Adult and Higher Education (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/aduhe.2023.051916
ISSN 2523-5826 Vol. 5 Num. 19

Utilization of Artificial Intelligence Technology in Higher Education Management

Longlong Wang[1,2,a]

[1]*School of Education, Philippine Women's University, Ermita Manila, Metro Manila, Philippine*
[2]*Department of Accounting, Shandong College of Economics and Business, Weifang, Shandong, China*
[a]*2022t1048@pwu.edu.ph*

Keywords: University Education Management, Artificial Intelligence Technology, Personalized Course Recommendation, C4.5 Decision Tree Algorithm

Abstract: Traditional university education management has issues such as low efficiency and lack of personalization. As artificial intelligence (AI) technology develops rapidly, its application in educational management in universities is increasingly becoming a focus of attention for academics and educational institutions. To explore the application of AI technology in higher education management, this paper focused on personalized course recommendations for students. The data from the 2010 KDD Cup Education Data Mining Challenge dataset was collected and cleaned using Talend and Apache Spark tools; information features were extracted using information gain, and finally the data was trained using the C4.5 decision tree algorithm to obtain a recommendation model. After experiments, the precision of this model for students' preferences in course selection reached 94%, and the F1 value of the model reached 0.93, indicating that the model had good precision and comprehensiveness. At the same time, the highest recommended course click through rate reached 0.39, indicating that the personalized recommendation ability of the model was excellent. This model improved the efficiency of students' course selection and the utilization of educational resources, exploring new ways for university education management.

1. Introduction

As various technologies have developed, today's society has entered the era of information technology. The management of higher education plays an essential role in the current education system, as it bears some of the responsibility of shaping the future society. The traditional management model of universities has some drawbacks in responding to the rapidly changing educational needs: it usually relies on tedious manual operations, and the decisions made are often subjective, making it difficult to meet the high efficiency and personalized needs of education management today.

To explore the utilization of AI technology in higher education management, this article first explored the methods explored by many scholars in higher education management, and then discussed examples of AI in education management. Subsequently, research was conducted on personalized course recommendation for students in higher education management. After collecting data, features were selected using the information gain method, and the C4.5 decision tree algorithm was used to train the data and construct a recommendation model. The experimental results showed that the model had good personalized recommendation performance, reducing the pressure on students to choose courses, improving the utilization of educational resources, and providing research ideas for the utilization of AI in university management.

2. Literature Review

In order to explore new ideas for higher education management, many scholars have begun to conduct a series of studies. Bhaskar P [1] and others discussed the utilization of blockchain technology in education management, and they believed that blockchain education technology is still a young discipline, but has great potential to benefit the entire education sector. Nurdiansyah N M [2] and others believed that education management should achieve educational quality goals through the participation of principals, teachers, students, stakeholders, and the community, rather than relying solely on managers. Scholars including Fearnley M R [3] used a technology acceptance model to investigate the factors influencing higher education teachers' adoption of learning management systems. The findings indicated that system quality and self-efficacy strongly affected perceived usefulness, which in turn indirectly affected attitudes and behavioral intentions towards technology. Song Yi [4] and others discussed the issue of safety management in teaching laboratories in higher education institutions. Scholars including Balayan A [5] discussed the evolution of graduate enrollment management and believed that graduate enrollment management should be given the same importance as undergraduate education. Ning Wang [6] believed that in the management of students in universities, counselors' guidance on students' psychological health and care for their daily lives are crucial for their healthy development. The level of counselor work and the construction of professional qualities are the key to determining educational outcomes. Scholars such as Nazem F [7] aimed to establish an employee performance model for Islamic Azad University that considers intellectual capital and knowledge management. Through a questionnaire survey, it was believed that the performance of university employees was correlated with their intelligence and knowledge management level, and it was believed that strengthening the quality improvement of employees can help better complete university education management. Scholars' research has discussed the important role of managers in the management of higher education, but the problems of relying on manpower, low efficiency, and strong subjectivity remain unresolved.

AI technology is a technology that studies how to enable computers to exhibit intelligent behavior. AI technology covers multiple sub fields, including machine learning, deep learning, natural language processing, computer vision, and more. AI technology is also widely used in university education management. Zhou Lin [8] linked AI technology with the ideological and political education of the Sports Talent Training Center, and believed that the assistance of AI technology in the education process has important practical significance. Alam A [9] discussed the potential and challenges of using composite AI in the field of education in India. Tian Xianpeng [10] conducted research on the governance transformation of educational data in the era of AI, and believed that the urgent challenge faced by educational data governance in the era of AI is to balance privacy protection and open sharing, and better utilize AI for the utilization of educational data. Scholars such as Renz A [11] explored the opposing relationship between traditional educational ideals and future education and knowledge transfer concepts, and the desire for flexibility and personalization has driven the debate on AI based learning systems. Scholars such as Cope B [12] proposed reflections on the limitations and potential of machine intelligence in education, and they believed that AI can make education more humane, but it would never replace human education. Numerous researchers have explored the relationship between AI and educational management, laying a solid foundation for the exploration of educational management, whether in terms of ideological education or data management [13].

3. Exploration of Methods

The content of higher education management covers all aspects, aiming to enhance the quality of education, support student development, and ensure the normal operation of higher education institutions. Its content includes multiple aspects: enrollment management, curriculum planning and teaching management, student service and management, property and asset management, scientific research and academic development, and faculty management, among others.

This study focuses on students' course planning issues and designs a personalized course recommendation system for students. It provides schools with more intelligent and higher education teaching services to meet the diverse learning needs of students.

3.1 Basic Process of Personalized Recommendation

The system involves multiple aspects such as data collection, feature extraction, and algorithm selection. The specific process is illustrated in Figure 1:

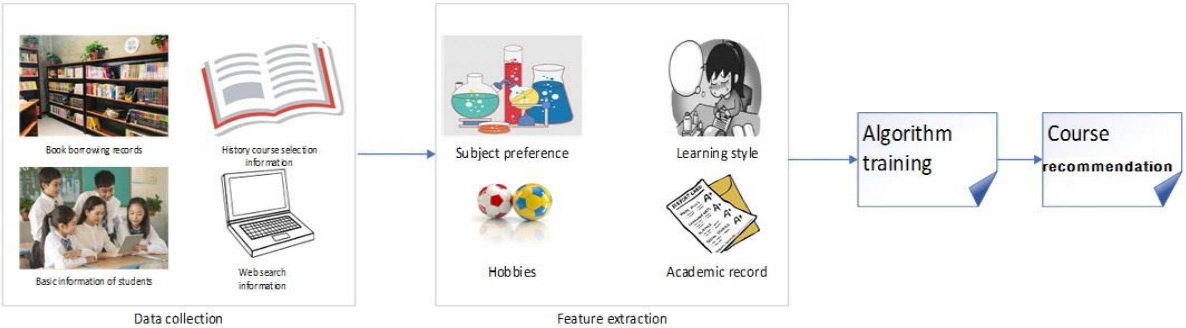

Figure 1: Personalized course recommendation algorithm flowchart

From Figure 1, it can be seen that when collecting data, it includes students' personal information, historical course selection records, records borrowed from the library, and records of daily search information. Through this information, students' subject preferences, learning styles, interests, and academic achievements are collected, and information features are input into appropriate algorithm models for training. Finally, the expected courses of the students are analyzed and recommended to them.

3.2 Information Data Collection and Cleaning

The first step in establishing a system is to collect data, which requires accuracy and completeness. The information in the process can be collected in the following ways: library borrowing records, historical course selection records, and basic student information can all be obtained through the school's management system; Students' online search information is obtained through a questionnaire survey out of respect for their privacy. The tool for collecting information uses Talend in the ETL (Extract, Transform, Load) tool, which is an open source ETL tool that provides rich data extraction, transformation, and loading functions. It supports multiple data sources and conversion operations, and can customize data processes according to the needs of the school management system.

After data collection, it is necessary to clean the data. Data cleaning can ensure the high quality, accuracy, and availability of the data, ensure data consistency, and lay the foundation for subsequent data feature extraction and analysis. Data cleaning includes the following steps: handling missing values, handling outliers, and formatting consistency.

Missing value processing is the process of dealing with possible missing parts in information, which can be filled in through the method of mean or median.

The handling of outliers is to identify outliers through statistical methods or domain expert knowledge. For example, the age data of a student is three digits, which is obviously unreasonable; it is also unreasonable for students to choose two courses that are conducted simultaneously in their history course selection information.

Format consistency is relatively simple, which involves processing different data formats in a consistent manner, such as date format, course format, etc.

The data cleaning tool uses the Apache Spark tool, which is a distributed computing framework that can be used for large-scale data cleaning.

3.3 Selection of Information Features

Feature selection is an important step before model training and has a significant impact on the precision of the final model. Common feature selection methods include Mutual Information (MI), Odds Ratio (OR), and Information Gain (IG). They use different methods for feature selection, and this study used the information gain method. Its core idea is to select features that can minimize information entropy. Entropy describes the distribution of energy in space, and the more uniform the energy distribution, the greater the entropy value. Information entropy is used to measure the purity of data and help select the best feature data.

For a discrete random sample a, the entropy is defined as $H(a)$, and its formula is expressed as:

$$H(a) = -\sum_{i=1}^{n} P_a \cdot \log_2 P_a \tag{1}$$

Among them, P_a represents the probability of variable a output, and n represents the number of categories of samples in the dataset. If there is only type a in the dataset, then n=1, $P_a = 1$, $\log_2(1) = 0$, and the information entropy is $H(a)$=0, indicating that the information distribution in this dataset is extremely uneven.

The conditional entropy $H(X, A)$ is defined as the entropy value after the occurrence of event A, and its expression is:

$$H(X, A) = -\sum_{i=1}^{n} P(X, A) \cdot \log P(X, A) \tag{2}$$

Before and after the occurrence of event A, the magnitude of conditional entropy $H(X, A)$ changes, and the expression for the difference between it and the original entropy is:

$$IG(X) = H(X) - H(X, A) \tag{3}$$

Then, $IG(X)$ is the information gain of event A. The larger the $IG(X)$, the greater the degree of entropy change after event A occurs, and the greater the ability of the event to reduce uncertainty. By analogy, in information feature extraction, the greater the information gain, the better the feature can distinguish differences between different categories and improve the fitting ability in model training. Figure 2 shows the fitting curve of information gain and fitting ability based on this experimental study:

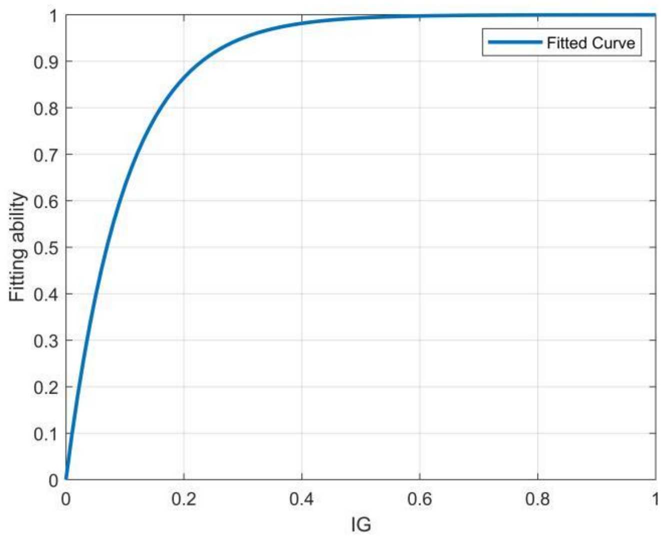

Figure 2: Relationship between information gain and fitting ability

From Figure 2, it can be seen that in this study, the improvement of information gain on model fitting ability is nonlinear, mainly in the range of 0 to 0.4. Further improvement of information gain on model fitting ability is limited. Therefore, when extracting information features, it is not necessarily that the greater the information gain of the features, the better it is. When reaching a certain threshold, it would increase the computational burden of the model and increase processing time.

3.4 Training of Decision Trees

Information gain is usually used for feature partitioning in decision trees. Decision Tree (DT) is a common machine learning algorithm commonly used for regression and classification problems. It constructs a tree like structure through recursive partitioning of the dataset from a root node to an internal node and then to a leaf node. Each internal node represents the testing of a feature, and each leaf node is a category. The structural diagram is shown in Figure 3:

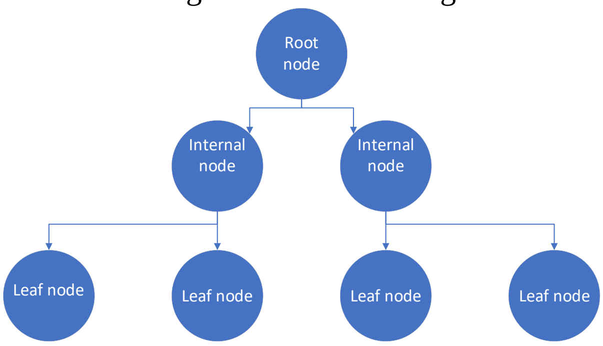

Figure 3: Decision tree structure diagram

This experiment uses the C4.5 decision tree algorithm in the decision tree, which is different from the simple decision tree algorithm. It only focuses on attributes with high information gain, but instead selects the feature attribute with the highest proportion of useful information for the next step of splitting, known as the Information Gain Ratio. The information gain rate solves the preference problem of information gain for features with a large number of values, and it punishes the number of feature values. The formula for calculating the information gain rate is as follows:

$$\mathrm{GainRatio}(S,A)=\frac{\mathrm{IG}(X)}{\mathrm{SplitInfo}(S,A)} \tag{4}$$

Among them, $\mathrm{IG}(X)$ is the information gain; A is the feature; S is the dataset; $\mathrm{SplitInfo}(S,A)$ is the split information of feature A, which is used to penalize cases with high feature values.

4. Personalized Recommendation Experiment

4.1 Datasets

This experiment uses the dataset from the 2010 KDD Cup Education Data Mining Challenge and selects 2000 student information for testing. Some of the experimental equipment used in this experiment is shown in Table 1:

Table 1: Partial equipment model diagram

Operating system	Window10 64-bit
Software environment	Java、C#
Central Processing Unit	AMD Ryzen 7 5800X
Graphics Processing Unit	AMD Radeon RX 6700 XT
Random Access Memory	8GB*2

4.2 Evaluation Indicators

In order to test whether the recommended courses can meet the preferences of the recommended candidates, this study uses Hit Rate (HR), P (Precision), and F1 Score to evaluate the performance of the model.

In students' personalized course recommendations, the click through rate reflects how many courses have actually been clicked on by students in the personalized course recommendation list generated by the recommendation algorithm. Its expression is:

$$\mathrm{HR}=\frac{n}{N} \tag{5}$$

Among them, n is the number of recommended courses that have been clicked on, and N is the total number of recommended courses.

Precision measures the accuracy of model predictions, reflecting the degree of accuracy recommended by the model to students. Its expression is:

$$P=\frac{TP}{TP+FP}\times 100\% \tag{6}$$

The F1 score is a comprehensive indicator for evaluating model performance, expressed as:

$$F1=\frac{2\times R\times P}{R+P} \tag{7}$$

Among them, P represents precision and R represents recall rate.

4.3 Personalized Recommended Course Experimental Results

Courses are classified based on the characteristics of different categories of courses. The classification method is shown in Figure 4:

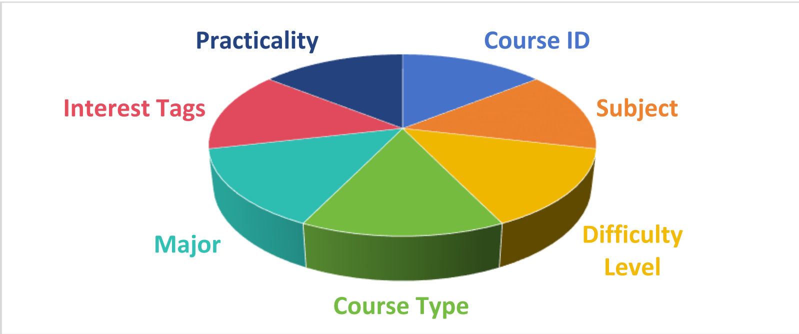

Figure 4: Schematic diagram of classification method

From Figure 4, it can be seen that the course of this experiment is classified based on the seven features shown in the diagram, which intersect and overlap with each other. Finally, 45 recommended types of results are obtained.

The extracted features are trained into the model and Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, and Naive Bayes algorithms are simultaneously introduced for comparative experiments. 2000 pieces of information are divided into training and testing sets in an 8:2 ratio, and precision and F1 values are calculated. The results are shown in Table 2:

Table 2: Training and test results of each algorithm

	Training set		Test set	
Algorithm	P(%)	F1	P(%)	F1
SVM	79	0.78	85	0.87
KNN	81	0.79	89	0.89
Logistic Regression	77	0.76	90	0.88
Naive Bayes	83	0.81	87	0.84
DT	87	0.85	94	0.93

From Table 2, it can be seen that regardless of the algorithm, it performs better in the test set after training than during testing. The decision tree algorithm used in this study performs well in both the training and testing sets. The precision rate in the training set is 87%, which is ahead of other algorithms and 10% higher than the lowest Logistic Regression algorithm. The F1 value reaches 0.85, which does not reach the ideal 0.9 or above. In the test set, the precision of the decision tree algorithm reached 94%, which is 7% higher than the training set, and the F1 value also reached 0.93, which is higher than 0.9, indicating that the model performs very well in terms of precision and comprehensiveness.

Figure 5 shows the results of click through rates for 45 recommendation types:

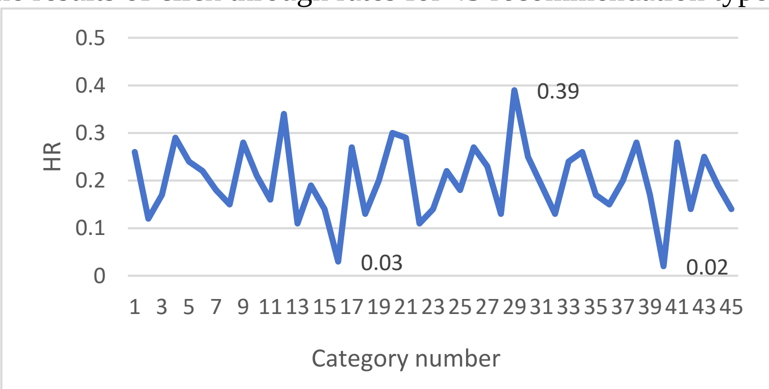

Figure 5: Schematic diagram of click through rates for different recommendation types

From Figure 5, it can be seen that the type with the highest click through rate is Type 29, with a click through rate of 0.39. The type with the lowest click through rate is Type 40, with a click through rate of only 0.02. Under normal circumstances, HR values between 0.1 and 0.3 are within a reasonable range. The HR values in this experiment are roughly within this range, with only the 16th and 40th types being lower than 0.1. This reflects the overall strong personalized recommendation ability of the algorithm.

5. Conclusions

AI technology plays an essential role in university education management, and personalized course recommendation is of great significance in promoting students' diversified learning and improving the efficiency of educational resource utilization. This article presented a personalized course recommendation model based on the C4.5 decision tree algorithm, which can accurately predict students' preferences for course selection. The recommended courses have also received good feedback, with excellent click through rates. A slight deficiency is that this experiment has not yet planned for the balance of school curriculum resources, which may lead to over recommendation of popular courses and neglect of unpopular courses. It is hoped to solve this problem in the future by combining more expert opinions.

References

[1] Bhaskar P, Tiwari C K, Joshi A. Blockchain in education management: present and future applications. Interactive Technology and Smart Education, 2020, 18(1): 1-17.

[2] Nurdiansyah N M, Hudriyah H. Policy and Implementation of Education Management Based on Madrasah. Al-Tanzim: Jurnal Manajemen Pendidikan Islam, 2021, 5(1): 14-27.

[3] Fearnley M R, Amora J T. Learning Management System Adoption in Higher Education Using the Extended Technology Acceptance Model. IAFOR Journal of Education, 2020, 8(2): 89-106.

[4] Song Yi, Gao Dongfeng. Practice on Improving the Safety Management System and Management Ability of Teaching Laboratories in Higher Education Institutions in the New Era. Experimental Technology and Management, 2020, 37 (1): 1-2.

[5] Balayan A, Connor C, LaFave J. The Evolution of Graduate Enrollment Management. Strategic Enrollment Management Quarterly, 2022, 9(4): 35-46.

[6] Ning Wang. Exploring the Methods of Student Management for College Counselors in the New Situation. Education Science Development, 2021, 3 (11): 213-215.

[7] Nazem F, Mazaheri S, Amanihamedani M. Performance Model of Islamic Azad University Staff Considering Intellectual Capital and Knowledge Management. Webology, 2022, 19(1): 6489-6503.

[8] Zhou Lin. The combination of ideological and political education and artificial intelligence in the cultivation of sports talents in Chinese universities. Journal of Shenyang Institute of Physical Education, 2021, 40 (5): 40-45.

[9] Alam A. Possibilities and challenges of compounding artificial intelligence in India's educational landscape. International Journal of Advanced Science and Technology, 2020, 29(5): 5077-5094.

[10] Tian Xianpeng. Privacy Protection and Open Sharing: The Transformation of Educational Data Governance in the Age of Artificial Intelligence. Research on Audiovisual Education, 2020, 41 (5): 33-38.

[11] Renz A, Hilbig R. Prerequisites for artificial intelligence in further education: Identification of drivers, barriers, and business models of educational technology companies. International Journal of Educational Technology in Higher Education, 2020, 17(1): 1-21.

[12] Cope B, Kalantzis M, Searsmith D. Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. Educational Philosophy and Theory, 2021, 53(12): 1229-1245.

[13] James Y. L. Teaching Methods of Power Mechanical Engineering Based on Artificial Intelligence. Kinetic Mechanical Engineering (2022), Vol. 3, Issue 3: 54-61.