The Construction and Application of Library Intelligent Acquisition Decision Model Based on Decision Tree Algorithm

Hong Pan
Library, Yanbian University, Yanji, 133000, China

***Abstract*—In today's digital age, libraries, as the core institutions of knowledge management and information services, are facing an increasing demand from readers. In order to provide more efficient, accurate, and personalized interview services, intelligent interview decision-making in libraries has become an important research field. Traditional manual interview services face challenges such as personnel training and knowledge updates, making it difficult to quickly adapt to new needs and changes. To address these issues, research is being conducted on using machine learning technology to perform post pruning on the basis of standard decision trees and combining it with fuzzy logic to design a fuzzy decision tree. The experimental results show that the F rejection rate (FN) of the model rapidly decreases to about 0.1 as the number of training iterations gradually increases, and stabilizes at around 0.05 after 210 rounds of training, which is 0.10 lower than the rule-based decision model FN. The intelligent acquisition decision-making model designed in this study has higher accuracy and stability, and has certain application potential in the field of intelligent acquisition decision-making in libraries.**

***Keywords*—Decision tree; machine learning; fuzzy logic; intelligent interview model; post-pruning**

I. Introduction

With the continuous development of information technology and the rise of intelligent applications, library management is facing the demand for greater efficiency, convenience, and intelligence. Among them, intelligent acquisition decision-making in libraries is an important link, which involves analyzing, answering, and recommending services to readers. In order to improve the accuracy and efficiency of intelligent acquisition decision-making in libraries, many researchers use collaborative filtering algorithms, which predict books that users may be interested in by analyzing their historical behavioral data. The disadvantage is that it requires a large amount of user behavior data, and the algorithm's recommendation efficiency is not high when new books or new users join (cold start problem) . Some scholars also use hybrid recommendation systems that combine collaborative filtering, content recommendation, and other recommendation methods to compensate for the shortcomings of a single algorithm. However, hybrid recommendation systems may become complex and difficult to manage, and parameter tuning is a challenge [1]. The decision tree model has higher interpretability compared to other machine learning models, such as neural networks or support vector machines [2]. They make decisions through a series of easily understandable rules, enabling library staff to understand the recommendation logic of the model. It achieves data classification and prediction by dividing the dataset into different subsets and continuing to recursively construct decision rules on each subset [3]. In intelligent interview decision-making in libraries, decision tree algorithms can construct corresponding decision rules based on the characteristics of the questions raised by readers, helping the library system respond quickly and accurately to reader needs. However, standard decision trees have defects such as poor processing of continuous data, sensitivity to changes in input data, and susceptibility to overfitting. To address these issues, a fuzzy decision tree (FDT) is designed based on standard decision trees and combined with fuzzy logic, and applied to intelligent acquisition decision-making in libraries. It is expected that optimizing the existing decision tree model will help improve the quality and accuracy of library system interview decision-making. The article mainly consists of four parts. The second part is a review of the current research status on decision trees and fuzzy logic both domestically and internationally. The third part establishes an intelligent acquisition decision model for libraries based on improved decision trees. The fourth part conducts comparative experiments and applicability experiments on the optimization effect of the model.

The novelty of the article lies in the following points. First, the model considers many factors, such as book value, purchase funds, readers' needs and collection structure, and makes acquisition decisions more consistent with the actual needs of library services by building a more comprehensive decision-making framework. Secondly, using information gain as the criterion for feature selection ensures that the model focuses on variables that have a significant impact on classification, such as author, category, price, publisher, and publication time, thereby improving the prediction accuracy of the model. Thirdly, the C4.5 method in the decision tree algorithm was used and pruned to avoid overfitting, thereby optimizing the model's generalization ability. Fourthly, by using fuzzy theory to deal with uncertainty and ambiguity problems, the traditional decision tree algorithm has been enhanced with the ability to handle fuzzy classification, enabling the model to more finely depict real-world situations, especially in situations where user needs are fuzzy or library resource descriptions are unclear. Fifthly, in the process of constructing a decision tree model, by designing fuzzy sets and membership functions, combined with fuzzy logic, the model can better handle the uncertainty in classification.

This article demonstrates the broad application prospects of artificial intelligence and machine learning technology in library work through the construction and application research of a library intelligent acquisition decision model based on decision tree algorithm. In the future, with the continuous progress of technology and the accumulation of data, intelligent interview decision-making models will play an increasingly important role in the actual work of libraries. Meanwhile, the research methods and achievements of this article can also provide reference and inspiration for intelligent decision-making in other fields.

II. Related Works

Recently, with the widespread application of machine learning in various industries, more and more people have begun to use machine learning to solve various difficulties in social learning. Charbuty et al. designed a decision tree based author information classifier to address the issue of low accuracy in traditional author topic classifiers. The experiment showcased that the classification accuracy of the model reached 98.65% [4]. Aldino and other researchers designed a classifier using the decision tree C4.5 algorithm to make it easier for management to determine who is the appropriate student to receive financial aid, and conducted tenfold cross validation on the classification results; The experiment showcases that the accuracy, precision, and recall of the model are all 87%, which means that the model can be well implemented in the system [5]. Tangirala et al. identified a mixture of attributes and minimum class labels for splitting conditions on each non leaf node of a decision tree to solve the problem of homogeneous fruit subsets, and proposed several splitting indices to evaluate splitting; The experiment demonstrates that applying two different segmentation indices, GINI index and information gain, gives the same accuracy [6]. Li and other researchers studied a practical federated environment with relaxed privacy constraints to address the issue of insufficient efficiency or effectiveness of gradient based decision trees for practical applications; the experiment showcases that compared to normal training using local data from each party, this method can significantly improve prediction accuracy [7]. Nancy and others proposed a new intrusion detection system to address the issue of intrusion detection systems neglecting the identification of new types of attacks; the system uses intelligent decision tree classification algorithms to detect known and unknown types of attacks; the experiment demonstrates that this method reduces false positives, energy consumption, and latency [8]. Ramya and other researchers have designed a detection and classification system that combines decision trees with intrusion detection systems (IDS) to enhance the energy security performance of the power grid; The experiment illustrates that the model can accurately classify and predict most attacks [9].

Elhazmi et al. designed a prediction model for mortality of severe adult COVID-19 patients admitted to ICU by combining decision tree and conventional model logic regression to predict the mortality of COVID-19 patients; the experiment showcases that the accuracy of the prediction model reaches 96.65% [10]. Mariniello and other researchers designed a method category on the ground of decision tree ensemble vibration to address the challenges of structural damage detection and localization in vibration data, and analyzed the dynamic characteristics of the structural system (i.e. pattern shape and natural frequency) to obtain a structural health assessment model; The experiment indicates that the accuracy, reliability of probability prediction, and positioning error of the model perform best when compared with multiple algorithms [11]. Li et al. proposed adaptive control of the gradient of training data for each iteration and leaf node pruning to improve the accuracy of the GBDT model while preserving strong guarantees of differential privacy, to tighten the sensitivity limit; the experiment demonstrates that this method can achieve better model accuracy than other baselines [12]. Sharma and other researchers used fuzzy logic methods to control the operation of ventilation systems that provide fresh air to the environment to solve the problem of insufficient ventilation in indoor environments that damages human health; The experiment showcases that the indoor ventilation system controlled by this model increases the ventilation rate by 15.6% [13]. Arji and others have designed a rule-based fuzzy logic, adaptive neuro fuzzy inference system (ANFIS) to address the significant impact of the spread of infectious diseases on global health and economy; the experiment illustrates that this technology has improved the accuracy of infectious disease identification by 31.34% [14].

In summary, decision trees and fuzzy logic play an increasingly important role in the development of machine learning, but there are few related studies that combine the two to assist in intelligent library acquisition decision-making. Therefore, this study combines fuzzy logic to design a library intelligent acquisition decision model on the ground of decision trees, to further improve the efficiency of library management.

III. Improvement of Decision Tree and Design of Library Intelligent Acquisition Decision Model Scheme

This chapter is separated into two sections. The first section first designs a library intelligent acquisition decision model scheme on the ground of the standard decision tree. The second section mainly addresses the shortcomings in the standard decision tree and combines it with fuzzy logic to make some improvements, designing a fuzzy decision tree.

A. *Optimization Strategy for Book Interview Based on Decision Tree*

To reasonably select interviewees and provide books and materials that meets the needs of readers in the limited resources of the library. This requires effective interview decision-making methods to improve service quality and promote the dissemination of knowledge and the development of academic research. At present, there are four main influencing factors involved in the decision-making of book acquisition in universities, including book value, procurement funds, reader needs, and collection structure. The collection structure of a library mainly affects the service level and overall level of literature resources. In different application fields, the acquisition concept of libraries also varies, resulting in similar collection structures. The structural framework of influencing factors for book acquisition decision-making is shown in Fig. 1.

To ensure the effectiveness of the decision model, it is also necessary to select more suitable feature variables, and the selection criteria mainly depend on the classification effect of the samples. The methods for selecting feature variables include information gain or information gain ratio, and this study mainly uses information gain to select feature variables. When learning decision trees, the five features with greater information gain selected include author, category, price, publisher, and publication time. Among them, the author's writing level determines the overall quality of the book content. The author of a book plays an important role in the procurement of books. Decision tree is a commonly used machine learning algorithm used to solve classification and regression problems. It is a tree based model that performs prediction by segmenting and judging the dataset [15-16]. The decision tree is composed of nodes and edges, with each node representing a feature or attribute, and edges representing the relationship between feature or attribute values [17]. The root node represents the most important feature, the internal node represents the intermediate feature, and the leaf node represents the final output or decision result. The advantages of decision trees include ease of understanding and interpretation, ability to handle discrete and continuous features, and robustness to outliers and missing data. The structural diagram of the decision tree is shown in Fig. 2.

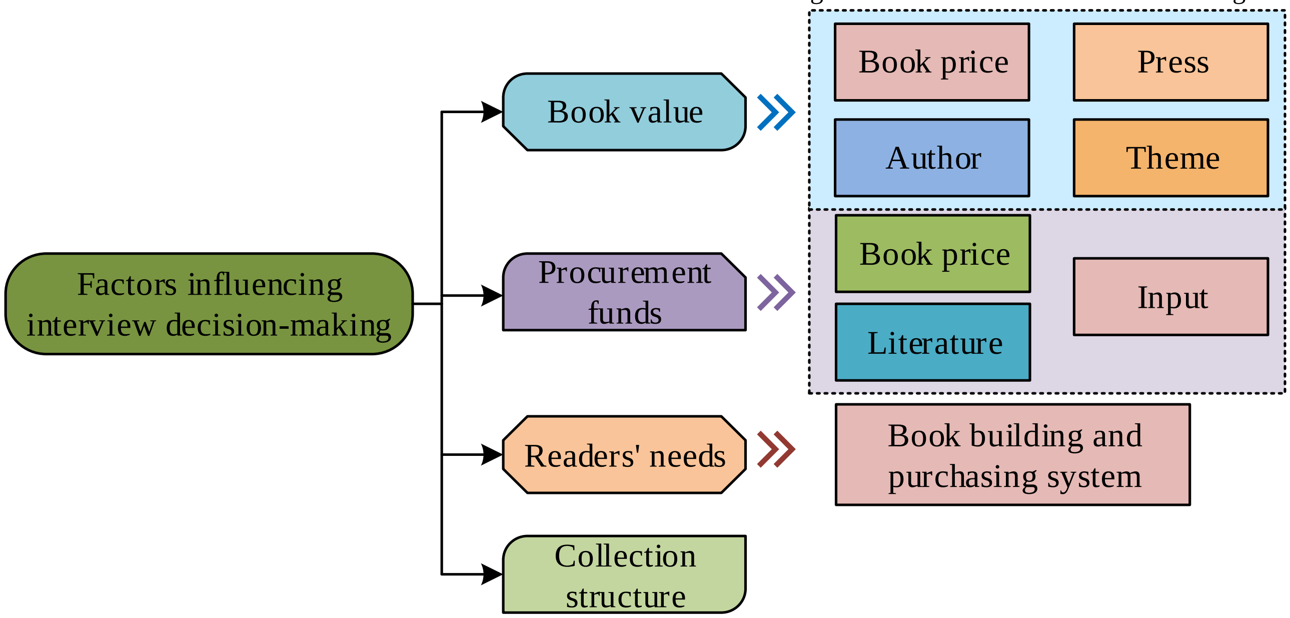


Fig. 1. The structural framework of factors influencing book acquisition decisions.

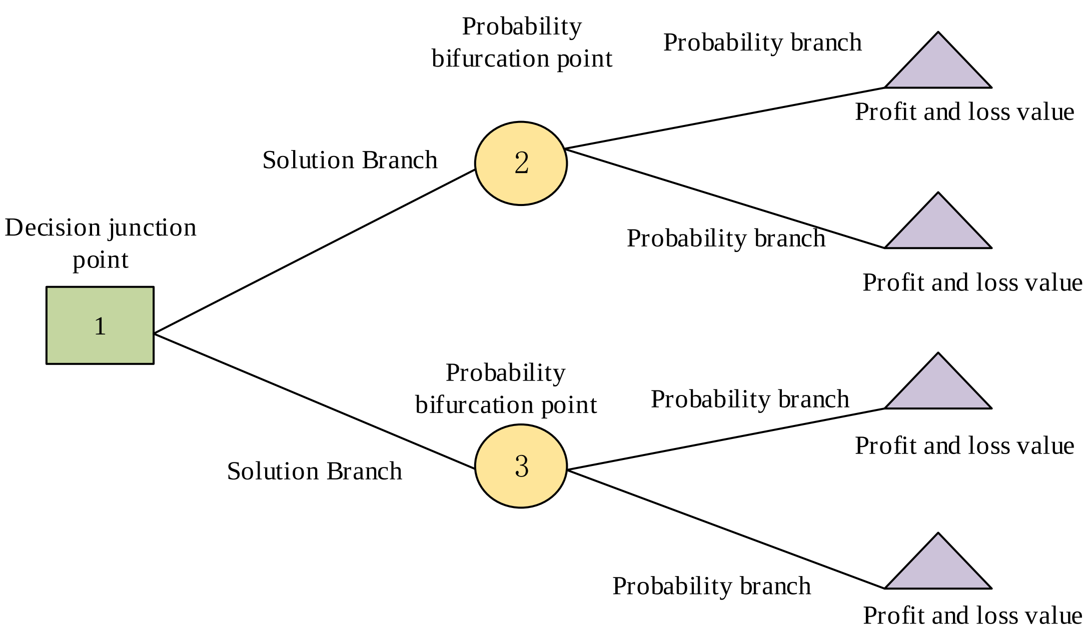


Fig. 2. Structure diagram of decision tree.

Fig. 2 shows a schematic diagram of the decision tree structure, which includes decision binding points, solution branches, probability branches, probability bifurcation points, profit and loss values, etc. Among them, decision junction points, also known as internal nodes or split nodes, represent the decision-making basis for dividing the current data. The decision binding point uses a certain feature or attribute and corresponding threshold to divide the dataset into two or more subsets. The solution branch refers to the edge of the decision junction point, which represents different decision paths or options. Each scheme branch corresponds to a specific feature value or attributes value, indicating that the dataset moves towards different sub nodes on the ground of the value of that feature value. The probability branch indicates the probability transition from one node to another, corresponding to the conditional probabilities of different categories. Profit and loss value is an indicator that measures the effectiveness of decision tree splitting. When selecting a certain feature as the decision binding point, the quality of partitioning is evaluated by calculating the profit and loss value after partitioning on that feature. The profit and loss value can be on the ground of different criteria, such as the Gini index or information gain. The Gini index in the decision tree is shown in Eq. (1).

$$Gini(p) = 1 - \sum (pi)^2 \tag{1}$$

In Eq. (1), pi represents the probability that the sample belongs to the i-th category. The formula for information gain in the decision tree is shown in Eq. (2).

$$IG(D,A) = H(D) - \sum (|\frac{Dv}{D}|) * H(Dv) \tag{2}$$

In Eq. (2), D represents the dataset; A represents a feature; $H(D)$ represents the entropy of the dataset; $H(Dv)$ represents the entropy of a subset. The information gain in C4.5 algorithm is shown in Eq. (3).

$$GainRatio(D,A) = IG(D,A) / SplitInfo(D,A) \tag{3}$$

In Eq. (3), $SplitInfo(D,A) = -\sum (\frac{|Dv|}{|D|}) * \log^2 (\frac{|Dv|}{|D|})$. Gini index and information gain are commonly used indicators to select the best features for node partitioning. The information gain ratio is an improved indicator introduced in the C4.5 algorithm, which avoids excessive preference for features with more values. Then, it is necessary to prune the decision tree, which is a technique used to reduce the complexity of the decision tree model. When constructing a decision tree, overfitting often occurs, where the model is too complex to generalize well to new data samples. Pruning is the process of reducing the risk of overfitting by pruning some branches or leaf nodes of a decision tree, thereby improving the generalization ability of the model. Decision tree pruning can be divided into two methods: pre-pruning and post-pruning. This study used post pruning, as shown in Fig. 3.

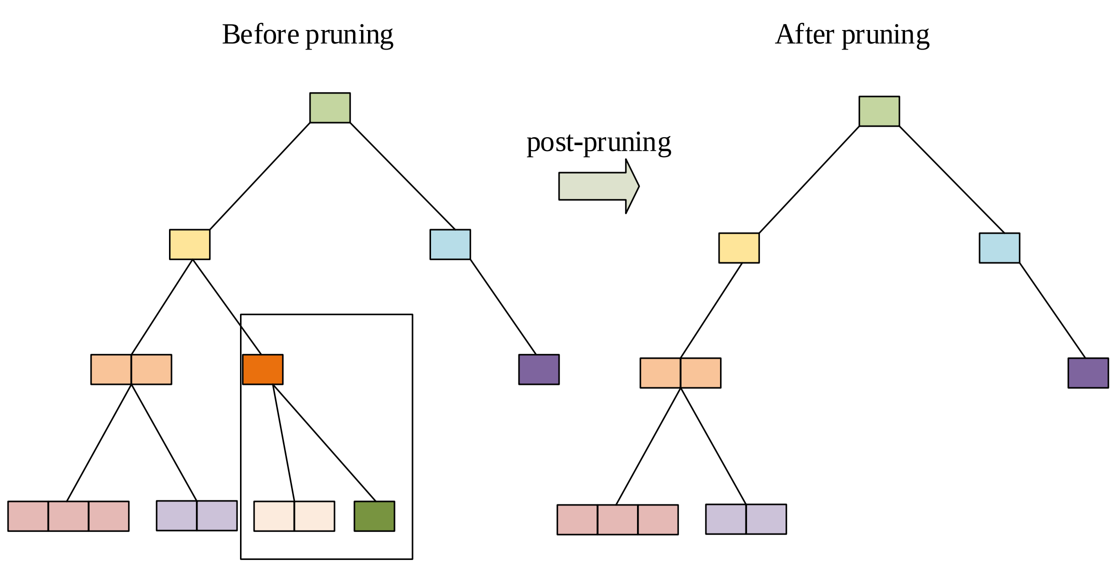


Fig. 3. Decision tree pruning process.

Fig. 3 shows the pruning process of a decision tree, which involves pruning an existing decision tree from the bottom up after it, is constructed. Firstly, this study evaluates each leaf node and calculates its performance indicators (such as accuracy, error rate, etc.) on the validation set. Then, it gradually tries pruning, replacing the leaf node with its parent node, and observes whether the model performance has been improved. If the performance of the model improves after pruning, perform pruning operations; otherwise, keep it as is.

B. An Intelligent Acquisition Optimization Model for Library Based on Fuzzy Decision Tree

On the ground of the basic principle of decision tree, establish a library intelligent acquisition decision model on the ground of decision tree. Firstly, it collects data related to library interviews, including user information, library resources, borrowing history, etc. Then there is data preprocessing, which involves cleaning and preprocessing the data, including handling missing values, outliers, and duplicate values. The third step is feature selection, which evaluates the importance of features through feature analysis and correlation detection. The fourth step is data partitioning, which divides the dataset into training and testing sets for model training and evaluation. Then it constructs a decision tree and uses the decision tree algorithm C4.5 to construct a decision tree model. It then prunes the decision tree and performs pruning operations on the constructed decision tree to avoid overfitting. The seventh step is model training and evaluation, evaluating the performance of the model on the ground of the performance of the test set. Then it uses the constructed decision tree model to predict and make decisions on new interview situations or user needs. Finally, on the ground of feedback and results from practical applications, the model is optimized and adjusted. The basic algorithm flow of the library intelligent acquisition decision model on the ground of decision tree algorithm is shown in Fig. 4.

Decision tree algorithms are usually able to generate models with good interpretability and comprehensibility, but when the problem has ambiguity, traditional decision trees may become complex and difficult to understand. Library interviews involve many ambiguous situations, such as the ambiguity of user needs and the fuzzy description of library resources [18-19]. Traditional decision trees can only handle discrete classification and attribute values, and cannot effectively handle ambiguity. To address this issue, this study introduces fuzzy theory to handle the uncertainty of data. It allows node partitioning to have fuzzy membership, rather than strict binary partitioning [20]. Such fuzzy partitioning can better handle data with uncertainty or fuzziness. Fuzzy theory can be roughly divided into the following types, as shown in Fig. 5.

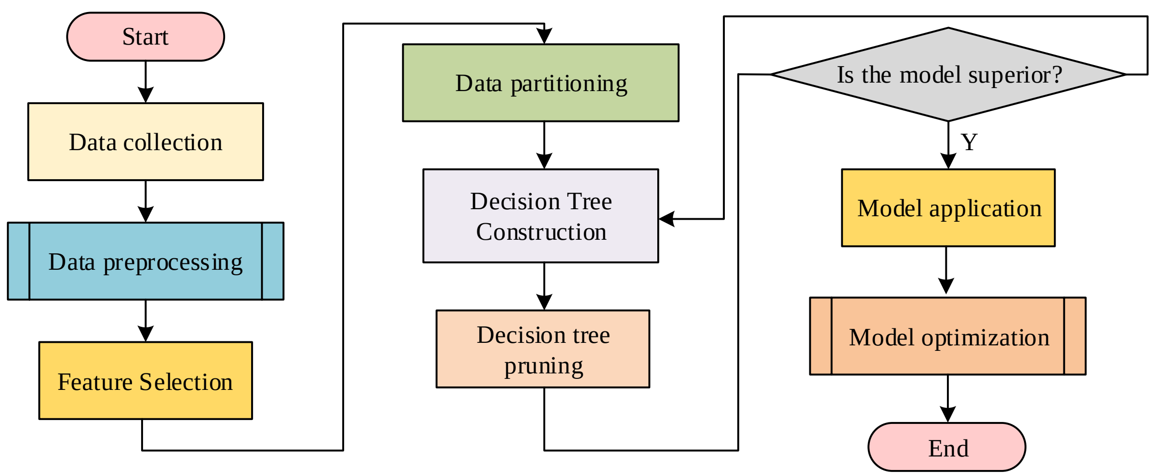


Fig. 4. The algorithm process of library intelligent acquisition decision model.

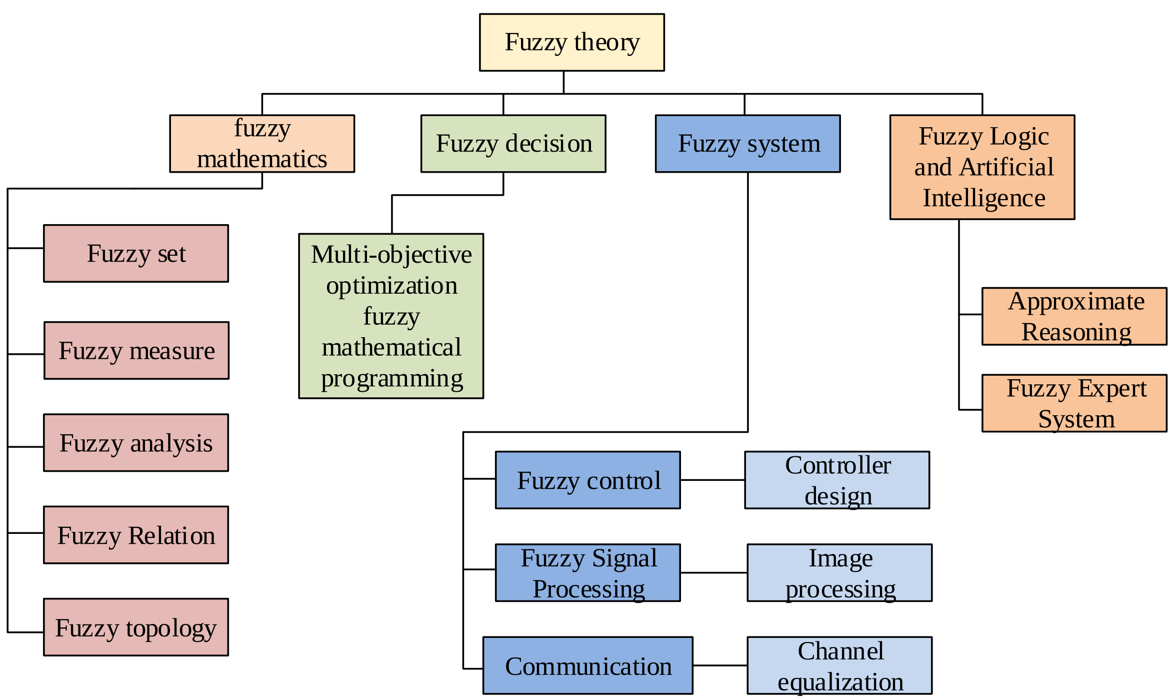


Fig. 5. Fuzzy theory classification.

As shown in Fig. 5, fuzzy theory is a very large concept with extensive applications in modern society. It mainly covers aspects such as fuzzy mathematics, fuzzy decision-making, fuzzy systems, uncertainty and information, as well as fuzzy logic and artificial intelligence. Fuzzy mathematics is the foundation of fuzzy theory, which is used to deal with problems that cannot be clearly classified as true or false. Fuzzy decision-making is dedicated to dealing with decision-making problems containing fuzzy factors. Fuzzy system is a method of applying fuzzy theory, which simulates human thinking patterns and generates fuzzy outputs by applying fuzzy rules to input data. Fuzzy logic and artificial intelligence are important components of fuzzy theory, which includes approximate reasoning and fuzzy expert systems. The application of fuzzy theory can better handle uncertainty and fuzziness, thus achieving more accurate and reliable results. Assuming X is a given finite set, the fuzzy subset is shown in Eq. (4).

$$A = \frac{A(e_1)}{e_1} + \frac{A(e_2)}{e_2} + \ldots + \frac{A(e_N)}{e_N} \tag{4}$$

In Eq. (4), the potential in the set is used to measure the size of the set, as defined in Eq. (5).

$$M(A) = \sum_{i=1}^{N} A(e_i) \tag{5}$$

In Eq. (5), A represents a fuzzy subset. The probability formula for samples belonging to a certain class in nodes in a fuzzy decision tree is shown in Eq. (6).

$$\beta_A^C = SIM(A,C) = \frac{M(A \cap C)}{M(A)} = \frac{\sum_{x\in X} \min(\mu_A(x), \mu_C(x))}{\sum_{x\in X} \mu_A(x)} \tag{6}$$

In Eq. (6), $M(A)$ represents the sum of all membership degrees representing the fuzzy set A. For any fuzzy subset, the relative frequency of the j-th fuzzy category of its non leaf node is shown in Eq. (7).

$$p_{ij}^{(k)} = \frac{M(T_i^{(k)} \cap T_i^{(n+1)} \cap X)}{M(T_i^{(n)} \cap X)} \tag{7}$$

In Eq. (7), $T_i^{(k)}$ represents a fuzzy subset; X represents a non leaf node, and the fuzzy classification entropy of each fuzzy subset of X is shown in Eq. (8).

$$FEntr_i^{(k)} = -\sum_{j=1}^{m} p_{ij}^{(k)} \log_2 p_{ij}^{(k)} \tag{8}$$

In Eq. (8), $1 \le k \le n$, $1 \le j \le m$, then the average fuzzy classification entropy of attributes on non leaf nodes is defined as shown in Eq. (9).

$$FEntr_k = -\sum_{j=1}^{m_k} \omega_i FEntr_i^{(k)} \tag{9}$$

In Eq. (9), ω_i represents the weight of the i-th attribute value. So the non assignability of classification on non leaf nodes is shown in the definition of Eq. (10).

$$Ambig_i^{(k)} = \sum_{j=1}^{m} (\pi_{ij}^{(k)} - \pi_{i,j+1}^{(k)}) \ln j \tag{10}$$

In Eq. (10), m represents the number of fuzzy categories, so the average classification of non leaf nodes cannot be specified as shown in Eq. (11).

$$Ambig_k = -\sum_{j=1}^{m_k} \omega_i Ambig_i^{(k)} \tag{11}$$

When obtaining classification rules from uncertain information, fuzzy membership functions are often used to describe uncertainty. The fuzzy membership function characterizes the degree of uncertainty of the membership function through its parameters. This method can help handle data or situations that cannot be clearly classified into a certain category. The decision tree algorithm flow combining fuzzy logic is shown in Fig. 6.

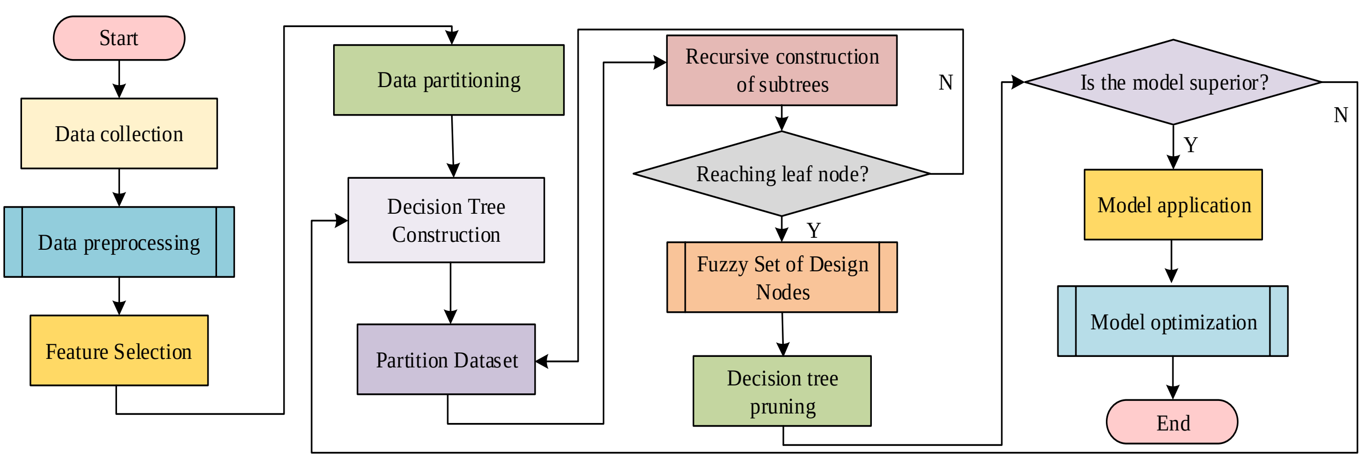


Fig. 6. Decision tree algorithm flow combining fuzzy logic.

Fig. 6 shows the flowchart of the fuzzy decision tree algorithm. Compared with the standard decision tree algorithm, the steps for building the model have slightly changed, while the rest are roughly the same. When building a model, the dataset is first divided into different subsets on the ground of the selected features, with each subset corresponding to a sub node. It recursively constructs a subtree, recursively executing the above steps for each sub node until it reaches the leaf node or cannot be further divided. It designs a fuzzy set of nodes, and on the ground of the partitioning results and category labels, designs a fuzzy set representation of the current node, including fuzzy membership functions and membership degrees. This study evaluates the model on the ground of indicators such as accuracy, recall, F1 value, etc. The accuracy formula is shown in Eq. (12).

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{12}$$

In Eq. (12), TP represents the true example; TN represents a true negative example; FP represents a false positive example; The recall rate formula is shown in Eq. (13).

$$Recall = \frac{TP}{TP+FN} \tag{13}$$

In Eq. (13), FP represents a false negative example; The recall rate represents the proportion of cases that the model correctly judges as positive, and can measure the model's ability to correctly recognize positive cases. The formula for F1 value is shown in Eq. (14).

$$F1 = \frac{2(Accuracy * Recall)}{(Accuracy + Recall)} \tag{14}$$

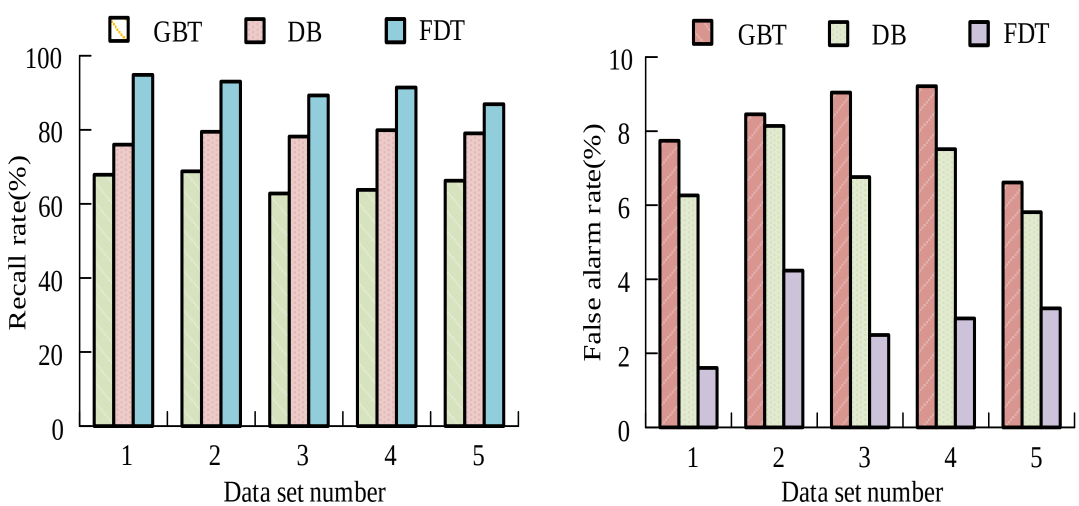
IV. Algorithm Performance Testing and Model Applicability Analysis

This chapter is separated into two sections. The first section mainly verifies the improvement effect of the decision tree algorithm and conducts comparative experiments with various similar algorithms. The second section mainly analyzes the applicability of the library intelligent acquisition decision-making model and applies it to actual library management.

A. Performance Evaluation and Comparative Study of Fuzzy Decision Trees in Library Acquisition Decision-making

The experiment combines fuzzy sets to design a fuzzy decision tree on the ground of a decision tree. To study the algorithm performance and superiority of the model, Gradient Boosting Tree (GBT) and AdaBoos algorithm (DB) were introduced in the experiment to compare with the fuzzy decision tree proposed in the study. This experiment used PyTorch 1.8 software on the Windows 10 system platform to train three models 500 times using five different datasets, as shown in Fig. 7.

Fig. 7 shows the evaluation results statistics of three models in terms of recall and error rates. Fig. 7 (a) shows that the FDT model performs best in terms of recall rate, with various evaluation indicators reaching 96.45%, 92.64%, 87.92%, 91.17%, and 86.65%, respectively. In terms of false alarm rate, the FDT model also performs well, with significantly lower error rates than the other two models, which are 1.88%, 4.48%, 2.53%, 2.61%, and 2.86%, respectively; this means that the FDT model has a good effect in reducing false positives and can more accurately control the situation of erroneous reports. In summary, the FDT model performs well in terms of recall and false alarm rates, with high accuracy. The F1 value is an indicator that comprehensively considers the accuracy and completeness of the classifier, providing a more reliable performance evaluation under imbalanced datasets. The F1 values of the three models are shown in Fig. 8.



(a) Comparison of recall rates

(b) Comparison of false alarm rates

Fig. 7. Results of recall rate and false positive rate evaluation.

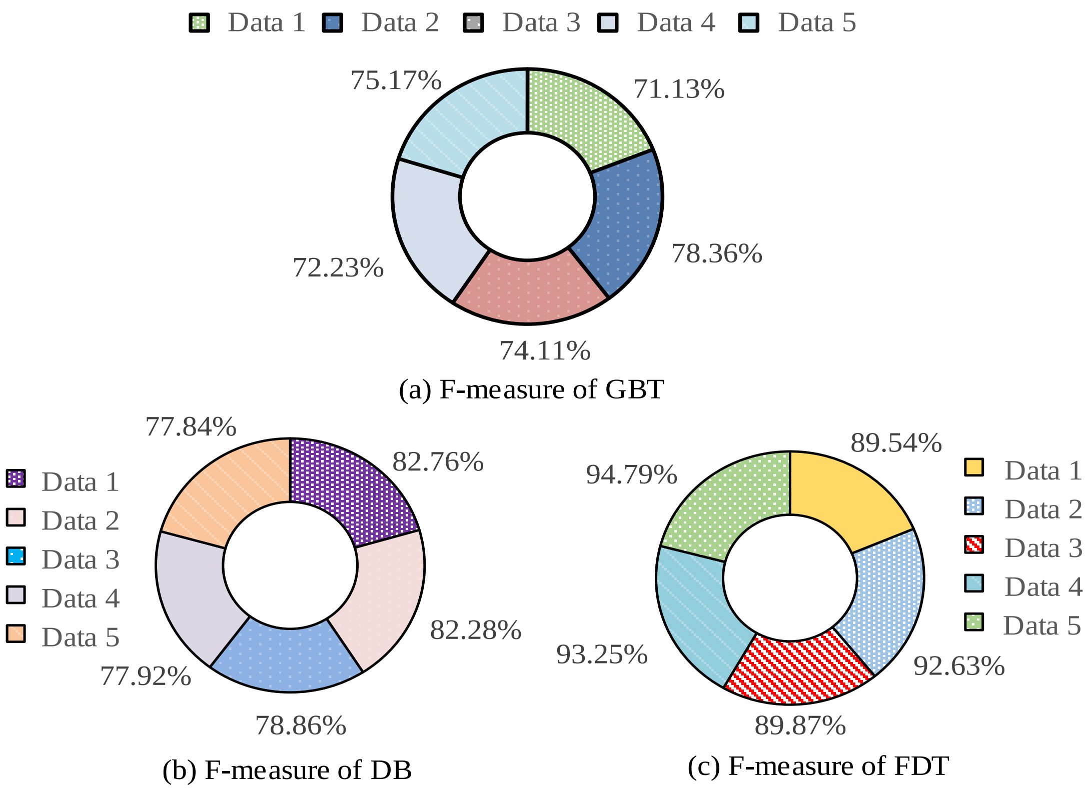


Fig. 8. F-measure for the three models.

Fig. 8 shows the results of training three models on five datasets. From the perspective of F1 values, the single GBT model is relatively low which is only 71.13% on dataset 1. Compared to this, the DB model exhibits good performance on most datasets, with a small and high average F1 difference, demonstrating its strong generalization ability. On the other hand, the F1 value of the FDT model is significantly higher than the other two models, which are 89.54%, 92.63%, 89.87%, 93.25%, and 94.79%, respectively. Overall, the average F1 value of the FDT model reaches 92.02%. The FDT model achieved the highest F1 value on all datasets, demonstrating its excellent performance in classification tasks. In addition, the AUC curves of each model were recorded in the experiment, as shown in Fig. 9.

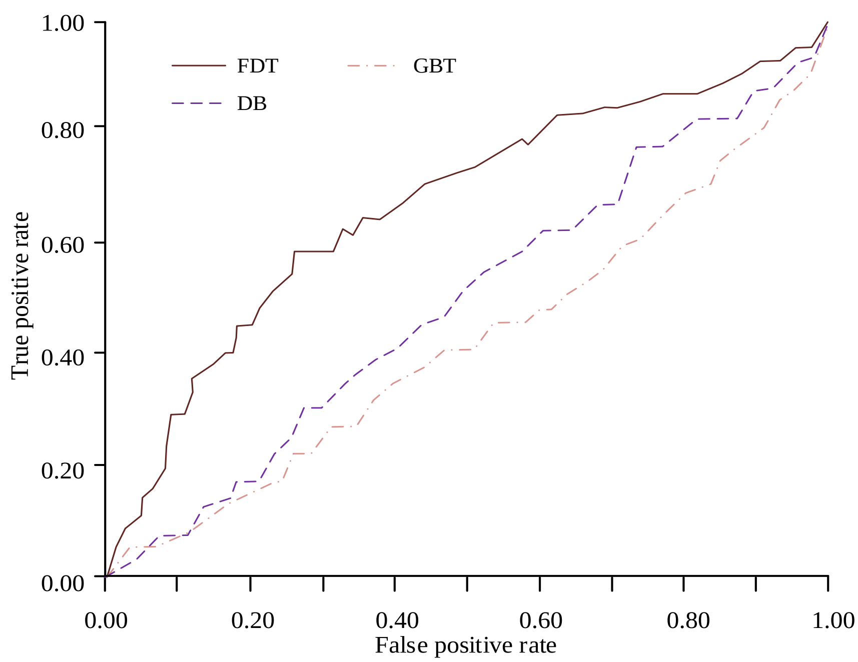


Fig. 9. AUC curves of three models.

Fig. 9 shows the statistical results of each model in terms of AUC values. In this figure, the horizontal axis represents the false positive rate, and the vertical axis represents the true positive rate. By observing Fig. 9, it can be observed that the ROC curve of the FDT model is always above the ROC curve of other models, indicating that its true positive rate performance is better under different false positive rates. In addition, on the ground of the enclosed area, i.e. AUC value, it can be concluded that the FDT model has the maximum AUC area of 0.681, while the AUC of the GBT and DB models are 0.517 and 0.463, respectively.

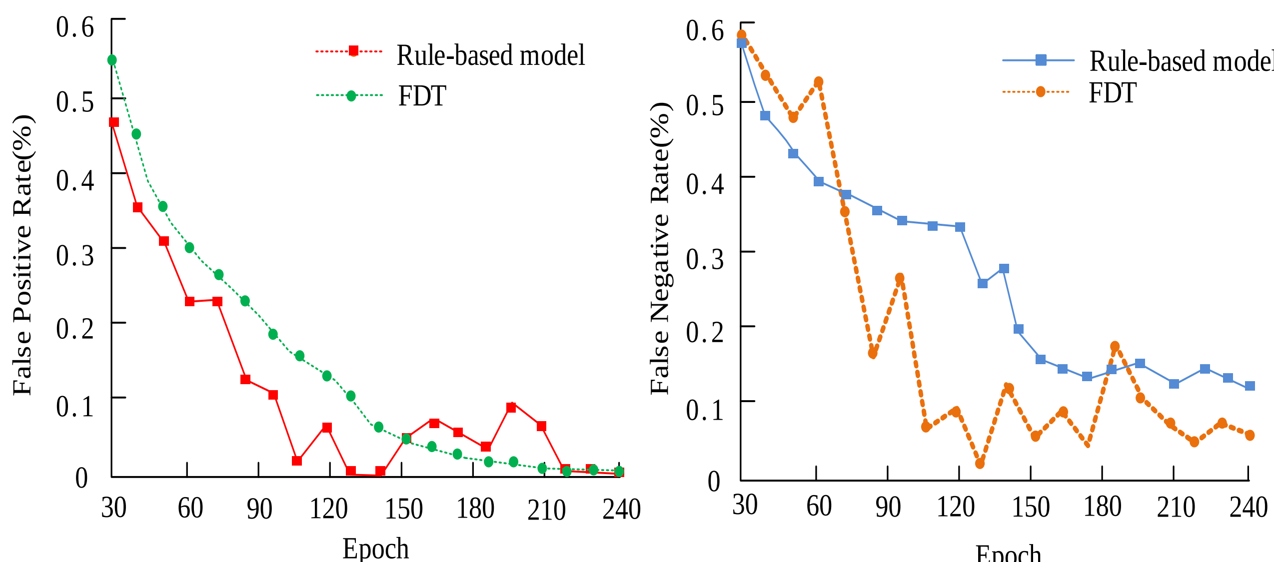
B. *Research on the Application of FDT Library Intelligent Acquisition Decision Model*

The experiment first conducted extensive algorithm performance testing and comparative analysis on the FDT library intelligent acquisition decision-making model, fully proving the superiority of the model. However, to demonstrate the practical application value of the model, further research on its applicability is needed. By deploying the FDT model in an actual library environment and comparing it with traditional methods, the efficiency, accuracy, and application cost of the model were evaluated. Some of the library's procurement data are shown in Table I.

The experiment used the data in Table I to train the FDT library intelligent acquisition decision-making model and rule-based model. A rule-based model uses predefined rules and conditions to make interview decisions. For example, on the ground of factors such as user borrowing history and needs, establish a set of rules to determine whether to recommend a certain book to the user. And it uses False Positive Rate and False Negative Rate as evaluation indicators. False positive rate refers to the model mistakenly determining the proportion of resources that do not require interviews as those that require interviews; the false negative rate refers to the proportion of resources that the model mistakenly judges as not requiring interviews. The training results are shown in Fig. 10.

TABLE I. PARTIAL PROCUREMENT DATA OF THE LIBRARY

Registration no	Book attributes	Warehousing time	Classification number	Rice ($)
BC99851	Storage	2019.08.06	B223.15	65.3
BC98765	Storage	2019.08.06	y	56
BC89732	Circulate	2019.08.06	H319.4	58
BC89875	Storage	2019.01.01	H314	32.1
BC89712	Circulate	2019.01.01	H319.9	36.3
BC88952	Circulate	2019.01.01	I214.5	32.3
BC84564	Storage	2019.09.01	K512.4	50.2
BC86832	Circulate	2019.01.01	K512.4	12.3
BC56465	Circulate	2019.07.01	Y	13.2



（a）The relationship between the number of training rounds and the variation of FP

（b）The relationship between the number of training rounds and the variation of FN

Fig. 10. The relationship between the number of training rounds and the variation of FP and FN.

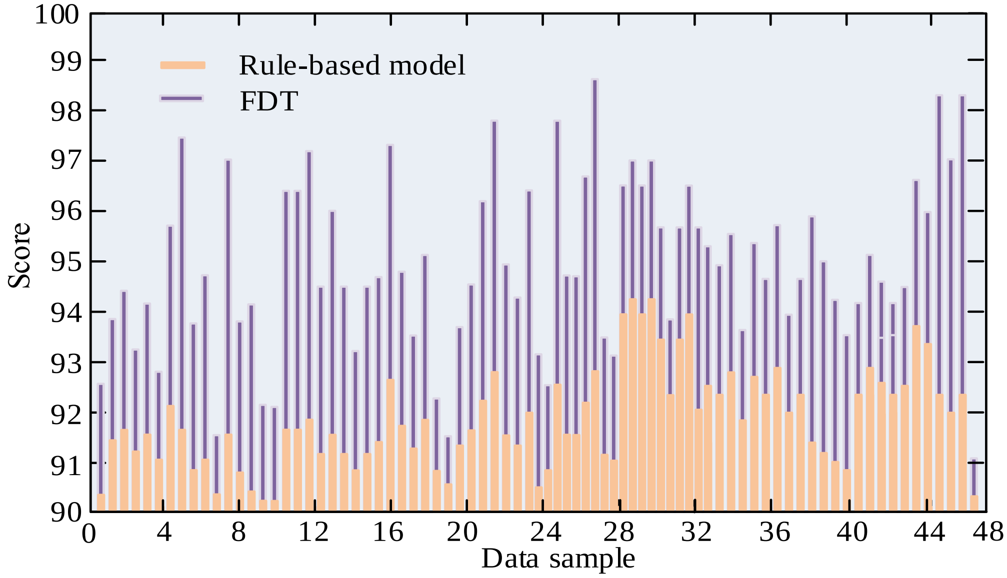


Fig. 11. Rating of the recommendation effect of the FDT model by 48 students.

Fig. 10 shows the changes in FP and FN values of the FDT model and rule-based model as the number of training rounds increases. According to Fig. 10 (a), it can be observed that the false positive case (FP) value of the FDT model continues to decrease and tends to stabilize as the number of training rounds increases. On the contrary, the FP values of rule-based models are not very stable and show an increasing trend after 150 rounds of training. Further observation of Fig. 10 (b) shows that the error rejection rate (FN) of the FDT model rapidly decreases to about 0.1 as the training frequency gradually increases, and stabilizes at around 0.05 after 210 rounds of training. However, the error rejection rate of the rule-based library acquisition decision-making model ultimately stabilized at around 0.15 levels. These results emphasize the advantages and effectiveness of the FDT model in this classification problem. The experiment also randomly invited 48 middle school students from the library to rate the recommendation effect of the FDT library intelligent acquisition decision-making model. The relevant results are shown in Fig. 11.

Fig. 11 shows the rating of 48 students on the recommendation effectiveness of the FDT model and the rule-based library intelligent acquisition decision-making model. The figure shows that most students are satisfied with the recommendation performance of these two models, and the scores given are above 90 points. However, students rated the FDT library intelligent acquisition decision-making model higher than the rule-based model, at least 1.2 points higher. Specifically, the average score of the FDT model is 94.3, while the average score of the rule-based model is 91.8. In summary, most students hold a satisfactory attitude towards the recommendation effectiveness of FDT models and rule-based models.

V. Results and Discussion

In the research of intelligent procurement decision-making systems in libraries, traditional manual decision-making methods rely on experience and lack the ability to handle large datasets, making it difficult to quickly adapt to new needs and changes. Faced with this challenge, a series of improvements have been made to the standard decision tree, and a new fuzzy decision tree (FDT) model has been designed and implemented. This model aims to better handle uncertainty and ambiguity, and improve the accuracy and efficiency of decision-making. When evaluating the FDT model, several different datasets were used and compared with two popular algorithm models - Gradient Boosting Tree (GBT) and Adaptive Boosting Algorithm (DB) - for testing. The test results show that the GBT model has a relatively low F1 value on dataset 1, only 71.13%, while the DB model performs well on most datasets, but its average F1 value difference is not significant and lower than the FDT model. Although these two models perform well in certain aspects, they still have limitations in dealing with ambiguity and uncertainty. In contrast, the FDT model performs significantly better than the GBT and DB models on all test datasets. Specifically, the F1 values of the FDT model on five datasets were 89.54%, 92.63%, 89.87%, 93.25%, and 94.79%, respectively, with an average F1 value of 92.02%. This result highlights the significant ability of the FDT model to ensure high recall and accuracy, which is particularly crucial for libraries as it can reduce the risk of missing important books and avoid purchasing books that do not meet demand. The ROC curves of the FDT model always lie above the ROC curves of the GBT and DB models, indicating that the FDT model can maintain higher true positive rates at different levels of false positive rates. As an indicator of the overall performance of the model, the AUC value of the FDT model is 0.681, significantly higher than GBT's 0.517 and DB's 0.463. This result further demonstrates the superiority of the FDT model in distinguishing between different categories (books that need to be purchased and those that do not). In terms of practical application, 48 students evaluated the effectiveness of the FDT model, and the results showed that students generally rated the FDT library intelligent interview decision-making model higher than rule-based models, with an average score of at least 1.2 points higher. The average score of the FDT model is 94.3 points, while the rule-based model is 91.8 points. This indicates that in practical applications, users are more satisfied with the recommendations provided by the FDT model, which may be because the FDT model can more accurately predict user needs and provide more personalized recommendations. However, despite the excellent performance of the FDT model in various aspects, research has also found a major drawback of this model: longer training time. This may be due to the model's need to evaluate and integrate a large number of fuzzy rules when processing data, resulting in an increase in computational complexity. The prolonged training process may limit the practicality of the model in application scenarios that require rapid updating of decision models to adapt to new situations. In the future, distributed computing resources can be utilized to allocate model training tasks to multiple computing nodes for parallel processing, significantly reducing training time. This requires the use of specialized distributed deep learning frameworks, such as the distributed version of TensorFlow, which is also an area that needs improvement in future research.

References

[1] Fang B, Jiang M, Shen J. Deep Generative Inpainting with Comparative Sample Augmentation. Journal of Computational and Cognitive Engineering, 2022, 1(4): 174-180.

[2] Nimrah S, Saifullah S. Context-Free Word Importance Scores for Attacking Neural Networks. Journal of Computational and Cognitive Engineering, 2022, 1(4): 187-192.

[3] Guo Y, Mustafaoglu Z, & Koundal D. Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms. Journal of Computational and Cognitive Engineering, 2022, 2(1), 5–9.

[4] Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2021, 2(1): 20-28.

[5] Aldino A A, Sulistiani H. Decision Tree C4. 5 Algorithm for Tuition Aid Grant Program Classification (Case Study: Department of Information System, Universitas Teknokrat Indonesia). Jurnal Ilmiah Edutic: Pendidikan dan Informatika, 2020, 7(1): 40-50.

[6] Tangirala S. Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm. International Journal of Advanced Computer Science and Applications, 2020, 11(2): 612-619.

[7] Li Q, Wen Z, He B. Practical federated gradient boosting decision trees//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(4): 4642-4649.

[8] Nancy P, Muthurajkumar S, Ganapathy S, Santhosh Kumar, S. V. N., Selvi, M., & Arputharaj, K. Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks. IET Communications, 2020, 14(5): 888-895.

[9] Ramya K, Teekaraman Y, Kumar K A R. Fuzzy-based energy management system with decision tree algorithm for power security system. International Journal of Computational Intelligence Systems, 2019, 12(2): 1173-1178.

[10] Elhazmi A, Al-Omari A, Sallam H, Mufti, H. N., Rabie, A. A., Alshahrani, M., ... & Arabi, Y. M. Machine learning decision tree algorithm role for predicting mortality in critically ill adult COVID-19 patients admitted to the ICU. Journal of infection and public health, 2022, 15(7): 826-834.

[11] Mariniello G, Pastore T, Menna C, Festa, P, & Asprone, D. Structural damage detection and localization using decision tree ensemble and vibration data. Computer-Aided Civil and Infrastructure Engineering, 2021, 36(9): 1129-1149.

[12] Li Q, Wu Z, Wen Z, & He, B. Privacy-preserving gradient boosting decision trees. Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(1): 784-791.

[13] Sharma S, Obaid A J. Mathematical modelling, analysis and design of fuzzy logic controller for the control of ventilation systems using MATLAB fuzzy logic toolbox. Journal of Interdisciplinary Mathematics, 2020, 23(4): 843-849.

[14] Arji G, Ahmadi H, Nilashi M, Rashid, T. A., Ahmed, O. H., Aljojo, N., & Zainol, A. Fuzzy logic approach for infectious disease diagnosis: A methodical evaluation, literature and classification. Biocybernetics and biomedical engineering, 2019, 39(4): 937-955.

[15] Wang F, Song G. Bolt-looseness detection by a new percussion-based method using multifractal analysis and gradient boosting decision tree. Structural Health Monitoring, 2020, 19(6): 2023-2032.

[16] Sun L, Li Q, Fu S, & Li, P. Speech emotion recognition based on genetic algorithm–decision tree fusion of deep and acoustic features. ETRI Journal, 2022, 44(3): 462-475.

[17] Zhang J, Jia H, Zhang N. Alternate support vector machine decision trees for power systems rule extractions. IEEE Transactions on Power Systems, 2022, 38(1): 980-983.

[18] Prasenjit C. Model for selecting a route for the transport of hazardous materials using a fuzzy logic system. Vojnotehnički glasnik, 2021, 69(2): 355-390.

[19] Sahu R, Dash S R, Das S. Career selection of students using hybridized distance measure based on picture fuzzy set and rough set theory. Decision Making: Applications in Management and Engineering, 2021, 4(1): 104-126.

[20] Xiao F. A distance measure for intuitionistic fuzzy sets and its application to pattern classification problems. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019, 51(6): 3980-3992.