Research Article

Construction and Application of Enterprise Internal Audit Data Analysis Model Based on Decision Tree Algorithm

Yuna Si [ID]

School of Accounting, Xijing University, Xi'an, Shaanxi 710123, China

Correspondence should be addressed to Yuna Si; 20140030@xijing.edu.cn

With the systematization, informatization, and intelligence of the management and operation of communication enterprises, the amount of audit content and business data is increasing. The traditional audit work is mainly offline audit, the task is heavy and temporary, there are few auditors, and there is a lack of unified data mart and analysis model support. In order to solve many problems in traditional auditing, this paper realizes the interface of data dictionary attributes through effective domain management of multisystem data, so as to achieve the purpose of system support and effective management to improve data quality. Furthermore, the platform is extensible, which can gradually expand the data coverage and support more management fields and business systems. A method of obtaining continuous attribute clustering center points by K-means algorithm and using triangular fuzzy numbers to fuzzify process continuous data is proposed. At the same time, a visual fuzzy decision-making system based on Fuzzy ID3 and Min-Ambiguity algorithm was designed. Adding the C4.5 and CART algorithms implemented in the Weka open source data mining software, through experimental analysis, the differences in the classification accuracy and the number of rules generated by the four decision tree algorithms are compared. The experiment found that the Fuzzy ID3 algorithm has a high accuracy rate on each dataset and the number of rules is large. The number of rules generated by the CART algorithm is the least because of the model characteristics of its binary tree and the characteristics of using the Gini index as the criterion for the selection of splitting attributes. Comparing Fuzzy ID3 and Min-Ambiguity fuzzy decision tree algorithms, it is found that the overall performance of the former is better than that of the latter, and the effect of authenticity on the two algorithms is analyzed experimentally. This paper constructs the internal audit flowchart in the big data environment and studies the optimization of the internal audit process in the big data environment. We evaluate traditional internal audit processes using process optimization theory. The internal audit process in this paper refers to the internal audit process under the risk-oriented model that is widely used at this stage. The optimization research on the internal audit process in this paper is mainly aimed at the internal audit process of general enterprises with a certain level of informatization, at a large scale, and certain foundation and requirements for the analysis and use of data. The optimization of internal audit process is feasible in research and practical application and has universal applicability and reference value.

1. Introduction

In the era of Internet 4.0, the amount and speed of data generation have grown exponentially, and the wave of informatization has swept across all walks of life [1]. At the end of the last century, McKinsey & Company first proposed the concept of "big data," and the era of big data began [2]. With the continuous updating and expansion of big data technology, the application scope and value of big data have been continuously enlarged [3]. Governments, corporate

executives, experts, and scholars in various fields have devoted themselves to the research of big data [4]. The big data audit is included in the focus of audit informatization work, which provides policy support and direction guidance for the promotion and development of big data audit work. With the continuous development and deepening of big data technology, data resources are constantly being explored. Massive data provides a broader vision for audit work, enriches audit data sources, strengthens auditing early warning and risk control functions, and improves enterprise

operation and management [5]. In terms of enterprise internal audit, the level of informatization is relatively low. Although there are enough theoretical foundations and practical cases, for various reasons, the road to enterprise informatization transformation is difficult. With the continuous expansion of the size of the enterprise, the scale of data has increased sharply, and the types of data that can be obtained have increased [6]. The traditional informatization method of audit work is increasingly unable to meet the needs of the company's internal audit, and it is difficult to collect and analyze the huge and complex data. The audit work cannot be carried out in depth, the audit efficiency is difficult to improve, the work results are affected, and it is difficult to meet the work functions of improving operational efficiency, reducing operational risks, and assisting the realization of organizational goals for the enterprise.

For enterprises, the application of big data is both an opportunity and a challenge. At present, enterprises have little experience in the application of big data [7]. The cost of building an informatization platform for internal audit is high; the quality of auditors is not high. It is difficult to advance, and many scholars have proposed ideas or construction ideas for the application of big data in internal audit [8]. Based on this, this article combines actual cases, from the design to construction of internal audit informatization, as well as the use effect and the next development trend. It is hoped that it can provide useful reference for enterprises that have not yet carried out informatization construction or data analysis platform in the construction of internal audit informatization system and data platform. At the theoretical level, by analyzing the impact of big data on internal audit informatization, the development direction and transformation path of internal audit informatization in the era of big data are clarified, and the theoretical research on big data auditing in the field of audit informatization is expanded. The research theories related to informatization keep pace with the times and develop continuously [9]. At the application level, by comprehensively considering the actual situation of the company, this paper guides enterprises to upgrade the original audit informatization model, realize large-scale data audit information transformation, and ultimately improve the efficiency and effectiveness of internal audit.

The audit data mart is the collection and storage center of audit data, providing data services and audit application support for groups, provincial companies, etc. It is the soul of the audit system, so it is particularly important to build an audit data mart, so as to achieve "early availability of data." This paper collects the required business, financial, engineering, and other data and realizes the interface of the data dictionary attributes through effective domain management of multisystem data, so as to achieve the purpose of system support and effective management to improve data quality. This paper describes the fuzzy decision tree in detail; explains the direction of its improvement; and focuses on the definitions of two fuzzy decision tree algorithms (Fuzzy ID3 algorithm and Min-Ambiguity algorithm), the selection rules for splitting attributes, and the algorithm flow. This paper mainly focuses on the experimental analysis of C4.5

algorithm, CART algorithm, Fuzzy ID3 algorithm, and Min-Ambiguity algorithm. Using the Weka software and the fuzzy decision-making system implemented in the Java environment, the experiment on the selected dataset in the UCI database is completed, and the differences in the classification performance of the four algorithms are summarized. It can be seen that the Fuzzy ID3 algorithm has great advantages in classification accuracy and number of rules. This paper studies the internal audit process in the big data environment and optimizes the traditional internal audit process to a certain extent based on the current development trend of enterprise internal audit. This paper mainly uses the ideas and methods of the process optimization theory to diagnose, optimize, and analyze the traditional internal audit process according to the optimization steps and finally constructs the internal audit flowchart in the big data environment. The application process and results of the traditional internal audit process and the internal audit process in the big data environment are compared and analyzed to evaluate the feasibility and value of the research results.

2. Related Work

The most influential decision tree method is the ID3 algorithm, and its core content is information entropy (the size of the information contained in the message) [10]. The division attributes are selected by calculating the difference of the two information amounts, i.e., the information gain, as a criterion. The ID3 algorithm considers that the attribute with high information gain is a good attribute, selects the attribute with the highest information gain as the division criterion for each division, and repeats this process until a decision tree that can perfectly classify the training set is generated. Through a large number of experimental studies, we have learned that the ID3 algorithm tends to select attributes with a large number of attribute values as the division attributes of the sample, but attributes with a large number of attribute values are not necessarily the best division [11].

Relevant scholars have conducted research on informatization data collection [12–14]. When internal auditors conduct audits, in order to avoid allowing test data to enter the system, parallel simulations are used for testing to avoid affecting the formal operating environment of the enterprise [15]. When auditing financial procurement, it is necessary to collect and clean financial and procurement data, enter the audit system environment, and summarize several common data collection and cleaning methods [16].

Thomas Porter, an American computer audit expert, became anxious and began to pay attention to the research on audit methods and methods in the environment of information development [17]. The content that needs to be audited inside the enterprise is becoming more and more complex. The management information systems used by enterprises are becoming more and more complex, so the internal auditors of enterprises are required to conduct system audits to ensure that the information-based management system can operate effectively [18, 19]. Relevant

scholars have summarized computer-aided tools and auditing techniques that can help auditors improve efficiency during auditing [20].

Relevant organizations and scholars in the United States started their research on internal audit informatization earlier and in-depth [21]. After the Enron incident, the United States attached great importance to the role of internal audit, and the norms and regulations for internal audit were also relatively complete [22]. With the help of information technology, internal auditing has developed rapidly. However, foreign research is more personalized [23]. Different companies develop their own internal auditing information systems, which do not have commonalities and lack clear information. For the reference of system design, domestic scholars refer to relevant foreign theories combined with the relevant knowledge of computer science to formulate relevant specifications for computer auditing process and information system auditing, and standardize the auditing process of the internal audit information system [24]. The design of audit informatization system provides a reference. Despite the in-depth theoretical research on audit informatization, there are still interdisciplinary boundaries in the combination of the two. Studies in the theoretical circle are currently based on the concept of big data and cloud computing [25]. Some ideas are put forward for the breakthrough of the internal audit information system, but there are few successful models in the practical field.

3. Methods

3.1. Enterprise Internal Audit Data Architecture Design.
Through the big data platform extraction, integration, processing, and collection of various business platform data items required for auditing, big data computing tools are used for market construction, and distributed MPP database is used for model processing and self-service analysis.

First, it realizes the heterogeneous data integration, cross-system process scheduling, and business integration required for auditing; second, it creates standardized data services and service orchestration capabilities to realize the open sharing of data capabilities; third, the idea of data-+process improves the capture and efficient processing of real-time data and breaks the barriers to data acquisition by auditing.

In order to better manage and use business data, an audit data mart is built based on the enterprise data mart. The enterprise data mart extracts the interface data uniformly, integrates the source data to support the scanning of the intermediate table, and processes and generates it on the intermediate table. The data architecture structure design description is as follows:

(1) Data source

The audited data mart mainly comes from the BSS, MSS, OSS and some manual data of the enterprise, and the data is extracted from these data across domains and systems.

(2) Audit data mart

It mainly includes the basic wide table and the integration model. The basic wide table is mainly the basic data table required for processing and integrating the data extracted by the interface and formal auditing. The integration model is based on the basic data table and business model; analyzes, processes, mines, and explores process data; and forms process data such as risk scanning model and management and responsibility risk.

(3) Audit market application

The audit marketplace is mainly used for providing application service result data, including risk scanning, thematic audits, audit marketplaces, off-site audits, and project auditing applications.

3.2. Audit Data Processing Method.
In order to centralize the data of different data sources, data processing can be better used after redefining and sorting out. Because the data is collected from different data sources, it involves the process of cross-database and cross-system data integration and processing. The audit business process flow is shown in Figure 1.

Data processing refers to extracting, cleaning, merging, and summarizing data from different systems and storing the final results in the database service library by business classification and hierarchical concentration, so as to facilitate the query and use of other systems. The data processing process designed this time includes the following:

(1) Market sorting: according to the needs of the audit business, a set of standardized data mart tables have been sorted out

(2) Data acquisition: according to business needs, different source data systems are opened up, and the original data is extracted from the data source to the local temporary database table

(3) Data integration: based on the time and geographical dimensions and combined with the sorted market, the local temporary data is cleaned, merged, aggregated, classified, etc., and the data market is output

(4) Data service: the big data technology is used to quickly respond to the support of the data, and the integrated data is automatically extracted through ETL tools; the data processing results are classified and stored in the formal database table in layers, so as to provide efficient auditing

3.3. Data Structure Design.
The audit data mart is the center of audit data storage and application. It is necessary to build an audit data mart to provide data services and intelligent support for audit applications for groups and provincial companies. In order to meet the needs of the group's cross-audit, the data mart principles designed this time are as follows: unified data specification (regular optimization), unified analysis model (unified management), and unified
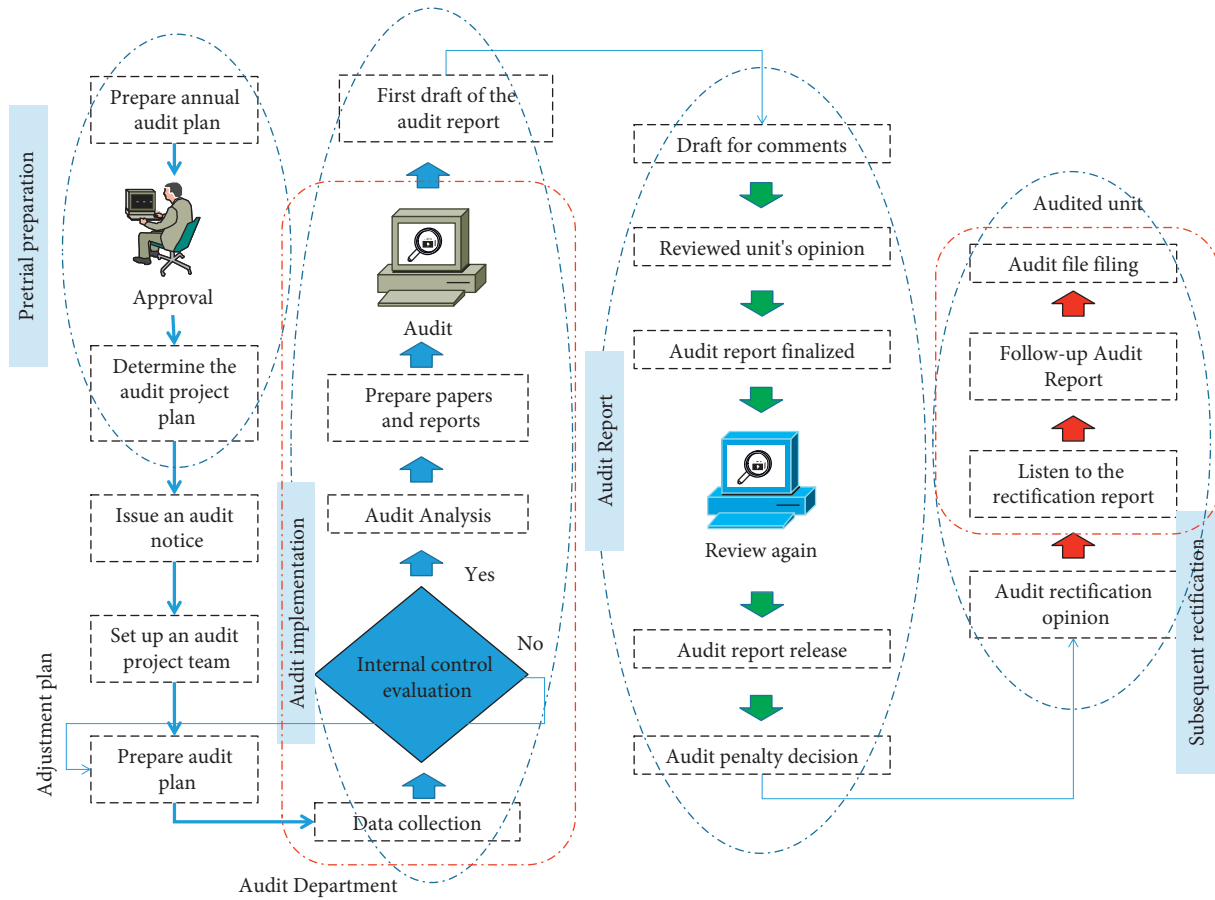
FIGURE 1: Audit business process flow.

interface interaction. Table 1 describes the structure of the audit market table and data table.

3.4. Fuzzy Decision Tree Analysis. A decision tree is constructed from a set of data that contains categorical attributes. Usually, the data that will be used for modeling will be divided into two groups. Part of it is used as training data to generate the tree model, while the other part is used as validation data to test the classification accuracy of the model. The nodes in the constructed model can be regarded as a collection of attributes, and the leaf nodes represent the final classification result of an instance. A decision tree model is essentially a set of rules. The path from the root node to each leaf represents a classification rule, and these rules are grouped together through a hierarchical relationship.

Decision tree algorithm is a heuristic greedy algorithm, and its tree-building process also shows a "divide and conquer" idea. When generating a decision tree, if all instances in a node belong to the same class, it is set as a leaf node; otherwise, an optimal attribute is selected as the node attribute according to a certain rule. Figure 2 shows the tree-building process of the fuzzy decision tree algorithm. The fuzzy improvement of the classical decision tree algorithm mainly includes the following contents.

The first is the preprocessing of continuous attributes, that is, how to fuzzify continuous attributes. For most fuzzy decision tree algorithms, it is necessary to fuzzify the continuous attributes before modeling, and a few algorithms fuzzify the data during the modeling process.

The second is the selection rule of splitting attributes. Compared with the clear decision tree split attribute selection rule, the fuzzy decision tree algorithm extends it so that it can adapt to fuzzy data.

The third is the matching rule of the decision tree. A fuzzy decision tree will give the degree to which the test data belongs to a certain classification, that is, the embodiment of attribute membership, rather than an absolute classification like a clear decision tree.

Through the fuzzy extension of the decision tree algorithm, its application range is extended from the field of precise data to the field of fuzzy data. For those problems with uncertainty and ambiguity, this improved algorithm's representation of knowledge is more realistic.

In most cases, the fuzzy decision tree considers two parameters, the significance level α and the true degree β, when constructing the tree model. The significant level α is the cut set of the fuzzy set A, which is generally used for data preprocessing before the algorithm is executed. The larger the α, the less the ambiguity of the data, and if it is too large, the training set will be empty. Usually, α is not needed if the

TABLE 1: Example of audit mart.

| Serial number	Audit name	Table name
1	Wide table of income details	SJ_BWT_ACCT_BALANCE_PAYOUT_M
2	List of arrears	SJ_BWT_FIN_OUT_INCOME
3	Write-off flow meter	SJ_BWT_SUBSIDY_SHLD_M
4	Terminal subsidy table	SJ_BWT_CUST_POINT_EXCHANGE_M
5	Balance expenses	SJ_BWT_CUST_POINT_M
6	Commission user level data	SJ_BWT_CUST_COIN_BALANCE
7	Package information	SJ_BWT_ACCT_WOFF_ACC_BILL_M
8	Statement of bills	SJ_BWT_PRD_PD_INST_M
9	Account information	SJ_BWT_CHARGE_LIST_M
10	Three-year change schedule of points	SJ_BWT_FIN_IN_INCOME
11	Customer points information	sj_bwt_cust_info_m
12	Points redemption schedule for three years	SJ_BWT_FIN_OWE_LIST_M
13	In-system income detailed table	SJ_BWT_SERV_SETTLE_DETAIL
14	Customer information	SJ_BWT_FIN_OWE_M
15	Arrears table	sj_bwt_acct
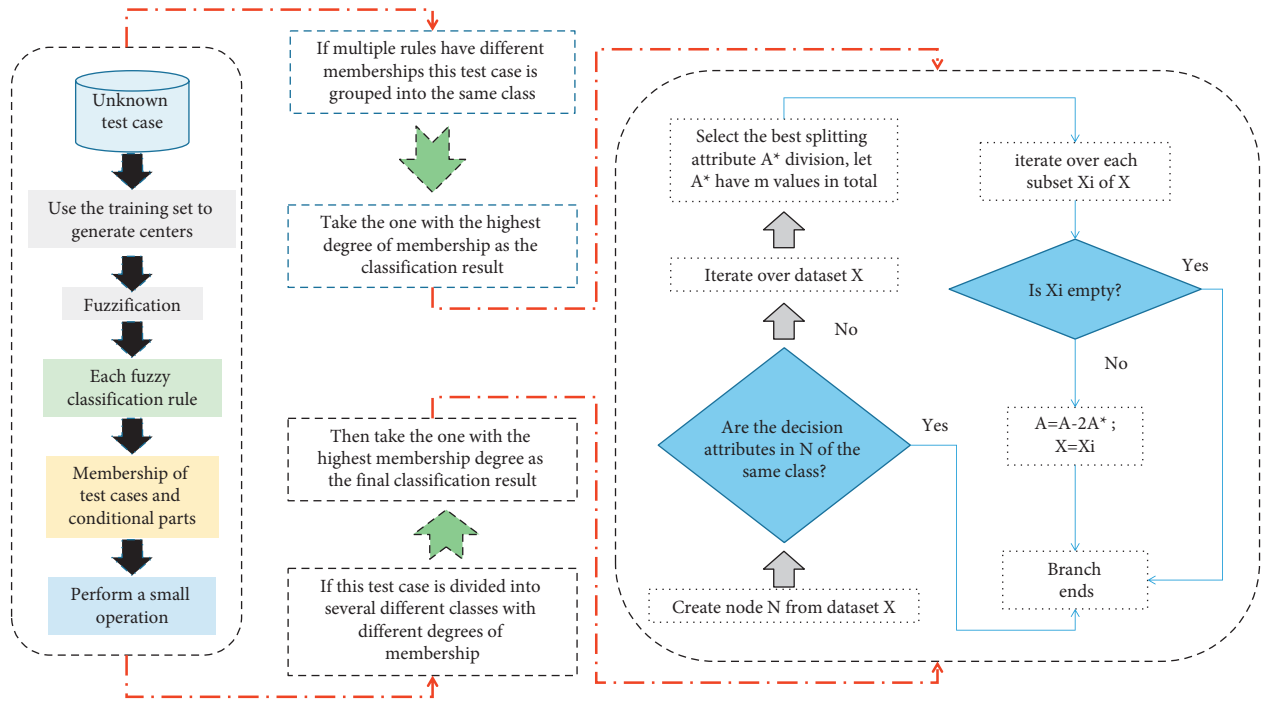
FIGURE 2: Flowchart of fuzzy decision tree algorithm.

model performs as expected. For node A, the truth degree of its attribute classification C is as follows:

$$\beta_{A,C} = \frac{\left[\prod u_A(x) \cap u_C(x)\right]}{\prod_{x \longrightarrow X} x \longrightarrow X}. \quad (1)$$

Here, M represents the cardinality measure, and the true degree β is used to terminate the generation of the tree. When the true degree of the node exceeds the given threshold β, the construction of this branch ends and leaves are generated.

The model constructed by the fuzzy decision tree algorithm is the same as the ordinary tree model. The branches from the root node to the leaf node represent the rules of classification. Different from the rules generated by the

classical decision tree, a rule in the rule set generated by the fuzzy decision tree may eventually divide the instances into different categories with different degrees of membership, and there may also be multiple rules matching the same instance. Therefore, special treatment should be taken when using the model for classification. This article uses the rule matching method based on the maximum and minimum operations.

3.5. Design of Fuzzy Decision Tree Algorithm. With the widespread application of fuzzy set theory in decision tree algorithms, many excellent algorithms have been proposed one after another. This section will introduce two representative fuzzy decision tree algorithms. No matter which

decision tree algorithm is used, the generated decision tree structure is generally similar, the nodes are composed of each attribute name, and the edges are composed of fuzzy subsets whose attribute values are fuzzified.

3.5.1. Fuzzy ID3 Algorithm.

The Fuzzy ID3 algorithm introduced in this paper takes the average fuzzy information entropy as the criterion for selecting splitting attributes. The overall tree-building process of the Fuzzy ID3 algorithm is similar to that of the ID3 algorithm. The core difference lies in the selection rules for splitting attributes. Compared with the concept of information entropy in ID3 algorithm, Fuzzy ID3 algorithm uses fuzzy information entropy of probability distribution as the selection criterion of splitting attribute. For each optional attribute value $T_{i,k}$, the relative frequency of the j-th fuzzy category $T_{j,n+1}$ on the non-leaf node S is defined as follows:

$$p_{ij,k} = \frac{M\left(S \cap T_i^k \cup T_{j-1}^n\right)}{M\left(T_{j-1}^n \cup S\right)}. \tag{2}$$

For non-leaf node S, the semantic value $T_{i,k}$ is defined as follows:

$$I\left(T_{i,k}\right) = \prod_{j=0}^{m-1} \left[p_{ij,k} \cdot \log\left(p_{ij,k}\right)\right]. \tag{3}$$

The average fuzzy classification information entropy of attribute A_k is defined as follows:

$$E(S, A_k) = \prod_{i=0}^{m_k} \left[p_i \cdot I\left(T_{i,k}\right)\right], \tag{4}$$

where

$$p_i = \frac{M\left(T_{i,k} \cup S\right)}{\prod_{i=0}^{m_k-1} M\left(T_{i,k} \cap S\right)}. \tag{5}$$

When the Fuzzy ID3 algorithm builds the tree model, it always selects the attribute with the smallest average fuzzy information entropy as the splitting attribute; that is, the k-th attribute is preferentially selected to satisfy

$$E(S, A_{k'}) = \min_{k \longrightarrow (1,n)} E(S, A_k). \tag{6}$$

In the same algorithm, the scale of the model is controlled by the degree of realism.

3.5.2. Decision Tree Algorithm Based on Min-Ambiguity.

The Fuzzy ID3 algorithm still uses the concept of entropy, but the original formula is improved, and the concept of fuzzy information entropy is proposed to select the split attribute, so the algorithm will generate a fuzzy tree model. The heuristic decision tree algorithm based on Min-Ambiguity, referred to as Min-Ambiguity algorithm, will be introduced in this section. This method uses classification uncertainty as the basis for the selection of split attributes, which is closer to the way of human thinking and has high comprehensibility. For each optional attribute value $T_{i,k}$, the

normalized representation of the j-th fuzzy category $T_{j,n+1}$ on the non-leaf node S is defined as follows:

$$\tau_{ij,k} = \frac{p_{ij,k}}{\prod_{j=0}^{m-1} p_{ij,k}}. \tag{7}$$

For a non-leaf node S, the classification uncertainty for each optional attribute value $T_{i,k}$ is defined as follows:

$$\text{Ambig}\left(T_{i,k}\right) = \prod_{j=0}^{m-1} \ln\left(\pi_{ij}^{k-1} - \pi_{i+1,j}^k\right). \tag{8}$$

For non-leaf nodes S, the average uncertainty of attribute A_k is defined as follows:

$$G(A_k) = \prod_{j=0}^{m-1} (1 - p_i) \cdot \text{Ambig}\left(T_i^{k+1}\right). \tag{9}$$

The Min-Ambiguity decision tree algorithm selects the attribute with the smallest average classification uncertainty as the splitting attribute each time. During data preprocessing, the significance level α is applied to filter the data, and the trueness parameter β is applied to determine the leaf nodes. The specific process of the algorithm is basically the same as that of the Fuzzy ID3 algorithm, but the criteria for selecting split attributes are different.

4. Results and Analysis

4.1. Algorithm Accuracy Comparison.

By applying the C4.5 algorithm, the CART algorithm, the Fuzzy ID3 algorithm, and the Min-Ambiguity algorithm to the 7 groups of preprocessed datasets, the correct rates of different algorithms in the classification of each dataset can be obtained as shown in Table 2.

The Fuzzy ID3 algorithm has the strongest adaptability to different data and achieves the highest accuracy rate in the dataset. The C4.5 algorithm and the Min-Ambiguity algorithm have the highest accuracy in a group of datasets. The CART algorithm has the worst average accuracy.

Although the average classification accuracy of the C4.5 algorithm and the CART algorithm on the 7 groups of datasets is relatively low, they still have 80.02% and 78.93% accuracy, indicating that the C4.5 and CART algorithms are constantly dealing with continuous attributes. The recursive greedy algorithm of bisection continuous data items with optimal information gain rate and minimum Gini index can also produce higher accuracy.

Although the fuzzy decision tree has an advantage in dealing with continuous attributes, the average accuracy of the Min-Ambiguity algorithm is not much higher than that of C4.5 and CART. In particular, the C4.5 algorithm is only 0.37% less accurate than the Min-Ambiguity algorithm on average. This shows that although fuzzy decision tree has certain advantages in dealing with continuous attributes, the heuristic of split attribute selection also has an important impact on the final result. Observing the 7 groups of datasets, it can be found that the four algorithms have achieved the highest accuracy, indicating that the attributes of the dataset are high cohesion and low coupling, the differences between

TABLE 2: Comparison of classification accuracy of the four algorithms.

Number of categories	CART algorithm (%)	C4.5 algorithm (%)	Min-Ambiguity algorithm (%)	Fuzzy ID3 algorithm (%)
4	78.32	79.33	87.44	90.40
5	74.41	75.82	86.93	91.71
8	77.94	80.52	83.21	90.22
3	72.53	81.40	88.61	93.54
5	73.82	74.91	84.13	96.10
6	76.20	71.24	87.82	94.21

different categories are more obvious, and the classification is clear.

4.2. Comparison of Algorithm Rules.

The model generated by the decision tree algorithm is actually a collection of rules, and the number of rules directly affects the efficiency of decision-making. With the construction of the decision tree, the number of leaf nodes will continue to increase, and more and more decision rules will be generated. In theory, the more the rules, the higher the classification accuracy. However, too many rules often cause overfitting, which means that the classification effect of the training set will be better, but the prediction accuracy of the test set will decrease. Figure 3 shows the classification accuracy of the training and test datasets under different numbers of rules for a decision tree.

The number of rules generated by the decision tree is an important indicator for judging whether the algorithm is good or bad, as is the accuracy of the model. A high-performance decision tree should use as few rules as possible to obtain a higher decision accuracy rate. The statistics of the number of classification rules generated by the four algorithms are shown in Figure 4.

Overall, it is clear that the CART algorithm produces the fewest number of rules in each dataset. There are two reasons through analysis. First, the model generated by the CART algorithm is a binary tree, so there are fewer leaf nodes; in addition, the CART algorithm uses the Gini coefficient as the index for the selection of the split attribute, which makes it easier to achieve the required purity standard when building a tree. For these two reasons, fewer rules are generated.

The number of rules generated by the four algorithms on the Balance-Scale dataset is significantly higher than the number of rules generated by these algorithms on other datasets, indicating that the dataset is more complex and the coupling between decision attributes is strong. That is, it is difficult to determine the classification result with fewer attributes.

4.3. Algorithm Authenticity Analysis.

The size of the real degree is closely related to the size of the decision tree, so choosing a suitable real degree is very important for the fuzzy decision tree algorithm. Through the experiments of the four decision tree algorithms on different datasets, it is found that the Iris dataset has strong categorizability. In this section, the Iris dataset is used as an example to illustrate the impact of realism on fuzzy decision trees.

By analyzing Figure 5, it can be seen that for the fuzzy decision tree, with the increase of authenticity, the correct rate of classification is not less than 85%. Moreover, the Fuzzy ID3 fuzzy decision tree algorithm is more accurate than the Min-Ambiguity fuzzy decision tree algorithm.

By analyzing Figure 6, it can be seen that with the increase of authenticity, the number of rules generated by the model is increasing. This is because the greater the degree of authenticity, the more the recursion times of each branch, so the scale of the spanning tree is larger, and more rules are obtained. Therefore, in general, the authenticity is very important for fuzzy decision trees.

4.4. Analysis of the Impact on Internal Audit Planning Phase.

The audit plan is a strategic plan that is in a guiding position in an audit project and has an impact on the audit efficiency and audit results. The implementation and management of the audit plan is an inevitable requirement for the sound development of the audit work, and it is also the realization of the audit objectives and the effect of the audit implementation. In the traditional internal audit process, due to the contradiction between the resources required by the internal audit to realize its functions and the cost of enterprise management, the importance of the audit plan as a guiding link in the audit activity is self-evident. Under the risk-oriented mode, the content of audit is not limited to the inherent risks in the financial statements of the enterprise and the internal control risk of the enterprise, but it also includes the audit of the overall risk management content of the enterprise. In the new development situation of the big data environment, it is also necessary to pay attention to the impact of big data on internal audit and the changes in audit risk.

In the current audit plan management, problems such as the duplication and lack of the coverage of the plan, and the lack of the overall grasp and allocation of resources and tasks in the plan are more prominent. In general, the audit plan management has certain randomness and subjectivity in the actual implementation process and lacks long-term and scientific unified planning. Not only has the establishment of the audit project not formed a perfect scientific system, but in actual operation, due to the lack of effective control in the formulation of the plan, it is difficult to meet the actual flexible needs during the implementation of the project. The proportion of various problems existing in audit plan management is shown in Figure 7.

The development of big data has made the problems in the management of the original audit plan more prominent, and the optimization of the audit plan has also received more
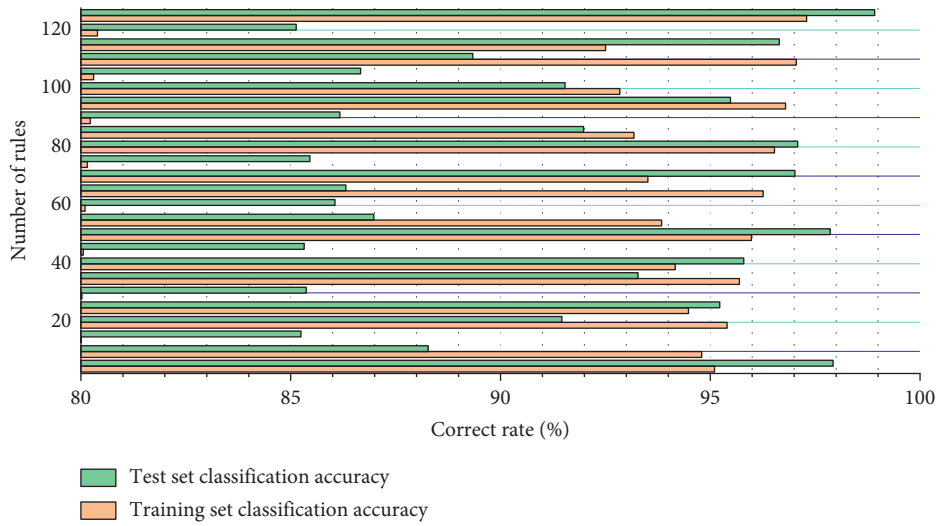
FIGURE 3: Comparison of classification accuracy of training and testing datasets under different rules of a decision tree.
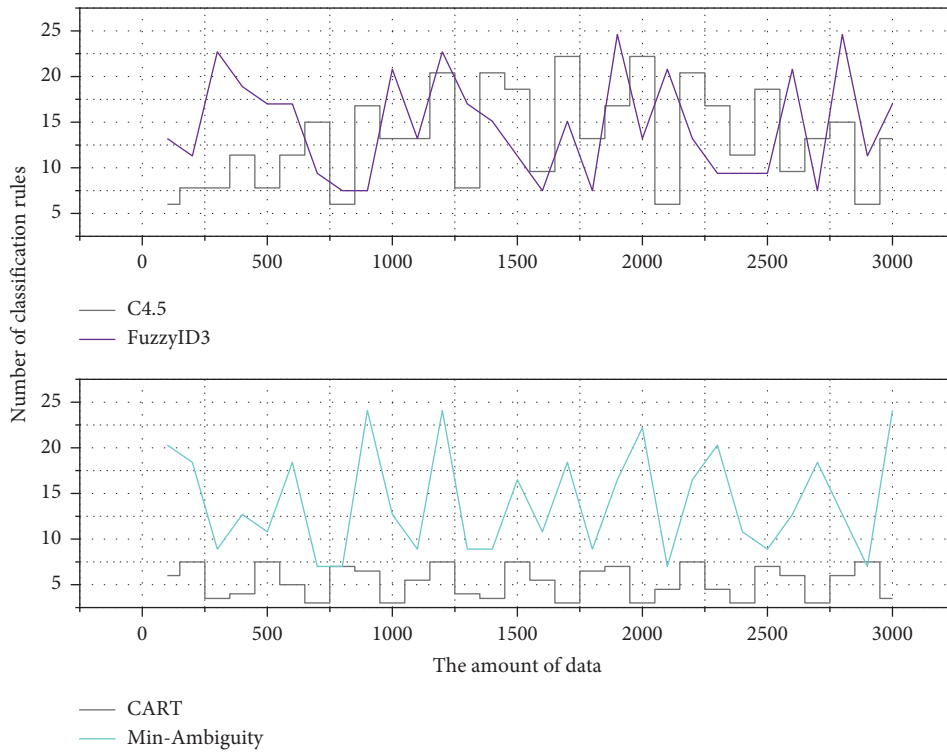

FIGURE 4: Statistics of the number of classification rules generated by four algorithms.

and more attention from the management. Changes in the environment and changes in the audit mode have prompted changes in audit objectives and processes. As the primary part of the audit process, it is more urgent to optimize the audit plan. In the big data environment, the problem solving in the existing audit planning stage should be combined with the current actual needs and the future development direction of the enterprise. The optimization of the internal audit process planning stage in the big data environment is not only to solve the current practical problems and to ensure the orderly and efficient implementation of the audit.

The method is also aimed at the development of the information age, meeting the practical needs of future auditing and achieving the basic requirements of the effectiveness of the audit function.

4.5. Analysis of the Impact on the Implementation Stage of Internal Audit. The implementation stage of traditional internal audit is mainly to carry out a series of tests and analysis on the audit project plan and collect audit evidence to provide material support and data support for the
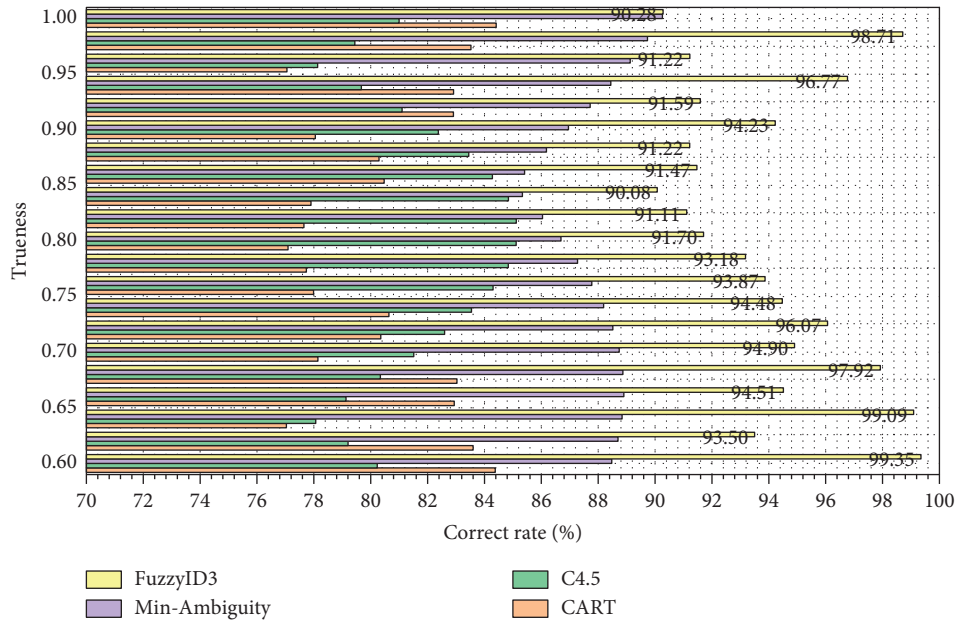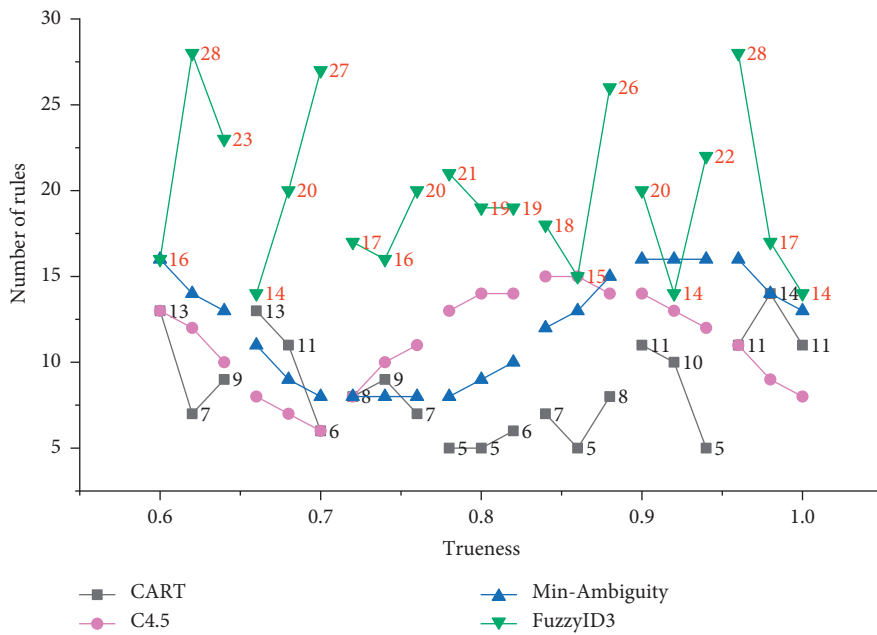
FIGURE 5: Influence of realism on accuracy.

FIGURE 6: The effect of authenticity on the number of rules.

preparation of audit working papers. Due to the limitation of audit resources and the limitations of audit technology and methods, one of the commonly used audit methods in the implementation stage of traditional audit process is sampling audit, which judges the evidence collection of auditors according to the control level of the audited unit and the acceptable audit risk level. The application of big data software and the establishment of data information platform make it possible to collect data more comprehensively and carry out multidirectional correlation analysis, and the breadth and depth of audit evidence can also be gradually enhanced with the development of information technology, which facilitates obtaining audit trails from a broader and comprehensive range of audit data.

It is difficult for the existing data platform functions to meet the data processing and analysis needs. The high cost of use and the difficulty in conducting a unified comparative analysis and correlation analysis without a unified data standard are two issues that experts agree on. The importance of algorithms to the internal audit implementation phase is shown in Figure 8.
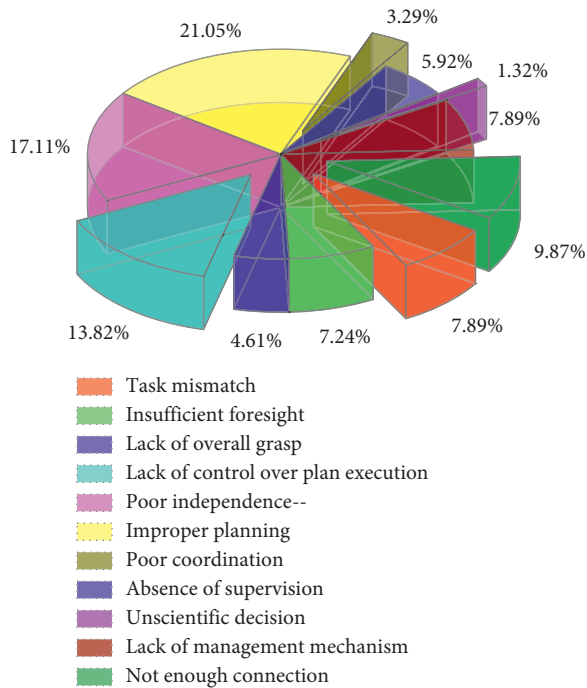
FIGURE 7: Proportion of various problems in audit program management.

FIGURE 8: Importance of algorithms to the implementation phase of internal audit.

In the big data environment, the emergence of new problems makes the original software development level and direction unable to meet the needs of internal auditing under the new trend, and it is difficult to carry out unstructured data analysis in the big data environment and expand the correlation analysis of business electronic data. In response to these questions, some practical suggestions are given in the questionnaire. Among them, the most recognized ones are the need to train auditors to improve their comprehensive ability, to introduce audit talents with computer professional background, and to standardize enterprise data resources with reference to national data standards or industry standards.

In the big data environment, the data volume of enterprises is more complex than before, and the cooperation of multiple departments and multiple business lines makes it more difficult to analyze related data. Especially for the internal audit of most private enterprises, due to the limitation of audit resources and the need for enterprise audit cost management, it is usually difficult to use the help of the enterprise internal audit cloud platform like government audit. The construction of enterprise information system and the use and upgrade of audit software all need to be carried out spontaneously by the enterprise, and the limitations brought by the cost and resource environment will be greater in comparison. It is generally difficult for the scale and level of internal audit departments of enterprises to meet the needs of data analysis in the big data environment. The lack of professional knowledge and technology of personnel and the lag of relevant standards and software development make enterprises have needs for the big data audit analysis stage.
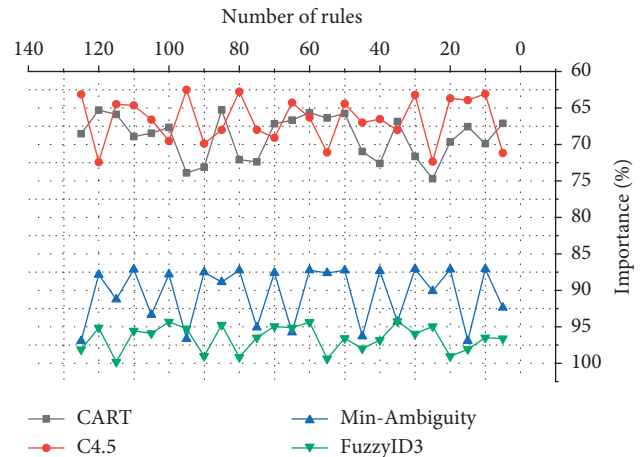
The improvement of the audit implementation stage should consider the actual needs of the current enterprise development, as well as the adaptability to changes in the internal and external environment of the enterprise, and the rapid development of information technology makes it necessary for enterprises to implement big data audits for risks and opportunities. The demand is placed on the long-term strategic development of the company. The optimization of the internal audit process in the big data environment can serve as a reference for the optimization of the internal audit process carried out by the company in practice, so that the internal audit can adapt to the actual business during implementation.

4.6. Analysis of the Impact on the Internal Audit Reporting Stage. After completing the audit on-site work, the internal audit team needs to communicate with the management of the audited unit according to the problems found and the corresponding countermeasures and suggestions put forward, and reach an agreement on the audit report, or get the audited unit for different opinions. The source of the content of the audit report is mainly based on the clues found in the audit implementation stage. The internal audit standards define the audit report as follows: after the internal auditors implement the necessary audit procedures for the audited unit according to the audit plan, they will audit the audited matters.

The impact of the big data environment on the audit report is reflected in the content and situation of the report results. In terms of the content of the audit report, since a large amount of data is used in the audit implementation stage and the data more represents the historical state of the audited unit, if the actual situation has a large unexpected change, then the conclusion of the audit report will be determined in the audit report. The formulation of the entire audit plan also depends on the big data environment. Due to the continuity of data generation and the unpredictability of events, the content of the audit report is likely to fail to cover the scope required by the audit objectives, resulting in the
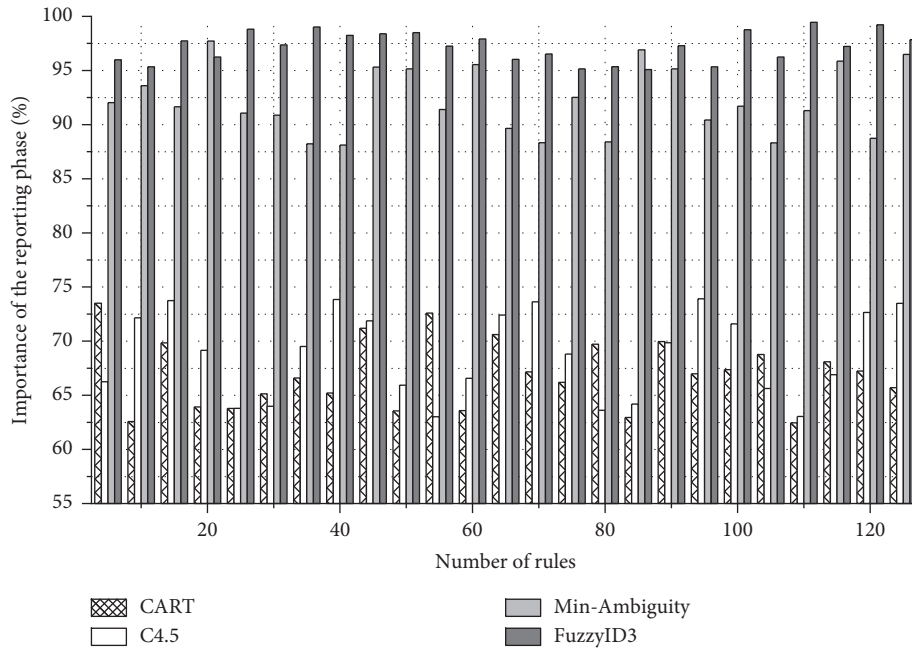
FIGURE 9: The importance of algorithms to the internal audit reporting stage.

distortion of the conclusions of the audit report. In addition, in the analysis stage of the audit process, sometimes when the requester of the report needs to read the content in detail, there will be interactive obstacles, making the information transmission insufficient.

Due to the application of information technology in the entire audit process, the preparation process of audit reports is also different from the simple situation in the past, and higher logic requirements and information system application requirements are put forward for the compilers and auditors. In addition, this fast and changeable information environment will inevitably increase the risk of information technology. This risk not only exists in the lack of audit platform settings and functional editing, but also in data leakage and data destruction during use. The importance of algorithms to the internal audit reporting stage is shown in Figure 9.

Technically, the implementation and transmission of audit data and continuous monitoring technology can also provide certain guarantees. The second is the use of visualization techniques in the report to express complex data models and data relationships in a clear and understandable way. Training auditors will help them to master the thinking and methods of big data auditing as a whole, help identify data relationships in auditing big data, and help auditors judge the laws of data and discover data trends, so as to further analyze deeper insights. In order to meet the needs of internal audit management, the internal audit department can also summarize some common and general report templates, flexibly meet the diversified needs of enterprises, and help improve the efficiency of auditing.

5. Conclusion

The audit data mart is the collection and storage center of audit data, providing data services and audit application support for enterprises, and is the soul of the audit system, so it is particularly important to build an audit data mart. Through the data mart research of telecommunication enterprises, the construction of audit data will need to integrate data resources as the basis, build a unified and standardized data mart, realize the unified aggregation of audit-related data, break down data barriers, and convert data collection to audit data. We build a complete set of audit data marts on the basis of basic data to provide services for analysis models, so as to realize timely and effective identification and prevention of enterprise risks, and provide guarantee for the effective implementation of enterprise strategies. The data source of the data mart is mainly based on the guidance of the group and the actual situation of the enterprise to collect the required business, financial, engineering, and other data. Two fuzzy decision-making algorithms, Fuzzy ID3 and Min-Ambiguity, are introduced. The difference between the classical decision tree algorithm and the fuzzy decision tree algorithm is summarized. We implement a fuzzy decision-making system, which has a good human-computer interface and integrates the mail classification model proposed in this paper. The differences in classification accuracy and number of rules of the four decision tree algorithms are compared through experiments, and the influence of authenticity on fuzzy decision trees is analyzed. Experiments show that the fuzzy decision tree algorithm, especially the Fuzzy ID3 algorithm, has certain advantages in dealing with

continuous attributes. In the audit planning stage, data analysis is carried out first, and then the audit plan is prepared; before the audit implementation stage, the risk assessment of the audited unit is carried out; in the audit implementation stage, the data analysis work is transferred to the off-site audit work. In the audit report stage, audit problems and audit suggestions are considered, monitoring data indicators are formed, and feedback is obtained in the follow-up audit stage with the help of the big data processing center, so as to strengthen the initiative of internal auditors to carry out work. In terms of organizational structure and human resource structure, the application of audit process optimization results requires a big data processing center as a data collection and analysis platform, as well as comprehensive auditors with a certain information technology foundation to implement and ensure the optimization process.

Data Availability

The data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

References

[1] B. F. Siggelkow, J. Trockel, and O. Dieterle, "An inspection game of internal audit and the influence of whistle-blowing," *Journal of Business Economics*, vol. 88, no. 7-8, pp. 883–914, 2018.

[2] R. A. Rahman, S. Masrom, S. Masrom, N. B. Zakaria, and S. Halid, "Auditor choice prediction model using corporate governance and ownership attributes: machine learning approach," *International Journal of Emerging Technology and Advanced Engineering*, vol. 11, no. 7, pp. 87–94, 2021.

[3] Y. Gholami, F. Mansouri, and H. Yazdifar, "Analyzing the relationship between social and professional identity characteristics of the audit committee and the steering system on the quality of financial reporting: a legal oriented artificial approach," *International Journal of Finance & Managerial Accounting*, vol. 6, no. 21, pp. 129–145, 2021.

[4] A. Ježovita, B. Tušek, and L. Žager, "The state of analytical procedures in the internal auditing as a corporate governance mechanism," *Management: Journal of Contemporary Management Issues*, vol. 23, no. 2, pp. 15–46, 2018.

[5] H. Kiral and H. Karabacak, "Resolution of the internal audit-based role conflicts in risk management: evidence from signaling game analysis," *Group Decision and Negotiation*, vol. 29, no. 5, pp. 823–841, 2020.

[6] M. P. Bach, K. Dumičić, B. Žmuk, T. Ćurlin, and J. Zoroja, "Internal fraud in a project-based organization: CHAID decision tree analysis," *Procedia Computer Science*, vol. 138, pp. 680–687, 2018.

[7] D. Appelbaum, D. S. Showalter, T. Sun, and M. A. Vasarhelyi, "A framework for auditor data literacy: a normative position," *Accounting Horizons*, vol. 35, no. 2, pp. 5–25, 2021.

[8] N. I. Mustika, B. Nenda, and D. Ramadhan, "Machine learning algorithms in fraud detection: case study on retail consumer financing company," *Asia Pacific Fraud Journal*, vol. 6, no. 2, pp. 213–221, 2021.

[9] N. Muñoz-Izquierdo, M. J. Segovia-Vargas, M. D. M. Camacho-Miñano, and D. Pascual-Ezama, "Explaining the causes of business failure using audit report disclosures," *Journal of Business Research*, vol. 98, pp. 403–414, 2019.

[10] A. M. Ali and J. Anis, "CEO's emotional commitment level and its firm capital structure choice: decision tree analysis," *Asian Journal of Economics and Empirical Research*, vol. 5, no. 1, pp. 65–78, 2018.

[11] N. Singh, K. H. Lai, M. Vejvar, and T. C. E. Cheng, "Data-driven auditing: a predictive modeling approach to fraud detection and classification," *Journal of Corporate Accounting & Finance*, vol. 30, no. 4, p. 121, 2019.

[12] D. M. M. Quintero and J. M. C. Flórez, "Model for evaluating the subjectivity of findings in audits of quality management systems," *Calitatea*, vol. 19, no. 167, pp. 36–42, 2018.

[13] D. A. Appelbaum, A. Kogan, and M. A. Vasarhelyi, "Analytical procedures in external auditing: a comprehensive literature survey and framework for external audit analytics," *Journal of Accounting Literature*, vol. 40, no. 1, pp. 83–101, 2018.

[14] C. Serbănescu and A. Vintilescu, "Efficiency and effectiveness in social assistance using corporate social responsibility and the economics of tagging," *Romanian Journal of Political Science*, vol. 18, no. 2, pp. 91–118, 2018.

[15] T. M. Alam, K. Shaukat, M. Mushtaq et al., "Corporate bankruptcy prediction: an approach towards better corporate world," *The Computer Journal*, vol. 64, no. 11, pp. 1731–1746, 2021.

[16] S. Y. Kim and A. Upneja, "Majority voting ensemble with a decision trees for business failure prediction during economic downturns," *Journal of Innovation and Knowledge*, vol. 6, no. 2, pp. 112–123, 2021.

[17] R. L. Sonza and G. M. Tumibay, "Decision tree algorithm in identifying specific interventions for gender and development issues," *Journal of Computer and Communications*, vol. 08, no. 2, pp. 17–26, 2020.

[18] Y. L. Hsu and G. C. Reid, "Two-stage decision-making within the firm: analysis and case studies," *Managerial and Decision Economics*, vol. 42, no. 6, pp. 1355–1373, 2021.

[19] G. Salijeni, A. Samsonova-Taddei, and S. Turley, "Big Data and changes in audit technology: contemplating a research agenda," *Accounting and Business Research*, vol. 49, no. 1, pp. 95–119, 2019.

[20] B. Xu, D. Huang, and B. Mi, "Research on E-commerce transaction payment system basedf on C4.5 decision tree data mining algorithm," *Computer Systems Science and Engineering*, vol. 35, no. 2, pp. 113–121, 2020.

[21] F. Yahya, K. Boukadi, Z. Maamar, and H. Ben-Abdallah, "A fuzzy-based approach for identifying candidate business processes for socialization," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 27, no. 1, pp. 19–39, 2019.

[22] J. Wang and H. Cheng, "Application of blockchain technology in the governance of executive corruption in context of

national audit," *Tehnički Vjesnik*, vol. 27, no. 6, pp. 1774–1780, 2020.

[23] C. Serrano-Cinca, B. Gutiérrez-Nieto, and M. Bernate-Valbuena, "The use of accounting anomalies indicators to predict business failure," *European Management Journal*, vol. 37, no. 3, pp. 353–375, 2019.

[24] A. Gepp, M. K. Linnenluecke, T. J. O'Neill, and T. Smith, "Big data techniques in auditing research and practice: current trends and future opportunities," *Journal of Accounting Literature*, vol. 40, no. 1, pp. 102–115, 2018.

[25] Y. Jiang and S. Jones, "Corporate distress prediction in China: a machine learning approach," *Accounting and Finance*, vol. 58, no. 4, pp. 1063–1109, 2018.