# Research on Design and Implementation of Enterprise Financial Data Risk Association System Based on Data Mining

**Luo Cui**

Haojing College of Shaanxi University of Science and Technology, Xi'an, 712046, China

**Keywords:** Enterprise financial, Internal control, Risk control, Data mining

**Abstract:** In this paper, the author researches on the design and implementation of enterprise financial data risk association system based on data mining. By using the proposed system based on data mining, the system indicates that the financial crisis is a gradual process, and the model has good use prospects in the forecast of 2-3 years. Based on this, this paper focuses on analysing the financial risk of enterprises, and using data mining technology to build an enterprise financial risk assessment and early warning model, so as to provide suggestions for the development of enterprises. The model established in this paper has a good effect on short-term financial risk prediction.

## 1. Introduction

Data mining technology is a frontier discipline in modern times, and many scientists have studied it. Nowadays, the technology is widely applied (Amran, Manaf & Che, 2008). The main content of this paper is to apply this technology to the prediction of corporate financial crisis, so as to establish a Logistic regression model, which is of great significance for today's corporate finance (Chen & Du, 2009). For an enterprise, it is possible to anticipate before the risk comes, so as to take corresponding measures in time, so that the loss of the enterprise will be reduced to a certain extent and improve the company's efficiency (Geng, Bose & Chen, 2015).

If the company's financial problems are not handled properly, it will cause huge losses to the company's operation (Gray & Debreceny, 2014). Every company is facing a potential financial crisis. If the company loses seriously, it will also impact the banks and other enterprises in the society and affect the development of market economy (Griffin & Mahon, 1997). How to prejudge the potential financial crisis is an urgent problem for a company. The application of data mining plays an important role in solving such problems. It is an interdisciplinary processing technology, mainly based on AI, statistics and database technology. The technology can extract the key data of the enterprise and deal with it, thus helping the decision-makers to adjust the coping strategies to reduce the effect of the risk. In our country, the prejudgment technology of financial risk is still in the early stage. Therefore, the study of this article is of great significance in two aspects of theory and application (Han & Kamber, 2011).

## 2. Financial risk analysis theory and development

### 2.1 The role of financial risk analysis

Under the establishment and improvement of China's socialist market economy system, the management of enterprises gradually realized the importance of risk analysis. Enterprises mainly carried out risk management and risk control through risk analysis, and applied it widely to other fields (Kirkos, Spathis & Manolopoulos, 2007). The financial risk analysis of enterprises mainly has the following functions: it is the need of the national economic decision (Koyuncugil & Ozgulbas, 2012). Under the market economy system, the economic activities of Chinese enterprises can run normally, but the state's macro regulation and control can guide enterprises to develop in the right direction. Every economic policy of the state has greatly influenced the business activities and economic benefits of the enterprises (Miller & Bromiley, 1990). If the state does not take into

  DOI: 10.25236/csam.2019.028

account the operational capacity and operation rules of the enterprises when making policies, it will be very dangerous to make macroeconomic regulation and control. This will cause the enterprise's economy to suffer the wave type floating will seriously affect the operation of the enterprise; the enterprise loan financing needs. The development of enterprises can't be separated from the activity of financing. When an enterprise makes a loan to a bank or someone else, the other needs to know about its financial situation. Understand the profitability, cash turnover, development potential of enterprises, their capital status, business risk and capital turnover, and analyze their financial status, asset flow status and debt repayment on time (Ngai, Hu, Wong, Chen & Sun, 2011). When the enterprise management is pursuing the maximization of the interests of the enterprise, the enterprise should bear the risk of the enterprise in the development that is the uncertainty. In the process of operation and management, enterprises need to compile relevant financial statements, analyze their financial status, and finally make the next development strategy. The interests of the enterprise and the investor are one, and the risk is one. Therefore, before the investment, the investors will first understand all aspects of the enterprise. For example, understand the business situation, profitability, growth speed and bonus, and evaluate the risk status of the enterprise (Ravisankar, Ravi, Rao & Bose, 2011). Figure 1 shows the data flow diagram of the EWS.

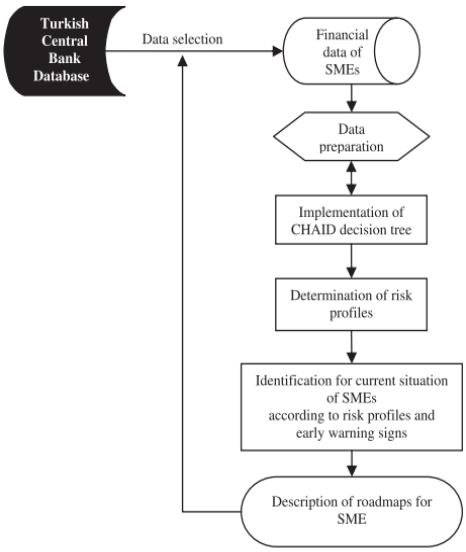


Figure 1. Data flow diagram of the EWS

## 2.2 Financial crisis warning

Early warning is a preventive mechanism, mainly in the early prediction, for example, the occurrence of financial risk deterioration in enterprises, which is usually caused by the financial crisis of enterprises (Spira & Page, 2003). When the financial risk of an enterprise can't be controlled in time and effectively, it will quickly transform into a financial crisis in the short term. Above this kind of situation in the middle and small start-up and listed companies is one of the most common financial phenomenon (Sung, Chang & Lee, 1999). Therefore, for such enterprises, building a crisis warning mechanism and models based on the actual situation is an essential part. Data mining phases is shown in figure 2.

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (1)$$

$$\sigma_x = \sqrt{np(1-p)} \quad (2)$$

$$x_{new} = a + bx \quad (3)$$

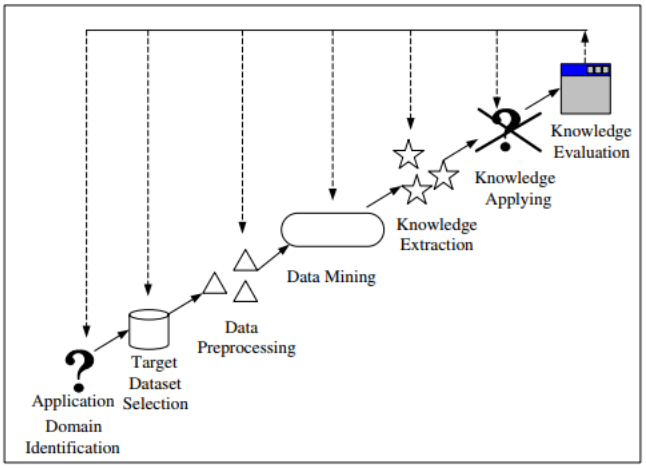


Figure 2. Data mining phases

### 2.3 Data mining technology

Data mining is a step of knowledge discovery in database, specifically, is refers to through the corresponding algorithm from a large number of data, to timely and accurately search out the value of information hiding or may be hidden in the data (Wang, 2017). Enterprises (users) can achieve this goal through statistics, online analysis and processing, information retrieval, machine learning, expert system and corresponding pattern recognition. This is mainly because the data mining thought, itself derived from statistics, with the aid of the modern artificial intelligence technology and other related theories. A large number of specially designed data analysis model, which is mainly due to the high performance computer and distributed computer processing technology (Witten & Frank, 1999). In practical applications, there are four common algorithms for data mining, classification and prediction, clustering, association rules, and time series. This paper focuses on the selection of decision tree classification methods. Conceptual framework for classifying the applications of data mining to FFD sis shown in figure 3.

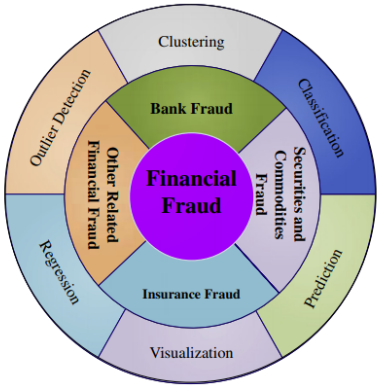


Figure 3. Conceptual framework for classifying the applications of data mining to FFD

## 3. Financial crisis early-warning of statistical data mining technology

### 3.1 Risk analysis

About the financial risk analysis and data mining concepts defined above, it has been briefly described. This paper has a brief to apply data mining technology to the daily work of enterprise financial risk analysis, implement steps include: the first is to determine the object of financial risk analysis, and prepare the financial risk analysis and related data, The second is the preparation of the object and the data preprocessing, again is the implementation of data mining, finally, the result of the test to evaluate and explain the problems and rules, and on the basis of the theory of knowledge and practical work related to the integration, which is the assimilation of knowledge.

The system structure and topology is shown in figure 4.

$$O_{pj} = f(NET_{pj}) = f(\sum_{i=0}^{1} W_{ji} O_{pi}), j = 1,2,...,m \quad (4)$$

$$O_{pk} = f(NET_{pk}) = f(\sum_{j=0}^{m} W_{kj} O_{pj}), k = 1,2,...,n \quad (5)$$

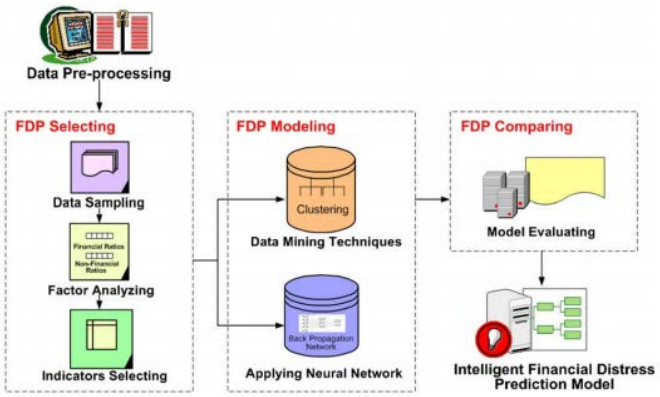


Figure 4. The system structure and topology

### 3.2 Preparation in advance

Preparation including identifying objects, and data preparation, although in the data mining technology, the enterprise financial risk analysis activity process becomes relatively simple, but its inherent thought still is more complex, the emphasis is on analysis, including the object of mining, data preparation, model, and the results of data mining to integrate processing analysis and so on, this is a whole process of unification. In this process, every link in the process of enterprises need to consider the stage what needs to be done, for example, what data mining, need how many data mining, data mining to what degree, in the process of mining knowledge which need to extract the key factors or indicators. Therefore, preparatory work can be summarized as two aspects, one is the data selection and determination, the second is accurate and processing, the purpose is to get the mining the data by packet processing, make its efficiency to achieve optimization, and finally realizes the simplify the actual effect of the mining model.

### 3.3 Build a model

Before the model is built, data needs to be analyzed in order to establish a suiTable model. For this reason, data analysis is the key to the success of data mining. Before establishing the model, we first need to analyze the data, select the variable factors, select the relevant factors from the variable factors, eliminate the unrelated factors, so as to shorten the modeling time, and further improve the use effect of the model. Then you need to select a sample from the data to build the model. From the analysis of many problems in data mining, it is more appropriate to set up multiple data samples by using selected data without information shortage. Setting up the model and selecting the appropriate data from it is a cycle process, and the aim is to find a model that can solve the problem. In the search process, the relevant personnel also need to modify the data and problems of the model. The accuracy of the model is very important, and from its accuracy, the statistical data should not be completely abandoned. In the actual process, it is necessary to use time to determine the accuracy of the data mining model. In statistical analysis, the association relationship found in data mining is easy to be omitted. For example, when using data mining tools, the question of "why a customer will react after 6 months of advertising" will not be explored. This is the case of the percentage of correctness within a given error range, which needs to be done by statistics. The speed of the model and the speed of the prediction result determine the performance of the model. In general, the speed of the neural network is relatively slow. The detected model can be excavated

by data. In the process of application, data mining is done automatically by mining, but its improvement and selection algorithm needs manual operation.

$$\mathrm{T}ypeI = \frac{Y_2}{Y_3} \quad (6)$$

$$TypeII = \frac{Y_4}{Y_6} \quad (7)$$

$$\mathrm{T}ypeIII = \frac{(Y_2 + Y_4)}{Y_9} \quad (8)$$

Enterprises need to indicators of financial risk and the factors of its existence, all need to be taken into account, based on risk control analysis on the construction of the early warning mechanism of financial crisis, the financial crisis early warning mechanism to promote the realization of the target control risk analysis. This paper mainly uses the classification method of decision tree, that is, a hierarchical structure composed of nodes and directed edges, which mainly contains three key nodes. The first is the root node. The node has no entry, mostly zero or multiple edges. The second is the internal node, and there are only (mostly) two or third entries. In the process of early warning model construction, we need to accurately control the function of early warning system of financial risk, for example, forecasting function, diagnosis function, health care function and so on. The forecast function to the listed company, the production process tracking of listed companies, the actual situation of the listed company's production and operation of listed companies with the intended target, matching each other, in-depth analysis of some problems may occur in the operation of enterprises, predict it, and find out valuable information from the deviation. The prediction mechanism is mainly found in the financial risk factors in economic activities, timely forecasting, pay attention to the effect, so you can make the management of the enterprise as soon as possible, develop solutions for better, and avoid or reduce the risks of loss of financial crisis. The second is the diagnosis function, diagnosis of this part is mainly based on the track, in front of the forecast results, analyzed the concrete, can make use of some advanced modern management mechanism, through the diagnosis of actual operating conditions to evaluate the company, find out the existing problems and disadvantages. After forecasting and diagnosis, the next step is to control and find the cause based on the existing problems, so that we can "suit the right medicine" and take relative measures to solve and deal with the deviation and fault in the development of business.

In this paper, the financial risk prejudgment model is established to illustrate the principle of Logistic regression analysis. Its specific description is as follows: A=0 refers to the financial crisis of the company, and A=1 means the company's financial normal. B represents the probability of the financial crisis of the A=0 company, and B1= has no financial crisis probability (of which B+B1=1).

Firstly, the parameters of the model are set, and the company's loss and non-loss are respectively represented by 0 and 1 respectively. The probability of the loss of the company was taken as the dependent variable, and the five factor variables selected by the factor analysis model were independent variables, and the Backward method was used to return 120 samples in 2,3 and 4 years.

$$\mathrm{F} = \frac{\sum_{j=1}^{J} N_j \left( \overline{X_J} - \overline{\overline{X}} \right)}{\sum_{j=1}^{J} \left( N_j - 1 \right) S_j^2 / \left( N - J \right)} \quad (9)$$

$$\mathrm{g}(X) = \beta_0 + \beta_1^0 X_{1j} + \ldots + \beta_k^0 X_{kj} \quad (10)$$

### 3.4 The results of the analysis

Relevant personnel will determine the specific interpretation and evaluation method based on the success or failure of the mining operation results. However, in the process of using data, analysts would like to be able to evaluate the results so that the data can be successfully applied. In the process of mining data, the relevant personnel should consider the following problems: the operational model is better than the model on the data set, and the accuracy of the model is better than the accuracy of other models. The model is mainly based on the sample model, but the actual process is far from the time difference.

### 3.5 Knowledge application

Before the application of data mining, it must be approved by the business personnel. The structure of data mining is combined with experts in various fields to get a program that is suiTable for anyone to apply. In order to get a program that is applicable to anyone, it is necessary to use the technology of application mining. Only in this way can the results of data mining be evaluated accurately. The structure obtained from the analysis is input into the business information system, so that data mining can be applied to the actual process, and it can be applied to the actual management decisions. Whether a model has meaning is judged from three aspects. It is easy for customers to understand and determine the effectiveness and potential value of new data and test data. In addition, there are some objective standards, and these standards are drawn up. In addition, there are some objective criteria. To achieve the above points, the validity of the model is available.

## 4. The case study

In the following two well-known Real Estate Company as an example, the two companies data from Chinese Securities Star Website and info website. In the evaluation of financial status indicators, it involves a comprehensive field, such as earnings per share, asset return, operating income growth rate, asset turnover ratio, liquidity ratio, quick ratio, cash liability ratio, market rate and market sales rate. In the selection of variable index, the company's debt paying ability, profitability, market operation ability and growth rate are the main factors. A total of five financial indicators, namely the company's solvency, profitability, operating cash flow, and the development ability, need to examine 14 items. In the selection of sample data, two typical real estate listed companies, A Company and B company, are selected. Compared with the indexes of earnings per share, net income and capital accumulation fund, B Company is higher than A Company. In addition, B Company is also significantly higher than A Company in the comparison of liquidity ratio, speed ratio, inventory turnover rate and the net profit rate of enterprises in the current period. In terms of the asset and liability ratio, A has reached 60%, while B is 40.3% (see Table 1).

Table 1. Financial indicators affected financial distress

| Code | Financial indicators | p |
|---|---|---|
| D1B | Profit before tax to own funds | <0.001 |
| D1A | Return on equity | <0.001 |
| D1F | Cumulative profitability ratio | =0.001 |
| B12 | Short-term liability to total loans | =0.001 |
| B1 | Total loans to total assets | =0.0230 |
| D2F | Interest expenses to net sales | =0.0011 |
| B9 | Fixed assets to long term loans + Own funds | =0.0027 |
| B5 | Long-term liabilities to total liabilities | <0.0001 |
| D2B | Cross profit to net sales | =0.0332 |
| B13 | Bank loans to total assets | =0.0012 |
| B6 | Inventory dependency ratio | <0.0001 |
| C7 | Own funds turnover | =0.0432 |
| A8 | Short-term receivables to total assets total assets | =0.0121 |
| D2E | Operating expenses to net sales | =0.0149 |
| C2 | Receivables turnover | <0.0001 |

Through correlation analysis, we can more accurately find various factors that really affect the financial risk of enterprises. By comparing with counterparts, we can find out the problems in time. For example, we analyze the company's internal financial information under the strategy of mining association rules, and set different support thresholds and confidence thresholds according to the actual situation of the company and industry, so we can find out the anomalies in financial indicators more accurately. By comparing with peers, enterprises have realized the reasonable classification of test data, which helps to improve the importance and feasibility of the whole early warning system in the classification and decision-making of financial status. It can be seen that the company turnover ratio, inventory turnover rate, asset liability ratio and net sales profit are all higher than 50%, indicating that the correlation between financial indicators is strong. The two companies have the highest support and confidence in inventory turnover, which shows that the emergence of the company's financial risk is closely related to the inventory turnover. According to the data from the inspection, the relevant personnel can make further decisions based on the success or failure of mining operation results, and make specific explanations and evaluations for them. There is a comprehensive and accurate evaluation of the results obtained, and it can be used to predict the trend of development in the future. In this part, if the financial risk analysis of data obtained the expected effect (success), for the company's financial department staff, the next step is to mining analysis, financial early warning model itself, namely the data inspection and applied to the operation of the model, discusses analyzes its really better than the data set on the model, at the same time also need to analysis from the test results, the accuracy of the model is better than other models.

## 5. Conclusion

This paper, by using data mining to anticipation of enterprise financial risk, got good results, prediction, and both the data accuracy is timeliness, distance nearer the time its accuracy is higher, conversely, the lower the accuracy. This indicates that the financial crisis is a gradual process, and the model has good use prospects in the forecast of 2-3 years. This also makes the company understand that financial crisis is a gradual process. Logistic regression model is adopted in this paper. The model has its own characteristics, and there is no multi-model or mixed use, so it has its limitations. This technology is widely used and has a wide range of theoretical knowledge. Therefore, many models and theories should be used in future research to obtain a better model.

Most of the existing property crisis early warning models are based on data mining. Data mining is to extract useful, unknown and potentially valuable information from a large number of uncertain, incomplete and clear actual data. When data mining is used to analyze financial risks, it will go through the following steps: mining objects, data preparation, building models, data mining and result analysis. Only by doing a good job, can the financial risk be accurately applied. In a word, the establishment of an early warning system can provide more predictability for investors and guide the scientific and feasible decisions of financial projects at a time. For the staff of the finance department, the most important thing is to reflect the real value information of the company, which is a warning for the leaders and investors to get into the crisis earlier.

## References

[1] Amran, A., Manaf Rosli Bin, A., & Che Haat Mohd Hassan, B. (2008). Risk reporting: An exploratory study on risk management disclosure in Malaysian annual reports. Managerial Auditing Journal, 24(1), 39-57.

[2] Chen, W. S., & Du, Y. K. (2009). Using neural networks and data mining techniques for the financial distress prediction model. Expert systems with applications, 36(2), 4075-4086.

[3] Geng, R., Bose, I., & Chen, X. (2015). Prediction of financial distress: an empirical study of listed chinese companies using data mining. European Journal of Operational Research, 241(1), 236-247.

[4] Gray, G. L., & Debreceny, R. S. (2014). A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. International Journal of Accounting Information Systems, 15(4), 357-380.

[5] Griffin, J. J., & Mahon, J. F. (1997). The corporate social performance and corporate financial performance debate: Twenty-five years of incomparable research. Business & society, 36(1), 5-31.

[6] Han, J., & Kamber, M. (2011). Data mining: concepts and techniques. Data Mining Concepts Models Methods & Algorithms Second Edition,5(4), 1 - 18.

[7] Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. Expert systems with applications, 32(4), 995-1003.

[8] Koyuncugil, A. S., & Ozgulbas, N. (2012). Financial early warning system model and data mining application for risk detection. Expert Systems with Applications, 39(6), 6238-6253.

[9] Miller, K. D., & Bromiley, P. (1990). Strategic risk and corporate performance: An analysis of alternative risk measures. Academy of Management Journal, 33(4), 756-779.

[10] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559-569.

[11] Ravisankar, P., Ravi, V., Rao, G. R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. Decision Support Systems, 50(2), 491-500.

[12] Spira, L. F., & Page, M. (2003). Risk management: The reinvention of internal control and the changing role of internal audit. Accounting, Auditing & Accountability Journal, 16(4), 640-661.

[13] Sung, T. K., Chang, N., & Lee, G. (1999). Dynamics of modeling in data mining: interpretive approach to bankruptcy prediction. Journal of management information systems, 16(1), 63-85.

[14] Wang, X. (2017). Research on the analysis of internet financial risk based on data mining. Revista De La Facultad De Ingenieria, 32(3), 720-729.

[15] Witten, & Frank, I. H. (1999). Data mining. Practical Machine Learning Tools & Techniques with Java Implementations, 13(4), 1-1.