Intelligent Transportation and Smart Cities
V. Mezhuyev et al. (Eds.)

doi:10.3233/ATDE240404

The Computer Audit for Health Affordable Housing Based on Data Mining

Zheng LI[a], Saisai YAN[a,1], Xinran CHENG[a] and Sachin KUMAR[b]
[a] College of Finance and Economics, Shanghai Lida University, Shanghai, China
[b] Department of Computer Science and Engineering, Galgotias College of Engineering and Technology, Greater Noida, India

Abstract. The study aims to improve the accuracy rate of data classification under computer audit. An appropriate data mining (DM) algorithm is employed to classify affordable housing data. First, the research status of the DM algorithm and affordable housing audits is introduced, and the process of computer audit is described. Second, after analyzing the advantages and disadvantages of well-known DM algorithms, a decision tree (DT) is constructed using C4.5. In the end, C4.5 is enhanced by the interest and simplified by the Taylor series. According to the findings, the naive Bayes tree (NBTree), the classification and regression tree (CART), and C4.5 have average accuracy rates of 85.4%, 85.7%, and 85.6%, respectively. This shows that C4.5 and NBTree have an excellent effect on classifying the sample set. CART and C4.5 have a faster modeling speed, and their classification accuracy on huge data sets is better than that of NBTree. Therefore, it can be concluded that C4.5 and CART have good scalability; C4.5 has the best performance, and its ratio of rules to leaf points is higher than that of the other two algorithms; compared to the original C4.5, the enhanced C4.5 has a greater accuracy rate and a rule-to-leaf node ratio, but their modeling time is the same; the simplified C4.5 has the best performance, its classification accuracy rate is 98.7%, its modeling time is 0, and its interpretability is 1.9. This study provides a reference for the information development of affordable housing audits.

Keywords. Data mining, Computer audit, Affordable housing, Data classification, C4.5 algorithm

1. Introduction

The technical departments of various industries have applied artificial intelligence (AI) and made outstanding achievements [1]. The audit department is one of them. The audit is defined by legislation as an impartial financial supervisory activity in which a specifically appointed body examines the principal projects, financial receipts and outlays, and operations of governments, financial institutions, businesses, and organizations [2]. Facing extensive data, traditional audit departments find it is difficult to extract and analyze them. With this regard, the information-based audit based on computer technology is used by enterprises, institutions, and national institutions, and the relevant cost is significantly reduced [3]. But it also brings some risks, such as accounting information distortion and economic crime. Therefore, the integration of audit and information technology needs to be strengthened [4].

[1] Corresponding Author: Saisai YAN, 857938157@qq.com

This study is based on the audit of affordable housing. The shortcomings in the current audit data analysis include the backward and passive development of traditional analysis models and the difficulty in tracking audit data [5, 6]. Data mining (DM) technology has a good application prospect in the audit industry, which can help auditors analyze massive data efficiently and effectively improve audit efficiency [7]. Affordable housing project is called social housing, public housing, housing and so on in foreign countries. Related research also started earlier, mainly focusing on the distribution mode, policy system and construction management of affordable housing. Public rental housing policies in western countries are mainly aimed at low-and middle-income people [8]. At the end of last century, Singapore applied the central provident fund to housing financing [9]. At the beginning of the 21st century, Japan applied national finance to housing security [10]. In the research on the construction and management of affordable housing, the mode of cooperative construction of affordable housing between real estate enterprises and the government has been relatively mature [11].

Recently, there has been a lot of interest in the application of AI and DM methods in the field of computer auditing. First, Wang et al. (2023) successfully classified audit data using deep learning methods, namely convolutional neural networks (CNNs). Their research highlighted the higher accuracy and efficiency of CNN in dealing with large-scale audit data, which provides strong support for computer audit [12]. Meanwhile, Alkahtani and Aldhyani (2022) focused on the application of recurrent neural network (RNN) in the audit field within the scope of deep learning, and further improved the accuracy of classification through the processing of time series data [13]. Together, these two lines of inquiry illustrated the potential benefits of deep learning for data classification. Sestino and De Mauro (2022) employed the K-means clustering algorithm in their DM research to categorize affordable housing data. Their research showed that clustering algorithm can identify potential patterns in data and help to better understand the structure of audit data [14]. Correspondingly, the research of Mahmoodi Khaniabadi et al. (2023) focused on the mining of association rules, which provided a deeper insight for computer auditing by discovering the correlation between data. These two research directions show the diversity and adaptability of DM algorithms when dealing with audit data [15]. It was worth noting that attempts to integrate multiple algorithms have emerged in recent studies. For instance, Sharma et al. (2022) developed a compound method to address the complex relationship in audit data by integrating the support vector machine and decision tree (DT) algorithms. Their study highlighted the value of combining several methods, which can boost classification robustness and accuracy while offering a novel approach to the field of computer auditing [16].

On the other hand, the research on algorithm application also includes many fields. Hua et al. (2023) realized efficient binary classification through logistic regression algorithm, which highlighted the excellent ability of logistic regression in handling audit exceptions [17]. Nti et al. (2022) revealed the potential of reinforcement learning in dealing with uncertainty and complex audit environment by introducing deep reinforcement learning [18]. In light of large-scale data, Kumar et al. (2023) implemented a distributed method and stressed the value of distributed computing to increase the effectiveness of computer auditing [19]. Ouyang et al. (2022) studied anomaly detection by generating adversarial network (GAN), and successfully revealed the effectiveness of GAN in discovering potential anomaly patterns [20]. Farina et al. (2022) used self-supervised learning method to extract features from large-scale audit data, which highlighted the potential of self-supervised learning in learning useful representations from unlabeled data [21]. Akgun and Greenhow (2022) introduced time series analysis and long and short-term memory (LSTM) networks. They also provided a more accurate categorization method for audit data that is time-sensitive [22]. These developments made it possible to carefully analyze the patterns and trends discovered in audit data.

In terms of graph-based method, Zan (2022) made remarkable achievements in the classification of audit network data, providing a novel modeling method for complex relationships [23]. Kuo et al. (2022) studied the classification method when dealing with uncertain data through fuzzy logic and fuzzy clustering algorithm, which provided a new perspective for dealing with fuzzy data in auditing [24]. The ensemble learning approach was used by Dwivedi et al. (2022) in conjunction with numerous classifiers, highlighting the significance of ensemble learning in enhancing algorithm robustness [25]. Bradshaw et al. (2022) introduced the concept of meta-learning and a method for classifying audit data based on a small sample size. In the situation of scarce data, this method provided guidance for computer auditing [26]. Under the context of big data, Abdullahi et al. (2022) embraced high-performance computing and parallel computing technologies, which enhanced the speed and effectiveness of audit data processing and offered a creative solution for large-scale data processing [27].

In addition, other related research has also made important contributions to the development of computer auditing. For example, Mendoncan et al. (2022) studied the demand for model interpretation in audit scenarios by introducing model interpretation technology, which provided a practical method for understanding the reasons behind model decisions [28]. This series of research together constitutes a rich and diverse research picture scroll, which deeply digs into the wide application and future development direction of AI and DM in computer auditing.

According to the previous research, the Chinese audit system based on DM technology is not intelligent enough and the data processing is not perfect enough, so the audit data analysis based on DM technology needs further research. However, there are also some shortcomings in the audit research of affordable housing in China. Enhancing the theoretical and practical investigation of affordable housing projects will be crucial in advancing social development. The study provides a novel and effective approach for affordable housing audit by selecting an appropriate DM algorithm and applying it to the classification of computer audit data of affordable housing.

The main task is to optimize the network security operation mechanism and environment and improve the existing audit level of affordable housing. First, the classical algorithms in DM are introduced and compared, and the DT algorithm is applied to the affordable housing audit. Secondly, the initial data are preprocessed, and trials are conducted to compare how well the classification algorithms work. After that, C4.5 is selected to perform algorithm modeling and is improved. The innovation is that the improved C4.5 can improve audit efficiency. Finally, the improvement suggestions for the current affordable housing audit are proposed, which provides practical guidance and a theoretical basis for the future affordable housing audit.

2. Theoretical Basis and Dm Algorithm

2.1. Theoretical Basis

The audit data come from the enterprise's financial departments. The characteristics of the audit data are seen in Figure 1.

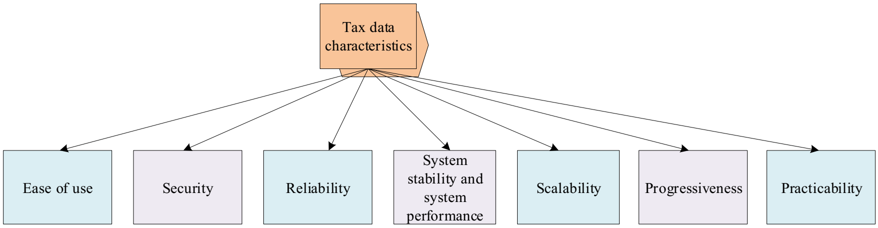

Figure 1. Characteristics of audit data.

Figure 1 shows that audit data are easy to use, safe, reliable, good system performance, stable, scalable, practical, and progressive. Based on these characteristics, appropriate DM technology is selected.

The audit is carried out by an information-based computer audit method. The requirements of the traditional audit include assurance, economic supervision, and evaluation [29]. Based on the traditional audit, computer audit includes another two requirements: a computer system and a computer. Computer audit is realized based on computer technology. The process of computer audit is displayed in Figure 2.

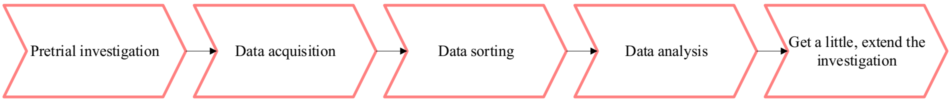

Figure 2. Process of computer audit.

Pre-trail inquiry, data collecting, data sorting, data analysis, and investigation extension are the five steps of computer audit. Pre-trail investigation and data collection are completed in the preparation stage. Data sorting converts and cleans the audited data to ensure data consistency and integrity. Audit analysis processes audit data by implementing an analysis model. Data sorting and model implementation are the focus of audit work [30].

DM technology is a branch of computer science. Its primary function is to extract data patterns and transform them into structures that human beings can understand. The processing flow of DM technology is seen in Figure 3.

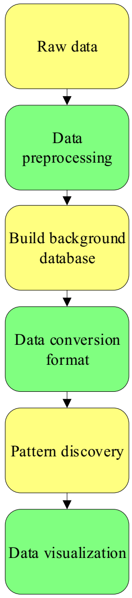

Figure 3. Processing flow of DM technology.

Figure 3 shows that the original data are collected and preprocessed, and a background database is established. After that, the data format is converted, and model implementation and data visualization are performed. Data preprocessing can unify the data format. The background database has the data with a unified format, and the overall batch conversion of the data format is carried out. DM technology includes many algorithms, such as the widely used DT and cluster analysis [31].

Predictive modeling of continuous and discrete parameters is the primary application of DT. The Iterative Dichotomizer-3 (ID3) algorithm is a classification algorithm. A DT is built by the selection window, and the node and branches are the attributes with maximum information in the database [32]. It is widely used for processing images, video data, and binary data. It has a fast classification speed and clear principle, but it also causes a bias problem and is sensitive to noise [33]. C4.5 is developed from ID3, and it is used to realize multi-objective classification. It is made by chopping down ID3's tree structure and adding the processing algorithm for continuous attributes and partial data. To choose qualities, it makes advantage of the information gain rate [34]. C4.5 has a high accuracy rate, easily comprehensible classification algorithms, and directly processable continuous characteristics. However, its program does not work when the training set is large. The Classification and Regression Trees (CART) develop from ID3. Its application scope is broader than that of ID3 and C4.5. It can solve the problem that ID3 cannot be generalized and deal with continuous data directly. Another commonly used DT is the Bayesian classification algorithm, which can improve generalization by learning naive Bayesian parameters [35].

The clustering algorithm is used for data preprocessing. Common clustering algorithms include the hierarchical clustering algorithm, K-means clustering algorithm, Fuzzy C-Means (FCM) clustering algorithm, and Self Organizing Maps (SOM) clustering algorithm. Their details are displayed in Figure 4.

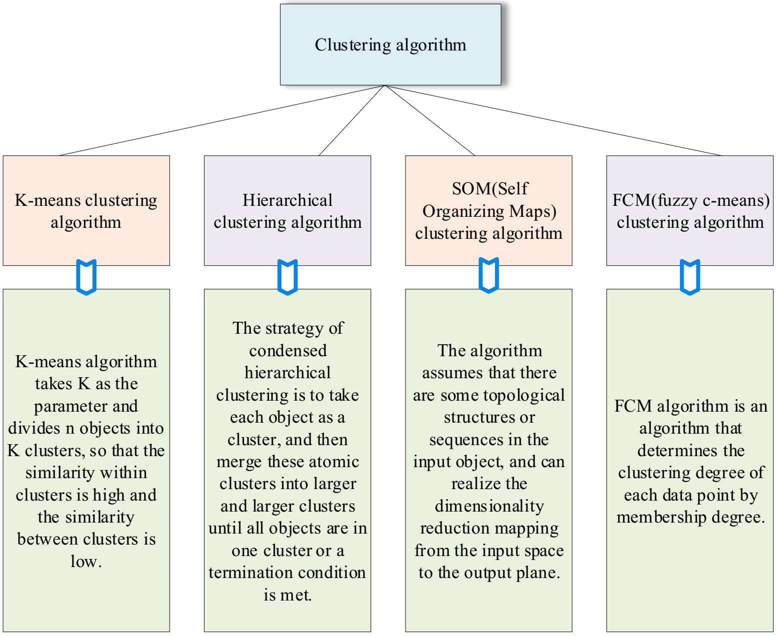

Figure 4. Clustering algorithms.

The effectiveness and widespread use of the K-means method in large-scale data clustering is demonstrated in Figure 4.

In DM, some point groups deviate from the normal trajectory, also known as outliers. Anomaly DM is outlier detection. The commonly used outlier detection algorithms are the outlier detection algorithm based on the regression model, density, nearest neighbor degrees, and distance clusters.

2.2. Computer Audit Based on Dt

The audit of affordable housing emphasizes the actual situation of affordable housing distribution. The audit information includes industrial and commercial registration information, provident fund loan information, real estate information, endowment insurance information, and vehicle information [36].

In data collection, the electronic data and target data are collected. Data collection methods include direct copy, data interface, and backup file recovery. The research object is some original data of the audited units in the research area and the relevant data scales. Data preprocessing includes data cleaning, conversion, integration, and verification. Audit data cleaning needs to clear irrelevant attributes, fill in blank values, and clear noise. Because the data are authentic and their attribute values are within a range, they may be illegal data when input manually. After these data are cleaned, the number of records obtained is 1210, including 855 training sample sets and 360 test sample sets. Their formats are unified through data conversion. C4.5 is used to discretize the continuous data because it can make discrete attributes transform into continuous attributes. Data integration is performed by selecting data attribute subsets. The DT is selected based on the evaluation indexes. There are five performance indexes: accuracy, robustness, computing speed, interpretability, and scalability.

2.3. C4.5 Modeling

After performance comparison, C4.5 is selected for modeling, and its flow is shown in Figure 5.

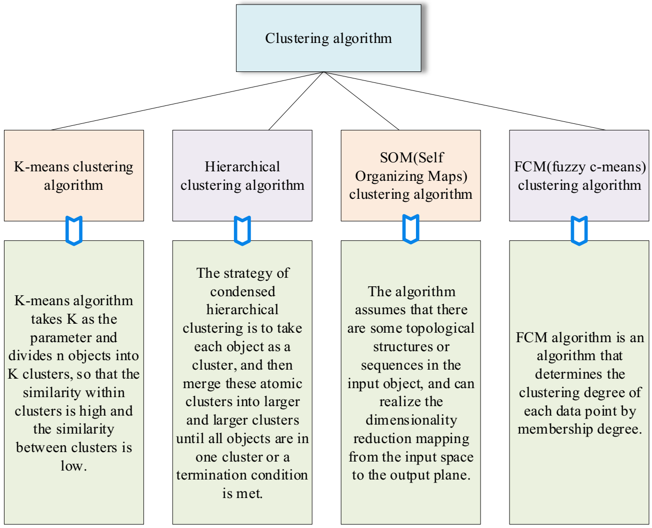

Figure 5. Modeling flow of C4.5.

In Figure 5, C4.5 chooses the information gain rate as both the test and decision attributes.

The computer category information entropy is shown in equation (1):

$$H(C)=-\sum_j p\left(C_j\right)\log_2 p\left(C_j\right)=\mathrm{Info}(F) \quad (1)$$

In equation (1), F is the training sample set, C is the category, p is the probability, and H is the information entropy.

nformation entropy is calculated by equation (2):

$$H(C/V)=-\sum_j p\left(v_j\right)\sum_j p\left(C_j/v_i\right)\log_2 p\left(C_j/v_i\right)=-\sum_{i=1}^{n}\frac{|F_i|}{|F|}\mathrm{Inf}_{o}(F_i)=\mathrm{Info}_v(F) \quad (2)$$

In equation (2), V is the attribute, and n is the number of values.

The information gain is calculated by equation (3):

$$I(C,V)=H(C)-H(C/V)=\mathrm{Inf}_{o}(F)-\mathrm{Info}_v(F)-\mathrm{gain}(v) \quad (3)$$

The splitting information of attribute V is calculated by equation (4):

$$H(V)=-\sum_i p(v_i)\log_2 p(V_i)=-\sum_{i=1}^{n}\frac{|F_i|}{|F|}\log_2\frac{|F_i|}{|F|}=\mathrm{split\_Info}(v) \quad (4)$$

The information gain rate is calculated by equation (5).

$$\text{gain\_ratio}(v)=I(C,V)/H(V)=\text{gain}(v)/\text{split\_Info}(v) \quad (5)$$

The evaluation methods of classification results include the k-fold cross-test method and retention method. The retention method divides the data into the test set and training set, of which two-thirds are allocated to the test set and one-third to the training set. The classifier is obtained by using the training set. The k-fold cross-test method divides the given data into subsets of equal size, takes the centralized data as the training set in turn, and finally obtains the classification model with the highest accuracy rate.

2.4. Improvement of C4.5

The traditional C4.5 has a bias problem in determining test attributes, and some attribute values are relatively concentrated. Therefore, it needs to be improved. First, interest degree γ is introduced into equation (2), and it is the user's interest in uncertain information. At this time, equation (2) transforms into equation (6).

$$H_\gamma(C/V)=-\sum_j p\left(V_j\right)+\frac{1}{|F|}\sum_i \left(C_j/V_i\right)\log_2 p\left(C_j/V_i\right)=\text{Info}_{vx} \quad (6)$$

Equations (7) and (8) are the deformations based on equations (3) and (5).

$$I_\gamma(C,V)=H(C)-H_\gamma(C/V)=\text{Info}(F)-\text{Info}_{v\gamma}(F)=\text{gain}_\alpha(v) \quad (7)$$

$$\text{Gain\_ratio }(v)=I_\alpha(C,V)/H(V)=\text{gain}_\alpha(v)/\text{spl it\_Info }(v) \quad (8)$$

Figure 6 demonstrates the enhanced C4.5's attribute selection procedure.

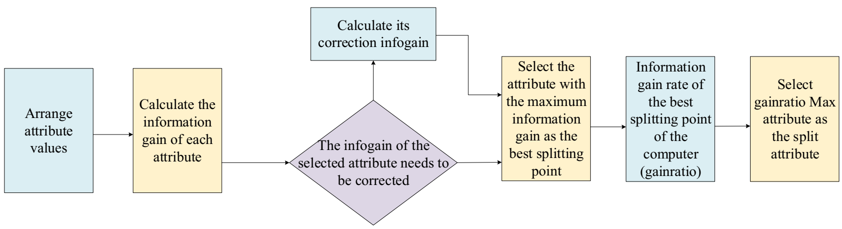

Figure 6. Attribute selection process of the improved C4.5.

The provident fund state attribute is added to the improved C4.5 after the communication with auditors.

Further improvements are necessary because the results indicate that the enhanced C4.5's modeling efficiency is low. The improved C4 5 is simplified to improve modeling efficiency and shorten the modeling time. Based on the previous research [37], the equivalent infinitesimal principle and Taylor series are used to improve the performance of C4.5. The improved C4.5 using logarithmic function operations is compared with that using four simple mixed operations.

2.5. Simulation Experiments

The data set used for performance comparison and experiment analysis comes from 9 data sets of the Waikato Environment for Knowledge Analysis (WEKA) website. The comparison algorithms include CART, C4.5, and the naive Bayes tree (NBTree). The attributes of data samples include continuous and discrete attributes to ensure the comprehensive detection of their performances. Table 1 displays the experimental data information.

Table 1. Information of experimental data

Samples	wisconsin	vowel	strike	servo	pollution	optdigits	meta	fruitfly	ecoli
Numbers	17	36	17	35	9	21	20	10	10
Number of recorded	435	683	57	351	768	1000	810	286	214

In Table 1, the sample sets with large data sets: pollution, optdigits, and meta. CART, C4.5, and NBTree are used to classify the above data samples. The comparison of experimental results includes accuracy rates, modeling speeds, robustness, scalability, and interpretability. Based on the experimental results, C4.5 is selected for modeling.

WEKA3.8.1 is the experimental environment. A test set of 30% of the data and a training set of 70% of the data are chosen. The eclipse is the programming environment used to implement the enhanced C4.5. Furthermore, a comparison is made between the enhanced C4.5's performance and the conventional C4.5's.

3. Analysis of Algorithm Performance

3.1. Comparison of Algorithm Performance

The accuracy rates of CART, C4.5, and NBTree are compared, as shown in Figure 7.

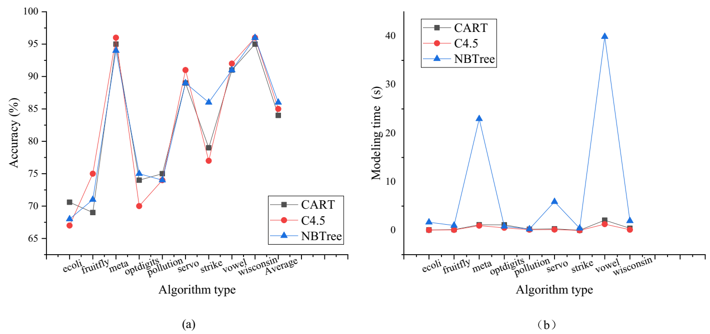

Figure 7. Comparison of the accuracy rates and modeling speeds. (a. comparison of the accuracy rates; b. comparison of the modeling speeds)

Figure 7(a) shows that the accuracy curves of CART, C4.5, and NBTree have different fluctuations, but the peak value is almost the same. The average accuracy rates of the three algorithms are discussed. The average accuracy rate of CART is 85.4%, that of NBTree is 85.7%, and that of C4.5 is 85.6%. This shows that C4.5 has a better classification effect.

The modeling speeds of the three techniques are displayed in Figure 7(b). With a modeling time of 39.9 seconds, NBTree has the slowest modeling speed. The modeling speeds of CART and C4.5 have no significant difference. The overall modeling speed of C4.5 is faster than that of CART. NBTree has the fastest modeling speed of 0.02 seconds, while C4.5 has the shortest modeling time of 0. C4.5 is relatively steady when the size of the data set is high, while CART requires a rapidly increasing amount of time. In terms of modeling speed and classification accuracy rate, CART and C4.5 outperform NBTree, as demonstrated in Figures 7(a) and 7(b). Additionally, they are highly scalable.

The robustness of the algorithms is compared based on the vowel test sample set and the vowel challenge sample set. Figure 8 presents a comparison of the modeling speed and accuracy rate between the two datasets.

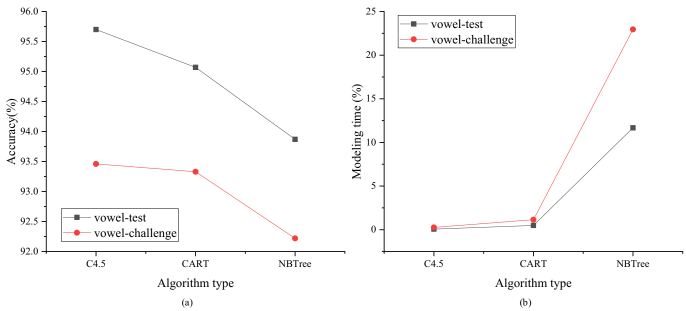

Figure 8. Robustness comparison. (a. comparison of mandatory classification accuracy rates; b. comparison of the modeling speed for robustness)

According to Figure 8, C4.5 performs the best in terms of modeling speed and classification accuracy rate. NBTree has the least loss in accuracy rate and the least increase in modeling time. Those of CART rank second. This shows that CART and NBTree have strong robustness.

The ratios of rules and leaf nodes of CART, C4.5, and NBTree are demonstrated in Figure 9.

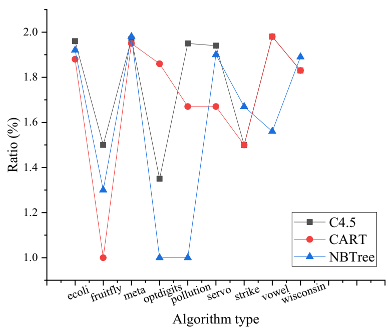

Figure 9. Comparison of the ratios of rules and leaf nodes.

Figure 9 shows that the ratio of rules to leaf nodes of C4.5 is higher than that of the other two algorithms. This proves that C4.5 has high interpretability. Compared with the leaf node ratio of NBTree, CART has no advantage.

3.2. Performance Comparison of the Improved Algorithm

The improved C4.5 model and traditional C4.5 are compared in the classification accuracy rate, modeling time, and model interpretability, and Figure 10 presents the findings.

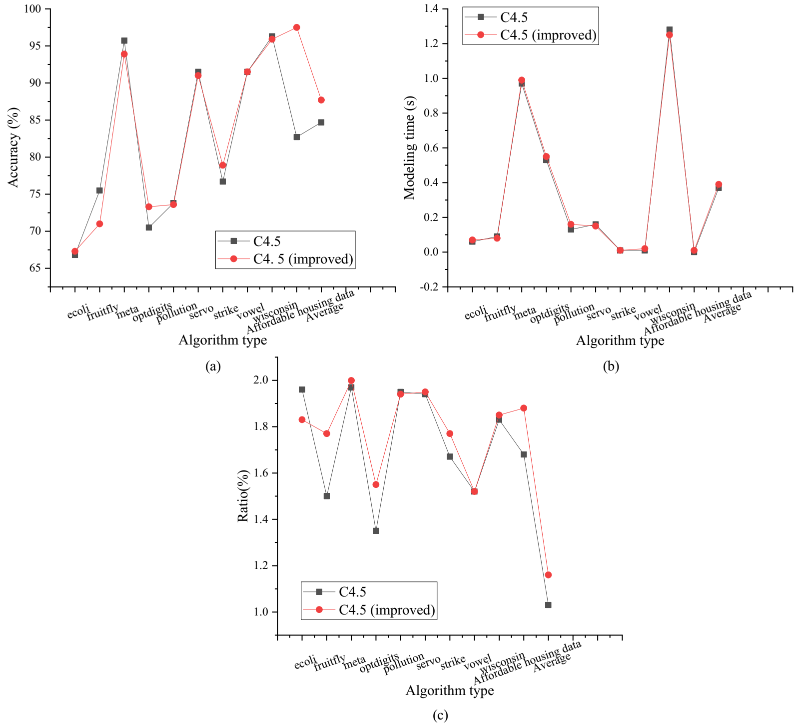

Figure 10. Comparison of the accuracy rates of the improved model and the traditional model. (a. comparison of the accuracy rate; b. comparison of the modeling time; c. comparison of the interpretability)

The accuracy rate of the upgraded C4.5 is higher than that of the conventional C4.5, as seen in Figure 10(a). When it comes to classifying affordable housing data, its accuracy rate is 97.5%. The enhanced C4.5 has an average accuracy rate of 87.7%, whereas the conventional C4.5 has an average accuracy rate of 84.7%. Figure 10 (b) displays the modelling time comparison between the modified algorithm and the conventional algorithm. There is hardly discernible variation in the modelling times. The enhanced C4.5 takes 0.39 seconds on average to model, while the conventional C4.5 takes 0.37 seconds on average. This demonstrates that the enhanced C4.5's modeling speed has not increased. Their comparison of the interpretability is displayed in Figure 10(c). Compared to the conventional C4.5 algorithm, the modified approach has a greater rules-to-leaf node ratio. The number of rules and leaf node ratio are used to determine interpretability. The interpretability of the model improves with a larger conversion ratio of a few leaf nodes. In contrast to the conventional C4.5, the enhanced C4.5 exhibits strong interpretability.

The performances of C4.5, the enhanced C4.5, and the C4.5 following four straightforward operations are compared using data on affordable housing. Table 2 displays the outcomes of the comparison.

Table 2. Performance comparison of different C4.5 algorithms

Model category/model performance	Accuracy rates	Modeling time	Interpretability
C4.5	84.7%	0s	1.7
Improved C4.5	97.5%	0.01s	1.9
C4.5(after four simple operations)	98.7%	0s	1.9

Table 2 shows that the average performance of the simplified C4.5 is the best, with an accuracy rate of 98.7%, its modeling time is 0, and its interpretability is 1.9. This shows that the research method improves the accuracy rate in a specific environment, and the equivalent infinitesimal and Taylor equation shortens the generation time of a DT.

4. Conclusion

The audit process of DM technology is introduced, the research state of DM technology and affordable housing is assessed, and the benefits and drawbacks of common DM technologies are contrasted in light of the rapid growth of computer technology. The C4.5 algorithm is used to classify the audit data of affordable housing. After the audit data are collected, classified, and preprocessed, a DT model is implemented by C4.5, and the classification results are evaluated. According to the evaluation results, C4.5 is improved. The classification accuracy rate of the simplified C4.5 is 98.7%. The research results provide a reference for the audit work of affordable housing. However, there are also some shortcomings. For example, the data have some objective and subjective deviations because they are the measured data in the study area, affecting the experimental results; using the Taylor equation to simplify C4.5 needs further discussion because the Taylor equation may still need to be improved. Therefore, the follow-up research will deeply study the data preprocessing technology, and the Taylor equation will be discussed further

References

[1] Putri DR, Ismanu S, Haris ZA. Computer Assistant Substantive Test Model In Audit Of Government Institution Financial Statements. J Ind Eng Manage Res. 2021 Jan; 2(4), 209-215, doi.org/10.7777/jiemar.v2i4.173

[2] Ali A, Ahmed M, Khan A. Audit logs management and security-A survey. Kuwait J Sci. 2021 Feb;48(3),211, doi.org/10.48129/kjs.v48i3.10624

[3] Dragonas E, Lambrinoudakis C, Kotsis M. IoT forensics: Exploiting unexplored log records from the HIKVISION file system. J Forensic Sci. 2023 Mar;68(6),2002-2011, doi.org/10.1111/1556-4029.15349

[4] Raynor K, Coenen L. Business model innovation and scalability in hybrid affordable housing organisations: empirical insights and conceptual reflections from Melbourne, Australia. J Hous Built Environ. 2021 Apr;12,20, doi.org/10.1007/s10901-021-09836-x

[5] Zvarikova K, Machova V, Nica E. Cognitive Artificial Intelligence Algorithms, Movement and Behavior Tracking Tools, and Customer Identification Technology in the Metaverse Commerce. Rev Contemp Philos. 2022 Dec;21,171-187, https://www.ceeol.com/search/article-detail?id=1071101

[6] Yang M, Kumar P, Bhola J, Shabaz M. Development of image recognition software based on artificial intelligence algorithm for the efficient sorting of apple fruit. Int J Syst Assur Eng. 2022 Jan;13(Suppl 1),322-330, doi.org/10.1007/s13198-021-01415-1 Oct.

[7] Rampini L, Re Cecconi F. Artificial intelligence algorithms to predict Italian real estate market prices. J Prop Invest Financ. 2022 Mar;40(6),588-611, doi.org/10.1108/JPIF-08-2021-0073

[8] Sohail A. Genetic algorithms in the fields of artificial intelligence and data sciences. Annals Data Sci. 2023 Jul;10(4),1007-1018, doi.org/10.1007/s40745-021-00354-9

[9] Rodgers W, Nguyen T. Advertising benefits from ethical artificial intelligence algorithmic purchase decision pathways. J Bus Ethics. 2022 May;178(4),1043-1061, doi.org/10.1007/s10551-022-05048-7

[10] Jiang Y, Li X, Luo H, Yin S, Kaynak O. Quo vadis artificial intelligence?. Disco Artif Intell. 2022 Aug;2(1),4, doi.org/10.1007/s44163-022-00022-8

[11] Sarker IH. Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. SN Comput Sci. 2022 Sept;3(2),158, doi.org/10.1007/s42979-022-01043-x

[12] Wang FY, Miao Q, Li X, Wang X, Lin Y. What does ChatGPT say: The DAO from algorithmic intelligence to linguistic intelligence. IEEE/CAA J Autom Sin. 2023 Feb;10(3),575-579, doi.org/10.1007/s42979-022-01043-x

[13] Alkahtani H, Aldhyani TH. Artificial intelligence algorithms for malware detection in android-operated mobile devices. Sensors. 2022 Oct;22(6),2268, doi: 10.4018/978-1-6684-6291-1.ch024

[14] Sestino A, De Mauro A. Leveraging artificial intelligence in business: Implications, applications and methods. Technol Anal Strateg. 2022 Nov;34(1),16-29, doi.org/10.1080/09537325.2021.1883583

[15] Mahmoodi Khaniabadi S, Javadpour A, Gheisari M, Zhang W, Liu Y, Sangaiah AK. An intelligent sustainable efficient transmission internet protocol to switch between User Datagram Protocol and Transmission Control Protocol in IoT computing. Expert Syst. 2023 Jun;40(5),e13129, doi.org/10.1111/exsy.13129

[16] Sharma A, Shoval S, Sharma A, Pandey JK. Path planning for multiple targets interception by the swarm of UAVs based on swarm intelligence algorithms: A review. Iete Tech Rev. 2022 Aug;39(3),675-697, doi.org/10.1080/02564602.2021.1894250

[17] Hua, H., Li, Y., Wang, T., Dong, N., Li, W., & Cao, J. (2023). Edge computing with artificial intelligence: A machine learning perspective. Acm Comput Surv. 55(9), 1-35. doi.org/10.1145/3555802

[18] Nti IK, Adekoya AF, Weyori BA, Nyarko-Boateng O. Applications of artificial intelligence in engineering and manufacturing: A systematic review. J Intell Manuf. 2022 Dec; 3(6),1581-1601, doi.org/10.1007/s10845-021-01771-6

[19] Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. J Amb Intel Hum Comp. 2023 May;14(7),8459-8486, doi.org/10.1007/s12652-021-03612-z

[20] Ouyang F, Zheng L, Jiao P. Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. Educ Inf Technol. 2022 Nov;27(6),7893-7925, doi.org/10.1007/s10639-022-10925-9

[21] Farina E, Nabhen JJ, Dacoregio MI, Batalini F, Moraes FY. An overview of artificial intelligence in oncology. Futur Sci Oa. 2022 Jan;8(4),FSO787, doi.org/10.2144/fsoa-2021-0074

[22] Akgun S, Greenhow C. Artificial intelligence in education: Addressing ethical challenges in K-12 settings. AI Ethics. 2022 Dec;2(3),431-440, doi.org/10.1007/s43681-021-00096-7

[23] Zan J. Research on robot path perception and optimization technology based on whale optimization algorithm. J Comput Cogn Eng. 2022 Feb;1(4),201-208, doi.org/10.47852/bonviewJCCE5978202055

[24] Kuo RY, Harrison C, Curran TA, Jones B, Freethy A, Cussons D, et al. Artificial intelligence in fracture detection: a systematic review and meta-analysis. Radiology. 2022 Nov;304(1),50-62, doi.org/10.1148/radiol.211785

[25] Dwivedi S, Vardhan M, Tripathi S. Defense against distributed DoS attack detection by using intelligent evolutionary algorithm. Int J Comput Appli. 2022 May;44(3),219-229, doi.org/10.1080/1206212X.2020.1720951

[26] Bradshaw TJ, Boellaard R, Dutta J, Jha AK, Jacobs P, Li Q, et al. Nuclear medicine and artificial intelligence: best practices for algorithm development. J Nucl Med. 2022 Mar;63(4),500-510, doi.org/10.2967/jnumed.121.262567

[27] Abdullahi M, Baashar Y, Alhussian H, Alwadain A, Aziz N, Capretz LF, Abdulkadir SJ. Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review. Electronics. 2022 Jul;11(2),198, doi 10.3390/electronics11020198

[28] Mendonca RV, Silva JC, Rosa RL, Saadi M, Rodriguez DZ, Farouk A. A lightweight intelligent intrusion detection system for industrial internet of things using deep learning algorithms. Expert Syst. 2022 Aug;39(5),e12917, doi.org/10.1111/exsy.12917

[29] Sembiring I, Iriani A, Ginting JVB, Ginting JA. A Novel Approach to Network Forensic Analysis: Combining Packet Capture Data and Social Network Analysis. Int J Adv Comput Sc. 2023 Jul;14(3),114, doi:10.14569/IJACSA.2023.0140353

[30] Guarnera L, Giudice O, Livatino S, Paratore AB, Salici A, Battiato S. Assessing forensic ballistics three-dimensionally through graphical reconstruction and immersive VR observation. Multimed Tools Appl. 2023 Feb;82(13),20655-20681, doi.org/10.1007/s11042-022-14037-x

[31] Ageed ZS, Zeebaree SR, Sadeeq MM, Kak SF, Yahia HS, Mahmood MR, Ibrahim IM. Comprehensive survey of big data mining approaches in cloud systems. Qubahan Acad J. 2021 Dec;1(2),29-38, doi.org/10.48161/qaj.v1n2a46

[32] Wang Z, Zhang Y, Chen Y, Liu H, Wang B, Wang C. A Survey on Programmable Logic Controller Vulnerabilities, Attacks, Detections, and Forensics. Processes. 2023 Nov;11(3),918, doi.org/10.3390/pr11030918

[33] Joseph DP, Viswanathan P. SDOT: Secure Hash, Semantic Keyword Extraction, and Dynamic Operator Pattern-Based Three-Tier Forensic Classification Framework. IEEE Access. 2023 Jun;11,3291-3306, doi: 10.1109/ACCESS.2023.3234434

[34] Regin R, Rajest SS, Singh B. Spatial Data Mining Methods Databases and Statistics Point of Views. Innov Inform Commun Tec S. 2021 Sept;10,103-109, doi.org/10.4108/eai.3-5-2021.169578

[35] Nakaji K, Yamamoto N. Quantum semi-supervised generative adversarial network for enhanced data classification. Sci Rep-Uk. 2021 Feb;11(1),1-10, doi.org/10.1038/s41598-021-98933-6

[36] Dubey AK, Kumar A, Agrawal R. An efficient ACO-PSO-based framework for data classification and preprocessing in big data. Evol Intell. 2021 May;14(2),909-922, doi.org/10.1007/s12065-020-00477-7

[37] Basiri A, Rahmani S, Mirkarim MB. Infinitesimal generators of Lie symmetry group of parametric ordinary differential equations. Int J Nonlinear Anal. 2021 Oct;12(1),877-891, doi.org/10.22075/IJNAA.2021.22387.2356