] Rane, N. L., Choudhary, S. P., & Rane, J. (2024). Artificial Intelligence-driven corporate finance: enhancing efficiency and decision-making through machine learning, natural language processing, and robotic process automation in corporate governance and sustainability. Studies in economics and business relations, 5(2), 1-22.

[30] Appelbaum, D., Kogan, A., Vasarhelyi, M., & Yan, Z. (2017). Impact of business analytics and enterprise systems on managerial accounting. International journal of accounting information systems, 25, 29-44.

[31] Qasim, A., & Kharbat, F. F. (2020). Blockchain technology, business data analytics, and artificial intelligence: Use in the accounting profession and ideas for inclusion into the accounting curriculum. Journal of emerging technologies in accounting, 17(1), 107-117.

[32] Nazarova, K., Zaremba, O., Nezhyva, M., Hordopolov, V., Harbar, V., & Gorovyj, V. (2021). Business analysis and audit of foreign economic activity of the enterprise. Studies of Applied Economics, 39(5).

[33] Ding, R. (2022). Enterprise intelligent audit model by using deep learning approach. Computational Economics, 59(4), 1335-1354.

[34] 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.

[35] Haihong Bian,Shengwei Bing,Quance Ren,Can Li,Zhiyuan Zhang & Jincheng Chen. (2025). Electric vehicle charging load forecasting based on user portrait and real-time traffic flow. Energy Reports2316-2342.

[36] Ni Li & Yinshui Xia. (2025). Corrigendum to "Movie recommendation based on ALS collaborative filtering recommendation algorithm with deep learning model" [Entertain. Comput. 51 (2024) 100715]. Entertainment Computing100835-100835.

[37] Zhi Hong Qi,Da Wei Zhou,Yiran Yao,Han Jia Ye & De Chuan Zhan. (2025). Adaptive adapter routing for long-tailed class-incremental learning. Machine Learning(3),68-68.

[38] Shuping Zhang. (2025). Integrating User Profiles and Collaborative Filtering for Smart Recommendation of Tourism City Cultural and Creative Products. International Journal of High Speed Electronics and Systems(prepublish).