

Research Article

Optimization of Management Mode of Small- and Medium-Sized Enterprises Based on Decision Tree Model

Yuzhu Diao¹ and Qing Zhang ²

¹School of Business, Linyi University, Linyi, Shandong 276000, China

²School of Business and Economics, University Putra, 43400 Seri Kembangan, Malaysia

Correspondence should be addressed to Qing Zhang; gs57469@student.upm.edu.my

Received 6 November 2021; Revised 29 November 2021; Accepted 30 November 2021; Published 17 December 2021

Academic Editor: Miaochao Chen

Copyright © 2021 Yuzhu Diao and Qing Zhang. 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.

Decision tree algorithm is a common classification algorithm in data mining technology, and its results are usually expressed in the form of if-then rules. The C4.5 algorithm is one of the decision tree algorithms, which has the advantages of easy to understand and high accuracy, and the concept of information gain rate is added compared with its predecessor ID3 algorithm. After theoretical analysis, C4.5 algorithm is chosen to analyze the performance appraisal results, and the decision tree for performance appraisal is generated by collecting data, data preprocessing, calculating information gain rate, determining splitting attributes, and postpruning. The system is developed in B/S architecture, and an R&D project management system and platform that can realize performance assessment analysis are built by means of visualization tools, decision tree algorithm, and dynamic web pages. The system includes information storage, task management, report generation, role authority control, information visualization, and other management information system functional modules. They can realize the project management functions such as project establishment and management, task flow, employee information filling and management, performance assessment system establishment, report generation of various dimensions, management cockpit construction. With decision tree algorithm as the core technology, the system obtains scientific and reliable project management information with high accuracy and realizes data visualization, which can assist enterprises to establish a good management system in the era of big data.

1. Introduction

With the continuous development and progress of computer technology and big data technology, it provides a lot of new ideas and methods for the management of Internet enterprises [1]. In today's rapid development of the Internet, the use of Internet technology and big data analysis for scientific and effective management of enterprise affairs, increase the core competitiveness of enterprises, promote the progress of the completion of enterprise projects, improve the efficiency and motivation of employees, reduce the cost of operation, and maintenance of enterprises, to build a characteristic management system of Internet enterprises has a vital role [2]. The application of computer and big data technology to enterprise management can fully improve the management efficiency of the enterprise, and more and more Internet enterprises are using various systems for enterprise

management, such as Tencent's agile R&D system TAPD and Zen Dao system [3]. These systems can help enterprises manage projects and tasks and assign work to employees. Through these systems, projects can be managed faster and more efficiently. In business management, employee performance appraisal and project cost analysis play a key role in the development of the company [4]. A fair and just evaluation system can improve the motivation of employees, enhance the quality of project completion, and increase the creativity and responsibility of employees. The use of data mining technology, combined with the enterprise project management system, based on the information reported by employees in the system to analyze and evaluate the performance of employees, can help employees to carry out self-planning and provide reference advice for management [5].

At the same time, using project management data and performance evaluation results, the cost of the enterprise can

be analyzed, and more scientific allocation can be made according to the ability of employees to promote the scientific and sustainable development of the enterprise. R&D project management system is to manage the projects of Internet enterprises engaged in software development [6]. Through this system, project managers plan and manage the tasks involved in software development projects, organize and coordinate the life cycle of a project, and make the project complete its final goal efficiently through an effective management system [7]. In this paper, we use the decision tree algorithm in data mining to analyze the performance appraisal results generated by employees in the system and generate a decision tree to analyze the decision tree, which can help employees understand the focus of performance appraisal, make further planning for their own work, and improve work efficiency [8]. At the same time, the decision tree can directly generate the performance appraisal results of employees, which can reduce the workload of enterprise managers in the future, and also make the performance appraisal more fair and transparent, and mobilize the motivation of enterprise employees [9]. Based on the performance appraisal results, employees are rewarded, punished, and rated, and their remuneration is determined and combined with the expenses in project management, the reports of each dimension of the enterprise are generated and displayed in a visual form, which makes the costs of the enterprise clearer and helps the leadership of the enterprise to understand the situation of the enterprise, so as to formulate the next strategy and reasonably allocate the work of employees according to their abilities, thus to improve the competitiveness of the enterprise market [10]. Decision tree algorithm, as a common algorithm in data mining, is also widely used in many fields. Its development process is very long from simple to complex, from simple to deep. In 1966, Hunt developed a concept learning system for learning individual concepts, which was an early inductive learning system for decision trees [11].

Research on decision tree algorithms is continuing with the aim of improving the accuracy of decision tree algorithms and combining them with related techniques in other fields to generate more benefits. Decision tree algorithms are widely used in education, performance appraisal, research, and other fields. In China, there is a need for deeper research and development in this area. In this paper, we use the lightweight MVVM framework Vue.js, combined with C# language, SQL Server database, JavaScript assembly language, and other development technologies to develop and design the R&D project management system. Based on the C4.5 decision tree algorithm, the data set generated in the system is trained and analyzed to generate a decision tree related to performance evaluation. Through testing and analysis of the decision tree, we find out the points that the project leader pays attention to when scoring the performance appraisal and form a complete performance appraisal system based on the performance appraisal attributes in the system and the decision tree. Visualization tools are used to generate enterprise-related cost reports. Through the management of projects, tasks, and employees, the system realizes project management functions such as project

creation and management, task flow, employee information filling and management, performance appraisal system establishment, report generation of each dimension, and management cockpit construction. This paper uses the C4.5 decision tree algorithm as the core technology to obtain high-precision performance assessment results and efficient project management system, visualize cost data based on system data, and assist enterprises to establish a good management system and performance assessment system. Through these systems, projects can be managed faster and more efficiently. In business management, employee performance appraisal and project cost analysis play a key role in the development of the company.

2. Related Work

The current project management level is not high, the procedure is complicated, difficult to operate, and far from the developed countries, there are many loopholes in the market management, and the quality of the system varies, usually using C/S mode, which cannot realize the project information network query; or the project management products are too targeted, not much innovative content, and require secondary development of customized methods to meet the individual needs of enterprise project management [12]. In the project management industry, Wenpu and Huateng are leading in China; however, in the face of increasing innovation and competition, they are not flexible enough to meet their needs, and there is a certain risk of project implementation due to the instability and low technical ability of the enterprise personnel in China. For example, the project management system of financial enterprises in China is outsourced for custom development or developed in-house, and there are few ways to purchase standard software [13]. At present, the application of China's project management system is relatively backward, the large-scale use of the Internet for system management is after 2010, and at the same time, China's existing major project management software are task-centered and rarely focus on the management of staff performance and the lack of cost analysis. Therefore, the research on the R&D project management system of Internet enterprises in China is inadequate, and the way of performance assessment in China is too backward [14].

The researcher quantified a set of performance appraisal systems based on the K-means algorithm by combining decision tree algorithm and cluster analysis to realize a quantitative performance appraisal system, which can achieve a scientific and objective evaluation of the performance appraisal of the enterprise employees. The researcher researched and developed the system by using the ID3 decision tree algorithm for employee performance evaluation and introduced the concept of decision scheduling in the ID3 decision tree algorithm, which effectively reduces the complexity of the algorithm [15]. At present, China is paying more and more attention to the application method of performance appraisal and is working together to realize a fair, scientific, and effective performance appraisal system. For data mining algorithms, although China started late, in

recent years, it has made very significant achievements, in the Internet industry, financial industry, meteorological analysis, e-commerce, and other fields [16]. And major universities have also invested a lot of energy to further explore the value of the method and more in-depth study of the principles of these algorithms. For example, researchers have applied the ID3 decision tree algorithm to the human resource system to support the company's strategic decision-making, and researchers have applied the ID3 algorithm to the performance evaluation of employees in research institutes, as well as the data mining algorithm mentioned in the previous paragraph to the generation of the performance evaluation system [17].

The application and research of these algorithms have accelerated the development of project management systems in China and provided a very scientific and effective method for generating a good performance system. Based on the analysis of the application of decision tree algorithm and performance appraisal system in China, data mining algorithms have been applied to various management systems in China [18]. According to the previous research, combined with the existing project management system and performance appraisal system in China, the application of decision tree algorithm to performance appraisal can improve the fairness and efficiency of performance appraisal evaluation and help enterprises to manage better. Foreign project management software started earlier and developed very quickly. The earliest U.S. Army managed the Manhattan Project through project management technology and achieved very good results. At present, project management software in foreign countries has been introduced through algorithms and other ways to improve the traditional management methods, and through practical proof and standardization of research, the establishment of a complete project management knowledge system [19].

At present, the more popular project management software abroad is Microsoft Project, with task allocation, progress tracking, budget management, and workload analysis. In the software development industry, project management systems have long been used by some of the larger software development companies in Europe and the United States [20]. The overall development of foreign performance appraisal system is earlier than China and ahead of domestic. Foreign countries have always attached great importance to the development of the performance appraisal system, and its scientificity, fairness, and reasonableness of evaluation have formed a large gap with China. Through the method of data mining, performance appraisal is analyzed and combined with big data and the Internet, and a typical example is PeopleSoft [21]. Through comparative research, it is found that a very complete performance appraisal system has been gradually generated abroad, which has improved the essentials of performance appraisal and made the development of enterprises gradually scientific and simple. The developed countries such as Europe and the United States are making use of the increasingly developed Internet technology and big data technology to improve the performance appraisal methods with the progress of the market and to continuously improve the performance

appraisal methods according to the current corporate strategies.

3. Management Model Optimization for Small- and Medium-Sized Enterprises

3.1. Data Mining. Data mining is to analyze the actual data to obtain the hidden data that people cannot see directly, which is large, random, fuzzy, and discontinuous, and the hidden data are unknown and useful. With the development of databases, the management of data has become more and more complex, and the amount of data generated has become larger and larger. In this context, data mining technology was developed to extract information from data that we need but are difficult to find. This method is now widely used in production management, scientific exploration, market analysis, and engineering design. Data mining is a cross-disciplinary discipline involving a variety of disciplines, mainly the integration of artificial intelligence, database, statistics, visualization technology, and other disciplines to collect and mine the data to obtain some useful information, which can help decision-makers to make the right judgment and reduce unnecessary risks. The main steps are shown in Figure 1.

Data preparation is to collect and organize the information to be mined and in practice to collect the data for your own purposes or to construct your own data from the collected data sets. Data integration is the processing of the collected data according to the user's needs and understanding the characteristics of the field, mainly the missing parts of the data and the cleaning of the dirty data in the data. The next step is data selection, i.e., the selection of data in the database and the identification of the set of data to be analyzed to narrow the scope of processing and improve the quality of data mining. Data preprocessing is to clean the data through statistics, algorithm analysis, and other methods, to remove unnecessary noise data, to get the valid and standardized data set we want, to ensure the integrity and consistency of the data set. Data mining firstly determines the target, that is, the type of knowledge to be discovered, then selects a suitable data mining algorithm according to the determined target, uses the algorithm to correlate or classify the data set, extracts the relevant knowledge, and finally expresses it in some form. Finally, the extracted knowledge is analyzed, and the useful information is extracted from the knowledge and displayed through visualization tools. The main methods of data mining are classification, valuation, prediction, association rules, and clustering, where classification, valuation, and prediction are guided data mining, which can build a model that can describe specific attributes through data, and association rules and clustering analysis are unguided data mining, which use all attributes to find a certain relationship. Different data mining methods have their own data mining algorithms, such as decision tree algorithm for classification, regression analysis for prediction, and K-means clustering for clustering, which are described in Figure 2.

The algorithm used in this paper is the decision tree algorithm in the classification algorithm. Among the data mining classification algorithms, the decision tree classification

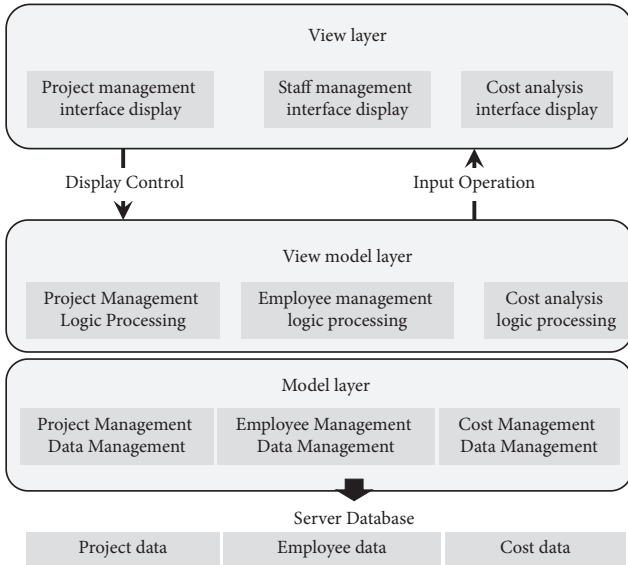


FIGURE 1: Data mining steps.

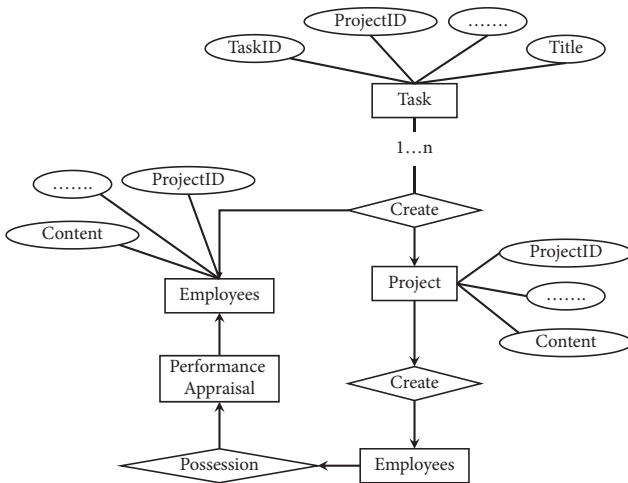


FIGURE 2: Data mining algorithm classification chart.

method has the advantage of being easy to understand and does not require too much specialized background knowledge compared to other classification methods. The results generated by decision trees are usually expressed in the form of if-then rules, which are clear and simple, and are widely used in many fields such as financial industry, meteorological analysis, and traffic management. The root node represents an attribute of the tree, the leaves represent the classification tokens, and the branches represent the output results. The method starts from the root node and iterates through the tree, assigning instances to its children based on the results of the test. Each child node takes a value for that feature and continues to test and assign instances through the recursive method until it reaches the leaf node, where the instances are finally assigned to the class of the leaf node. In the decision tree, there are two kinds of data sets: sample data set and test data set. The sample data set is a collection of data in which the attributes and classifications are known, and the algorithm is used to train the sample data set to produce the corresponding decision tree. The test dataset is

used to test the generated decision tree, bring the data into the decision tree, derive the final categories, compare with the actual types, and measure the accuracy of the decision tree. A decision tree algorithm is efficient, easy to understand, computationally small, and good at handling discrete data. A decision tree can be judged based on its correctness, its effectiveness after testing the sample data set, and its complexity, indirectness, and scale. Information entropy is a key element of decision tree algorithms, where the word “entropy” is a word used in thermodynamics and represents a measure of the degree of chaos in a system in physics. In 1948, Shannon, the father of information theory, borrowed the concept of entropy and introduced information entropy, which is defined as the probability of occurrence of discrete random events. Generally speaking, the higher the probability of a message appearing, the more it is cited, and the lower the information entropy, the higher the probability of the message appearing. The formula for calculating information entropy is defined as follows:

$$G(x) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{-q_i}{\ln q_i}, \tag{1}$$

$$q_i = \frac{\pi}{\sqrt{x+y}}$$

Decision tree algorithms usually consist of three steps: feature selection, decision tree generation, and decision tree pruning. Commonly used decision tree algorithms include the ID3 algorithm and C4.5 algorithm based on maximum information gain rate which is improved from the ID3 algorithm, and CART based on Gini index.

3.2. Model Optimization of Decision Tree Model. The algorithm is based on the concept of information gain for the selection of classification attributes of the decision tree. The value of information gain is the difference between the impurity level of the sample data before classification and the impurity level after classification. The ID3 algorithm is based on information theory and uses information entropy and information gain as criteria to classify existing data sets. When constructing the decision tree, the information gain is taken into account for each branch node selection. The information gain of all attributes is calculated and compared, and the attribute with the largest information gain is taken as the splitting attribute, and the subsequent leaf nodes continue to cycle through this operation to generate the decision tree. ID3 algorithm can only work on discrete attributes, and the first step is to calculate the information entropy of the target attribute of the dataset using formula $H(X)$, and the second step is to calculate the expected information of a certain attribute of the dataset, assuming that the total classification is A , which can be divided into n categories, and the probability of each category is $P(C_1), P(C_2), \dots, P(C_n)$, and then, the information entropy of attribute A after partitioning is as follows:

$$P(C_i) = \frac{\sqrt{\lim_{n \rightarrow \infty} \sum_{i=1}^n c_i^2}}{\lim_{n \rightarrow \infty} \sum_{i=1}^n c_i^2},$$

$$G(x) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(C_i), \quad (2)$$

$$G(x) = \frac{\sqrt{c_1 + c_2 + \dots + c_n}}{\lim_{n \rightarrow \infty} \sum_{i=1}^n c_i^2}.$$

The third step requires the information gain, which is the difference between the information entropy obtained in the first step and the information entropy obtained in the second step for attribute C. The formula is as follows:

$$G(A) = Q(x) - Q_i(x),$$

$$Q(x) = \lim_{n \rightarrow \infty} \sum_{i=1}^n q_i,$$

$$Q_i(x) = \lim_{n \rightarrow \infty} \frac{-\sum_{i=1}^n q_i}{n}, \quad (3)$$

$$G(A) = \lim_{n \rightarrow \infty} \sum_{i=1}^n q_i + \frac{\lim_{n \rightarrow \infty} \sum_{i=1}^n q_i}{n}.$$

The higher the information gain, the more suitable the attribute is for classification, and the attribute is selected as the current node. The columns of the attribute are eliminated from the list, and the rest of the data are iterated from the first step. When there is only one value of the target attribute in the classification, or the proportion of all values of the attribute reaches a threshold, the iteration is finished. The final decision tree is generated when the iteration is completed.

The C4.5 algorithm discretizes the continuous data as follows: First, all the values of the attribute A are sorted in ascending order to obtain the sequence of attribute values $(X_{A1}, X_{A2}, \dots, X_{AN})$, in which there are $N-1$ dichotomous methods, i.e., there are $N-1$ separation thresholds, and the dichotomous method is used to divide this data series into two parts, i.e., two subdata sets (X_{A1}, \dots, X_{AI}) . The calculation steps of the C4.5 algorithm are generally the same as those of the ID3 algorithm, but the information gain rate of the attributes is added to the information gain, and the process is as follows: first, the splitting information is calculated by assuming that the training data set X is divided into i subdatasets by the value of A attribute, X_j denotes the number of samples in the j th subdataset, and X is the number of samples in the original dataset, and the splitting information of attribute A can be obtained as follows:

$$S(X) = - \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{\{X_i\}}{X} \ln \frac{\{X_i\}}{X}, \quad (4)$$

$$\{X_i\} = x_1 + x_2 + \dots + x_n,$$

Then, the information gain of the sample set after splitting according to attribute A is given by

$$\text{Im for } (x, y) = G(x) - G(y). \quad (5)$$

Information gain rate of the sample set after attribute A splitting is

$$G(x) - G(y) = \frac{\text{Gain}(x, y)}{\text{Split}(x, y)}. \quad (6)$$

After the C4.5 algorithm is executed, the result of the information gain rate of all attributes is obtained, and the attribute with the highest information gain rate is selected as the split attribute of the current node. The other attributes will continue to be computed recursively. As the attributes are gradually computed, the information gain rate becomes smaller and smaller, and the attribute with the relatively larger information gain rate is selected as the classification attribute. The flowchart of C4.5 is shown in Figure 3.

3.3. Decision Tree Pruning. The training sample is a key factor in decision tree building, and when the training sample is too small or there are problems with the data, the resulting decision tree may be anomalous, which may lead to inaccurate decision trees and may make the decision tree very complex. According to the research, not all complex and large decision trees result in more accurate rule sets. Therefore, complex decision trees need to be simplified, which is called pruning. Prepruning occurs during the construction of the decision tree when the growth is terminated early in the computation process, and pruning is performed. Prepruning is simple, but it is difficult to determine the timing of termination during the growth of the decision tree, which makes it not very practical. The general decision tree algorithm uses postpruning. Postpruning is performed on a fully grown decision tree in a bottom-up manner, replacing node subtrees with leaf nodes that do not meet the confidence level, and labeling the class with the most frequently used class in the node subtree. The pruning step is repeated until all the nodes satisfy the condition, and the final generated decision tree is more reliable than the previous decision tree. Compared to prepruning, postpruning reduces many perceived interventions. The common postpruning methods are CCP (cost complexity pruning), REP (reduced error pruning), PEP (pessimistic error pruning), and MEP (minimum error pruning). In the comparative study of several pruning methods, it is found that PEP pruning is a top-down pruning method, which has the highest accuracy among several pruning methods and does not require a separate pruning data set.

At present, most of the existing project management information systems are developed and designed using B/S architecture, which is a browser-server architecture model, and users generally access the system through a browser. B/S system has the advantages of easy maintenance and upgrade, low cost, security, etc. The front end uses Vue.js, a bottom-up progressive MVVM framework for building user interfaces, and the back end uses C#, an object-oriented

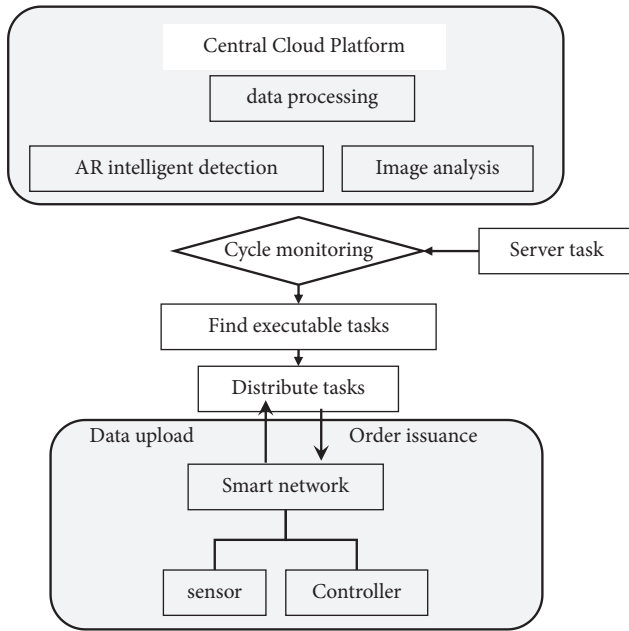


FIGURE 3: Algorithm flowchart.

development language derived from C++. MVC is an abbreviation for Model-View-Controller, which is a framework for layering systems, separating business logic, data, and display interfaces. The framework is layered on the system, and the business logic, data, and display interface is separated, to more clearly delineate their respective functions. The model layer is usually used to handle the logical part of the application data, generally responsible for access to the database. The view layer is the display layer, which handles the display part of the data and is created based on the model data. The controller layer is the user interaction part, which reads data from the view layer, controls user input, and sends it to the model layer.

MVC framework has the advantages of low coupling, high reusability, low life cycle cost, high maintainability, and fast deployment. However, it increases the complexity of the system structure and implementation, and for some simple pages, using MVC framework may lead to the reduction of its operational efficiency; at the same time, there may be inefficient access of view to model; the close connection between view and controller will lead to their independent reuse. MVVM framework is the abbreviation of Model-View-ViewModel, which is an improved version based on MVC framework and MVP framework. This makes it very difficult to maintain the MVC pattern. This is because the controller for processing logic and data transformation in complex projects becomes very large and difficult to maintain, so in order to change the limitations of this model, the logic processing and data transformation of the controller are stripped out from it, and a special object, ViewModel, is created to manage these operations. This approach makes the controller code become very small and easy to manage. This makes the MVVM framework as the mainstream framework used in software development at this stage.

4. Optimization Analysis Based on Decision Tree Algorithm

A fair and responsible appraisal system will stimulate the motivation of the employees, improve the efficiency and motivation of the employees, and enhance the competitiveness of the enterprise. Today, with the rapid development of the Internet, performance appraisal through data mining will become a new way. When an employee submits a task using this system, the system will record the time of task completion and the number of codes submitted by the employee, and the project manager will evaluate the employee based on the information submitted by the employee. Using these data, we analyze the employee's performance through decision tree algorithm to get the employee's performance appraisal results, so that the company can guide the employee's work and help the employee to improve his or her quality, and also improve the technical strength of the company. Before data mining, we first determine the sample data, and the data are selected from the system project management function in the company online, the information filled in by employees and the staff assessment scores entered by managers. As the company is a software development company, the data were mainly collected from R&D workers in the company. 270 data were collected, 180 of which were used as the training data set for the decision tree algorithm, and the rest were used as the test data set. The data set is mainly from the system's employee information table, the employee's work situation according to the task, and the manager's evaluation of the employee during the time period. Employee information table: it is mainly the information of the employees when they registered in the system, among which the performance evaluation criteria are as follows: name, department, major, degree, etc. Work status: this data set is determined by the submission of employee task cards, such as the number of codes submitted, the filling of hours, the completion of work, and the efficiency of work completion, etc. These data are automatically calculated by the system after the employee fills in the data and stored in different tables of the system, and summarized by SQL statements. Employee evaluation: These data are evaluated by the project management according to the performance of the subordinate employees. This process of integrating data is called data integration. The data are aggregated by the unique ID of the employee in the system and stored in the performance appraisal table with the structure as shown in Figure 4.

In the process of getting the data set, there are many data with null value. For these data, we judge the data before entering the variable value into the system database and prohibit the data from appearing as null. Many of the above variable values are extracted from other database tables in the system, and when we get these data, we will fill the null values before entering, and the filling rules are as follows: code amount: these data are null, which means the employee has no code submission record in this time period, and it is recorded as 0 value. The data are empty, which means the employee did not fill in the hours during this time period. Since the filling rate is related to the employee's attendance,

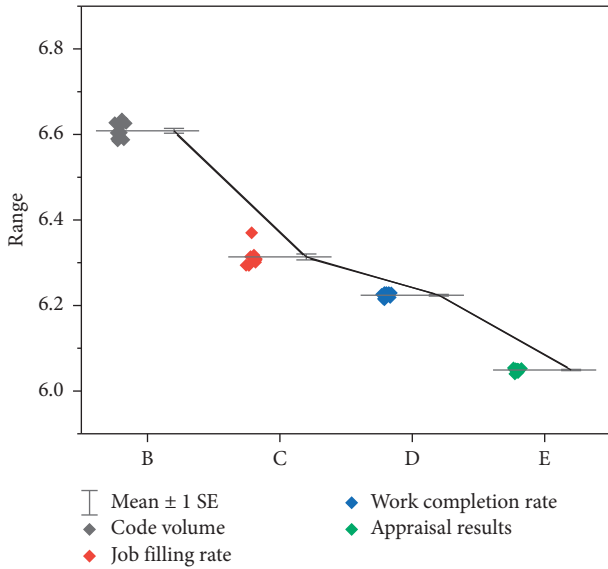


FIGURE 4: Performance appraisal changes.

if it is empty, the employee can find his attendance records in the attendance sheet and calculate the percentage to be filled in. Learning ability, technical ability, work efficiency: if these data are empty, the employee’s supervisor has not evaluated him/her, and the default evaluation is middle-grade C. In the process of data acquisition, there will be some random errors or mistakes, which we call noisy data. The appearance of noisy data can cause great errors in the data mining results. Therefore, we need to preprocess the noisy data; usually, the methods to deal with noisy data are as follows: regression, boxing, human-computer combination check, clustering, etc. For example, when the code volume data are abnormal, the average value of the data can be used to replace the noise data. In addition, we can also use the principle of multiple regression algorithms to smooth the noise data. In this paper, we will use different methods to deal with the noise data according to the attribute values. Since the decision tree algorithm cannot handle continuous values, it is necessary to discretize the continuous variables before data mining, as shown in Figure 5. Through testing and analysis of the decision tree, we find out the points that the project leader pays attention to when scoring the performance appraisal, and we form a complete performance appraisal system based on the performance appraisal attributes in the system and the decision tree.

In this paper, we take the report of the company’s employees using this system in a certain month of 2019, and according to the above data processing process, we get the data set in Figure 6, and the data in the table are taken from the performance appraisal scores of 30 employees in the company’s software development department in January 2019, and in the actual calculation process, we will take the data of 30 employees in this department in the first half of 2019 for the decision tree algorithm. In the actual calculation process, we will take the data of 30 employees in the department in the first half of 2019 for the calculation of the decision tree algorithm and 90 data in the third quarter of 2019 for testing the results of the decision tree algorithm. The system can automatically generate decision

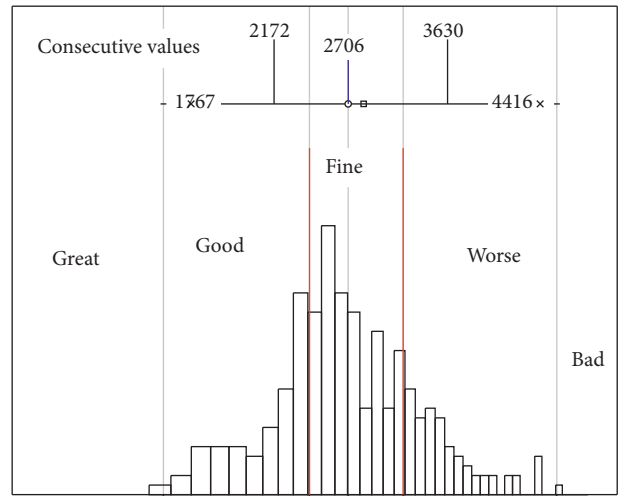


FIGURE 5: Consecutive values of the work-filling rate.

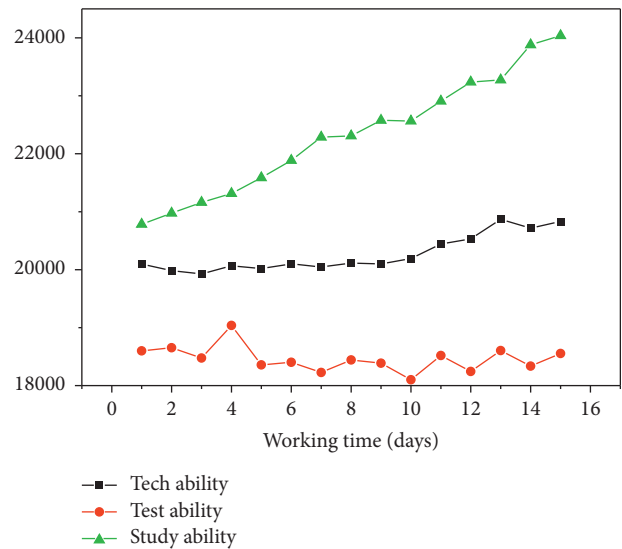


FIGURE 6: Algorithm training data set.

trees based on the data set, and through the analysis of the decision trees, we can find out the modules that managers are concerned about in the performance appraisal—the amount of code and the completion of work. These data are derived from the information reported by employees in the system, which requires employees to plan their tasks in their daily work, complete them on time, and report their costs to the system in a timely manner. The generated decision tree algorithm is recorded by the system and used for the evaluation of the performance appraisal afterward.

The objective of this example is the final evaluation result of the performance appraisal, and the objective class is “evaluation result” with 4 values: A, B, C, and D. In the 180 extracted data sets, the 4 values are as follows: 59, 92, 25, and 4.

In this example, there are 6 attributes as splitting possibilities, and the information entropy of each attribute can be calculated by equation (2). $(40, 81, 20, 3) = 1.4922$ bits. When the value is master’s degree, there are 36 data, as shown in Figure 7, among which the samples belonging to

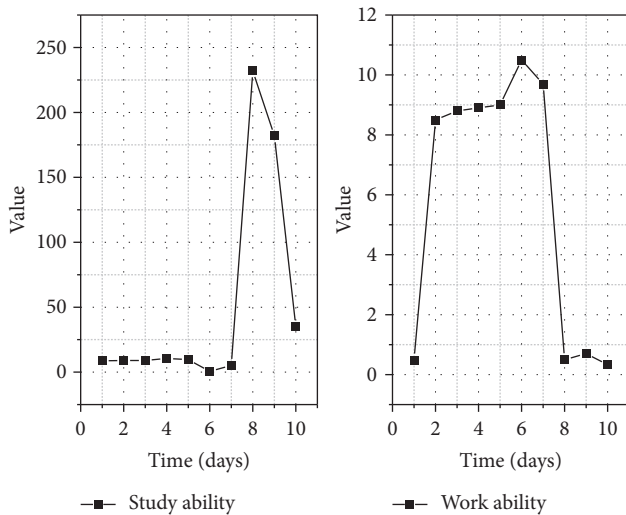


FIGURE 7: Decision tree test dataset.

the four types are as follows: 19, 11, 5, and 1, then $I(19, 11, 5, 1) = 1.5484$ bits, from which we can get its information entropy when the attribute is education: $H(\text{education}) = 144/180*$. When the attribute is code, there are 4 kinds of values, when the value is A, and there are 36 data, and the samples are 24, 12, 0, and 0. Then, we have $I(24, 12, 0, 0) = 0.87218$ bits and 0. If the value is C, there are 60 data with samples of 6, 48, 6, and 0. Then, there is $I(6, 48, 6, 0) = 1.2954622$ bits. If the value is D, there are 24 data items with samples of 0, 12, 12, and 0. Then, we have $I(0, 12, 12, 0) = 1.0000$ bits.

The evaluation results of 90 test samples are compared with the evaluation results generated by the C4.5 decision tree algorithm and ID3 decision tree algorithm, of which 30 test samples are listed here for space reasons. After each project is created, members of the project need to create multiple tasks within the project. The types of tasks can be created by the system administrator. The task package also needs to estimate the cost to be consumed under this module. After the task is created, the project team members who need to complete the task should create task cards under the task, which are defined as short-term plans and goals for the project members in the process of completing the task. Employee management module is the core of this paper, and the key function of the research is the employee performance appraisal system based on decision tree calculation. In the employee management module, firstly, the employees create their own accounts and fill in the information. Then, the system administrator fills in the employee's organization and other information of the employee.

There are two kinds of staff management, one is the project manager's management of subordinate staff, and the other is the staff's self-management. First of all, employees can assign their own tasks and establish their own cards in the system, and there is a personal workbench in the system, where employees can view their own tasks and cards and self-manage the status of the task cards. Under the workbench, employees can fill in their working hours and view their

weekly and monthly reports. The performance appraisal module is a submodule of the staff management module, which generates the three key information of code volume, work completion, and work completion rate based on the information reported by the staff in the system, and the project manager needs to evaluate the learning ability and technical ability of the staff based on their usual performance in the system. These indicators are used to generate the performance evaluation decision tree. The system requires the project leader to give the final rating of the employee's performance appraisal. After the raw data are available, the system generates a decision tree related to the performance appraisal based on the data set, which helps the employee understand the areas that need to be improved. It will determine the employee's salary.

5. Conclusion

In this paper, from the perspective of project management, the C4.5 decision tree algorithm, which is an outstanding data mining technique, is used for performance evaluation analysis. Firstly, we collect and process the information related to the performance of software R&D project leaders and employees to form a performance dataset, and then generate a decision tree model related to performance appraisal based on this dataset through data preprocessing, attribute splitting, model building, pruning operation, and evaluation analysis. The decision tree model was tested with the test dataset generated by the system, and the accuracy rate reached more than 90%, which is higher than the decision tree result of the ID3 algorithm. Based on the analysis of the decision tree, we can identify the areas that need to be improved in the software development work and provide a reference for the performance evaluation of project managers. On the other hand, the information about project management and the evaluation of employee performance in the system will have an impact on the cost of the company. Based on this information, cost reports can be generated for the company. The reports are visualized in a multi-dimensional way to show the cost of the company. The reports can be used by project leaders and relevant company leaders to understand the company's operation clearly and can be used to guide the company's next planning. This paper is based on Vue.js, a lightweight MVVM framework, and uses C# language, SQL Server database, and other tools to design the system. The system is based on the C4.5 decision tree algorithm to mine and analyze the data set generated by the system, generate decision trees related to performance evaluation, and use Echarts and other visualization tools to generate related cost reports. The system is developed in B/S architecture, and a R&D project management system and platform that can realize performance assessment analysis are built by means of visualization tools, decision tree algorithm, and dynamic web pages. In the future, members of the project need to create multiple tasks within the project, and the types of tasks can be created by the system administrator.

Data Availability

All data, models, and codes generated or used during the study are available within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] M. D. A. Al-Falahi, K. S. Nimma, S. D. G. Jayasinghe, H. Enshaei, and J. M. Guerrero, "Power management optimization of hybrid power systems in electric ferries," *Energy Conversion and Management*, vol. 172, pp. 50–66, 2018.
- [2] R. Bourbon, S. U. Ngueveu, X. Roboam, B. Sareni, C. Turpin, and D. Hernandez-Torres, "Energy management optimization of a smart wind power plant comparing heuristic and linear programming methods," *Mathematics and Computers in Simulation*, vol. 158, pp. 418–431, 2019.
- [3] M. Bray, W. Wang, M. A. Rees et al., "KPDGUI: an interactive application for optimization and management of a virtual kidney paired donation program," *Computers in Biology and Medicine*, vol. 108, pp. 345–353, 2019.
- [4] R. H. Byrne, T. A. Nguyen, D. A. Copp, B. R. Chalamala, and I. Gyuk, "Energy management and optimization methods for grid energy storage systems," *IEEE Access*, vol. 6, pp. 13231–13260, 2018.
- [5] X. Dong, X. Li, and S. Cheng, "Energy management optimization of microgrid cluster based on multi-agent-system and hierarchical stackelberg game theory," *IEEE Access*, vol. 8, pp. 206183–206197, 2020.
- [6] S. Feng, X. Shen, and B. Hu, "Optimization of traffic demand management policy in China: towards a sustainable mode split," *Transportation Planning and Technology*, vol. 41, no. 2, pp. 198–210, 2018.
- [7] Y. Han, Q. Zeng, Z. Geng, and Q. Zhu, "Energy management and optimization modeling based on a novel fuzzy extreme learning machine: case study of complex petrochemical industries," *Energy Conversion and Management*, vol. 165, pp. 163–171, 2018.
- [8] H. Karami, M. Ehteram, S.-F. Mousavi, S. Farzin, O. Kisi, and A. El-Shafie, "Optimization of energy management and conversion in the water systems based on evolutionary algorithms," *Neural Computing & Applications*, vol. 31, no. 10, pp. 5951–5964, 2019.
- [9] S. Khayyam, N. Berr, L. Razik, M. Fleck, F. Ponci, and A. Monti, "Railway system energy management optimization demonstrated at offline and online case studies," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 11, pp. 3570–3583, 2018.
- [10] M. Lagouir, A. Badri, and Y. Sayouti, "Multi-objective optimization dispatch based energy management of A microgrid running under grid connected and standalone operation mode," *International Journal of Renewable Energy Development*, vol. 10, no. 2, pp. 333–343, 2021.
- [11] Y. Li, Y. Du, T. Xu et al., "Optimization of thermal management system for Li-ion batteries using phase change material," *Applied Thermal Engineering*, vol. 131, pp. 766–778, 2018.
- [12] D. D. Lieira, M. S. Quessada, A. L. Cristiani, and R. I. Meneguette, "Algorithm for 5G resource management optimization in edge computing," *IEEE Latin America Transactions*, vol. 19, no. 10, pp. 1772–1780, 2021.
- [13] X. Meng, Q. Li, G. Zhang, T. Wang, W. Chen, and T. Cao, "A dual-mode energy management strategy considering fuel cell degradation for energy consumption and fuel cell efficiency comprehensive optimization of hybrid vehicle," *IEEE Access*, vol. 7, pp. 134475–134487, 2019.
- [14] H. Moradi, M. Esfahanian, A. Abtahi, and A. Zilouchian, "Optimization and energy management of a standalone hybrid microgrid in the presence of battery storage system," *Energy*, vol. 147, pp. 226–238, 2018.
- [15] J. W. Ryan, A. S. Murray, P. J. Gilligan et al., "MRI safety management in patients with cardiac implantable electronic devices: utilizing failure mode and effects analysis for risk optimization," *International Journal for Quality in Health Care*, vol. 32, no. 7, pp. 431–437, 2020.
- [16] Z. Sai, C. Lu, S. Jiang, L. Shan, C. James, and N. N. Xiong, "Energy management optimization of open-pit mine solar photothermal-photoelectric membrane distillation using a support vector machine and a non-dominated genetic algorithm," *IEEE Access*, vol. 8, pp. 155766–155782, 2020.
- [17] N. Sulaiman, M. A. Hannan, A. Mohamed, P. J. Ker, E. H. Majlan, and W. R. Wan Daud, "Optimization of energy management system for fuel-cell hybrid electric vehicles: issues and recommendations," *Applied Energy*, vol. 228, pp. 2061–2079, 2018.
- [18] V. Tarzиеv and M. Georgiev, "Efficient management as optimization of the organization," *IJASOS-International E-journal of Advances in Social Sciences*, vol. 4, no. 10, pp. 6–15, 2018.
- [19] W. Wang, L. Liu, J. Liu, and Z. Chen, "Energy management and optimization of vehicle-to-grid systems for wind power integration," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 1, pp. 172–180, 2020.
- [20] W. Wu, W. Wu, and S. Wang, "Thermal management optimization of a prismatic battery with shape-stabilized phase change material," *International Journal of Heat and Mass Transfer*, vol. 121, pp. 967–977, 2018.