

Research Article

Female Employment Data Analysis Based on Decision Tree Algorithm and Association Rule Analysis Method

Hong Liu ¹ and Junxia Liu²

¹School of Economics and Finance, Xi'an Jiaotong University, Xi'an 710061, Shaanxi, China

²Business School, Xi'an Fanyi University, Xi'an 710105, Shaanxi, China

Correspondence should be addressed to Hong Liu; liuhong918@stu.xjtu.edu.cn

Received 11 October 2021; Revised 27 October 2021; Accepted 21 December 2021; Published 15 February 2022

Academic Editor: Rahman Ali

Copyright © 2022 Hong Liu and Junxia Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Improving the employment rate and quality of women's employment has been a hot spot of social concern in recent years. Traditional female employment information has not been fully utilized; only simple storage and query functions have been completed. In order to mine and use female employment information, we find the rules that exist in it, and to better understand the current female employment problems, this article focuses on the data mining of female employment information data. This paper introduces in detail the current employment trends of women and the advantages of decision tree algorithms, the hot content and key technical points of data mining technology, the research progress of this technology at home and abroad, etc. Decision tree analysis algorithm for association rules: According to the formula calculus, the advantages of the decision tree algorithm and the association rule analysis method are displayed more clearly. The model of female employment information is mainly constructed, and the process of data collection and data model establishment is introduced. According to a set of female employment information data, the algorithm is analyzed through the decision tree analysis algorithm based on association rules, and the reliability of the algorithm is verified. Finally, the whole article is summarized, and its basic contents are explained. The decision tree analysis method based on association rules can effectively dig out the information hidden in the female employment information data, and it has an important role in guiding the self-planning of college women. This algorithm has the characteristics of high operating efficiency and stable results, and it also has a positive effect on other types of data mining.

1. Introduction

In the past few decades, with the development of social economy, the transformation of the industrial structure, and the awakening of female consciousness, more and more women have participated in the labor market, as shown in the "2012 World Development Report: Gender Equality and Development." A young woman in Rafah said: "I think a woman must receive education and must work so that she can prove herself to the society and become a better mother" [1]. In terms of the world, women have been made in employment with great progress, but despite this, gender discrimination in employment is still widespread. To this end, the author commented on the "The World's Women 2015: Trends and Statistics" published by the United Nations

Statistics Division [2], the WDI database published by the World Bank [3], and the "Women at Work: Trends 2016" published by the World Labor organization. Relevant data have been statistically analyzed to explore the current status and trends of women's employment worldwide, to better understand the status quo of women in the world labor market, and to provide references for promoting women's social labor participation.

This article intends to construct a female employment data analysis system based on a large amount of actual female personal information and employment data using the decision tree analysis algorithm and the association rule analysis method in traditional data mining methods. This article intends to use the decision tree algorithm because it has the following advantages:

- (1) The analysis results are easy to present and understand. Because the final analyst of the system often does not have a deep understanding of the meaning of the data, it is difficult to understand the results of the analysis through professional vocabulary. Therefore, the presentation form of data mining analysis results is very important. The decision tree represents the final classification result in a graphical tree structure, and the generated rules are described in the form of IFELSE, which is in line with people's logical thinking for analyzing problems.
- (2) The amount of calculation is small, and the efficiency is high. The decision tree method mainly uses operations such as search, filter, addition, and comparison, and its computational complexity is relatively small compared to other methods. In this way, the efficiency of the entire system will be higher.
- (3) The role of key attributes can be presented. In the constructed decision tree system, a high-level decision tree node means that the attribute represented by the node has higher importance. On the contrary, the attribute represented by a low-level node has a lower effect on classification. Two sections at the same level have the same importance.

Based on the above advantages, in the female employment data analysis system based on the decision tree algorithm and the association rule analysis method studied in this article, decision tree technology and the association rule analysis method will be used to establish an analysis model and mine employment-related data for women. By studying data such as women's personal information and the attributes of the final employment unit, it is possible to dig out important information and information with high guiding value.

2. Related Work

Data mining is a nontrivial process of extracting effective, novel, and potentially useful knowledge from complex data. In actual situations, data have the characteristics of large quantity, noise, incompleteness, vagueness, and high randomness. Data mining, also known as "Knowledge Discovery in Database," is a multidisciplinary field. Data mining integrates the research results of the latest technologies such as database technology, artificial intelligence, machine learning, statistics, knowledge engineering, and information retrieval, and its applications are very wide. As long as there is an analytically valuable database, data mining can be used to mine useful information [4]. At present, the latest research on foreign data mining technology focuses on discovering knowledge. Bayesian analysis methods and boosting methods are getting more and more attention; at the same time, the algorithm combining KDD is also widely used [5]; the statistical regression method is a traditional data mining method, and the effect is obvious after being applied in the KDD analysis method [6]. In terms of application, it is mainly reflected in the conversion of the KDD method from solving a single problem to solving a class of

problems. The United States is a leader in global data mining technology research and occupies the core technology of research [7]. Since data mining software has driven the development of the market and unearthed the needs of the market, the need for data mining software is also very large. Many well-known international companies have set up data mining departments one after another, increasing their investment in the field of data mining. So far, a series of data mining software with mature technology and high application value has appeared in the market. The following are currently the most important data mining software:

2.1. Knowledge Studio. A database mining tool was developed by the Angoss Software Company, which can import external models and rules into the system. The biggest advantages: fast response, fast calculation, high readability of models and supporting documents, and easy to add new algorithms.

2.2. IBM Intelligent Miner. The software developed by IBM can complete the selection of data, transform according to requirements, and dig out the hidden information of the data according to the purpose.

2.3. SPSS Clementine. SPSS has a long history and is an earlier analysis software. It can convert data into an easy-to-understand graphical interface, and it can also be stably combined with neural network algorithms [8]. The software introduces many new concepts of data mining, laying the foundation for the process of data mining [9].

Since the domestic research on data mining started late, it is not very mature and is still in the developing stage. The latest developments include as follows: in the research of classification and recognition, trying to establish the theory of classification according to the set to realize the processing of large databases; fusing rough set and fuzzy theory together, which is very suitable for discovery recognition [10]. In China, major universities, research institutes, and scientific research companies are the main drivers of data mining technology development, and they have made great progress in many fields [11]. From the current point of view, several research hot spots of data mining mainly include Website Data Mining, Bioinformatics/Genomics data mining, and textual data mining (Textual Mining) [12]. In short, data mining is a powerful tool. It needs to generate a data model under the guidance of existing data, and the obtained model must be verified in real life [13]. Data analysts must know what the principles of the selected data mining algorithm are and how they work, and must have a deep understanding of the domains that are expected to solve the problem, understand the data, and understand the process. Only in this way can they finally explain the results obtained and make data mining truly meets the needs of people in the information age and serves the society [14]. Data Mining has actively been used since long in different fields, e.g., healthcare [15, 16], research [17], social communication [18], and education [19]. This article focuses on

the analysis of female employment data. Due to the high correlation of female data, the risk of adopting the Bayesian analysis method is higher, so the decision tree algorithm is adopted. But for each sample of female data, there are many attributes. How to choose these attributes for decision making is the focus of analysis. In this paper, a decision tree algorithm based on association rules is used. On the basis of the decision tree algorithm, the association rules are used to select the attributes of female employment data for decision making.

3. Methods

3.1. Main Methods of Data Mining. Data mining technology has developed rapidly in recent years, and the corresponding technical methods have also been expanded and improved in many ways, forming various data mining methods. Various data mining methods can be divided into the following categories: clustering algorithms, classification algorithms, decision trees, and so on. Table 1 describes the functional characteristics and application fields of various data mining algorithms. This section focuses on clustering analysis algorithms and association rule algorithms [20].

3.2. Decision Tree Algorithm. A decision tree algorithm is an efficient and accurate classification method. It can construct a decision tree representation and corresponding classification rules from a large amount of unordered and irregular data, and is widely used in the field of data mining [21]. The origin of the decision tree method is the concept learning system (CLS), which was subsequently developed to the ID3 method through the improvement of scholars in the corresponding field and finally evolved into the C4.5 algorithm, which has the ability to deal with continuous attributes [22]. Other typical decision tree methods include CART, SLIQ, and SPRINT, and have a certain degree of application in other fields [23]. This article focuses on the CLS algorithm and the ID3 algorithm, and compares and analyzes the advantages and disadvantages of these algorithms. And for the data mining system, the appropriate decision tree algorithm is selected [23].

3.3. CLS Algorithm. The CLS algorithm is the original algorithm for decision tree learning, and other subsequent decision tree learning algorithms are developed on the basis of the CLS algorithm. The idea of the CLS algorithm is as follows: the database to be processed is selected, and one of the attributes is selected as the test attribute. This attribute is mapped to the decision node in the decision tree [24]. Different attributes can correspond to their respective subsets. If the subset has no content, or the samples of the subset belong to the same category, then it is taken as a leaf node. Otherwise, a classification attribute is reselected, and the subset is divided [24]. The sign of the end of the decision tree construction is as follows: all the subsets are empty; all the subsets have belonged to the same category. For example, the relationship between a person's hair color, eye color, and ethnicity is shown in Table 2.

According to the information provided in Table 2, if "eye color" is the test attribute, the subset that can be obtained by applying the decision tree algorithm is shown in Figure 1.

Furthermore, "hair color" is selected as the test attribute, and the subset is obtained, as shown in Figure 2. Since the final samples are all of the same type, the decision tree is constructed.

3.4. ID3 Algorithm. The ID3 algorithm is a widely used decision tree algorithm, which has the characteristics of high efficiency and accuracy [25]. That is, on the basis of CLS, the step of decision attribute selection is added. In projects with many attributes, it can be more accurate to make a decision, and measure and calculate the amount of information. The concept of the amount of information was proposed by Shannon in 1948. Its formulas are introduced as follows [25]:

The amount of information Ia_i of an event a_i can be measured using the following formula:

$$I(a_i) = p(a_i) \log_2 \frac{1}{p(a_i)}, \quad (1)$$

where $p(a_i)$ is the probability of event a_i .

If there are n irrelevant events represented as a_1, a_2, \dots, a_n , then the average amount of information is represented as follows [26]:

$$\begin{aligned} I(a_1, a_2, \dots, a_n) &= \sum_{i=1}^n I(a_i) \\ &= \sum_{i=1}^n p(a_i) \log_2 \frac{1}{p(a_i)}. \end{aligned} \quad (2)$$

In the decision tree classification system, suppose S is the training sample set and $|S|$ is the number of samples in the training sample set. The sample is divided into n different classes C_1, C_2, \dots, C_n , and the sizes of these classes are marked as $|C_1|, |C_2|, \dots, |C_n|$. Then, we get the probability that the sample S is of the C_i class, expressed by the following formula [26]:

$$p(S_i) = \frac{|C_i|}{|S|}. \quad (3)$$

Attribute A contains a number of attribute values, which can be described as $X(A)$ in the form of a set. We take the sample subset whose value is v and mark it as S_v . Then, on the branch node after the attribute A is selected, the entropy of the node's sample set S_v classification is $E(S_v)$ [26]. In order to obtain the expected entropy value caused by A , the weighted sum of the entropy of each subset S_v is calculated. The weight value belongs to the proportion of S_v in the original sample S , that is, $|S_v|/|S|$. Therefore, the expected entropy of A can be described as follows:

$$\text{Entropy}(S, A) = \sum \left(\frac{|S_v|}{|S|} \right) * \text{Entrpoy}(S_v). \quad (4)$$

Then, the information gain value $\text{Gain}(S, A)$ of attribute A for the original sample set S is defined as follows [27]:

TABLE 1: Comparison of various data mining algorithms.

Technical methods	Main functions and characteristics	Application areas
Decision tree	Inductive classification; comprehensibility	Telecommunications, medicine, retail, etc.
Genetic algorithm	Clustering and optimization; efficiency	Finance, insurance, agriculture, etc.
Rough set	Uncertainty classification	Retail, financial, manufacturing, etc.
Neural networks	Prediction, classification, and clustering; poor interpretability	Telecommunications, insurance, manufacturing, etc.
Bayesian network	Classification, clustering, and prediction; easy to understand	Medicine, manufacturing, telecommunications, etc.
Association rules	Classification and clustering	Retail, insurance, manufacturing, etc.
Statistical analysis	Clustering; accurate results; easy to understand	Finance, manufacturing, medicine, etc.
Support vector machines	Classification; small error	Medicine, telecommunications, and financial industries, etc.

TABLE 2: Person's eye color, hair color, and race.

Personnel	Eye color	Hair color	Race
1	Black	Black	Asian
2	Blue	Gold	White people
3	Grey	Gold	White people
4	Blue	Red	White people
5	Grey	Red	White people
6	Black	Gold	Mixed race
7	Grey	Black	Mixed race
8	Blue	Black	Mixed race

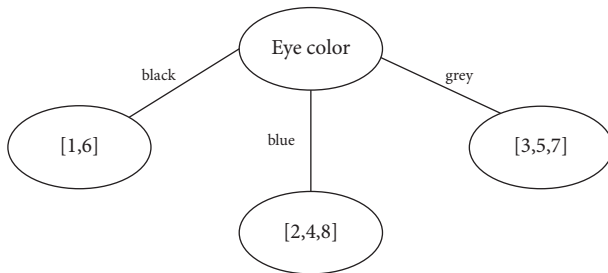


FIGURE 1: Preliminary construction of decision tree.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \text{Entropy}(S, A). \quad (5)$$

Among them, $\text{Gain}(S, A)$ refers to the expected compression of entropy due to the selection of attribute A [27]. The greater the $\text{Gain}(S, A)$, the more information provided by the selection of test attribute A for classification. The ID3 algorithm is to select the attribute with the largest information gain $\text{Gain}(S, A)$ for each node as the test attribute.

3.5. Association Rule Analysis Algorithm. Association rules can dig out the connections in the data, which is an important analysis method in data mining algorithms and is widely used in various fields [28]. Shopping basket analysis is a case of the outcome of association rules. That is, a successful case of customer shopping habits is obtained by mining the relevant data of customer shopping habits [26]. The basic concepts of association rules are as follows:

Suppose I is a collection of items, defined as $I = \{I_1, I_2, \dots, I_n\}$. Task-related data D is a collection of transactions (or tuples). Transaction T is a collection of

items, and each transaction has a transaction identifier $T_i d$. If the item set A is a subset of T , and $T_i d$ is a transaction, then the related concepts are defined as follows [28]:

Support count: the number of occurrences of an item set is the number of transactions that contain the item set in the entire dataset [28].

The association rule is defined as follows: if $A \subset I$, $B \subset I$, $A \cap B = \phi$, $A \Rightarrow B$, then A and B are considered to have an association relationship [28].

Support: in D , the support of an association rule refers to the probability that D contains A and B [29].

$$\begin{aligned} \text{Support}(A, B) &= P(A \cap B) \\ &= \text{support_count}(A \cap B) \\ &= \text{count}(T). \end{aligned} \quad (6)$$

Confidence degree c : it refers to the conditional probability of both A and B [29].

$$\text{confidence}(A, B) = P(B|A). \quad (7)$$

The most commonly used algorithms for association rules are the a priori algorithm and FP-Tree algorithm.

This article focuses on the analysis of female employment data. Due to the high correlation of female data, the risk of adopting the Bayesian analysis method is higher, so the decision tree algorithm is adopted. But for each sample of female data, there are many attributes. How to choose these attributes for decision making is the focus of analysis. In this paper, a decision tree algorithm based on association rules is used. On the basis of the decision tree algorithm, the association rules are used to select the attributes of female employment data and are used in decision making.

4. Experiments and Discussions

4.1. Data Model. The data model we used in this research work consists of the attributes of the selected female colleges that include the following, and the experimental data are shown in Table 3.

- (1) Degree, which represents the level, e.g., undergraduate and masters
- (2) Type of school, i.e., up to which level the specific school provides education

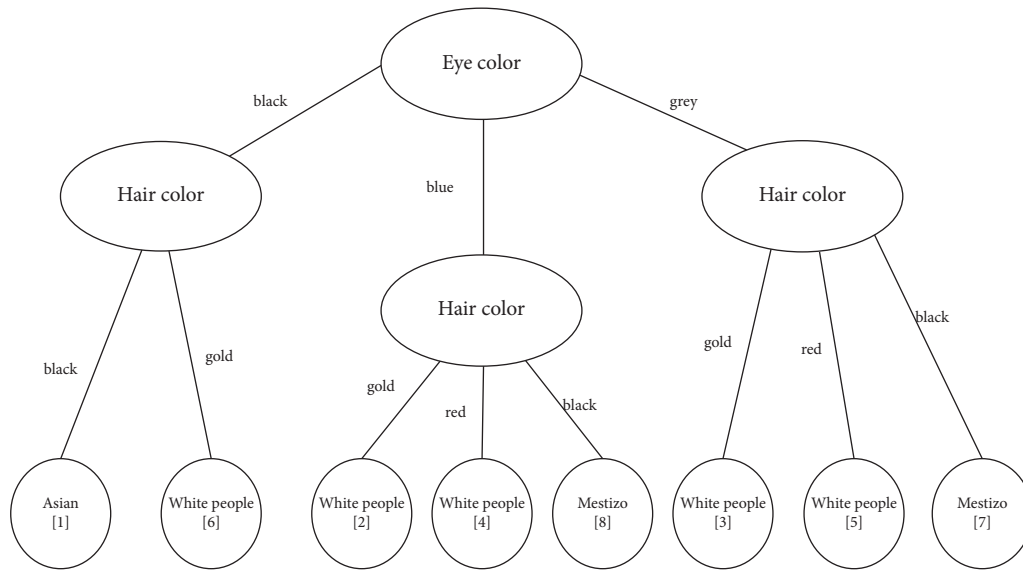


FIGURE 2: Decision tree.

TABLE 3: Experimental data.

Serial number	Degree	Type of school	Professional type	Learning situation	Type of company	Salary grade	Type of job
1	Master	985	Telecommunication	Excellent	Private enterprise	1	Research
2	Bachelor	985	Telecommunication	Excellent	Foreign company	2	Maintenance
3	Bachelor	211	Telecommunication	General	Private enterprise	2	Sales
4	Bachelor	Undergraduate	Telecommunication	General	Private enterprise	3	Sales
5	Master	Undergraduate	Civil engineering	Excellent	Private enterprise	1	Research
6	Master	Undergraduate	Medical	General	Foreign company	1	Research
7	Master	985	Medical	General	State-owned enterprise	2	Maintenance
8	Bachelor	985	Telecommunication	General	Foreign company	1	Sales
9	Bachelor	985	Medical	General	Private enterprise	3	Research
10	Master	Undergraduate	Medical	Excellent	Foreign company	1	Sales
11	Bachelor	Undergraduate	Civil engineering	Excellent	Foreign company	2	Sales
12	Bachelor	985	Civil engineering	General	State-owned enterprise	3	Maintenance
13	Master	Undergraduate	Telecommunication	General	State-owned enterprise	2	Research
14	Bachelor	Undergraduate	Civil engineering	General	Private enterprise	4	Maintenance
15	Bachelor	Undergraduate	Medical	General	Private enterprise	4	Maintenance
16	Bachelor	Undergraduate	Civil engineering	General	Foreign company	4	Research
17	Bachelor	985	Telecommunication	Excellent	State-owned enterprise	3	Sales
18	Bachelor	985	Telecommunication	Excellent	State-owned enterprise	3	Research
19	Master	211	Civil engineering	General	State-owned enterprise	2	Maintenance
20	Bachelor	Undergraduate	Medical	General	Private enterprise	3	Maintenance
21	Bachelor	Undergraduate	Civil engineering	General	Foreign company	4	Research
22	Master	Undergraduate	Civil engineering	Excellent	Private enterprise	1	Sales
23	Master	Undergraduate	Telecommunication	Excellent	Private enterprise	1	Sales
24	Master	211	Civil engineering	Excellent	State-owned enterprise	2	Research
25	Bachelor	Undergraduate	Medical	General	Foreign company	1	Research
26	Bachelor	211	Civil engineering	Excellent	Foreign company	3	Research
27	Bachelor	985	Telecommunication	General	State-owned enterprise	3	Maintenance
28	Bachelor	Undergraduate	Medical	Excellent	Foreign company	2	Sales
29	Bachelor	Undergraduate	Telecommunication	Excellent	Private enterprise	4	Research
30	Bachelor	Undergraduate	Civil engineering	Excellent	Foreign company	4	Research

(3) Professional type, which refers to the specialization area

(4) Learning situation, which could be excellent, good, ordinary, or poor, etc.

(5) Salary grade, which refers to the pay-scale distribution of the employees. Higher grade employee will get higher salaries.

- (6) Type of job, which may be teaching, research, office work, or engineering and designing, etc.
- (7) School area
- (8) Academic performance, which may be excellent, ordinary, or poor
- (9) Scientific research ability, which may be excellent, ordinary, or poor
- (10) Community ability
- (11) Interview skills, i.e., how well a candidate can face interviews
- (12) Family background

In female attributes, it includes “school type,” “professional type,” “study performance,” and “society performance.” Enterprise attributes include “enterprise type,” “salary grade,” and “position type.” If we think positively, we are concerned about the following questions: what kind of influence will some attributes of women, such as “school type,” have on their employment? Conversely, if the goal of women is to enter a certain attribute unit, such as entering a state-owned enterprise, it is necessary in what ways? Regarding these two issues, a conclusion can be drawn through the decision tree analysis algorithm based on association rules.

4.2. Experiment Procedure

4.2.1. Data Preprocessing Algorithm. The database we got is based on subjective description, close to human-thinking mode, but it is not conducive to algorithm processing and application. Therefore, we must first undergo preprocessing to convert it into a data form that can adapt to the algorithm model. Note that under normal circumstances, the data also need to be cleaned and denoised, as it is not the focus of this article, so it is omitted, which is shown in Tables 4 and 5.

In this way, we converted the original database into a data format that is easy for algorithmic analysis and completed the preprocessing of the data.

4.2.2. Data Integration. In the decision tree about the attributes of the enterprise type, it is mainly applied to I1, I3, and I4. Through the preprocessing method, we can use its inverse process to understand that I1, I3, and I4 actually describe two attributes: degree information and school attributes. Then, I1 indicates whether the degree is an undergraduate or a master; I3 and I4 describe whether the school type is an ordinary undergraduate or a 211 undergraduate. In the same way, I12 describes the ability of the community, and I10 describes the academic performance.

According to the decision tree, we can generate classification rules:

- (i) Regarding the type of business, the rules are as follows:
- (ii) IF “Degree” = “Master” THEN
- (iii) IF “School Type” = “211 Undergraduate” THEN
- (iv) “Unit Type” = “State-owned Enterprise”
- (v) ELSE “School Type”-“General Undergraduate” THEN
- (vi) “Type of Unit”-“Private Enterprise” or “State-owned Enterprise”
- (vii) ELSE
- (viii) IF “School Type” = “211 Undergraduate” THEN
- (ix) “Type of Unit”-“State-owned Enterprise”
- (x) ELSE
- (xi) “Type of Unit,” “Private Enterprise” or “Foreign Enterprise”

4.2.3. Applying Decision Tree Algorithm. For the attribute of enterprise type, the decision tree is shown in Figures 3–5:

4.3. Experimental Results. Through the classification rules generated above, the following experimental conclusions can be obtained:

- (1) First, the types of enterprises employed by women are mainly related to the types of degrees and schools. Specifically, graduates from master’s degrees and 211 undergraduate schools have a greater chance of working in state-owned enterprises.
- (2) Those who graduated from undergraduate degrees and ordinary undergraduate schools are more likely to be employed in private enterprises and foreign companies.
- (3) Looking at it conversely, women’s degrees, school types, academic performance, and club abilities have a greater impact on employment. Among them, state-owned enterprises and foreign companies pay more attention to the type of graduate schools.
- (4) The academic performance and the ability of the club have a greater impact on the job category.
- (5) Through experimental analysis and experimental results, we can see that the analysis method based on decision trees and association rules has completed the analysis of female employment data and has unearthed the more valuable logical relationships within the data. Through the discovery of the results, colleges and universities can plan their teaching priorities, and women can shape themselves in a targeted manner. It can be seen from the experimental results that the experimental results are consistent with the actual phenomenon.

TABLE 4: Valid data when party R1R2R3 = 100.

Number	Degree			Type of school			Professional type			Learning situation			Comprehensive ability			Type of company		
	11	12	13	14	15	16	17	18	19	110	111	112	R1	R2	R3			
7	0	1	0	1	1	0	0	1	1	0	0	1	1	0	0			
12	1	0	0	1	1	0	1	0	0	1	0	1	1	0	0			
13	0	1	0	1	0	1	0	0	0	1	0	1	1	0	0			
17	1	0	0	0	1	1	0	0	0	1	1	0	1	0	0			
18	1	0	0	1	1	1	0	0	1	0	1	0	1	0	0			
19	0	1	0	1	0	0	1	0	0	1	0	1	1	0	0			
24	0	1	0	1	0	0	1	0	0	1	1	0	1	0	0			
27	1	0	0	1	1	1	0	0	0	1	0	1	1	0	0			

TABLE 5: Valid data when party R1R2R3 = 001.

Number	Degree			Type of school			Professional type			Learning situation			Comprehensive ability			Type of company		
	11	12	13	14	15	16	17	18	19	110	111	112	R1	R2	R3			
2	1	0	0	1	1	1	0	0	1	0	0	1	0	0	1			
6	0	1	1	0	0	0	0	1	0	1	0	1	0	0	1			
8	1	0	0	0	1	1	0	0	1	0	0	1	0	0	1			
10	0	1	1	0	0	0	0	1	0	1	1	0	0	0	1			
11	1	0	1	0	0	0	1	0	0	1	1	0	0	0	1			
16	1	0	1	0	0	0	1	0	1	0	0	1	0	0	1			
21	1	0	1	0	0	0	1	0	1	0	0	1	0	0	1			
25	0	1	1	0	0	0	0	1	0	1	0	1	0	0	1			
26	1	0	0	1	0	0	1	0	1	0	0	1	0	0	1			
28	1	0	1	0	0	0	0	1	1	0	1	0	0	0	1			
30	1	0	1	0	0	0	1	0	1	0	0	1	0	0	1			

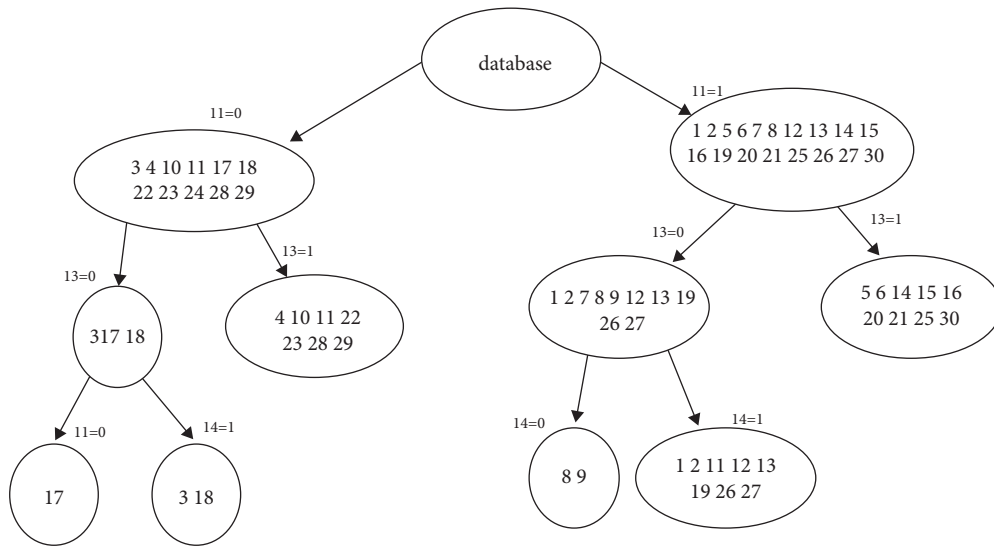


FIGURE 3: Company type decision tree.

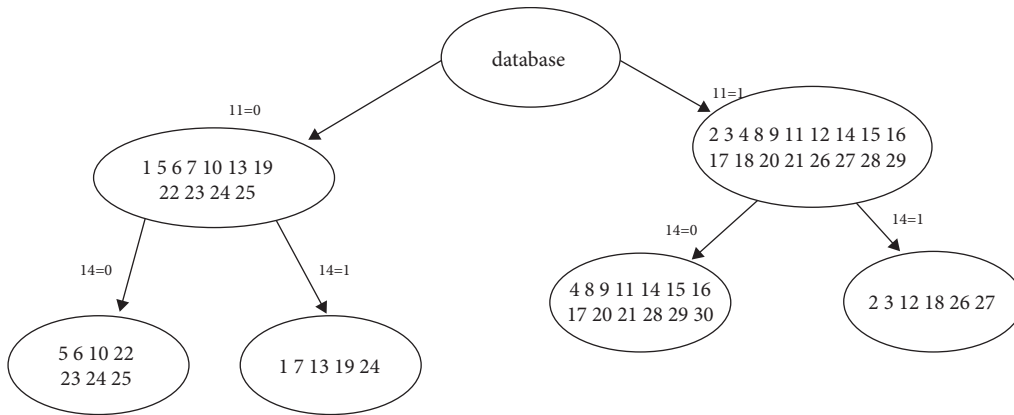


FIGURE 4: Salary grade decision tree.

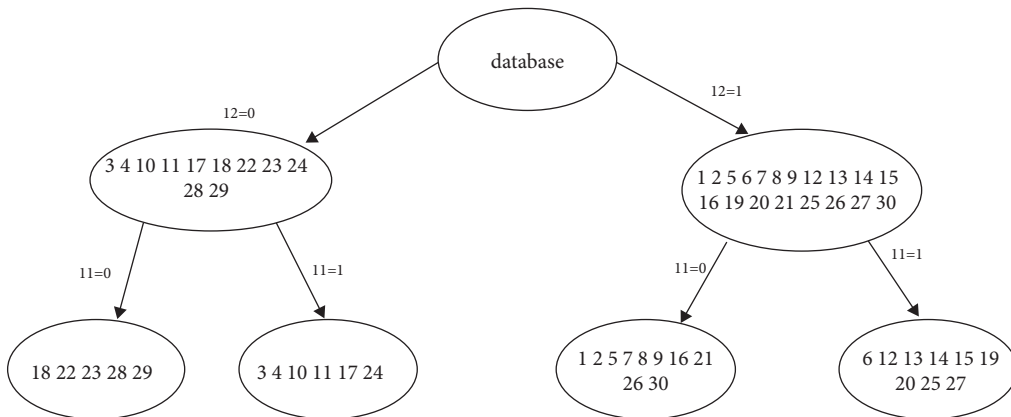


FIGURE 5: Job type decision tree.

5. Conclusions

With the transition from elite education to mass education in my country's higher education, the number of college enrollment is huge, and the corresponding number of fresh graduates is increasing. It is increasingly difficult for women in colleges and universities to obtain employment, and the employment situation is very severe [27]. How to guide women to understand the employment situation and master relevant skills in advance is very important. However, the processing of these data in major universities is still at the stage of primary data backup, query, and simple statistics, and there is no in-depth analysis of a large amount of score data, and there is no in-depth digging to find useful value to guide future education through teaching. This is a great waste of teaching information resources. The data mining technology of the decision tree analysis algorithm based on association rules is a feasible and effective method to solve this problem [28].

In this context, this article briefly describes the current status of data mining research at home and abroad and explains the importance of the selected mining method in the research data mining and completes the following content:

- (1) Data mining technology has developed rapidly in recent years, and the corresponding technical methods have also been expanded and improved in many ways, forming various data mining methods. Various data mining methods can be divided into the following categories: clustering algorithms, classification algorithms, decision trees.
- (2) Aiming at the characteristics of the women employment information database with more attributes and more complex results, this article combines association rules and decision tree analysis methods to effectively solve the problem of information mining in large databases. In this part, the algorithm principle and detailed steps of the decision tree analysis algorithm based on association rules are analyzed in detail.
- (3) A complete database model and an algorithm model are established. The steps and key points of the algorithm are elaborated in detail. And through a set of real employment information data, the algorithm's processing flow is introduced in detail. Finally, the experimental results are explained, which fully verified the reliability of the algorithm.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The study was supported by (1) Shaanxi Social Science Fund, the Research on the Integration and Innovation of "Rural-commercial-Culture-Sports and Tourism" in the Northern Foot of Qinling Mountains under the Rural Revitalization, (2020D041), (2) The Youth Innovation Team of Shaanxi Universities, Research on Modern Service Trade and Shaanxi Portal Economy Development under the "One Belt and One Road" (Initiative, No. 201990), and (3) the Research Team of High-quality Development and High-level Opening of Modern Service Industry in Shaanxi Province (XFU21KYTDA01).

References

- [1] S. Q. Shan, *Data Mining-Concepts, Models, Methods and Algorithms*, Tsinghua University Press, Beijing, China, 2004.
- [2] X. B. Zhang, "Decision tree algorithm and its core technology," *Computer Technology and Development*, vol. 23, no. 2, pp. 22–29, 2007.
- [3] M. Fan, *Data Mining---Concept and Technology*, Machinery Industry Press, South Norwalk, CT, USA, 2001.
- [4] W. D. Li, "Use data mining methods to analyze customer career value," *Computer Engineering and Applications*, vol. 34, no. 2, pp. 23–45, 2005.
- [5] H. Q. Dong, *Research on Multidimensional Data Visualization in Data Mining*, Wuhan University of Technology, Wuhan, China, 2006.
- [6] L. He and W. Lingda, "Summary of clustering algorithms in data mining," *National Defense University*, vol. 1, no. 2, pp. 12–23, 2007.
- [7] H. Scott, *Build an efficient ERP System*, Mechanical Industry Press, Beijing, China, 2009.
- [8] H. Luo and Z. M. Wang, *ERP Principle, Design And Implementation*, Publishing House of Electronics Industry, Beijing, China, 2005.
- [9] H. Ye and H. J. Chen, "Analysis on the functional features and strategic advantages of ERP system," *China Management Information*, vol. 12, no. 3, pp. 12–22, 2007.
- [10] X. M. Wen, *Data ETL Research and Prospect*, Guangzhou University of Technology, China, Guangdong, 2005.
- [11] L. J. Guan, "Research and application of data mining technology oriented to ERP," *Database and information management*, vol. 67, no. 1, pp. 67–78, 2007.
- [12] J. Y. Han, *Overview of Data Mining Technology*, Mechanical Industry Press, Beijing, China, 2005.
- [13] G. Q. Wang and D. Huang, "Data mining: concepts and techniques," *Computer Application Technology*, vol. 45, no. 2, pp. 12–23, 2007.
- [14] R. Wng, D. T. Ma, and C. Chen, "Data warehouse and data mining technology," *Computer Application Technology*, vol. 56, no. 2, pp. 52–56, 2007.
- [15] R. Ali, M. H. Siddiqi, M. Idris, B. H. Kang, and S. Lee, "Prediction of diabetes mellitus based on boosting ensemble modeling," in *Proceedings of the International Conference on Ubiquitous Computing and Ambient Intelligence*, pp. 25–28, Belfast, UK, December 2014.
- [16] R. Ali, J. Hussain, M. Siddiqi, M. Hussain, and S. Lee, "H2RM: a hybrid rough set reasoning model for prediction and management of diabetes mellitus," *Sensors*, vol. 15, no. 7, Article ID 15921, 2015.

- [17] R. Ali, S. Lee, and T. C. Chung, "Accurate multi-criteria decision making methodology for recommending machine learning algorithm," *Expert Systems with Applications*, vol. 71, pp. 257–278, 2017.
- [18] I. Ahmed, R. Ali, D. Guan, Y.-K. Lee, S. Lee, and T. Chung, "Semi-supervised learning using frequent itemset and ensemble learning for SMS classification," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1065–1073, 2015.
- [19] C. Romero and S. Ventura, "Data mining in education," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 12–27, 2013.
- [20] B. H. Wang and Z. B. Wu, "Data mining technology and its application status," *Information Technology*, vol. 33, no. 2, pp. 4–9, 2004.
- [21] W. J. Lou, F. S. Kong, and Y. S. Cao, "Summary of knowledge discovery in database," *Information Technology*, vol. 77, no. 2, pp. 25–31, 2003.
- [22] K. S. Qu and W. L. Chen, "Research and improvement of decision tree classification algorithm," *Computer Engineering and Applications*, vol. 2, no. 25, pp. 104–105, 2003.
- [23] Q. Zhang, *An Improved Algorithm of ID3 Algorithm*, Zhengzhou University, Zhengzhou, China, 2002.
- [24] Y. H. Cao, *Research on Classification Algorithms in Data Mining*, South China University of Technology, Guangzhou, China, 2002.
- [25] X. L. Sun, *Research on Decision Tree Technology Combined with Rough Set Theory*, Jilin University, Changchun, China, 2002.
- [26] Y. Long, *Pattern Recognition*, University of Science and Technology Beijing, Beijing, China, 2002.
- [27] M. Tom, *Mitchell. Machine Learning*, Mechanical Industry Press, Beijing, China, 2012.
- [28] L. Wang, "Research and application of ID3 algorithm," *Fujian Computer*, vol. 65, no. 1, pp. 55–60, 2010.
- [29] Q. Yang, "Learning algorithm based on decision tree," *Journal of Xiangtan Normal University*, vol. 78, no. 3, pp. 14–20, 1999.