

The Path Analysis of Data Mining Technology in Internet Financial Asset Management

Zhenshan Li

*School of Economics, Shanghai University, 200444 Shanghai, China
lizhenshantrend@163.com*

Abstract: *As a modern information processing technology, data mining technology has been widely used in all walks of life. The role played by great data analysis technologies cannot be replaced by other information technologies. It opened up new markets and created new functional models. It provides a strong guarantee for the investment of online financial asset management and reduces the investment risk of online financial asset management. This article focuses on the research of path analysis and DM technology to manage financial assets on the Internet. First, it uses bibliographic research methods to explain the functions of DM and the application direction of DM technology in Internet financial asset management. Finally, it uses DM technology to analyze financial assets on the Internet. Through data analysis, it can be obtained that potential buyers have the highest score for real estate project land with a score of 3.98; followed by transportation, public institutions, and public facilities. And property management, the scores are 3.87, 3.43, 3.23, 3.24 and 2.90; and the resident quality score is the lowest, only 1.21.*

Keywords: *Data Mining, Internet Finance, Asset Management, Path Analysis*

1. Introductions

With the continuous development of science and technology in our country, various industries have also developed rapidly [1-2], but Internet financial asset management is facing more and more problems, and even faces unprecedented financial asset management risks. This gives Internet finance asset management brings difficulties [3-4]. Because this kind of financial asset management on the Internet lacks relevant professional risk management and professional control of financial investment risks, and cannot really reduce financial investment risks, many Internet financial asset managers urgently need solutions [5-6]. Big data applications can effectively reduce investment failures in corporate financial management and reduce the difficulty of Internet financial asset management [7-8].

In view of the research of DM technology in Internet financial asset management, some researchers explained the necessity of applying DM technology to financial management investment [9]. In financial management, DM technology is used for financial investment. With the help of the function of predicting the results of data analysis, it is possible to accurately analyze the risk data of financial investment, provide an investment plan that maximizes the benefits of financial investment, and indicate some beneficial directions for financial investment [10]. The activities of some companies are interdependent, which will lead to increased financial risks. The application of big data technology can realize the exchange of data and information, effectively reduce the risk between companies, improve the financial efficiency, and make each party profitable. At present, the company is in a modernized operation process, facing the diversification of the industry and the trend of informatization development, which will cause the company's behavior and risks to become greater and greater [11]. Through the application of data technology in the company, a large amount of company data can be effectively obtained. Accurate and scientific assessment to effectively solve the risk management in the business [12].

This article focuses on the path analysis and research of DM technology in Internet financial asset management. First, it uses the literature research method to explain the function of DM and the application direction of DM technology in Internet financial asset management. Finally, use data mining technology to analyze real estate investment in Internet financial asset management.

2. Research on DM Technology and Internet Financial Asset Management

2.1. The Function of DM.

(1) Automatically predict trends and behaviors.

The DM system searches for the required prediction data in real-time in a large-scale database. Now, it is possible to quickly draw all the problems that require a large amount of mobile analysis and draw conclusions directly from the data itself. A good typical case is the problem of market predictability. The data found that using the promotional data recommended in the past to find the users with the highest return on investment in the future. There are also other uncertain issues, including anticipating bankruptcy or identifying people who are more likely to respond to certain events.

(2) Association analysis

Data association refers to the type of important knowledge discovery that is ubiquitous in the database. If there is mutual regularity between the values of two or more variables, this value is collectively called correlation. Correlation can be divided into simple correlation, time correlation and causal correlation. The purpose of correlation analysis is to find the relevant network in the hidden database through the network method. Sometimes the correlation function of all data in the database is unknown, even if it has been determined, it cannot be determined, so the model created by the correlation analysis is reliable and accurate

(3) Clustering

A database can be divided into many important subgroups, namely clusters. Grouping can promote people to improve their understanding of objective reality, and is an important prerequisite for the description of scientific concepts and the analysis of deviations. The main content of grouping technology includes two traditional pattern recognition methods and mathematical grouping. One of the main points is that when dividing objects or objects called models, not only need to fully consider the distance between the object and the divided object, but also the divided objects are considered comprehensively or called models to describe their meaning in detail, thus avoiding the use of traditional single-sided grouping techniques.

(4) Concept description

The main purpose of conceptual description is to describe the meaning and meaning of a specific type of object, and to summarize its relevance to this type of object. The description of the concept can be divided into two characteristic descriptions and cautious descriptions. The first is to describe the common characteristics between specific objects and types, and the second is to describe the common characteristics between different types of objects and objects difference. The typical description of creating a class only contains the similarity between all objects in the class and other objects, so there are many ways to create a unique description.

(5) Deviation detection

Every data in the database often has some abnormal entries, and these abnormal deviations can be detected from the entire database. Bias involves a lot of possible knowledge, such as division anomalies, a certain special state that does not fully meet the criteria, the similarity difference between observation results and model predictions, and changes in size over time. The basic method of deviation detection is to find a significant difference between the observation result and the reference value.

2.2. Application Direction of DM Technology in Internet Financial Asset Management

(1) Industry information.

Industry information includes: main business areas (province), company type (short description), characteristics and main business activities. The use of DM technology can analyze the main business conditions of similar companies in the industry, and then control the company's business plan in a targeted manner, investigate the configuration of investment projects and provide decision support.

(2) Economic and environmental information.

Information about the economic environment includes: economic development level, industrial structure, labor structure, material resource status, consumption level, consumption structure, and the

social and economic conditions faced by the company and its operating conditions, growth trends, industrial structure, and transportation business areas (Using DM technology to analyze the information of transportation, resources and other information in the province) can quickly and effectively obtain relevant links in a specific type of area, a specific type of population, etc. It also provides an objective basis for determining investment growth trends and main growth directions.

(3) Analysis of factors affecting investment

The analysis of external factors affecting investment mainly includes: external factors include macroeconomic environment, market economic environment, financial environment and economic growth environment; internal factors mainly include enterprise financial risks, cost management, brand competitiveness and talent barriers. DM technology can effectively find the reasons, and provide a basis for it to propose a solution and problem, and then it can solve these problems in a more targeted manner.

(4) Enterprise cost-benefit analysis

The existing DM technology can effectively summarize the data of each subsystem of the investment management system, create a data warehouse and use the data warehouse to conduct a comprehensive analysis of the cost-benefit status of the business, so as to truly understand the actual business status.

2.3. DM Algorithm

(1) Clustering algorithm based on partition.

The partition-based clustering algorithm divides a data set of a given size n into partitions k , where each partition represents a cluster, and $k \leq n$. This type of algorithm usually randomly selects the cluster center k , and then assigns the remaining objects to the cluster represented by the nearest cluster center according to a division criterion (such as a distance-based distance function). Then, iteratively find new cluster centers by optimizing the object function, and redistribute the remaining objects until the cluster centers no longer change or the number of iterations reaches the limit.

(2) Hierarchical clustering algorithm

The hierarchical grouping algorithm divides existing data sets into hierarchical clusters. This algorithm can be divided into two methods: aggregation and separation. Compression and separate hierarchical grouping use bottom-up and top-down strategies respectively to organize objects into a hierarchical structure. The aggregation method first treats each data object as a separate class, and then merges small classes into larger classes until the number of classes is met or other termination conditions are met. The separation method first treats all data objects as one class, and then subdivides the large class into many small classes in successive iterations until the termination condition is met. Different algorithms will choose different similarity or dissimilarity measures as the basis for merger or disintegration.

The basis of merging or splitting in hierarchical clustering algorithm is the similarity or dissimilarity between groups. This measurement criterion can be quantified as the distance between groups. Commonly used distances between clusters include minimum distance, maximum distance, mean distance, and average distance B . The distance is defined in the following formula, where $|p_i - p_j|$ is the distance between data objects p_i and p_j , m is the mean value of C_i , and n is the number of objects in the group G_i .

$$\text{dist}_{\min}(C_i, C_j) = \min_{p_i \in C_i, p_j \in C_j} \{|p_i - p_j|\}. \quad (1)$$

$$\text{dist}_{\max}(C_i, C_j) = \max_{p_i \in C_i, p_j \in C_j} \{|p_i - p_j|\}. \quad (2)$$

(3) Clustering algorithm based on density

It is difficult to find non-spherical groups in partition and hierarchical methods, while density-based methods can find clusters of any shape. Density-based methods use object density limits as the grouping criteria. High-density areas form clusters, and low-density areas form clusters or intervals between extreme values, and are not sensitive to abnormal data.

3. Use DM Technology to Analyze Real Estate Investment in Internet Financial Asset Management

3.1. Data Mining

(1) DM environment

DM refers to the complete process of extracting previously unknown, effective and practical information from many databases, and then using this information to make decisions or improve knowledge reserves. The DM environment can be shown in Figure 1.

(2) DM process

The application of DM technology in different fields, according to the characteristics of field data, the mining process is different. The process of DM it can be divided into five stages: data collection, data preprocessing, data conversion, and DM and knowledge generation. The DM process is shown in Figure 2.

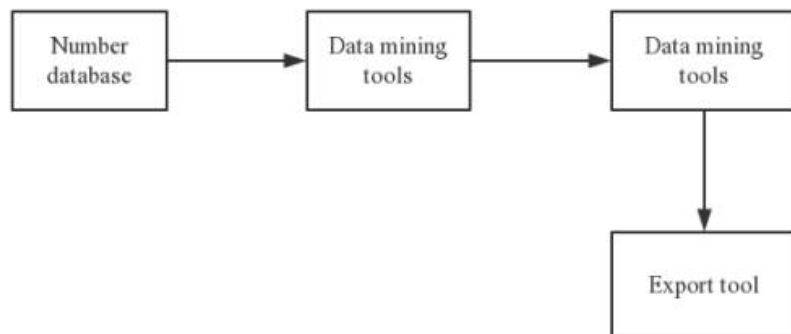


Figure 1: DM environment block diagram

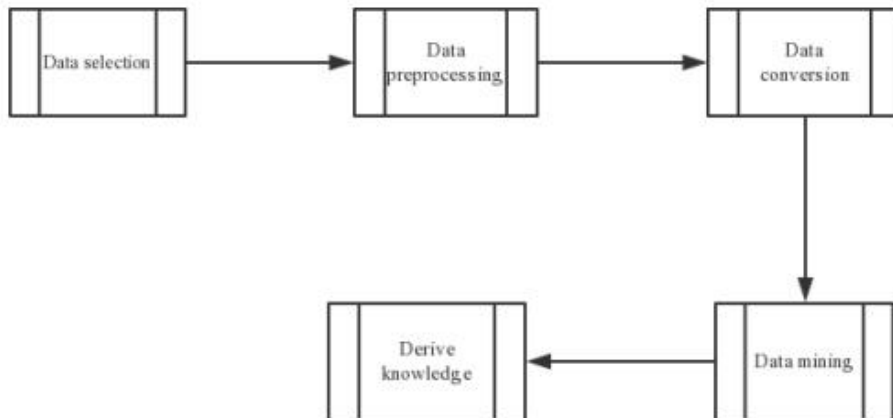


Figure 2: DM flowchart

3.2. Data Sources

According to the mining of the customer group information in the province, randomly select the demand satisfaction degree of the real estate investment project of the real estate company in the province to obtain the optimal project with high return on investment, so as to determine whether the company’s decision-making level should carry out the later stage of the project, provide a basis for investment decisions. Conduct customer return visits to buyers of houses in the province, and collect data on consumers' consumption concepts and lifestyles in the form of questionnaires.

3.3. Questionnaire Survey

The questionnaire survey is completed in two stages. The first stage is the distribution of the questionnaire, which is distributed according to the number of customers. The number of questionnaires distributed this time is 1267. The second stage is the questionnaire collection phase, which will be collected after 6 days. 1256, the response rate of this questionnaire was 95%

3.4. Data Processing

(1) Enter the questionnaire data into the EXCEL software, and establish a data table about the real estate plot elements. In order to obtain the real questionnaire data, the data must also be tested. In EXCEL, the Grubbs method is used to eliminate abnormal data.

(2) Import the processed data in EXCEL into SPSS software, and use SPSS software to analyze the data.

4. Result Analysis

4.1. Real Estate Plot Elements

According to the needs of customers, comprehensive evaluation of real estate plot elements is carried out to find out the weighting and scoring factors suitable for investment management projects. As shown in Table 1, the real estate element score table;

Table 1: Real estate plot elements

Real estate scoring elements	Mean score
Traffic	3.87
Commercial trade	3.98
Public institution	3.43
Quality of residents	1.21
public Utilities	3.23
Property management	3.24
Greening	2.90

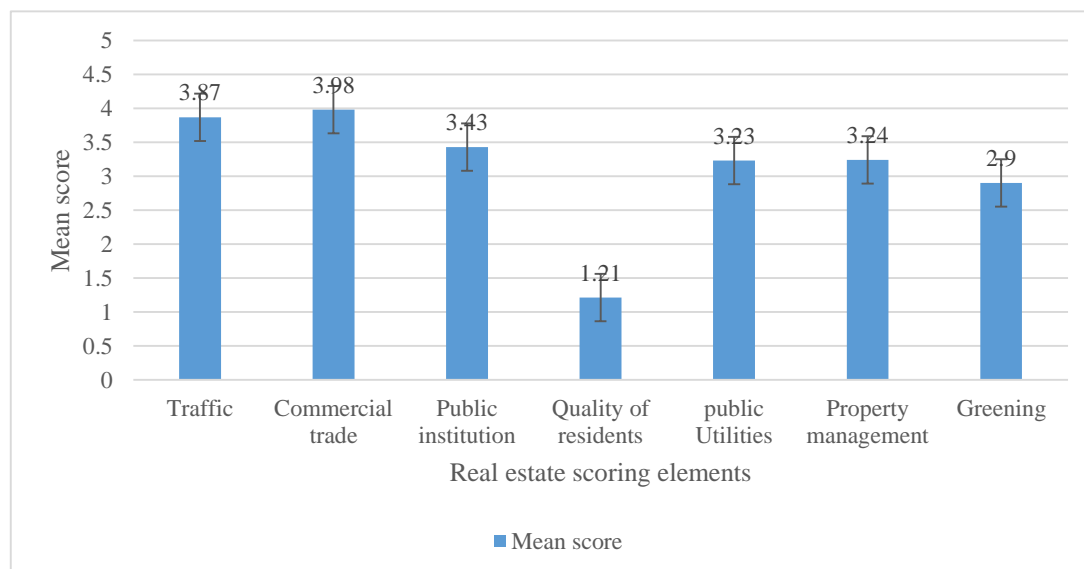


Figure 3: Real estate plot elements

It can be seen from Figure 3. Potential buyers rated the real estate project land with the highest score in commercial trade, with a score of 3.98; followed by transportation, public institutions, public facilities and property management, with scores of 3.87, 3.43, 3.23, 3.24 and 2.90; and the quality of residents. The lowest score is 1.21.

5. Conclusions

At present, DM has been widely used in all social classes, and its advantage is that other information technologies cannot replace big data technologies. It opened up a new situation for our country and established a new functional model. The application of DM reduces the risk of enterprises managing investment in financial assets on the Internet and provides effective protection for them. Big data technology is used for scientific analysis of company data to effectively control the investment risk of Internet financial services. Assets and supporting Internet financial assets have improved its financial performance. Therefore, it is inevitable and necessary to apply big data technology to the investment management of Internet financial assets. The forecast function of DM can be used to analyze data related to financial investment risks, so that the financial industry can effectively reduce financial investment risks.

References

- [1] Wang X. *Research on the Analysis of Internet Financial Risk based on Data Mining [J]. Revista De La Facultad De Ingenieria*, 2017, 32(3):720-729.
- [2] Kong Y, He W. *The path analysis of carbon emission reduction: A case study of the Silk Road Economic Belt [J]. Environmental engineering research*, 2020, 25(1):71-79.
- [3] Daewon, Kim, Seongcheol, et al. *Factors to Influence Switching Intention to Internet-only Bank from Legacy Bank: Focused on Financial Consumers' Asset Management [J]. Information Society & Media*, 2017, 18(1):105-134.
- [4] Seok-Chan, Jeong, Kyu-Lock, et al. *Development of IT Asset Management System for Financial Industry [J]. The Journal of Internet Electronic Commerce Research*, 2016(6):1-17.
- [5] Nosihana A, Yaya R. *Internet Financial Reporting dan Faktor-Faktor yang Mempengaruhinya Pada Pemerintah Kota dan Kabupaten Di Indonesia [J]. Jurnal Dinamika Akuntansi dan Bisnis*, 2016, 3(2):89-104.
- [6] Sinaki R Y, Sadeghi A, Lynch D S, et al. *Financial Asset Management Using Artificial Neural Networks [J]. International Journal of Operations Research and Information Systems*, 2020, 11(3):66-86.
- [7] Chaudhuri R, Ivkovic Z, Trzcinka C. *Cross-Subsidization in Institutional Asset Management Firms [J]. Review of Financial Studies*, 2018, 31(2):638-677.
- [8] Masih, A, Mansur, et al. *What can wavelets unveil about the vulnerabilities of monetary integration? A tale of Eurozone stock markets [J]. Economic modelling*, 2016, 52(Jan. Pt.B):981-996.
- [9] Pan H, Wang H, Wang Y, et al. *Rules of acupoint selection for diabetic peripheral neuropathy based on data mining technology [J]. Zhongguo zhen jiu = Chinese acupuncture & moxibustion*, 2016, 36(10):1111-1114.
- [10] Zhang H, Liao H, Wu X, et al. *Internet Financial Investment Product Selection with Pythagorean Fuzzy DNMA Method [J]. Engineering Economics*, 2020, 31(1):61-71.
- [11] Widyasari P A, Kurniawan E C. *Bank Financial Distress Condition Based on Ownership Structure, Audit Committee Characteristic, and Internet Financial Reporting Effect [J]. Jurnal Akuntansi*, 2020, 10(2):165-182.
- [12] Pang S, H Xian, Li R. *A default penalty model based on C2VP2C mode for internet financial platforms in Chinese market [J]. Electronic Commerce Research*, 2020(1):1-27.