International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-13

https://doi.org/10.70917/ijcisim-2026-1784
Article

Exploring the Process Re-engineering and Effectiveness Enhancement Path of Academic Affairs Management in Colleges and Universities Empowered by Digital Transformation

Yuqiu Wang[1] and Aida Hanim Binti A. Hamid[2,*]

[1] Nanjing Audit University Jinshen College, Nanjing, Jiangsu, 210023, China
[2] Senior Lecturer in Educational Management at the National University of Malaysia(UKM), 43600 Bangi, Selangor, Malaysia
* Correspondence author: P147317@siswa.ukm.edu.my

Abstract: Academic affairs management is the core of all management work in colleges and universities, and all colleges and universities are actively promoting the scientific and standardized management mode, and the management means are also moving towards digitalization and automation. In this paper, an efficient and accurate transformation prediction method for the digital transformation of academic affairs management is constructed based on the XGBoost algorithm, and the results of the students' academic early warning are divided into three categories. At the same time, feature selection as well as uneven data processing are performed on the data. The accuracy rate and F1 score were used as evaluation indexes to compare the XGBoost prediction model with SVM, RF and other algorithmic models, and SHAP was used to analyze the interpretability of the model. Data analysis showed that XGBoost predictive performance was good (F1 value of 0.879) and better than SVM, RF and other models. The influence of data accuracy, first semester grades, and educational support on the energy efficiency of digital transformation of academic affairs management is more obvious, and there is a significant difference in the influencing factors for the digital transformation of academic affairs management at different levels of transformation effectiveness.

Keywords: XGBoost algorithm; SHAP; feature analysis; digital transformation

1. Introduction

With the rapid development of information technology and the advancement of global informatization wave, digital transformation has become an inevitable choice for all industries to enhance competitiveness and achieve connotative development. Academic affairs management, as a core link in the operation of colleges and universities, has a direct relationship between the smoothness of the process and response efficiency to the quality of teaching, allocation of educational resources and the quality of talent cultivation [1-2]. Digital transformation, on the other hand, brings great technical support to improve the efficiency of academic management in colleges and universities, and the process of academic affairs can be made more automatic and intelligent by digital means such as upgrading the academic affairs system and improving the information technology facilities [3-4].

In this regard, Khatir and Madani, through structural equation modeling of 206 employees in educational administration, found that digital transformation in infrastructure, human and cultural dimensions can effectively improve educational management performance, but there is an unexpected negative correlation between transformation strategy and performance excellence, which needs to be alerted to the potential contradictions in the implementation of the strategy [5]. Harini et al. found that digital transformation significantly improves the operational efficiency and responsiveness of university academic management through e-learning platforms and administrative management systems, but the digital divide, human readiness and data security remain constraints [6]. Dhameria et al. structural equation modeling analyzed data from 250 academic staff and found that digital transformation significantly improves the efficiency and accessibility of academic management in higher education through the mediation of online learning integration and educational applications, emphasizing that digital platforms and system integration strategies are the key to success [7]. Baijun found through theoretical models and system analysis that digital transformation improves the effectiveness of university academic management by optimizing decision-making efficiency, resource allocation, and teaching quality, but technological adaptability, data security, and privacy are still the core challenges [8]. Through a case study, Yuliantari et al. found that the implementation of a school management information system improved administrative efficiency and data transparency, but was limited by power outages, network quotas, and infrastructure stability, and that the success of the digital transformation required synergistic support from infrastructure, digital literacy, and policy [9]. Belya and Hoi found through a systematic analysis that digital transformation improves the effectiveness of public management in general secondary education institutions by optimizing the management process and decision-making efficiency, and that its innovative direction needs to be centered on the integration of digital tools, personnel capacity building, and stage-by-stage implementation models [10]. Olawumi and Onasanya, through a descriptive survey, found that digital transformation was significantly and positively correlated with both teacher training and student engagement, emphasizing the need to prioritize the strengthening of teacher digital training and student engagement strategies in the enhancement of the efficacy of digital tools in educational management [11]. Nursalim and Darmawan found that the evaluation mechanism as a strategic tool in digital transformation, providing continuous feedback and evidence-based decision-making through evaluation models such as CIPP and goal orientation, is a core driver for improving the adaptability of education management and the quality of services [12]. Sari and Nuryati found that digital leadership, technology integration and institutional readiness are the core success factors for improving quality management in secondary and higher education, emphasizing the need for quality management systems to move from a compliance orientation to a dynamic technology support model [13]. Osorio and Banzato point out that digital technologies have extended from pedagogical tools to the core of educational management, and that in the future, a balance needs to be sought between technological implementation and social goals in order to reconfigure the value of the educational system [14]. Through a qualitative case study, Abibin et al. found that current universities and colleges have implemented digital tools such as academic information systems and learning management systems, but face barriers such as limited infrastructure and low digital literacy among faculty, and need to move forward with transformation through adaptive management, capacity building, and sustainable investment [15].

However, the traditional university academic affairs management is faced with a number of deep-seated dilemmas: the “data silos” between the systems, resulting in difficulties in collaboration and inefficiency [16]; The phenomenon of “focusing on construction but not application” is prominent, and advanced technology has not been effectively transformed into the enhancement of user experience [17]; The problem of “old wine in new bottles” is significant, and the management model and organizational processes have not been innovated in tandem with technological upgrades. Therefore, the digital transformation must adhere to the service orientation centered on teachers and students, through process reengineering and contact point optimization, to promote the academic management from the “controller” to the “service provider” role change, to build a new paradigm of humane, intelligent and efficient service [18-19]. In this context, the systematic reengineering of university teaching management processes with the help of digital technology and the exploration of effective paths to enhance its effectiveness are the iteration of tools at the technical level, as well as the deepening of the reform of university management concepts and service models.

The data for this paper was obtained from the publicly available dataset provided by the Portuguese Polytechnic Institute for the construction of a predictive model for the effectiveness of the digital transformation of Academic Affairs Management (XGBoost). Data preprocessing is performed on the dataset, and a data enhancement strategy based on the sampling method of SMOTE is used for the data imbalance problem that exists in the dataset, applied to the two categories in the dataset. At the same time, XGBoos is compared with random forests, SVM and other models by combining accuracy and F1 score evaluation metrics to prove the excellent effect of XGBoost prediction model in terms of evaluation metrics. The model is then combined with the SHAP method to conduct interpretability analysis, rank the importance of the features of digital transformation of academic affairs management, and analyze the existence of dependencies between the features.

2. Methods for predicting the effectiveness of digital transformation of university academic affairs management

2.1. XGBoost model

1) Decision tree model

There are many types of decision tree models, such as CART, ID3, C4.5, etc., and different decision tree models segment data with different evaluation methods. The common evaluation criteria are Gini coefficient and information gain.

Gini impurity is a measure of the degree of misclassification in a subset, assuming that the dataset D contains k different classes of samples, then the likelihood of randomly selecting i classes of samples is p_i, then the Gini impurity is:

$$Gini(D) = 1 - \sum_{i=1}^{k} p_i^2 \tag{1}$$

The higher the Gini impurity, the closer the number of each class in the dataset is represented. When comparing the segmentation effect of a certain feature on the dataset, if the smaller the Gini gain $GiniGain(D) = Gini(D) - \frac{N_1}{N_1+N_2}Gini(D_1) - \frac{N_2}{N_1+N_2}Gini(D_2)$. indicates the worse the segmentation effect, where D_1, D_2 is the subset after segmentation and N_1, N_2 represents the number of samples in the data set.

Information gain is very similar to Gini impurity, the difference is that its evaluation is information entropy:

$$Entropy(D) = -\sum_{i=1}^{k} p_i \log_2 p_i \tag{2}$$

Similarly, the smaller the information entropy is, the less chaotic the dataset is. In the process of decision tree training, the effect of the node splitting the data set can be judged by calculating the information gain before and after splitting the data set. This information gain is $InfoGain(D, A) = Entropy(D) - \sum_{v \in A} \frac{|D_v|}{|D|} Entropy(D_v)$, where A is the feature of the dataset after segmentation, and D_v is the subset of the data after segmentation by feature A .

Decision tree models are prone to overfitting problems when the tree model is deeper due to the fact that the model is essentially a greedy algorithm, and although various parameters can be adjusted to control the segmentation process, and pruning can also be performed, its effect may be limited. In addition, the decision tree model is not as stable as other models, and small changes in the training data may lead to huge changes in the model structure, which is not suitable for dynamic data that changes over time.

In order to circumvent the shortcomings of the decision tree model, many studies have proposed a variety of integrated learning models based on decision trees, such as GBDT, Random Forest, XGBoost, LightGBM, and so on. The basic idea of integrated learning model is to train multiple models and set the prediction structure of different models into the final prediction structure, so as to obtain a strong generalization ability.

2) XGBoost

XGBoost modifies the objective function of GBDT by adding a regularization boosting module, and its regularization objective function consists of two items:

$$obj(\varnothing) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{3}$$

where $\varnothing$ is the parameter to be optimized by the XGBoost model; its first term, l, is a differentiable convex function, also called a loss function, which measures the error between the predicted value, $\hat{y}_i$, and the true value, y_i; and its second term is a regularity term, $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|\omega\|^2$ is a function that measures the complexity of the tree model f_k, where γ and λ are the weight coefficients, T is the number of leaf nodes in the tree model f_k, and ω is the weight of the leaf nodes in the tree model f_k, the regular term is capable of smoothing out the weights that are eventually learned by the model. thereby avoiding overfitting.

As the above formula can not be directly derived, and XGBoost and GBDT as belong to the Boosting method, the training process can be equivalent to iteratively optimize each tree individually, and its objective function is:

$$obj^{(k)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(k-1)} + f_k(x_i)\right) + \Omega\left(f_k\right) \tag{4}$$

To reduce the computational effort, XGBoost performs a second-order Taylor expansion of the error term $f_k(x_i)$, and the approximate objective function after the expansion is:

$$\begin{aligned} \widetilde{obj}^{(k)} &= \sum_{i}^{n} l\left[(y_i, \hat{y}_i^{(k-1)}) + g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i) \right] + \Omega\left(f_k\right) \\ &= \sum_{i}^{n} l\left[g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i) \right] + \Omega\left(f_k\right) \end{aligned} \tag{5}$$

where g_i, h_i are the first order partial derivatives and second order partial derivatives of the objective function. Defining $I_j = \left\{ i \middle| q(x_i) = j \right\}$ as the samples divided on the leaf node j, where q represents the mapping of the samples to the leaf node with weight ω, the objective function is:

$$\widetilde{obj}^{(k)} = \sum_{j=1}^{T} l\left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i \right) \omega_j^2 \right] + \gamma T + \frac{1}{2} \lambda \omega_j^2 \tag{6}$$

The objective function takes a partial derivative of ω, and the optimal weight of leaf node j can be derived $\omega_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$, which is brought into the loss function to obtain the classification effect under the current tree structure q of the decision tree k:

$$\widetilde{obj}^{(k)}(q) = -\frac{1}{2} \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \tag{7}$$

This function is similar to the Gini impurity and information entropy, and can be used to verify the segmentation effect of the XGBoost function with a segmentation gain h function of:

$$SplitGain = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \tag{8}$$

where I_L and I_R are the samples that are partitioned into left and right subtrees, respectively, and $I = I_L \cup I_R$. During the training process, XGBoost selects the feature with the largest segmentation gain to segment the dataset, thus changing the structure of q.

2.2. TPE-based hyperparameter optimization methods

Bayesian optimization is a machine learning optimization method that uses Bayesian inference techniques to find optimal parameters. It takes the uncertainty of the parameters as a premise and converts the optimization problem into a probabilistic model so as to obtain the optimal solution by Bayesian inference.

SMBO is a specific implementation of Bayesian optimization, which is a process of continuously going through the process of finding the optimal solution by means of iteration. In each iteration, based on the existing tuning history $H = \left(x_1, f(x_1), \cdots, x_n, f(x_n)\right)$, SMBO builds a probabilistic model, and then chooses a new combination of parameters x^* based on some specific criteria. Typically, both GPR strategies are used to model the probability distribution of H as a way to approximate our optimization function f .

Here, the focus will be on the TPE method. If we want to find the best combination of parameters x^* , we need to maximize the acquisition function EI (Expected Improvement), whose expression is shown in Equation (9):

$$x^* = \arg\min_{x\in\chi} EI_{y^*}\left(x\right) \tag{9}$$

where $EI_{y^*}(x)$ is the acquisition function that represents the expected gain after sampling at the current optimal point. The expression for the expected improvement acquisition function is shown in Equation (10):

$$EI_{y^*}(x) = \int_{-\infty}^{+\infty} \max\left(y^* - y, 0\right) p_M\left(y\left|x\right.\right) dy \tag{10}$$

where y^* is a threshold value used to compute the EI acquisition function, which represents the currently optimal metric result. In Bayesian optimization, this Expected Improvement Acquisition Function is solved for and the value of x^* that maximizes it is chosen as the value of the hyperparameters to be taken for the next round of sampling.

This is because TPE models the likelihood function in the form of Bayes' theorem used in the probabilistic modeling, specifically by constructing two kernel density functions $l(x)$ and $g(x)$ to model excellent and poor samples, respectively:

$$p\left(x\left|y\right.\right) = \begin{cases} l(x), & y < y^* \\ g(x), & y > y^* \end{cases} \tag{11}$$

When the threshold y^* is selected, a parameter γ will be set which serves as the discretization point of y , $p(y < y^*) = \gamma$, and thus can be obtained:

$$p(x) = \int_R p\left(x\left|y\right.\right) p(y) dy = \gamma l(x) + (1-\gamma) g(x) \tag{12}$$

Substituting Eq. (12) into Eq. (10) yields:

$$EI_{y^*}(x) = \frac{\int_{-\infty}^{y^*} (y^* - y) p(y) dy}{\gamma + (1-\gamma)\dfrac{g(x)}{l(x)}} \propto \left(\gamma + (1-\gamma)\frac{g(x)}{l(x)}\right)^{-1} \tag{13}$$

As can be seen from equation (13), the EI_{y^*} 's are proportional to the inverse of the denominator. And after γ is determined, the size of the denominator depends only on the ratio $\dfrac{l(x)}{g(x)}$ of the two probabilities of x , where $l(x)$ stands for the maximum probability, and $g(x)$ stands for the minimum probability, and thus, in each iteration, an x^* that makes the value of the EI the largest is returned.

2.3. Methodology for Characterizing the Digital Transformation of Academic Affairs Management

The SHAP algorithm is an interpretable machine learning algorithm for solving the problem of “black-box” modeling of machine learning algorithms, a framework for interpreting the contribution of a sample to the prediction of an outcome in a model, which can be applied to the interpretation of the results of a variety of machine learning algorithms.HAP is a concept used in game theory to measure the contribution of each participant to the victory of a game. contribution value to the victory of the game. In machine learning, the model prediction result can be viewed as the victory of a game, and each feature can be regarded as a participant. Therefore, the SHAP algorithm defines the concept of SHAP value to measure the contribution value of each feature to the model prediction result.The core idea of the SHAP algorithm is to measure the impact of each feature on the prediction result by calculating the SHAP value of each feature among all possible feature combinations, i.e., the SHAP value of each feature for each sample $\varnothing_j$ and the mean value of the target variable for all the samples

$\varnothing_0$ form the corresponding sample model output estimates $g(z')$ as shown in equation (14):

$$g(z') = \varnothing_0 + \sum_{j=1}^{M} \varnothing_j \tag{14}$$

In the specific implementation, SHAP uses an approximation algorithm to compute the Shapley value for each feature based on a subset of the samples. SHAP can be used to explain the impact of each feature on the prediction results, which leads to a better understanding of the model's decision-making process. At the same time, SHAP can also help to discover outliers and feature values in the data, as well as to perform tasks such as feature selection and model optimization.

3. Data acquisition and pre-processing

3.1. Data sources and description

The dataset studied in this paper comes from the Zenodo database of the faculty dataset of the Polytechnic Institute of Portalegre, Portugal. This dataset provides more accurate assistance to teachers by analyzing their data in order to assess the risk of successful faculty management. This dataset records information on faculty members enrolled between the academic years 2012/2014-2022/2024 from 17 undergraduate colleges with different specializations such as Agronomy, Design, Education, Nursing, Journalism, Management, Social Services and Technology. This dataset contains a total of 4424 data, 13 feature attributes and 3 category labels and the dataset details are shown in Table 1.

Table 1. Dataset details

Attribute Category	Feature Attributes
Basic feature	Marital status($X1$)
	Sex($X2$)
	Age($X3$)
	Specialty($X4$)
	Household registration status($X5$)
	Educational Support($X6$)
Academic record	First semester grades($X7$)
	Second semester grades($X8$)
	Study attendance time($X9$)
Effectiveness of Digital Transformation in Academic Management	Transaction processing duration($X10$)
	Data accuracy rate($X11$)
	Abnormal alert capability($X12$)
	Faculty and Student Satisfaction($X13$)

3.2. Data pre-processing

Dealing with features that are not related to teachers' academics: there are feature attributes such as course credit score, number of course units, and total final assessment in teachers' academic performance attributes at the end of the first semester and at the end of the second semester, and some of these feature attributes are not related to teachers' assessment, so this part of the features is directly deleted, and only the final total final grade is kept. At the same time, the features are correlated and analyzed, as shown in Figure 1. Through the feature correlation, it can be found that there is a significant correlation between age, specialty, and domicile status, indicating that there are obvious differences in the age structure of different domicile and specialty groups among the samples, while gender is only correlated with age and specialty, and has almost no correlation with the core business indexes, such as the data accuracy rate and the abnormality warning ability, indicating that gender is not a key factor influencing the digital transformation of faculty affairs management.

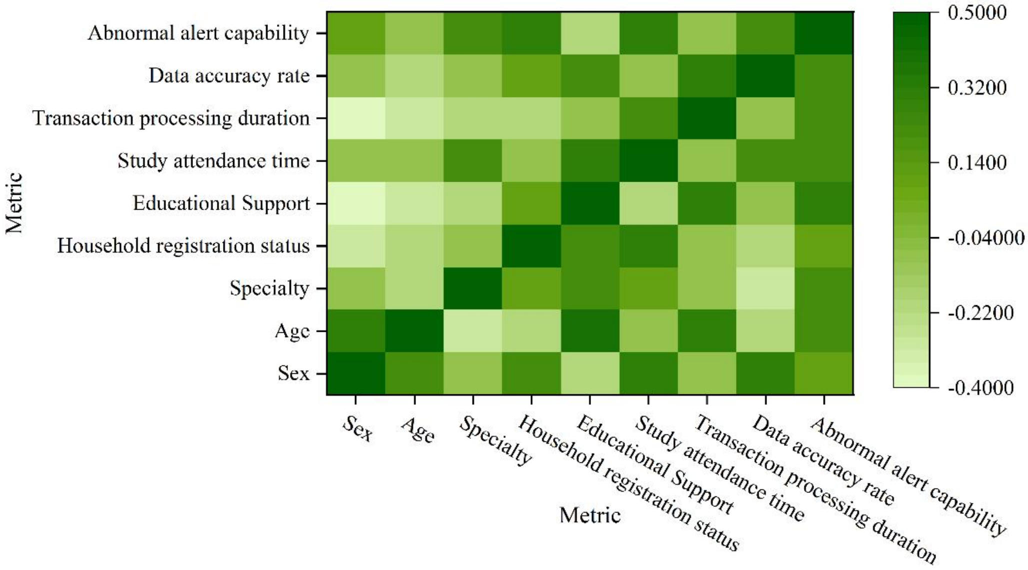


Figure 1. Characteristic correlation coefficient plot

In the energy efficiency of digital transformation of academic affairs management, the distribution of the three categories of the results of the energy efficiency of digital transformation of academic affairs management in this dataset is shown in Figure 2. The proportion of high efficiency of digital transformation is 96.04%, the proportion of medium efficiency is the lowest 34.70%, and the proportion of offsetting energy is 61.34%, which indicates that in the dataset, the digital transformation of academic affairs management shows the trend of “high efficiency”, and the sample transformation effect reaches a high level, which is effective. Data imbalance is a common problem, which may lead to accuracy in the case of most classes at the expense of the performance of a few classes. In order to solve this problem, some measures need to be taken to balance the distribution of data between different classes.

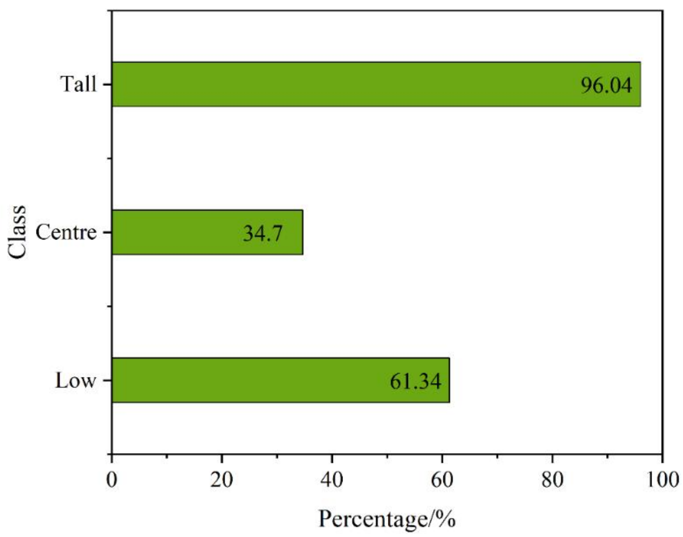


Figure 2. Digital Transformation Categories

SMOTE based oversampling technique is chosen to solve the data imbalance problem.The SMOTE method generates new data samples by synthesizing samples from a small number of class groups and adding these synthesized new samples to the dataset. This method not only increases the amount of data, but also generates new samples that differ from the original data, which helps to improve the generalization ability of the model. To ensure the real validity of the data, the imbalance of the data is eliminated by using SMOTE oversampling only on the training set, and the dataset is randomly divided into the training set and the test set according to 4:1, and there are 3156 data in the training set and 833 data in the test set, and the results of the processing are shown in Table 2.

Table 2. SMOTE oversampling results

Class	Low transformation	Transitional transformation	High transformation	Sum
Test set	214	149	470	833
Training set	982	575	1599	3156
SMOTE	1624	1633	1605	4862

4. Model evaluation and interpretation

4.1. Feature Selection and Hyperparameter Optimization

In this paper, recursive feature elimination (RFE) is firstly used for feature selection before formally starting the model training.RFE is a simple inverse selection method that uses repeated multiple validation method to fit the model. The results of recursive feature elimination in this paper are shown in Fig. 3. When the number of input features is less than 7, the prediction score of the model improves with the increase of input features, and the X7 prediction score is 0.66. When the number of input features is greater than 7, the prediction score of the model does not improve with the increase of the number of input features, but on the contrary, it decreases, which indicates that at this time, the model has been noisy variables. From this, the optimal number of input variables for the model was determined to be 7.

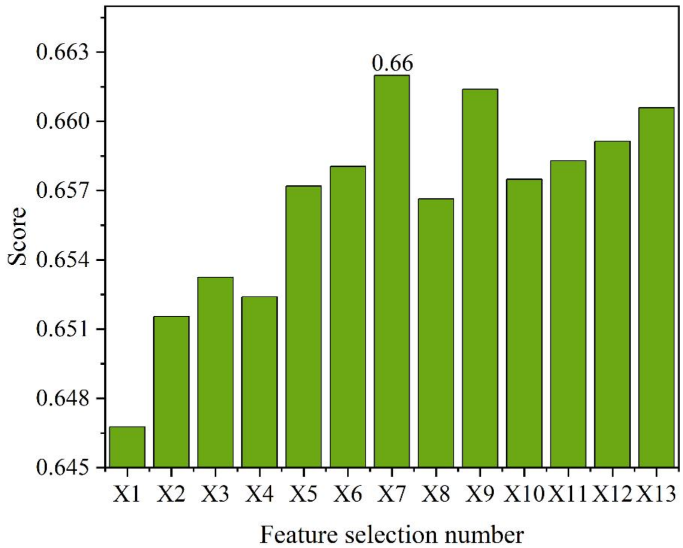


Figure 3. RFE feature selection results of XGBoost

The model performance was evaluated using the F1-Score method as shown in equation (15), where the final model score is the average of the F1 values for all categories:

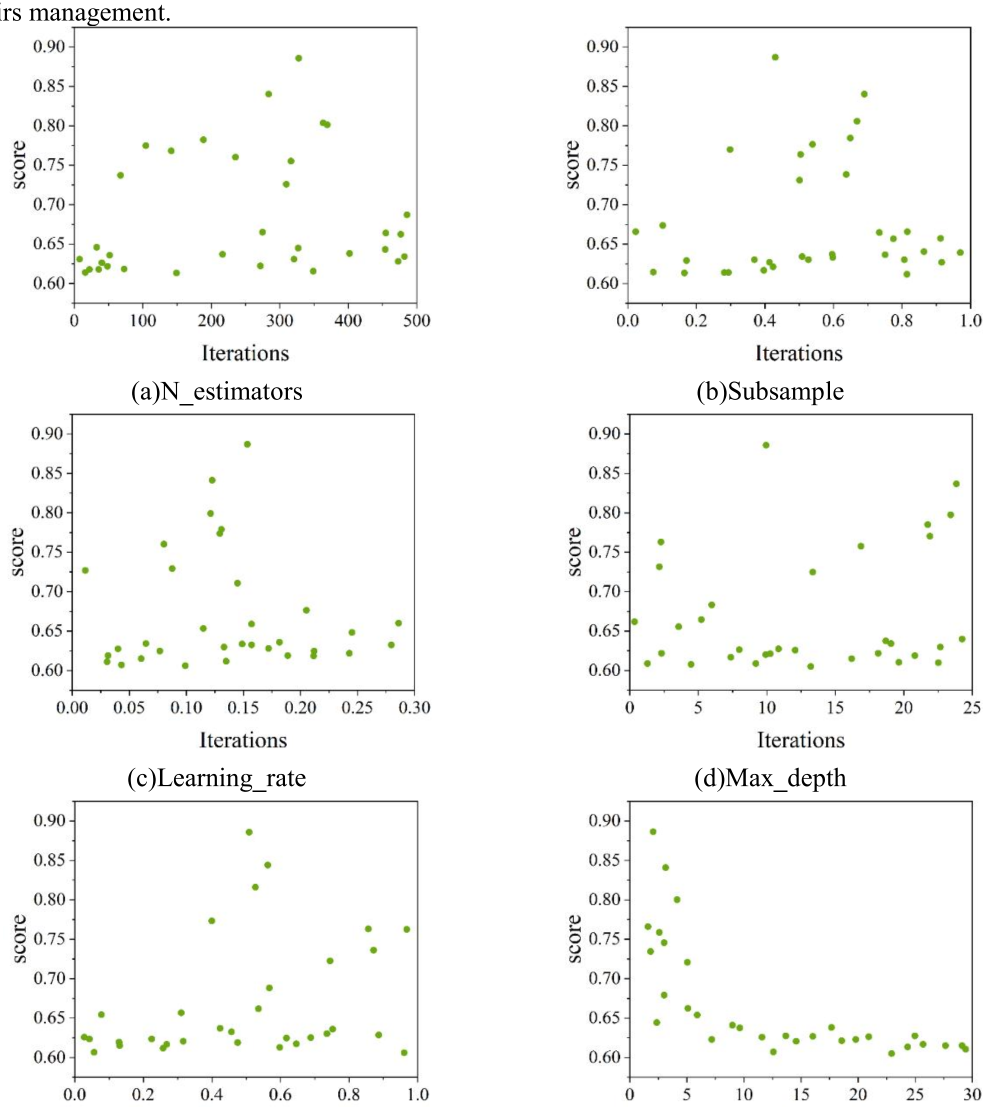
$$F1 = \frac{2 \times P \times R}{P + R} \tag{15}$$

where P denotes the precision rate and R denotes the recall rate, the formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{16}$$

$$R = \frac{TP}{TP + FN} \tag{17}$$

Generally speaking, the larger the F1 value of the model, the better the predictive performance of the model; on the contrary, the smaller the F1 value, indicating that the constructed model can not be well adapted to the research problem of this paper, and it is necessary to consider re-constructing the input features or changing to a more suitable model.The optimization process of the TPE hyper-parameters is shown in Fig. 4, and the optimal F1 value reached during the parameter optimization process is 0.887, which has a better prediction effect, indicating that the The model constructed in this paper can better predict the digital transformation energy efficiency of academic affairs management.



(a)N_estimators (b)Subsample

(c)Learning_rate (d)Max_depth

(e)Clsaample_bytree (f)Grama

Figure 4. TPE optimization process

4.2. Model Comparison

In order to validate the experimental results and provide more numerical insights, the XGBoost method was modeled against four prediction methods, namely Decision Trees, Support Vector Machines, Random Forests, and Light GBMs, which are currently being applied to the state-of-the-art in this field, and the results of the experiments are shown in Table 3, in which XGBoost is the baseline for the paired t-tests.The negative performance of F1 indicates that the method performs poorly compared to the XGBoost . The experiment proves that the XGBoost model has the best performance among all models with a precision of 0.898 and a recall of 0.861, which provides better prediction results.

Table 3. Comparisons of proposed method with other mainstream methods

Model	**P**	**R**	**F1**	**Performance comparison (%)**
Decision Tree	0.81	0.819	0.814	-7.565%**(0.037)
SVM (Support Vector Machine)	0.798	0.795	0.796	-9.629%*(0.075)
Random Forest	0.854	0.831	0.842	-4.354%***(0.001)
Light GBM	0.896	0.853	0.873	-0.791%(0.307)
XGBoost	0.898	0.861	0.879	/

Note: * at the 10% level, ** at the 5% level, *** at the 1% level. P-values are in parentheses.

4.3. Model interpretation

The XGBoost model was used to implement TreeExplainer to calculate the SHAP value of each feature, and the absolute value ranking of features is shown in Figure 5. As can be seen from the figure, the degree of influence of the features shows a clear echelon structure, data accuracy, first semester grades, and educational support have a more obvious influence on the energy efficiency of digital transformation of academic affairs management, the study attendance time and the ability to warn of anomalies as well as second semester grades also have an impact on the energy efficiency of digital transformation of academic affairs management, and gender, age, marital status, and profession have a more obvious influence on the energy efficiency of digital transformation of academic affairs management .

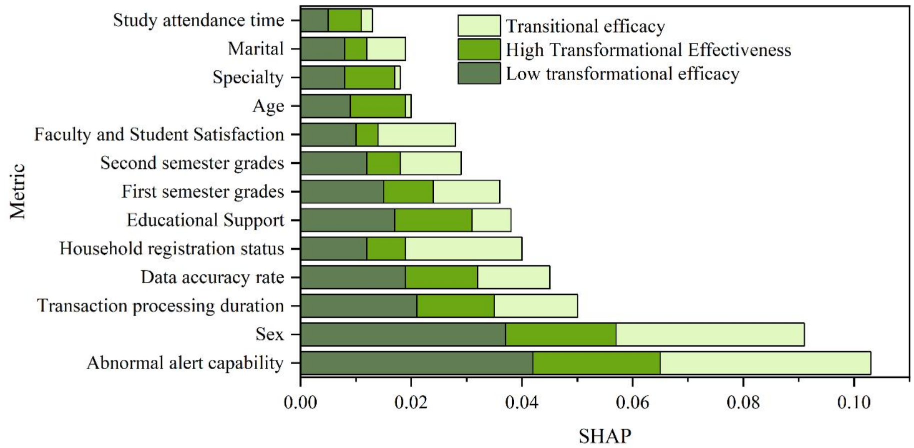
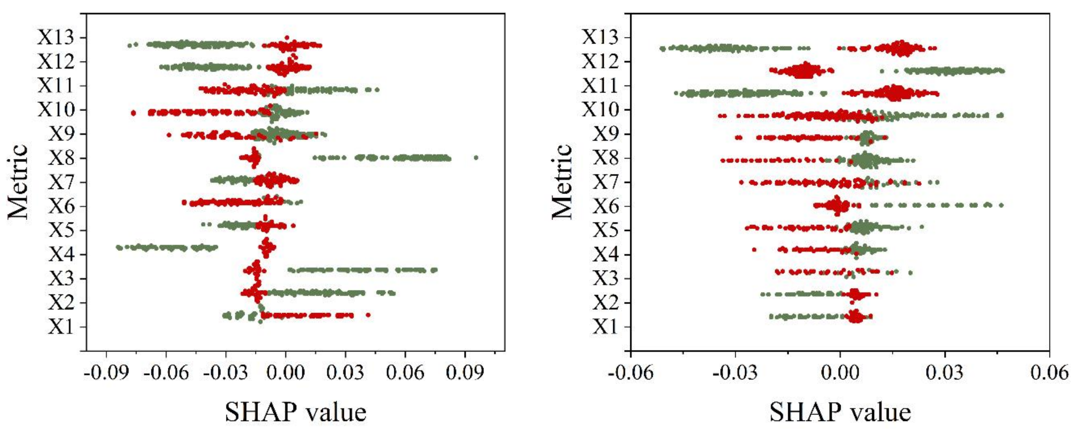


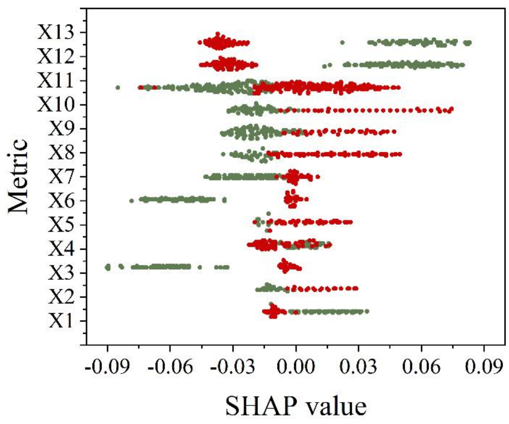
Figure 5. Absolute Value Sorting Chart of Features

The impact of each sample in the test set on the output results was calculated using the SHAP algorithm, which was used to assess the contribution of each feature to the model. The distribution of SHAP values for high, medium, and low transformational efficacy is shown in Figure 6. A point in the SHAP summary plot represents a feature, and a positive SHAP value means that the feature's impact on the model is positive, while a negative SHAP value means that the feature's impact on the model is negative, where green indicates that the feature's contribution is negative and red indicates that the feature's contribution is positive. The influencing characteristics of medium and high transformational efficacy are similar, the length of transaction processing and faculty and student satisfaction have a greater influence on the model of the medium transformational efficacy group, whereas the study attendance time positively influences the model to a lesser extent than than the length of than the length of transaction processing, but the age does not have a significant influence on the digital transformation of academic management in the medium transformational efficacy group.



(a)High Transformational Effectiveness

(b)Transitional efficacy



(c)Transitional efficacy

Figure 6. Distribution of SHAP values for digital transformation

4.4. Path to Improve the Effectiveness of Digital Transformation of Academic Affairs Management in Higher Education Institutions

1) Policy leadership to clarify the direction of digital transformation

The government and education authorities should formulate forward-looking and operable policy documents to provide a clear direction and target guidance for the digital transformation of university teaching services. Clearly stipulate that colleges and universities need to achieve the digital academic affairs management system coverage, data security protection level and other quantitative indicators within a specific period of time, so that colleges and universities in the process of promoting the digital transformation of the rules to follow.

2) Diversified funding to ensure construction investment

Establish a diversified funding mechanism to provide adequate financial support for the digital transformation of university teaching affairs. Government finance should continue to increase direct investment in the digitization of university teaching, set up special allocations, focusing on supporting the construction of basic projects. At the same time, social capital should be attracted to participate in university teaching digitalization projects through tax incentives and policy support. Colleges and universities themselves should also rationally plan their budgets, and set aside a certain percentage of tuition income and research fund balance for the construction of academic affairs digitization.

3) Deepen university-enterprise cooperation to promote the integration of production, learning, research and utilization

Strengthen the in-depth cooperation between universities and enterprises, and explore the development mode of integration of production, learning, research and utilization. Relying on the advantages of enterprises in technological innovation and market experience, we will jointly develop teaching management systems and teaching resource platforms that meet the actual needs of universities, and realize the digitalization and intelligent upgrading of teaching resources. Through the joint construction of training bases and scientific research platforms, we can promote the transfer of technology and the transformation of achievements, and enhance the ability of universities to serve the society. For example, a university has cooperated with a technology enterprise to jointly develop a teaching quality assessment system based on artificial intelligence, which analyzes multi-dimensional data in the teaching process to provide teachers with accurate teaching improvement suggestions, and has achieved good application results.

5. Conclusion

This paper explores the influencing factors of digital transformation of university academic affairs management using XGBoost model, and explains the degree of contribution of model features based on SHAP method. Through the above analysis, it is found that the characteristics of the tendency indicators affecting the digital transformation of university teaching management are complex and summarized as follows:

1) Constructing a prediction model for digital transformation of faculty management, the preprocessing of faculty data, classifying the category of faculty management prediction model into three categories (high transformation effectiveness, medium transformation effectiveness, and low transformation effectiveness), and combining the accuracy and F1 score evaluation indexes to compare the XGBoost model with other models, such as decision tree, SVM, etc., the prediction performance of

the XGBoost model is good, and the F1 value is 0.879, which is better than other models.

2) Interpretability analysis of the model using SHAP. The effects of data accuracy, first semester grades, and educational support on the energy efficiency of digital transformation of academic affairs management are more obvious. Compared to the high, medium, and low transformation effectiveness groups, there is a greater risk of difficulty in academic management, and its influence characteristics are recognizable.

References

1. Gong, X. (2024). A Study on the Balance between Efficiency and Effectiveness in Tertiary Education Administration Management. information technology, 7(9), 53-58.
2. Polishchuk, S., & Horbatiuk, O. (2023). The problem of quality and efficiency of educational institution management. Journal of Vasyl Stefanyk Precarpathian National University, 10(1), 197-204.
3. Fernández, A., Gómez, B., Binjaku, K., & Meçe, E. K. (2023). Digital transformation initiatives in higher education institutions: A multivocal literature review. Education and information technologies, 28(10), 12351-12382.
4. Abad-Segura, E., González-Zamar, M. D., Infante-Moro, J. C., & Ruipérez García, G. (2020). Sustainable management of digital transformation in higher education: Global research trends. Sustainability, 12(5), 2107.
5. Khatir, N., & Madani, H. (2024). The impact of digital transformation on achieving outstanding performance in educational administration. International Journal of Educational Management, 38(7), 1821-1838.
6. Harini, H., Ripki, A. J. H., Sulistianingsih, S., Herlina, H., & Putri, A. (2024). Digital transformation: the utilization of information and communication technology to enhance educational management efficiency in the modern era. Jurnal Minfo Polgan, 13(2), 1668-1674.
7. Dhameria, V., Muazeib, A. I. M., Blhaj, K. M. S., Sugiyarsih, S., & Rosadah, R. A. (2025). The impact of digital transformation in higher education management: Integrating online learning and educational applications for efficiency and accessibility. International Journal of Educational Qualitative Quantitative Research, 4(1), 15-24.
8. Baijun, H. (2024). Impact of Digital Transformation on Higher Education Management: A Theoretical Analysis. Journal of Industry and Engineering Management, 2(3), 35-39.
9. Yuliantari, B. T., Mustari, M., & Syarifuddin, S. (2025). Digital Transformation in School Administration and the Implementation of School Management Information Systems at Madrasah Aliyah Negeri 1 Mataram. Indonesian Journal of Educational Innovation, 1(3), 47-54.
10. Belya, H., & Hoi, N. (2025). INNOVATIVE DIRECTIONS FOR IMPROVING MECHANISMS FOR PUBLIC MANAGEMENT OF EDUCATIONAL INSTITUTIONS IN THE CONTEXT OF DIGITAL TRANSFORMATION. Philosophy and Governance, (11 (15)).
11. Olawumi, B. M., & Onasanya, H. I. (2025). Impact of digital transformation in educational management: A study on teacher training and student engagement. Interdisciplinary Journal Of Lifelong Learning, 1(1), 141-151.
12. Nursalim, A. M., & Darmawan, R. (2025). Evaluation as a Catalyst for Change: Building Superior Education Administration in the Era of Digital Transformation. Jurnal Office: Jurnal Pemikiran Ilmiah dan Pendidikan Administrasi Perkantoran, 11(1), 113-119.
13. Sari, N. T. W., & Nuryati, N. (2025). Educational Quality Management in the Era of Digital Transformation: A Literature Review of Secondary and Higher Education Institutions. Asian Journal of Multidisciplinary Studies, 8(1), 157-168.

14. Osorio, J., & Banzato, M. (2021, August). Digital transformation of education and learning through information technology in educational management. In Open Conference on Computers in Education (pp. 286-295). Cham: Springer International Publishing.
15. Abibin, Z., AL-Momani, M. O., & Ali, A. O. (2025). Digital transformation in the management of private Islamic higher education institutions: Opportunities and challenges. Zabags International Journal of Education, 3(2), 71-83.
16. Parry, G. (2013). Colleges and the governance of higher education. Higher Education Quarterly, 67(4), 315-339.
17. Guan, W. (2023). Analysis of music education management mode in colleges and universities in China. Frontiers in Educational Research, 6(5), 127-148.
18. Chen, X. Q., & Wang, Z. A. (2019). Research on educational administration management of colleges and universities: Based on big data perspective. DEStech Transactions on Computer Science and Engineering, 1-5.
19. Chen, N. (2021). The status quo and prospects of the information construction of teaching management in colleges and universities. Journal of Higher Education Research, 2(4), 194-196.