The Application of Data Mining Techniques in Employee Performance Assessment

ZHAO ZHENG
Lyceum of the Philippines University Manila Campus,
Manila 1002,
PHILIPPINES

Abstract: - Employee performance assessment is a powerful standard for measuring talent, and many companies pay more attention to the assessment of employee performance. Currently, there are many kinds of methods for employee performance evaluation. This leads to deficiencies in the data accuracy and data mining of current performance research. Therefore, to enhance the deep-level mining of performance data, the advantages of using methods are emphasized. This research uses data mining technology to measure employee performance and builds an improved ID3 decision tree algorithm model based on data mining technology, which can measure deeper employee performance. The experimental results show that the algorithm model is able to measure employee performance well, the accuracy of the decision tree algorithm is 93.2%, and the accuracy of the improved algorithm is 95.3%, so the improved algorithm is 39 ms shorter than the traditional algorithm in building the decision tree, and the algorithm accuracy is 2.1% higher. This shows that the improved decision tree algorithm of data mining technology can improve the precision and accuracy of employee performance evaluation.

Key-Words: - Employee Performance Measurement; Data Mining; Decision Tree Algorithm; Improvement; Improved algorithm; Algorithm comparison; Accuracy.

Received: October 25, 2022. Revised: September 9, 2023. Accepted: November 7, 2023. Published: December 4, 2023.

1 Introduction

In the emphasis on the development of talent today performance evaluation as each enterprise must examine the content, is the current enterprise screening talent, attracting talent, the use of talent is an important means, [1]. The rapid development of information data makes the information data in many fields explode, how to make good use of this information has become a problem worth thinking about, [2]. Although information mining technology can complete the processing and analysis of data through many algorithms, the current existence of many data mining algorithms is not perfect, [3]. Currently, there are many performance evaluation methods, but most of them can only obtain some surface data information and cannot explore the hidden relationships between data, so they cannot achieve more ideal results, [4]. Based on this, to improve the accuracy of employee performance evaluation and the efficiency of data utilization, the current data mining algorithms are improved, and a new data mining improvement decision tree algorithm employee performance evaluation model is constructed to deeply mine employee performance data. This can greatly compensate for the shortcomings of traditional methods. This experiment is divided into four parts, the first part is the elaboration of domestic and international scientific research results; the second part mainly

introduces the basic concepts of the research problem and the construction process of the algorithm model; the third part proves the feasibility of the algorithm through experiments; and the fourth part summarizes the whole research process.

2 Related Works

With the continuous development of science and technology, data mining technology is more and more widely used in various fields. Among them, the decision tree algorithm (DTA), as an excellent algorithm in the field of data mining, has the advantages of a simple model, fast computation speed, accurate classification, etc. [5], tried to quantify the confidence level of the decision tree algorithm by combining the regression tree algorithm with the DTA to further improve the confidence level of the DTA in the hypothetical dataset. The experimental results show that the combined algorithm avoids the problem of overfitting, and the generalization performance is significantly improved. [6], proposed an improved DTA to evaluate the skin-wise sensitivity in the cosmetic safety index. The experimental results showed that the algorithm was more efficient than the traditional assessment model, with better clarity of judgment and more reliable predictability. Vanveller found that the transformation process is extremely complex when students design the complete synthesis of a molecule. Therefore, to simplify the problem, the scholar proposed to integrate the decision tree algorithm into the step classification and developed a set of flowcharts. The experimental results showed that the incorporation of the algorithm greatly simplified the students' steps and reduced the difficulty of the production process, [7]. [8] proposed a modeling framework based on gradient-enhanced decision trees to better detect the residual capacity of lithium-ion batteries for electric vehicles. The framework experiments with an accurate assessment of the battery's health capacity by extracting effective indicators from the original battery. The model has better detection efficiency and higher validity of the remaining battery life than the traditional method.

In modern business management, performance measurement is widely used to assess the performance of employees and their contribution to organizational goals. Performance measurement is a systematic process designed to quantify and assess the competence, skills, and responsibility demonstrated by employees in their work, [9] To expand the performance assessment of professionals in academic libraries, a structural equation model of performance appraisal was introduced to 339 librarians from 190 universities in Pakistan. The experimental results showed that the performance model utility improved managerial competence and decision-making accuracy and enhanced human resource organization and knowledge management techniques. [10], to examine the employees' perceptions of a high-performance work system, 524 employees of three Greek manufacturing companies were subjected to an experimental performance measurement over a period of three months. The results of the experiment showed that employees' job satisfaction was positively related to job stress, role ambiguity, and role conflict. Thus this performance measure provides ideas for better regulation of high motivation of employees at work. [11], found that employees misuse social media which affects their job performance and thus proposed to incorporate the rules of social media in performance appraisal to regulate the relationship of sharing and accessing of information between the individual and the job. The experiment collected a 1200-point questionnaire from the public and private sectors of Pakistan. The findings indicated that personal and work-related text message use can improve employee performance through knowledge exchange. [12], investigated the mechanism of employee non-ethical organizational behavior on employees and supervisors, 304 employees and 94 supervisors in several manufacturing industries in China were used as research subjects to test the hypotheses using a hierarchical linear model. The experimental results show that employee non-ethical

organizational behavior is positively related to supervisors' performance appraisal, and at the same time supervisors' employee non-ethical organizational behavior indirectly affects employees' performance appraisal.

Combining the above various types of research and experimental results, it is not difficult to find that there have been a large number of studies focusing on the field of employee performance measurement, but there are still some challenges and problems. For example, the traditional assessment model is based on qualitative evaluation, lacks quantitative indicators and data support, and has limitations in the dimension of performance. There are few applications of decision tree algorithms for employee performance assessment. Therefore, this study aims to improve the decision tree algorithm, and in-depth discussion of the optimization method of employee performance evaluation, to provide a theoretical basis and new ideas for the research in this field.

3 Model Algorithm Design of Data Mining Techniques in Employee Performance Measurement

This chapter begins with a brief overview of data mining techniques and employee performance measurement, and then uses data mining techniques to measure employee performance, by improving the DTA in data mining techniques, a new decision tree improvement algorithm employee performance measurement model is obtained.

3.1 Research on the Application of Data Mining Techniques in Employee Performance Assessment

Performance assessment is an essential task in the enterprise, which can test and evaluate the salary status of employees, so the performance assessment system needs to have several characteristics. The first is systematic, performance evaluation needs to have a complete set of systematic methods. At the same time performance assessment system must be the need to have clarity of purpose to ensure the reasonable operation of employee performance assessment, [13]. Therefore, using data mining technology is a relatively simple method to obtain employee performance data, and the main process of data mining is shown in Figure 1.

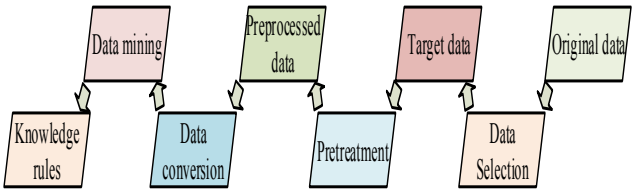

Fig. 1: Main process of data mining

As shown in Figure 1 the process of data mining first determines the data source, and then through the capture of the target data for data processing, through the processed data, and then data conversion to achieve the mining of data, and finally the mined data through the evaluation of patterns to achieve the discernment of useful knowledge. The function of data mining is to utilize the initial characteristics of the data, including data noise incompleteness, ambiguity, randomness, etc. Meanwhile, the content of data mining can be some semi-structured forms of data such as text and images. Currently used data mining methods are capable of data mining in different directions for a variety of fields. As shown in Figure 2 is an illustration of cross mining of different domains of data mining.

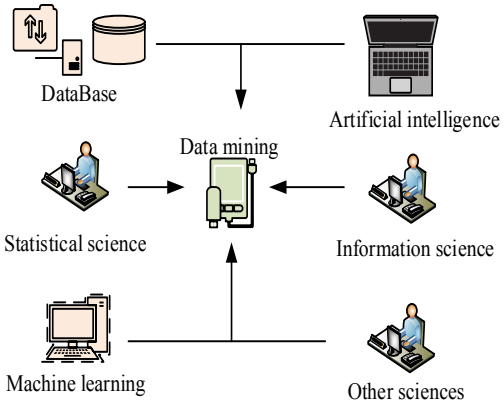

Fig. 2: Cross-mining in different fields of data mining

As shown in Figure 2, data mining can connect different fields through machine learning, statistics, and information science can achieve the analysis and utilization of data when using data mining, and database technology is used to mine data through artificial intelligence techniques. From the figure, it is clear that many fields can use data mining techniques to analyze data to solve multiple real-world problems. In general, the main function of data mining is to characterize the data ensemble of the object of study, and through the hidden relationships and laws of these data predict the future development of the data, [14].

The functions of data mining are, automatic prediction function is to be able to predict the information in the database autonomously; correlation analysis is to analyze the correlation information in the database to analyze the value of the variables; cluster analysis function is to be able to analyze the subset of a variety of data to cluster; descriptive concepts are to conceptually describe some things, and at the same time is divided into characteristic description and distinguishing description; bias detection refers to the existence of the database to data present in the database; bias detection refers to the detection of the data present in the database by measuring its bias value and true value. The general structure of a data mining system is the input data part, the data processing part, and the user interface part. As shown in Figure 3.

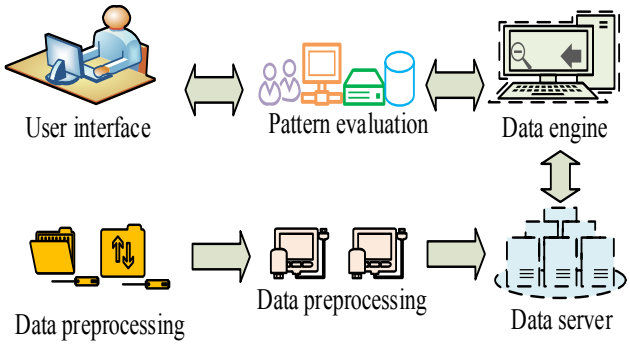

Fig. 3: Analysis of data mining processing structure

In Figure 3, the input part of the data mining system includes the storage of data and the pre-processing part of different data, which is pre-processed before the next operation. Data mining processing is mainly for the data objects that need to be processed, using data algorithms to mine the relevant data, [15].

3.2 Optimized Design of Performance Measurement based on Decision Tree Algorithm

The commonly used algorithmic models for data mining are the Decision Tree Algorithm, Bayesian Classification Algorithm, Support Vector Machine Algorithm, Genetic Classification Algorithm, Artificial Neural Network Algorithm, K Nearest Neighbors Classification Algorithm, Rough Classification Algorithm, and Fuzzy Classification Algorithm. From the process of data mining, the decision tree algorithm should be selected as the optimal data mining algorithm. The decision tree algorithm can classify a large amount of data with a purpose, mining useful hidden data information, and can be used for data prediction models, not only does it have a high efficiency at the same time for the construction of the algorithm is relatively simple, so it applies to the construction of the algorithm model of data mining, [16]. Decision Tree Algorithm is a predictive algorithm model that can be learned through the data training set as a regression, which is passed in a top-down manner, similar to a tree structure, usually, there are tree nodes, trunk nodes, and leaf nodes. This makes the algorithmic model have the following characteristics, accuracy, validity of data, complexity, simplicity, robustness of the decision tree, and scale of the decision tree. The main flow of the decision tree algorithm is shown in Figure 4.

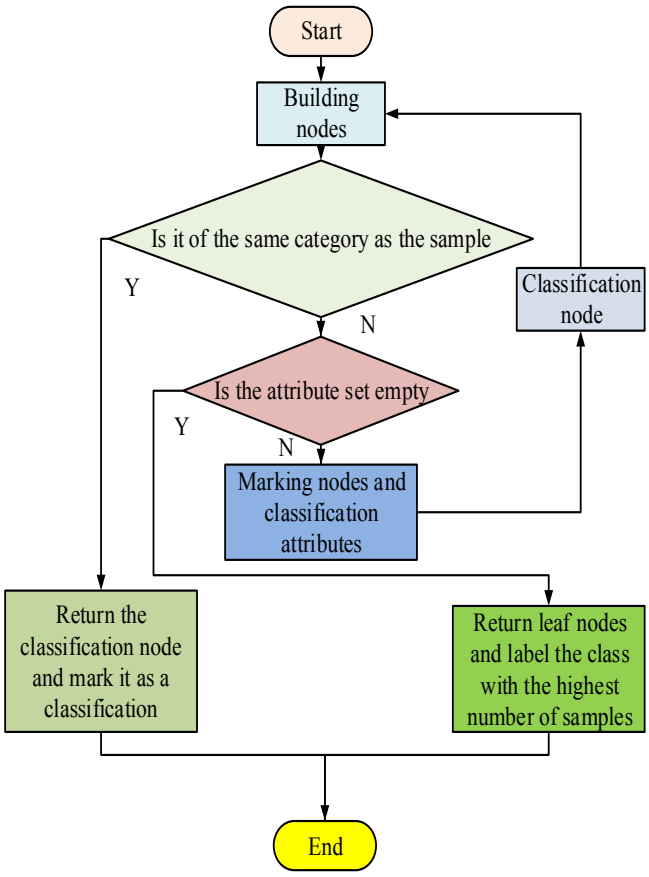

Fig. 4: Decision tree algorithm flowchart

As in Figure 4, the DTA model begins the phase of reality on the decision-making nodes to create, and then identify the data samples belonging to the same type of samples are returned to the leaf nodes, if not, then determine whether the selected subset is empty, the empty set is labeled with the set and then the output of the node classification to re-create the node, is not the empty set is returned to the leaves of several points to mark the samples clustered. Finally, end the algorithm. In the DTA, the use of information entropy can judge the information of the node through the construction of the DTA to complete the analysis. Let the data set of the sample be M The sample is n, then the correct information entropy formula is shown in Equation (1), [17].

$$I(M_1, M_2, \cdots, M_s) = -\sum_{i=1}^{s} P_i \log_2 P_i \qquad (1)$$

Equation (1), P_i indicates the probability of the sample set; M_i indicates the number of samples of the set C_i, if the type of occurrence of no sample number then $I = 0$, there is a sample number and the same case $I = 1$. Set a different category data set T which has u values, so through the data set can be divided into u subsets, at this time by the data set T information entropy formula for the Equation (2).

$$E(T) = \sum_{j=1}^{u} \frac{x_{1j} + x_{2j} + \cdots + x_{xj}}{n} I(x_{1j} + x_{2j} + \cdots + x_{xj}) \qquad (2)$$

Equation (2), $E(T)$ denotes the information entropy; $(x_{1j} + x_{2j} + \cdots + x_{xj})$ denotes the subset of the dataset; $\frac{x_{1j} + x_{2j} + \cdots + x_{xj}}{n}$ and denotes the weights of the subset of the dataset $I(x_{1j} + x_{2j} + \cdots + x_{xj}) = -\sum_{i=1}^{s} P_{ij} \log_2 P_{ij}$. The attribute value T is presented as the gain function of information as shown in Equation (3), [18].

$$Gain(T) = I(M_1, M_2, \cdots, M_s) - E(T) \qquad (3)$$

Equation (3), $Gain(T)$ denotes the reduced range of information entropy produced by the known function parameter T. The gain effect at this point can be expressed by Equation (4).

$$GainRatio(T) = \frac{Gain(T)}{SplitI(T)} \qquad (4)$$

Equation (4), $GainRatio(T)$ represents the size of information gain efficiency, $SplitI(T) = -\sum_{j=1}^{u} P_j \log_2 P_j$, information gain represents a measure of uncertainty in the matter of

information entropy, able to obtain different function datasets through the unit cost. SLIQ algorithm is an algorithmic model for high-speed decision-making in decision tree algorithms. SLIQ algorithm can replace the information entropy of the algorithm by $Gini$ where $Gini$ is shown in Equation (5).

$$Gini(t)=1-\sum_{i=1}^{s}\left[p(i/t)\right]^{2} \tag{5}$$

Equation (5), $p(i/t)$ stands for the occurrence probability at t, when the value of $Gini(t)$ is the smallest, all the node types belong to the same node, and when the value of $Gini(t)$ is the largest, the number of samples of all the nodes is distributed among the various data sets, so at this time, the expression of $Gini$ is shown in Equation (6).

$$Gini_{split}(l)=1-\sum_{j=1}^{m}\frac{m_j}{m}Gini(i) \tag{6}$$

Equation (6), m_j denotes the record at j; k denotes the nodes, and m denotes the quantity at l.

Decision algorithms are algorithms that classify data with fuzzy information entropy, and the ID3 algorithm is a kind of advancement of decision algorithms, in constructing the decision tree, it will preferentially fuzzy the set of attributes, and this fuzzy of attributes can avoid too many attributes to be interfered in the classification of continuous attributes. The relative frequency of the fuzzy leaf nodes in this case is shown in Equation (7).

$$P_{ij}^{k}=\frac{M(T_i^k\bigcap D^j\bigcap S)}{M(T_i^k\bigcap S)} \tag{7}$$

Equation (7), S represents the non-leaf nodes in the non-decision algorithm and T_i^k represents the classification of fuzzy attributes. Therefore, the fuzzy information entropy is defined as shown in Equation (8).

$$I(T_i^k)=-\sum_{j=1}^{m}\log(p_{ij}^k) \tag{8}$$

The average fuzzy information entropy in Equation (8) for the fuzzy attribute A_i is shown in Equation (9).

$$E(A,S)=\sum_{k=1}^{m_i}p_k I(T_i^k) \tag{9}$$

Equation (9), p_k stands for the weight of the fuzzy attribute A_i in the decision algorithm in the first k definition. In this case, the definition formula for the fuzzy attribute A_i is shown in Equation (10).

$$Gaini(A_i)=M(S)-E(A,S) \tag{10}$$

During the construction of the DTA, the largest fuzzy information entropy as the information entropy fuzzy attribute for the current classification, by using filters that increase the level of filtration when classifying the information of the data, and then by setting the real data parameter β to determine the conditions of the current leaf node, [19].

To determine the uncertainty, for any selected attribute value $T_i^K(1\le i\le n,1\le k\le m_i)$, categorize the j confusing classifications on the non-leaf node S D^j, at which point the defining formula for normalization is shown in Equation (11).

$$\tau_{ij}^{k}=\frac{p_{ij}^{k}}{\vee_{j=1}^{m}p_{ij}^{k}} \tag{11}$$

Equation (11), τ_{ij}^k denotes the normalized value, and the defined uncertainty for the value of the to-be-selected attribute T_i^k that can be determined

is shown in Equation (12).

$$Ambig(T_i^k) = \sum_{j=1}^{m} (\pi_{ij}^k - \pi_{i,j+1}^k) \ln j \quad (12)$$

In Equation (12), π_{ij}^k the order of arrangement $\pi_{i,j+1}^k$ representing the lift order, and the attribute uncertainty for non-leaf nodes is defined as shown in Equation (13).

$$G(A_i) = \sum_{j=1}^{m} p_i Ambig(T_i^k) \quad (13)$$

Equation (13), $G(A_i)$ denotes the uncertainty of the algorithm, for the uncertainty of the decision tree algorithm needs to preprocess the known data and finally generate the generating conditions of the algorithm parameters.

3.3 Performance Measurement based on Optimized Decision Tree Algorithm

From the construction process of the algorithm, construction process of the decision tree algorithm using ID3 is relatively complex, during which there are a large number of repetitive operations, and at the same time, due to the algorithm there are several logarithmic exponential operations made, the calculation of the need to mobilize the database to increase the computation time, and therefore the need to ID3 algorithm to improve the process of improving the algorithm and thereby reducing the cost of improving the efficiency of the operation, [20]. The decision tree model is usually constructed as shown in Figure 5, taking weather data as an example.

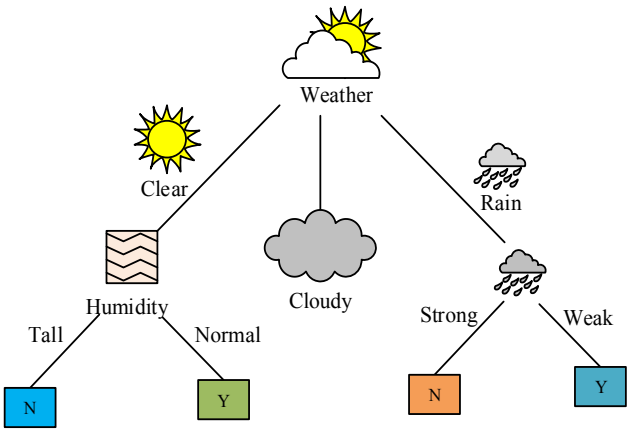

Fig. 5: Decision tree model

In Figure 5, the decision tree firstly makes a decision judgment on the wind data and temperature and humidity in the weather, whether the wind is strong or weak, whether the temperature and humidity belong to the normal range, and whether it is on the high side, and then through the decision judgment on the temperature and humidity and wind whether the weather belongs to the sunny or rainy days, while the cloudy days are listed separately, which is the general decision process of the decision tree.

The improvement of the decision-making algorithm process is, first of all, calculating sample data for all attributes of the decision-making scheduling, if the maximum value and other values can be calculated in the decision-making scheduling of the deviation, and then the maximum value for the only attribute maximum value. At the same time, due to the existence of a variety of decision-making attributes in the sample data, there will be several calculations that appear similar or equal, at this point in the calculation can no longer be decided to split the attribute, it is necessary to another calculated attribute gain and select the largest value of which is the split attribute value. At this point the decision function of the sample data belongs to the same category of sample attributes and can be used as a decision function of the leaf nodes, then the decision-making algorithm to improve the algorithm is completed. Because of the size of the true value of the decision function scale and the accuracy of

the classification, it is necessary to switch the maximum value in the DTA and the minimum value of the fuzzy information entropy. Therefore, the data mining process for employee performance is shown in Fig. 6, [21].

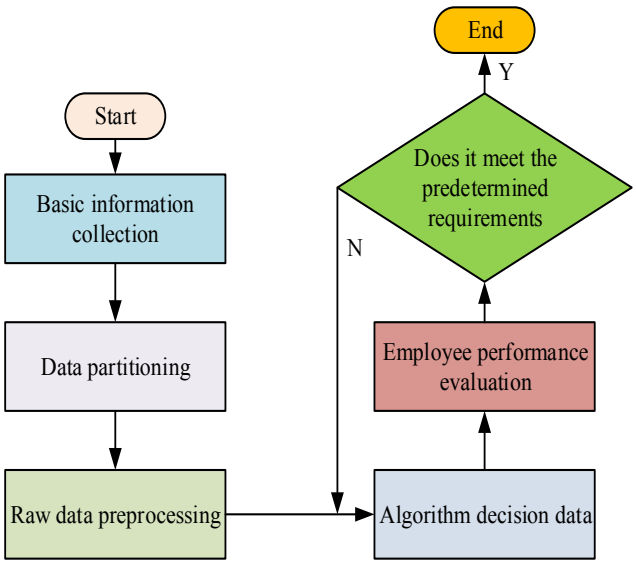

Fig. 6: Eployee performance data mining process

In Figure 6, firstly, the basic information of the employee is collected such as name, work number, department, etc. Secondly, the measured attributes of the employee performance information are divided into statistics, and then the raw data are processed such as data conversion, data integration, data screening, etc. After the data are processed, the improved decision tree algorithm is used to classify the data and make decisions. After the data processing, the improved DTA is used to classify the data, and the algorithm calculates the final employee performance assessment to determine whether it meets the predetermined needs and outputs the data if it meets the requirements, and re-assesses and calculates if it does not meet the requirements.

In the whole construction process, the experimental data of a company is firstly selected as the performance evaluation data, and the performance information of employees is collected during the data collection. Due to the large number of collected data attributes and quantities, it is necessary to pre-process the data for some data redundancy and information anomalies, including data integration, screening, conversion, and so on. Finally, a decision tree is constructed by improving the decision tree algorithm to evaluate the performance assessment data, [22].

The results of the performance evaluation will be divided into four respectively A \ B \ C \ D, the number of values of the four grades are 2, 7, 8, and 3, through Equation (1) to calculate the known

$$I(2,7,8,3)=-\frac{2}{20}\log_2\frac{2}{20}-\frac{7}{20}\log_2\frac{7}{20}-\frac{8}{20}\log_2\frac{8}{20}-\frac{3}{20}\log_2\frac{3}{20}=1.8016bits$$

. Then the information entropy of each performance data is calculated by Equation (1), and different levels of performance data can get different sizes of information entropy values. Finally, recursive operations are carried out on the split points of each piece of information until the complete decision tree generation can be realized. In the data fuzzy processing, in the data conversion process, it is difficult to speak about the progressive nature of data changes reflected, for example, when the data information and the grade division are only one point different, 59 points will be classified as D, 60 points will be classified as C, the results of this assessment is unfair. Calculate the degree of membership for each level of performance, and fuzzify the data to weaken the unfairness brought by the data, thus achieving a transition from quantitative to qualitative changes in continuous data. Therefore, the affiliation function is added to the improved algorithm, and the affiliation degree of different grades is calculated to be confused, to reduce the uncertainty in the data analysis and evaluation.

4 Analysis of Test Results of Data Mining Techniques in Employee Performance Assessment

In measuring employee performance, performance is measured using the hold-up method. The data is randomly divided into a test set and a training set. This experiment is to compare two algorithms one is

a decision tree algorithm and one is a decision tree improvement algorithm. The test results obtained using number sets with different sample sizes are shown in Figure 7.

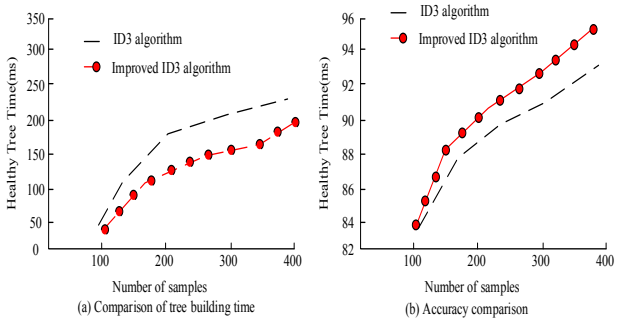

Fig. 7: Algorithm sample test results

As shown in Figure 7, when comparing the decision tree-building time of the algorithms, it is found that when the number of samples increases, the decision tree-building time increases. When the number of samples reaches the maximum value the decision tree building time reaches the maximum value. At this time, the building time of the decision tree algorithm is 229 ms, and the building time of the improved decision tree algorithm is 190 ms, with a difference of 39 ms between the two maximum values. When comparing the accuracy of the algorithms, it is found that the accuracy of the algorithms increases when the samples increase. When the samples reach 400, the accuracy of the algorithm reaches the maximum value. At this time, the accuracy of the decision tree algorithm is 93.2%, and the accuracy of the improved algorithm is 95.3%. The improved algorithm is 39 ms shorter than the traditional algorithm in the construction time of the DTA, and the accuracy of the algorithm is 2.1% higher. The fuzzy information of the decision tree is compared with the data as shown in Table 1.

As shown in Table 1, the assessment results such as realism and skill level can react to the fuzzy information of the current decision-making algorithm, when the skill level and the degree of enthusiasm for the work are A level, the assessment results obtained are also A level, and the assessed realism is also 100%. From the several samples whose truthfulness is less than 100%, there are two assessment results of grade A, one assessment result of grade B, and one assessment result of grade C.

Table 1. Comparison of fuzzy information in decision algorithms

Job Number	Level	Positive level	Growth ability	Degree of completion	Evaluation results	Realism (%)
1	A	A	/	/	A	100
2	A	B	/	/	B	100
3	A	C	/	/	B	100
4	A	D	/	/	B	100
5	B	/	A	/	A	76
6	B	/	B	/	B	100
7	B	/	C	/	B	92
8	C	A	A	/	A	81
9	C	B	A	/	B	100
10	C	C	A	/	B	83
11	C	D	A	/	C	100
12	C	/	B	/	C	94

The assessment results can reflect the current improved decision-making algorithm. The evaluation results can reflect the decision-making ability of fuzzy information of the current improved decision tree algorithm, and when the degree of truth is 100%, it means that the decision-making ability and evaluation ability of the decision tree algorithm at this time has reached the optimal value. Improved decision tree algorithm ID3 and the traditional three algorithms for comparison, Support Vector Machine (SVM) algorithm, Genetic (Genetic Algorithm, GA) algorithm, Convolutional Neural Network (CNN) to obtain as shown in Figure 8 Comparison graph of algorithm prediction performance accuracy.

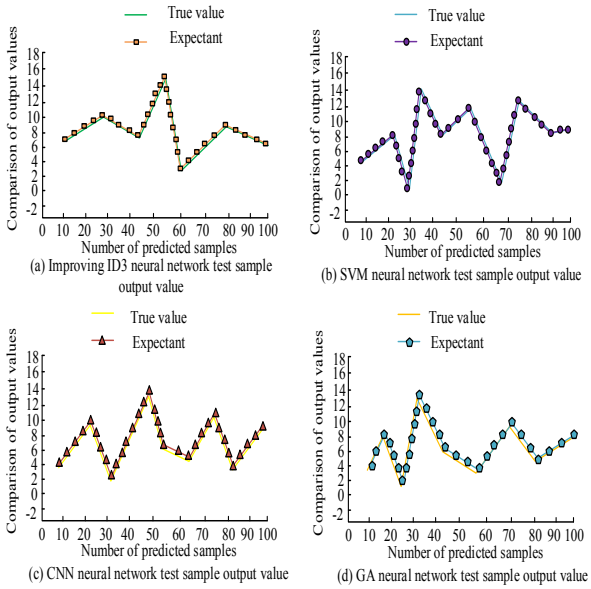

Fig. 8: Accuracy comparison of four algorithms

As shown in Figure 8, the four algorithms in the GA algorithm have the largest deviation value of the true value and the expected value, from the change curve of the true value and the expected value of the change trend is the same, but from the numerical point of view of the two deviations are the four algorithms with the largest deviation of the algorithm, the accuracy of its algorithm is 82.31%. Compared with the other algorithms, the improved ID3 algorithm is the four algorithms to find the one with the smallest deviation between the true value and the expected value, and its accuracy is 95.63%, of which the CNN algorithm's accuracy is 86.32%, and the SVM algorithm's accuracy is 90.23%. Thus the Improved ID3 algorithm has the highest accuracy among the four algorithms, which is 5.4% higher than the SVM algorithm, 9.31% higher than the CNN, and 13.32% higher than the GA algorithm. The employee learning ability and skill level derived from the algorithms were compared and obtained as shown in Table 2.

Table 2. Comparison of employee learning ability and skill level

Serial Number	Education		Learning ability				Skill			
	Undergraduate course	Master	A	B	C	D	A	B	C	D
1	1	/	/	/	1	1	0	1	/	0
				1	5					
2	1	0	0	/	/	0	0	0	0	1
				6	6					
			1	5						
3	0	1	/	/	0	0	1	0	1	0
			6	6						
4	1	0	0	0	1	1	0	1	0	0
5	1	0	0	0	1	0	0	0	1	0
6	1	0	1	0	1	0	1	0	0	0
7	1	0	1	0	0	0	0	0	1	0
8	0	1	0	0	0	1	1	0	0	0
				5	1					
9	1	0	0	/	/	0	0	0	1	0
				6	6					
								5	1	
10	0	1	1	0	0	0	0	/	/	0
								6	6	

As shown in Table 2, the number of postgraduate students in the entire workforce is low, the number of employees with strong learning skills is insufficient, and the majority of employees have a medium level of performance. There are also many employees with insufficient skill levels among them. It can be seen that the use of the algorithm to mine the skill level and learning ability of the employees can find out the deficiencies of the employees to improve in time. To compare the accuracy of the algorithms during test training compare the test accuracy of the four algorithms. As shown in Figure 9.

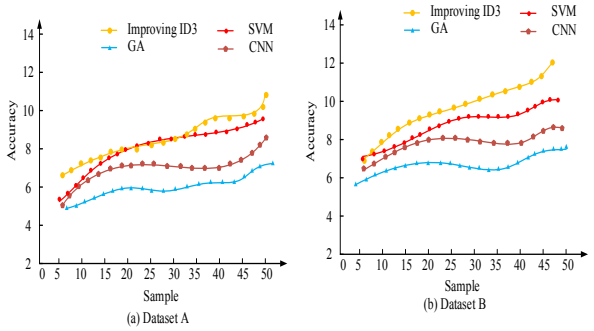

Fig. 9: Accuracy comparison of four algorithms

As shown in Figure 9, the accuracy of the four algorithms increases with the increase in the number of samples, but for the comparison of the two datasets it is found that the accuracy of dataset B is higher relative to the accuracy of dataset A. The accuracy of the four algorithms is relatively high. Also, the trend of accuracy change is relatively higher for the improved ID3 algorithm among the four algorithms and the GA algorithm has the lowest accuracy among the four algorithms. From the number of samples all four algorithms change in accuracy varying between 4-12, only the improved algorithm has the largest accuracy value. It can be seen that by comparing the four algorithms with different datasets, only the improved algorithm has the highest accuracy most suitable for measuring the samples. To test the error values of the algorithms, the error values of the improved algorithm are compared with the three algorithms obtained as shown in Figure 10.

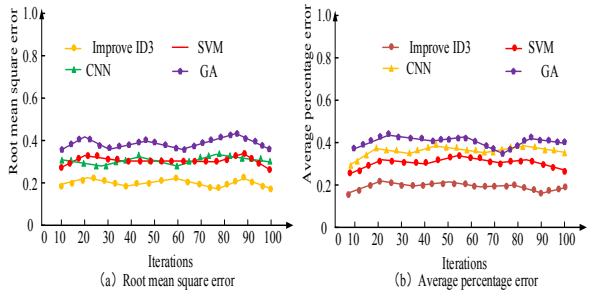

Fig. 10: Comparison of four algorithm errors

As shown in Figure 10, the root mean square error of the four algorithms comparing the improved algorithm has the smallest value of root mean square error, which varies between 0.15-0.25, where the maximum value of error occurs at the sample number of 60, when the value of error is 0.23. The GA algorithm has the largest value of error of the four algorithms, which varies between 0.35 and 0.44, where the maximum value occurs at the sample number of 85 at the value of 0.44. The error values compared to the other two algorithms are between the improved algorithm and the GA algorithm. Comparing the mean percentage error values of the algorithms, the smallest error is also in the improved algorithm, where the maximum value of 0.21 occurs at a sample size of 20, and the maximum value of 0.45 occurs at a sample size of 25 for the GA algorithm, which shows that the improved algorithm has the smallest error value of the four algorithms, with the maximum value of the root-mean-square error being lower than that of the GA algorithm by 0.21, and the maximum value of the mean percentage error being lower than that of the GA algorithm by 0.24. algorithm 0.24. To compare the stability of the algorithms, the four algorithms are compared in terms of loss function and number of iterations, which are obtained as shown in Figure 11.

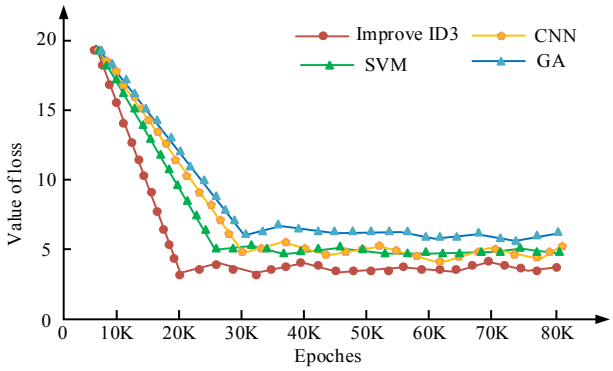

Fig. 11: Stability comparison of four algorithms

Figure 11 shows that the loss function value of the four algorithms decreases with the increase in the number of iterations and varies in that range when a certain value is reached. Among the four algorithms only the improved algorithm has the smallest loss function value in the number of iterations is 20 k when the loss function value reaches the minimum value of 3, GA algorithm reaches the minimum value of 7 when the number

of iterations is 30 k, SVM algorithm reaches the minimum value of 5 when the number of iterations is 25 k, and the CNN algorithm reaches the minimum value of 5 when the number of iterations is 30 k. It can be seen that the improved algorithm's loss function value of the improved algorithm is lower than the GA algorithm by 4 and lower than the SVM algorithm and CNN algorithm by 2. Therefore, the improved algorithm is more stable.

5 Discussion

Today, with the rapid development of information, there is a huge amount of data and information in all industries, and how to effectively use the information and data is an important means for the development and progress of enterprises. Employee performance evaluation data is related to the efficiency and enthusiasm of employees, which requires a more perfect and more effective evaluation system. Data mining technology has become a more widely used data information processing technology, the use of data mining technology can effectively mine employee performance data. In the current method, most of the existing employee performance evaluation methods are designed to mine the surface data, and cannot realize the deep data analysis, so this study uses the improved decision tree algorithm to mine the employee performance data at a deeper level.

As can be seen in Figure 7, when the number of samples of the decision tree algorithm increases the construction time of the decision tree increases, which indicates that the decision tree model of the decision tree algorithm will follow the change in the number of samples, and the improved decision tree algorithm builds the time faster in graphical representations, which may be due to the improved performance of the improved algorithm. In Table 1, it can be seen that some of the performance evaluation results are lower which may be due to the poor performance data of some employees, but from the evaluation of the degree of realism can be reacted to the current algorithm for the evaluation of employee performance, when the higher the degree of realism indicates that at this time the data of the affiliation value of the data is greater, the higher the degree of ambiguity of its data, the data embodied in the unfairness is smaller, so the algorithm to get the realism of the better the ability to deal with the data the more powerful!. From Figure 8 it can be seen that the accuracy of different algorithms for employee performance assessment is different, but the improved decision-making algorithm has the highest accuracy of assessment, while the accuracy of its algorithm changes shows a curve change, which may be caused by the stability of the algorithm at the same time as the number of samples is different and there are deviations in the stability of the algorithm. From Figure 9 it can be seen that comparing the algorithm prediction accuracy of the two datasets, the improved decision-making algorithm has the highest accuracy, which indicates that the improved decision-making algorithm has a better effect on the assessment of employee performance, while the accuracy change of the algorithm is increased with the increase in the number of samples, which may be due to the increase in the number of samples after the algorithm's decision tree construction is completed to make the algorithm's accuracy improve. As can be seen from Figure 10, the number of iterations of several algorithms increases the error value of the algorithm showing the form of curve changes, which may be due to the change in the number of iterations that will cause the sample fluctuations in the situation. As can be seen from Figure 11, the loss function of the algorithm tends to stabilize after the loss function decreases when the number of iterations increases, which may be due to the fact that the stability of the algorithm is lower at the beginning of the iteration, the loss function decreases more, and then the stability tends to a relatively stable state.

From the above results, the performance of the whole improved algorithm has better stability and higher accuracy than some other traditional methods. Therefore, the improved algorithm model can

enhance the ability to measure employee performance to a certain extent.

6 Conclusion

This research mainly uses the technology of data mining to carry out experiments on employee performance assessment, firstly, the data mining technology is outlined, then the decision tree algorithm is used to carry out employee performance assessment through the analysis of the data mining technology, then the decision tree algorithm is algorithmically improved, so that the assessment ability is improved, and finally the feasibility of the improved algorithm and the accuracy of the assessment are proved through the experiments. The experimental results show that the accuracy of the decision tree algorithm is 93.2%, and the accuracy of the improved algorithm is 95.3%, so the improved algorithm is 39 ms shorter than the traditional algorithm in building the decision tree, and the algorithm accuracy is 2.1% higher. When the algorithms were compared in the experiment, the improved ID3 algorithm had the highest accuracy, which was 5.4% higher than the SVM algorithm, 9.31% higher than the CNN, and 13.32% higher than the GA algorithm. The improved algorithm has the smallest error value, the maximum value of root mean square error is lower than the GA algorithm by 0.21, and the maximum value of average percentage error is lower than the GA algorithm by 0.24. The value of the loss function of the improved algorithm is lower compared to the GA algorithm by 4, and is lower than the SVM algorithm and CNN algorithm by 2, so that the improved algorithm is more stable. This shows that the improved decision tree algorithm is more suitable for employee performance assessment in terms of precision accuracy and stability. Although the algorithm was improved and employee performance was evaluated in this experiment, there are still problems in many aspects. First, the data samples used in analyzing and judging the algorithm are small, and more sample data will be studied in the future. Secondly, the parameters of employee performance in the experiment were not comprehensively analyzed, and more parameters will be analyzed in the future.

References:

[1] Yu X, Ding C, Zhu Z, Liu Z L, Li L, Zhao Z M, A Novel Method for RFID Reading Performance Measurement Based on the Effects of Exposure to Salt Mist, *Journal of Testing and Evaluation: A Multidisciplinary Forum for Applied Sciences and Engineering*, Vol. 50, No. 1, 2022, pp. 66-79.

[2] Cavicchi C, Vagnoni E, The Role of Performance Measurement in Assessing the Contribution of Circular Economy to the Sustainability of A Wine Value Chain, *British Food Journal*, Vol. 5, No. 124, 2022, pp. 1551-1568.

[3] Mio C, Costantini A, Panfilo S, Performance Measurement Tools for Sustainable Business: A Systematic Literature Review on the Sustainability Balanced Scorecard Use, *Corporate Social Responsibility and Environmental Managemen*t, Vol. 29, No. 2, 2021, pp. 367-384.

[4] Simone F, Frattini F, Landoni P, Performance Measurement of Collaborative Research and Development: An Exploratory Analysis, *International Journal of Innovation and Technology Management*, Vol. 17, No. 6, 2020, pp. 2-22.

[5] Yu S, Li X, Wang H, Zhang X, Chen S, C_CART: An Instance Confidence-Based Decision Tree Algorithm for Classification, *Intelligent Data Analysis*, Vol. 25, No. 4, 2021, pp. 929-948.

[6] Imai N, Takeyoshi M, Aizawa S, Tsurumaki M, Kurosawa M, Toyoda A, Sugiyama M, Kasahara K, Ogata S, Omori T, Hirota M, Improved Performance of the SH Test as An in vitro Skin Sensitization Test with a New Predictive Model and Decision Tree, *Journal*

of Applied Toxicology, Vol. 42, No. 5, 2022, pp. 1029-1043.

[7] Vanveller B, A Decision Tree for Retrosynthetic Analysis, *Journal of Chemical Education*, Vol. 98, No. 8, 2021, pp. 2726-2729.

[8] Li X, Hu Y, Xue B, Zhang Z, Li L, Huang K, Wang Y, Yu Y, State-of-Health Estimation for the Lithium-Ion Battery Based on Gradient Boosting Decision Tree with Autonomous Selection of Excellent Features, *International Journal of Energy Research*, Vol. 46, No. 2, 2022, pp. 1756 -1765.

[9] Rafi M, Ming Z J, Ahmad K, Estimation of the Knowledge Management Model for Performance Measurement in University Libraries, *Library Hi Tech*, Vol. 40, No. 1, 2020, pp. 239-264.

[10] Kloutsiniotis P V, Katou A A, Mihail D, Examining the "Dark-Side" of High Performance Work Systems in the Greek Manufacturing Sector, *Employee Relations*, Vol. 43, No. 5, 2021, pp. 1104-1129.

[11] Jafar R M S, Geng S, Ahmad W, Niu B, Chan F, Social Media Usage and Employee's Job Performance: The Moderating Role of Social Media Rules, *Industrial Management and Data Systems*, Vol. 119, No. 9, 2019, pp. 1908-1925.

[12] Zhan X, Liu Y, Impact of Employee Pro-Organizational Unethical Behavior on Performance Evaluation Rated by Supervisor: A Moderated Mediation Model of Supervisor Bottom-Line Mentality, *Chinese Management Studies*, Vol. 16, No. 1, 2022, pp. 102-118.

[13] Tafuro A, Dammacco G, Esposito P, Mastroleo G. Rethinking Performance Measurement Models Using A Fuzzy Logic System Approach: A Performative Exploration on Ownership in Waste Management, *Socio-Economic Planning Sciences*, Vol. 79, 2022, pp. 10-23.

[14] Garengo P, Sardi A, Nudurupati S S, Human Resource Management (HRM) in the Performance Measurement and Management (PMM) Domain: A Bibliometric Review, *International Journal of Productivity and Performance Management*, Vol. 71, No. 7, 2022, pp. 3056-3077.

[15] Xiao H, Shi J, Yang Z, Wang B L, Shi W X, Shi W X, Li M, Precision Improvement Method for Onsite Performance Measurement of Variable Refrigerant Flow System, *Building and Environment*, Vol. 208, 2022, pp. 5-21.

[16] Korsen E B H, Ingvaldsen J A, Digitalisation and the Performance Measurement and Management System: Reinforcing Empowerment, *International Journal of Productivity and Performance Management*, Vol. 71, No. 4, 2022, pp. 1059-1071.

[17] Park H J, Lee C W, Shin T, Roh B H, Park S B, Choi J, Implementation of Embedded Testbeds Using USRP and GNU-Radio for Performance Measurement and Analysis of PPS and PCO-Based Time Synchronizations, *International Journal of Interdisciplinary Telecommunications and Networking*, Vol. 13, No. 1, 2021, pp. 25-35.

[18] Nallusamy S, Majumdar G, Performance Measurement on Inventory Management and Logistics Through Various Forecasting Techniques, *International Journal of Performability Engineering*, Vol. 17, No. 2, 2021, pp. 216-228.

[19] Dey P, Jana D K, Evaluation of the Convincing Ability Through Presentation Skills of Pre-Service Management Wizards Using AI via T2 Linguistic Fuzzy Logic, *Journal of Computational and Cognitive Engineering*, Vol. 2, No. 2, 2022, pp. 133-142

[20] Neto G C D O, Tucci H N P, Filho M G, Performance Evaluation of Occupational Health and Safety in Relation to the COVID-19 Fighting Practices Established by WHO: Survey in Multinational Industries, *Safety Science*, Vol. 141, No. 1, 2021, pp. 10-22.

[21] Hou S, Research on the Application of Data

Mining Technology in the Analysis of College Students' Sports Psychology, *Hindawi Limited*, Vol. 2021, No. 10, 2021, pp. 8-16.

[22] Woo H, Kim I, Choi B, Computer Vision Techniques in Forest Inventory Assessment: Improving Accuracy of Tree Diameter Measurement Using Smartphone Camera and Photogrammetry, *Sensors and Materials*, Vol. 33, No. 11, 2021, pp. 3835-3845.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The author stated that the project was done independently by the author from the first draft to the final draft.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This research received no external funding

Data Availability Statement

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.