International Journal of Religion
2024
Volume: 5| Number 8 | pp. 463 – 474
ISSN: 2633-352X (Print) | ISSN: 2633-3538 (Online)
ijor.co.uk
DOI: https://doi.org/10.61707/mcbxxk41

Construction and Application of a Prediction Model for Students' Satisfaction in Chinese-Foreign Cooperatively-Run Educational Programmes Based on Machine Learning

Na Zhao[1], Norazah Abdul Aziz[2]

Abstract

The increasing popularity of educational programs by Chinese and foreign partners in China's universities presents challenges in administration due to differences in organization, social expectations, and academic requirements. This study presents a prediction model using Machine Learning techniques to assess student satisfaction levels. The Gradient Boosting Machine (GBM) framework, enabled by Principal Component Analysis (PCA), was used to collect data from 1,237 students. The approach was compared to other ML models like Random Forest, Linear Regression, Naïve Bayes, and Support Vector Machine and trained and validated using 5-fold cross-validation. The PCA+GBM model scored most effectively of significant metrics, with 93% accuracy, 92% precision, 91% recall, and 91.5% F1-score. Policymakers and educational administrators can use the outcomes as an incentive to enhance the strategy and standard of Higher Education (HE) that are run in collaboration between China and foreign countries.

Keywords: *Chinese-Foreign Cooperatively-Run Educational Programs, Machine Learning, Principal Component Analysis, Accuracy, Gradient Boosting Machine.*

INTRODUCTION

In the past few years, Chinese higher education has experienced an increasing number of effective Higher Education (HE) between Chinese and foreign institutions. Students may engage in a global HE while still based in China through these programs and partnerships between Chinese and foreign universities. Through these collaborations, both learning systems hope to provide students with a complete HE, more significant opportunities to learn about other cultures, and access to instructors worldwide. Students' academic and employment prospects are frequently enhanced when they earn degrees or certificates from Chinese and global partner institutions.

Global relationships in HE greatly enhance partnerships across international boundaries and allow the exchange of cultures and data. To help students succeed in today's internationally connected world, they expose them to distinct cultures and educate them on essential life skills. By bringing together various academic backgrounds and pedagogical opinions, such partnerships improve the HE process for all students. They also help universities' globalization attempts, improving scientific collaboration and general international competitiveness. These types of partnerships are also compatible with the more significant objectives of the Chinese government, which are to enhance HE and turn out graduates who are ready to compete internationally.

Contradictions in administrative processes, social conventions, and HE expectations among collaborators create problems for program directors supervising Chinese foreign partnerships.

Key challenges include:

Curriculum Alignment: Harmonizing their HE can achieve each institution's educational needs.

[1] Faculty of Education, Universiti Teknologi MARA (UiTM), Kampus Puncak Alam, 42300 Bandar Puncak Alam, Selangor Darul Ehsan, Malaysia, E-mail: hauezhao@outlook.com

[2] Faculty of Education, Universiti Teknologi MARA (UiTM), Kampus Puncak Alam, 42300 Bandar Puncak Alam, Selangor Darul Ehsan, Malaysia

Quality Assurance: Given differing accreditation standards, ensuring consistent Quality of Education (QoE) across the program.

Language Barriers: Managing language differences in instructional delivery and student support.

Cultural Integration: Developing an accepting environment that recognizes the distinctive perspectives of both domestic and international students.

Considering these challenges, there are significant possibilities presented by joint programs:

Curriculum Innovation: Developing programs in multiple fields integrating national and regional learning.

Research Collaboration: Increase the reputation of interdisciplinary research initiatives that address local and global issues.

Global Networks: Developing sustainable connections among alumni, which allow global professional collaborations.

The long-term sustainability of global collaborations is contingent upon how happy the students are with them. When students are satisfied with their studies, they are more inclined to do correctly, continue in school, and even support their programs. However, numerous chances for development remain undetected because standard assessment techniques fail to provide comprehensive and practical insight into students' experiences.

Predictive models that are motivated by data can provide helpful data about how students feel in this context by:

Assessing the key elements that impact levels of student satisfaction.

Providing predictions about student satisfaction results employing questionnaires, socioeconomic, and educational information.

The rule of helpful recommendations to administrators for enhancing the standard of QoE and the student experience.

In order to investigate the proportion of student satisfaction in programs run collaboratively by Chinese and foreign universities, this study proposes a predictive model that depends on ML techniques. The present research aims to shed light on the factors that improve student satisfaction and provide feasible options to improve international HE universities through the Gradient Boosting Machine (GBM) model, which is made possible by Principal Component Analysis (PCA). In order to determine the level of satisfaction students have with their HE at domestic and foreign universities, the study proposal provides a predictive approach. This study uses the ML approach. Relations between Chinese universities and foreign institutions are the primary topic of the study, which employed structured surveys and educational information to collect data from 1,237 students. A PCA-enabled GBM model has been built to predict happiness levels based on academic achievement, management help, and integration of cultural facilities and overall happiness. The approach was assessed with other ML methods, such as Random Forest (RF), Linear Regression (LR), Naïve Bayes (NB), and Support Vector Machine (SVM), and it was trained and verified utilizing 5-fold cross-validation. The proposed PCA+GBM approach had the highest levels of accuracy (93%), precision (92%), recall (91%), and F1-Score (91.5%) for predicting students' happiness. Academic policymakers and leaders may apply the findings to improve the framework and QoE of foreign HE in China.

LITERATURE REVIEW

A Graph Convolutional Network (GCN) framework was built by [1] in order to enhance the accuracy of their predictions about the academic achievement of students in CFCRS. In order to assess the degree of distinct resemblance, an undirected graph was developed, relating students who reached similar levels of academic achievement via the use of the Pearson correlation coefficient. Compared to the accuracy of SVM and RF models, the model's success rate of 81.5% was significantly greater.

[2] provided an approach for assessing the QoE using 18 factors distributed over four levels. These variables comprised material data, instructor nature, learning technique, and learning result types. Academic assistance, instructional materials, and management have been identified as significant factors using the DANP and TOPSIS methods. It is recommended that learning experiences, teacher training, and curriculum administration be improved to improve the QoE.

Through classroom interviews and observations, [3] studied the lessons learned by mentors and students participating in Sino-Foreign Cooperative Education (SFCE) programs. Factors like cultural disparities and challenging program timetables have been emphasized in their research. Some of the recommendations consisted of setting up a Program Management Committee, improving interpersonal interaction, and establishing training spaces that are easier to find.

A qualitative study on CCA among Chinese-Foreign Higher Education Cooperation Program (CFHECP) participants was conducted by [4]—the interviews named academic, psychological, and cultural adaptation as impediments. Society gains when CCA techniques improve, and students can adapt to different HE and cultural environments.

A group of 3,783 students from a Chinese global university was examined by [5] to find the association between virtual learning participation, academic achievement, and face-to-face class attendance. There was an improvement between virtual learning engagement and learning outcomes, as well as between physical participation and academic performance, based on the study's use of Spearman and Pearson correlation values. The findings demonstrate that students do superior academically when regular classroom instruction is paired with virtual learning experiences.

The causes driving graduate school in the United Kingdom were investigated by [6-10] among Chinese graduates of a Sino-British partnership university. HE, professional growth, development of oneself, and social expectations were the broad trends that emerged from the study's informal interviews with ten students obtaining a global master's degree. In addition, students' decision to study in the UK was impacted by five main themes: HE factors, social context, understanding and awareness, economic expenditure, and opportunities for professional growth. Those factors have been significantly affected by the time spent learning at the Sino-British university.

After the pandemic finished, [11-13] looked at how learning internationally improved. The studies demonstrated an increasing interest in learning internationally among Chinese citizens, which presents significant possibilities for developing China's international HE system. The study recommended that system improvement, HE system enhancement, HE framework optimization, and the fusion of international and humanities education are the four strategic approaches to improving international HE. For China's HE system to maintain its status as an internationally recognized institution, it must conform to the goals and objectives at the 5th Plenary Session of the 19th CPC Central Committee.

STUDY AREA AND PARTICIPANT DEMOGRAPHICS

China-Foreign cooperatively-run education initiatives are the focus of the present study. These programs combined Chinese universities with comparable institutions in different countries' HE systems. These programs are designed to provide students with international exposure and educational experiences that combine the strengths of Chinese and foreign HE systems [14-15]. The programs selected for this study are located in major Chinese cities known for their HE infrastructure, including Beijing, Shanghai, and Guangzhou.

Participant Demographics

The participants in this study were students enrolled in these programs during the academic years 2022-2023. A total of 1,237 students participated in the survey, comprising domestic Chinese students and international students from various countries. The breakdown of the participant demographics is as follows:

Table 1. Participant demographics.

Category	Total	Percentage
Total Participants	1,237	-
Gender		
Male	618	50%
Female	619	50%
Age Range		
18-20 years	448	36.2%
21-23 years	492	39.8%
24 years and above	297	24%
Nationality		
Chinese	775	62.6%
European	198	16%
North American	149	12%
Others (Asia-Pacific, Africa, South America)	115	9.3%
Field of Study		
Business and Management	347	28.1%
Engineering and Technology	305	24.7%
Humanities and Social Sciences	198	16%
Science and Mathematics	146	11.8%
Arts and Design	151	12.2%
Other (Health Sciences, HE)	90	7.3%

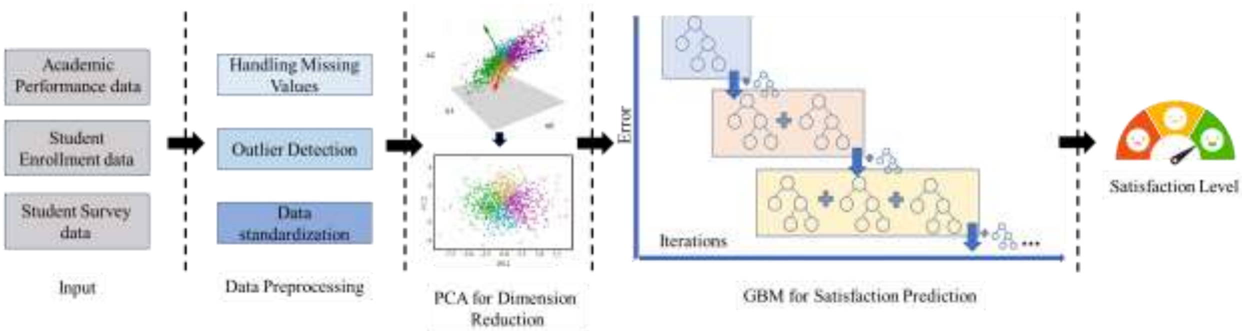

Figure 1. Proposed model.

PROPOSED MODEL

The proposed model has multiple layers with inputs collected from different sources in the data collection layer, then they are preprocessed at the next layer using multiple models; the preprocessed data is then handled through the PCA layer for dimension reduction, and the reduced feature set is processed using the GBM to predict the satisfaction level, the proposed architecture is presented in Fig. 1, and the following section details it:

Data Collection and Preprocessing

i) Data Collection

The data for this study were collected through a structured online survey distributed to students enrolled in Chinese-Foreign cooperatively-run educational programs during the academic year 2022-2023. The survey targeted 1,500 students at universities in Beijing, Shanghai, and Guangzhou.

Of the 1,500 students contacted, 1,237 completed the survey, resulting in an 82.5% response rate. This rate indicates the relevance of the survey to the participants' experiences. During September and November of 2022, undergraduates were requested to assess their level of

happiness with multiple facets of their HE, help with administration, integration into society, resources, and services, and general happiness using Likert-scale questions. Factors impacting student satisfaction in these educational contexts can be investigated by analyzing the collected responses, which create a data set. The research collected enrollment records and data on learning outcomes from the partner universities' administrative departments. Possessing the permissions we needed to meet data privacy standards allowed us to access student grades, program participation information, and graduation levels.

The survey included questions grouped into five main categories:

Academic Experience

1. Are students happy with the standard of teaching that you get as a result?
2. What degree of connection does the HE have to the job objectives of the students?
3. How acceptable are the learning resources that have been provided for the course?
4. How correctly do the tutorials measure the students' grasp of knowledge?
5. In what ways do you feel impressed by the adaptable nature of the courses given by the HE?

Administrative Support

1. What level of happiness do you think you have with joining?
2. What type of comments do you experience from the management team when you ask it questions?
3. How effective is the communication that comes from the leadership of the program?
4. To what degree do you feel fulfilled by the HE assistance services that are readily accessible?
5. Describe the level of dedication that the university's administration conducts with your educational records.

Cultural Integration

1. How happy are you regarding the social events that can be enjoyed through the program?
2. What is the degree of success rate the course has in resolving language challenges?
3. How easy is it to befriend other students from different cultural backgrounds?
4. How well are international students integrated into campus life?
5. How satisfied are you with the opportunities for social interaction?

Facilities and Resources

1. How satisfied are you with the accessibility of learning resources?
2. How adequate are the library services?
3. How satisfied are you with the QoE of the IT support?
4. How appropriate are the study areas and facilities?
5. How satisfied are you with the availability of research equipment and materials?

Overall Satisfaction

1. Overall, how satisfied are you with your experience in the program?
2. How likely are you to recommend this program to others?
3. How well has the program met your expectations?
4. How valued do you feel as a student in this program?
5. How likely are you to continue your studies in this program?

ii) Data Preprocessing

The dataset obtained from the online survey and academic performance records underwent systematic preprocessing to prepare it for analysis. The process began with the identification and handling of missing values. Mode imputation was used for missing Likert-scale responses, reflecting the most common response for those items. Outliers in academic performance data were identified and adjusted using administrative records, correcting grades outside the typical 0-100% range. All Likert-scale responses were verified to ensure consistency, adhering to a uniform scale of 1 to 5. Academic grades from various grading systems were converted to a standardized GPA format using a conversion table on a 0-4 scale.

Categorical variables such as nationality and field of study underwent one-hot encoding, transforming these categories into a binary format suitable for machine learning algorithms. Similarly, Likert-scale responses were converted from qualitative descriptions to quantitative values ranging from 1 to 5. The dataset integration involved merging survey data with academic records using student IDs as a unique identifier. This ensured the accurate matching of survey responses with academic and enrollment information. The data were then segmented into training and testing sets, allocating 80% for training the model and 20% for testing.

Table 2. Dataset Description

Attribute	Description	Type	Example Values
Student ID	Unique identifier for each student	Integer	101, 102, 103, ...
Gender	Student's gender	Categorical	Male, Female
Age	Age of the student	Integer	20, 22, 24, ...
Nationality	National background of the student	Categorical	Chinese, European, North American, Other
Field of Study	The academic discipline of enrollment	Categorical	Business, Engineering, Humanities, Science
Academic Experience Score	Average satisfaction with academics	Integer (1-5)	3, 4, 5
Administrative Support Score	Satisfaction with administrative support	Integer (1-5)	2, 3, 4
Cultural Integration Score	Satisfaction with cultural integration	Integer (1-5)	1, 2, 3
Facilities Score	Satisfaction with facilities	Integer (1-5)	3, 4, 5
Overall Satisfaction	Overall program satisfaction	Integer (1-5)	3, 4, 5
GPA	Grade point average on a 0-4 scale	Float	2.5, 3.2, 3.8
Enrollment Status	Current enrollment status	Categorical	Enrolled, Discontinued
Retention Rate	Percentage indicating retention	Percentage	85%, 90%, 95%

Supervised Machine Learning (SML)

SML is typically approached as either a classification or a regression problem. In regression, the goal is to predict a continuous outcome variable $y \in \mathbb{R}$, whereas classification focuses on predicting a categorical outcome variable $y \in \mathbb{N}$. Assumed the definite nature of predicting student satisfaction levels, this study focuses on classification. Over the years, numerous algorithms have been developed to address the challenges inherent in SML, which can be summarized as follows:

Given a dataset $\mathcal{D} = \{(X_1, y_1), \dots, (X_n, y_n)\}$, where each $X_i = (x_{i1}, \dots, x_{im})$ represents a vector of features, SML seeks to derive a function $h(x) = y$ that maps a feature vector $x \in \mathcal{X} \subseteq \mathbb{R}^d$ to a label $y \in \mathcal{Y} \subseteq \mathbb{R}$. The goal is to accurately predict the level of satisfaction (label) based on a set of input features (vector x) from the student data.

Dimensionality Reduction Using PCA

PCA is a statistical technique used for dimensionality reduction while preserving as much variability in the data as possible. PCA reduces the dimensionality of the survey data collected. This includes responses from multiple Likert-scale questions across categories such as academic experience, administrative support, and cultural integration. PCA identifies directions, known as principal components, where the variance of the data is maximized. The simple fact that these parts are perpendicular implies that they collect on various types of unpredictability.

First, we standardize the data 'X'. Each feature is ensured to contribute equally to the analysis by subtracting its mean and dividing by its standard deviation. The next step is to calculate the standardised data's covariance matrix. The EQU (1) defines the covariance matrix '£'.

$$\Sigma = \frac{1}{n-1} X^T X \tag{1}$$

where X^T is the transpose of X, and n is the number of data points. Variables' interrelationships with one another in terms of their standard deviations can be better explained using this matrix. Covariance matrix eigenvalues and eigenvectors are subsequently computed. In terms of directions of maximum variance, the eigenvectors are the PCA. The eigenvalues demonstrate how much variance each PCA. A model of the decomposition is provided by EQU (2).

$$\Sigma v = \lambda v \tag{2}$$

'λ' represents the eigenvalues and 'v' represents the eigenvectors. If we require information on how much deviation the first *'k'* components recorded, we can use the total variation explained ratio to select how many principal components to maintain. The total number of dimensions required to define the data correctly is selected in this phase. The selected principal components define a new subspace, and the original data has been transformed into that space. The process is performed by superimposing the initial data onto the region enclosed by the eigenvectors corresponding to the highest eigenvalues. EQU (3) describes the transformed data *'Y'*.

$$Y = XV_k \tag{3}$$

where V_k is the matrix containing the first $'k'$ eigenvectors.

GBM for Satisfaction Prediction

GBM is an ensemble learning method that progressively increases prediction accuracy by building models and addressing errors from earlier iterations. Predicting satisfaction with learning, requiring multiple variables, is a complex task that requires this method. In GBM, several steps are performed to build an effective model from several weak ones, frequently using Decision Trees (DT). A loss function measures the variance between the actual and predicted outcomes, and the objective is to limit that function.

The standard model that GBM uses to start predicts the target variable's mean. This initial model serves as the baseline for subsequent improvements. In each iteration, GBM fits a new decision tree to the residual errors identified by the aggregate of all previous models. This step involves calculating the gradient of the loss function. Residuals r_i for each data point i are calculated as EQU (4).

$$r_i = y_i - \hat{y}_i \tag{4}$$

where y_i is the actual outcome and $\hat{y}_i$ is the predicted outcome from the existing ensemble of trees.

A new DT is trained to predict the residuals rather than the actual outcomes. Each new tree corrects the errors from its predecessors. The influence of each new tree on the final model is regulated by a parameter known as the learning rate. The model update equation, incorporating the learning rate η, is EQU (5).

$$\hat{y}_i^{(\text{new})} = \hat{y}_i + \eta \cdot h(x_i) \tag{5}$$

where $h(x_i)$ represents the output of the new tree for observation x_i.The process of adding trees stops when the improvement in the loss function is below a set threshold or when a predetermined number of trees have been added.

Algorithm: PCA-Enabled GBM for Satisfaction Prediction

Input:

D: Dataset with features X and target y (student satisfaction levels)

Output.

$\hat{y}$: Predicted satisfaction levels

Procedure:

Preprocess Data:

Normalize each feature x_i in X to zero mean and unit variance.

Impute missing values in X using median or mode imputation.

Encode categorical features in X using one-hot encoding.

Apply PCA:

Compute the covariance matrix Σ of the normalized dataset X.

Perform eigenvalue decomposition on Σ to obtain eigenvectors V and eigenvalues λ.

Select the top k eigenvectors V_k that capture the desired cumulative variance (95%).

Project X onto the selected eigenvectors to transform into the new subspace: $X_k = X \times V_k$

Train GBM:

Initialize GBM parameters: number of stages N, learning rate η, and other parameters.

For each stage from 1 to N :

Fit a decision tree to the residuals of the model.

Update the model using the tree output adjusted by the learning rate η.

Optimize and Validate Model:

Perform k-fold cross-validation to determine optimal k and GBM parameters.

Use a grid search over parameter ranges to refine selections.

Evaluate the model using metrics such as RMSE, R^2, or accuracy.

Make Predictions:

Train the GBM model on all available data transformed by PCA using optimal parameters.

Predict satisfaction levels $\hat{y}$For new data, they first apply PCA transformation.

Assess model performance using a separate test set or out-of-sample validation.

End of Procedure

Return $\hat{y}$

5. Experimental Setup

The experiment was carried out in a hardware setup including an Intel Xeon CPU with 16 cores, 64 GB of RAM, a 2 TB SSD, and an NVIDIA Tesla V100 GPU. The operating system was Linux Ubuntu 20.04, utilizing Python 3.8. Key libraries included Scikit-learn for machine learning. The dataset was split into 80% for training and 20% for testing. Model parameters were optimized using 5-fold cross-validation and the Tab. 3 displays the parameters used for training. The GBM model was trained sequentially, adjusting the learning rate to balance learning speed and overfitting risk.

Table 3. Hyperparameters for the GBM Model.

Hyperparameter	Value
Number of Trees (N)	300
Tree Depth	5
Learning Rate (η)	0.1
Min Samples Split	10
Min Samples Leaf	2
Max Features	'sqrt'
Subsample	0.9

Model Assessment Metrics

1 *Accuracy: Measures the proportion of correctly predicted instances over the total number of instances, EQU (6),*

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \tag{6}$$

2 *Precision: Indicates the proportion of correct identifications, EQU (7).*

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{7}$$

3 *Recall (Sensitivity): Measures the proportion of actual positives that were identified correctly, EQU (8).*

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{8}$$

4 *F1-score: Combines precision and recall into a single metric by taking their harmonic mean, EQU (9).*

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

5 *Area Under the Receiver Operating Characteristic Curve (AUC-ROC): Measures the ability of the model to discriminate between classes. An AUC of 1 represents a perfect model, while an AUC of 0.5 represents a worthless model, EQU (10).*

$$AUC\text{-}ROC = \int_0^1 TPR(t)dt \text{ where } t \text{ varies over all possible threshold values} \tag{10}$$

6 *Mean Squared Error (MSE): Measures the average of the squares of the errors-that is the average squared difference between the estimated values and* the actual value, EQU (11).

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{11}$$

The proposed model was evaluated using the above metrics and compared against RF, LR, NB, and SVM models. The results from the evaluation are discussed below:

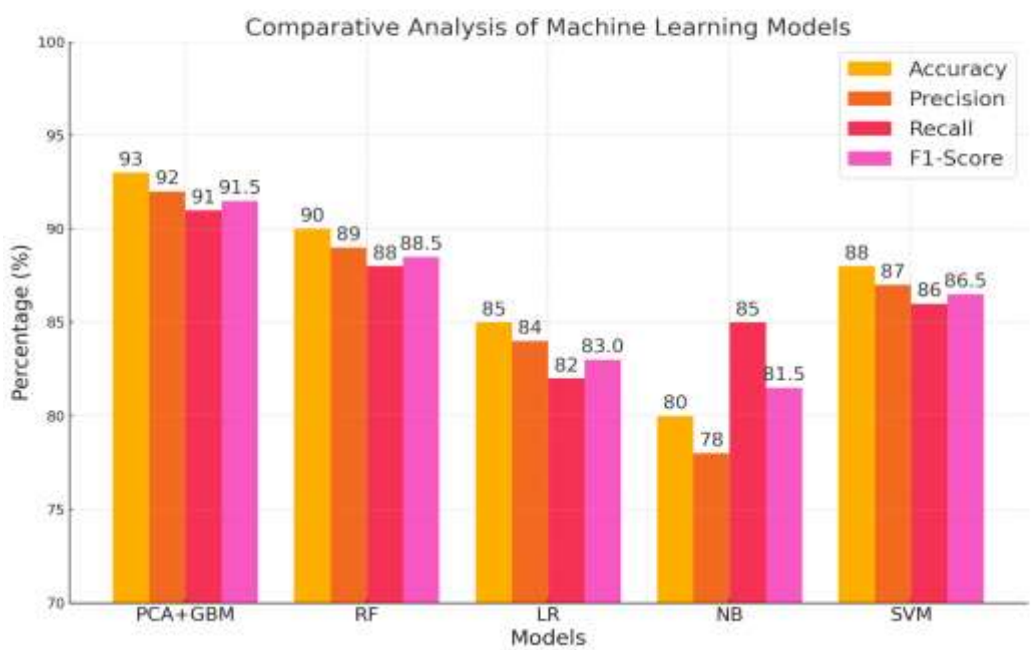

Figure 2. Performance comparison against different metrics

The analysis of the performance metrics for different ML models used in predicting student satisfaction is highlighted in Fig. 2. The PCA-enhanced GBM model outperforms the alternatives across all key metrics: it achieves an accuracy of 93%, precision of 92%, recall of 91%, and an F1-Score of 91.5%. These high scores demonstrate its superior ability to predict satisfaction levels correctly, precisely identify true positives, and maintain a balanced sensitivity and positive predictive value. In comparison, the RF model, which generally performs well in similar settings, posts lower scores with an accuracy of 90%, precision of 89%, recall of 88%, and an F1-Score of 88.5%. Although robust, RF falls short of the PCA with GBM regarding both effectiveness and efficiency in handling complex data.

Here, LR and NB significantly fail. LR has an issue managing the dataset's complexities, demonstrated by its 85% accuracy, 84% precision, 82% recall, and 83% F1-Score. NB's F1-score of 81.5% results from its slightly superior recall of 85% but a lesser precision of 78% and accuracy of 80%. Those results suggest that datasets with excess complex and non-linear relationships are unsuitable for a combination of models, specifically NB. The SVM model records the best scores with 88% accuracy, 87% precision, 86% recall, and an F1-score of 86.5%. Although SVM typically proves good in high-dimensional spaces, the PCA with the GBM model is to be more successful overall.

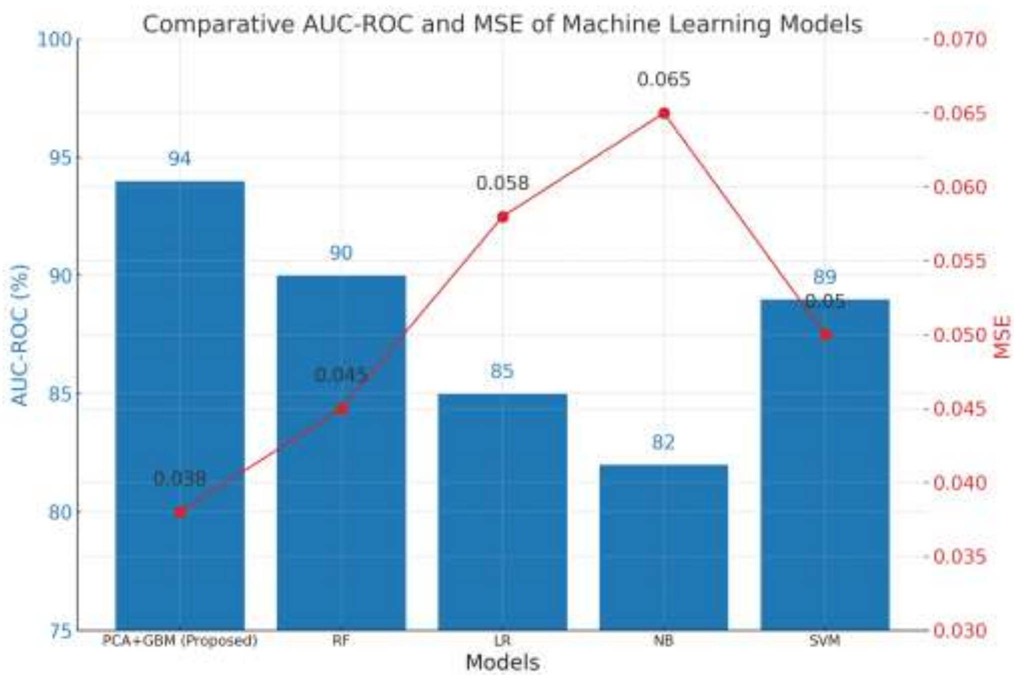

Figure 3. Performance comparison against AUC-ROC and MSE.

Fig. 3 displays the findings of evaluating the AUC-ROC and Mean Squared Error (MSE) for various models of ML. At 94% AUC-ROC and 0.038 MSE, the PCA-enhanced GBM (Proposed) model is a significant step above the rest of the field. The model's high AUC-ROC demonstrates that it can accurately classify results for various thresholds, which is especially valuable for identifying between happy and unsatisfied students. With such a small MSE, it is evident that this model is highly accurate at predicting future results; in other words, it results in exact predictions. RF comes in second with a 90% AUC-ROC and a 0.045 MSE. Although RF has significant classification abilities, it drops short compared to the PCA with the GBM model regarding discriminative power and precision. Still, it maintains a respectable level of performance, which indicates that it is an acceptable model—albeit one that loses the improvement of the PCA with GBM.

LR and NB exhibit lower performance in both metrics. LR has an AUC-ROC of 85% and an MSE of 0.058, indicating moderate classification ability and higher prediction errors than the leading models. NB shows the weakest performance with an AUC-ROC of 82% and the highest MSE of 0.065, suggesting significant limitations in discriminating between classes and predicting with precision. SVM has an AUC-ROC of 89% and an MSE of 0.050, placing it in the mid-range among the models analyzed. While SVM is generally effective in handling high-dimensional data, it appears to fall short of the PCA with GBM in terms of both classification effectiveness and prediction accuracy.

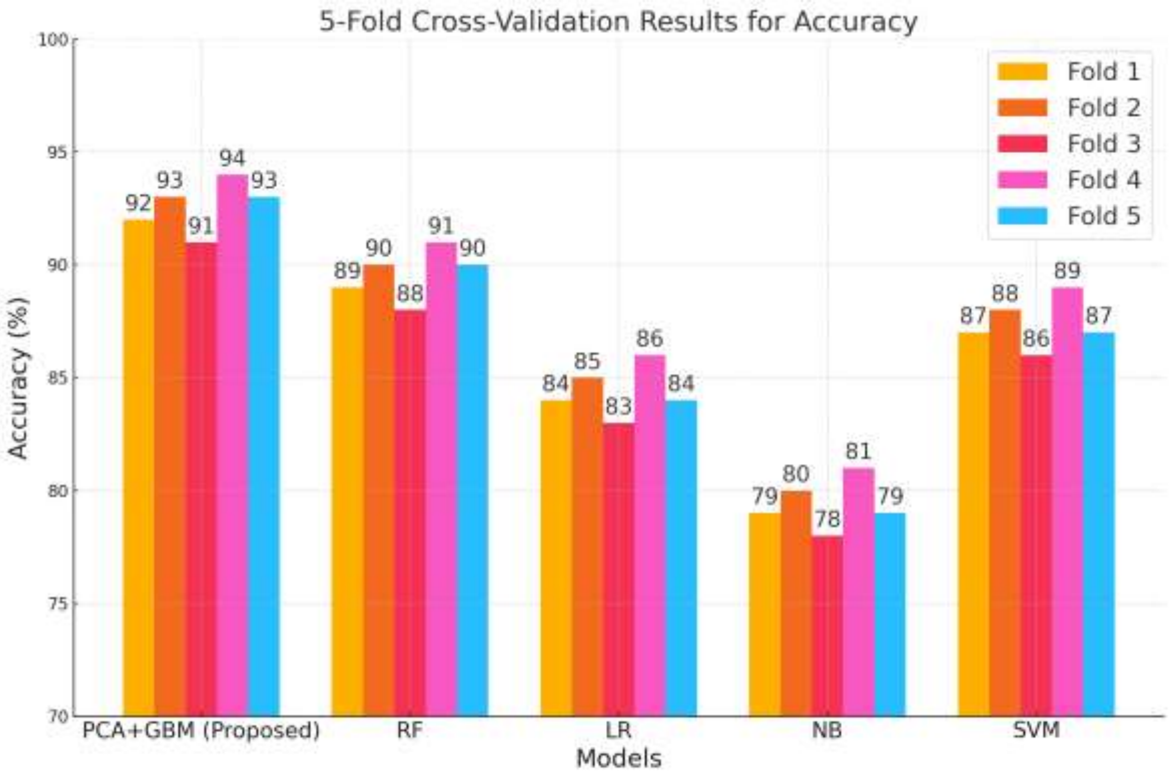

Figure 4.5-Fold Cross-validation results

The 5-fold cross-validation results of various ML models in predicting student satisfaction are shown in Fig. 4. The PCA with GBM model consistently outperforms others, showing accuracies between 91% and 94%, with an average of 92.6%, indicating stable and reliable predictive power. RF follows with an average accuracy of 89.6%, showing less consistency than PCA with GBM. LR and NB demonstrate lower and more variable accuracies of 84.4% and 79.4%, respectively, suggesting they are less suited for handling complex datasets. SVM maintains moderate consistency with an average accuracy of 87.4%, positioned between the higher performance of GBM and the lower outcomes of LR and NB. The results of this study support PCA with GBM as the best model for educational uses requiring reliable data-driven insights. It integrates high accuracy with coherence, which makes it perfect.

CONCLUSION AND FUTURE WORK

The research's predictive model provides an understanding of the factors related to student satisfaction in Higher Education (HE) programs collaboratively managed by Chinese and foreign entities. In order to provide educational administrators and policymakers with an approach, the present study uses a Principal Component Analysis (PCA)-enabled Gradient Boosting Machine (GBM) model to identify the most significant variables controlling happiness level. Across the Machine Learning (ML) models examined, the PCA+GBM model achieved the best findings (93% accuracy, 92% precision, 91% recall, and 91.5% F1-Score), outperforming RF, LR, NB, and SVM. Regarding general satisfaction among students, variables like learning experiences, administrative assistance, integration of cultures, and resources serve an essential role. Policies and administrators in the discipline of HE can benefit from these outcomes. The model can determine actions to improve support for administrative tasks and integration of culture and resources. Students are inclined to be happy and achieve superior outcomes when the curriculum connects to their planned career and HE objectives.

Recognizing similarities and resolving challenges is vital to integrating ML models into the decision-making procedure.

This study has limitations. Data were collected through surveys and academic records, which may not capture all dimensions of student satisfaction. Future studies could include data from social media and student support logs. Expanding the geographical scope could provide broader insights.

Acknowledgment: 2021 "Foreign Language Education in the New Era" Research Projects Project name: Research on the Construction of Teaching Teams in Chinese-foreign Cooperative Education Programs in Universities. Project number: 202113

REFERENCES

Hai-tao, P., Ming-qu, F., Hong-bin, Z., Bi-zhen, Y., Jin-jiao, L., Chun-fang, L., ... & Rui, S. (2021). Predicting academic performance of students in Chinese-foreign cooperation in running schools with graph convolutional network. *Neural Computing and Applications*, *33*, 637-645.

Huang, L., Zhang, W., Jiang, H., & Wang, J. L. (2023). The teaching quality evaluation of Chinese-Foreign cooperation in running schools from the perspective of education for sustainable development. *Sustainability*, *15*(3), 1975.

Tang, Y., & Eng, L. S. (2023). Mentors' and Mentees' Experiences in a Sino-Foreign Cooperative Education Programme: A Phenomenological Study in a Public University in China. *International Journal of Learning, Teaching and Educational Research*, *22*(7), 566-585.

Wang, Y., & Devarajoo, K. 2022 Status of Cross-Cultural Adaptation among Students in the US on the Chinese-Foreign Higher Education Cooperation Program from Anhui Province of China. *Education*, *4*(7), 46-54.

Li, N., Wang, J., Zhang, X., & Sherwood, R. (2021). Investigation of face-to-face class attendance, virtual learning engagement, and academic performance in a blended learning environment. *International journal of information and education technology*, *11*(3), 112-118.

Chen, S. (2019). *Motivations for pursuing postgraduate study in the UK: Perspectives of Chinese graduates of a Sino-British cooperative university* (Doctoral dissertation, University of Oxford).

Li, J. (2022). Exploring the Status Quo of Studying Abroad Development in the Post-Epidemic Era. In *Shaping Education Policy Discourse: Insights From Internationalization of Education Development in China* (pp. 429-440). Singapore: Springer Nature Singapore.

T. -T. Huynh-Cam, L. -S. Chen, T. Van Ho, D. Van Nguyen and T. -C. Lu, "Identifying Crucial Factors for E-students' Satisfaction in Vietnamese Emengency Remote Learning Using Machine Learning Methods," 2023 12th International Conference on Awareness Science and Technology (iCAST), Taichung, Taiwan, 2023, pp. 16-21, doi: 10.1109/iCAST57874.2023.10359269.

Jam, F. A., Khan, T. I., Zaidi, B. H., & Muzaffar, S. M. (2011). Political skills moderates the relationship between perception of organizational politics and job outcomes. Journal of Educational and Social Research, 1(4), 57-70.

C. Verma, Z. Illés and D. Kumar, "Using Machine Learning Predicting Satisfaction of Students towards Hybrid Learning," 2023 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Sakheer, Bahrain, 2023, pp. 88-93, doi: 10.1109/3ICT60104.2023.10391397.

P. Khongchai and P. Songmuang, "Improving students' motivation to study using salary prediction system," 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Khon Kaen, Thailand, 2016, pp. 1-6, doi: 10.1109/JCSSE.2016.7748896.

Z. Huang et al., "Predictive Nomogram for Depressive Symptoms in Postgraduate Medical Students: A Case-Control Study," 2021 International Conference on Information Technology and Biomedical Engineering (ICITBE), Nanchang, China, 2021, pp. 364-368, doi: 10.1109/ICITBE54178.2021.00085.

N. Rjaibi, L. Ben Arfa Rabai and M. Limam, "Modeling the prediction of student's satisfaction in face to face learning: An empirical investigation," International Conference on Education and e-Learning Innovations, Sousse, Tunisia, 2012, pp. 1-6, doi: 10.1109/ICEELI.2012.6360608.

A. Jain, K. shah, P. Chaturvedi and A. Tambe, "Prediction and Analysis of Student Performance using Hybrid Model of Multilayer Perceptron and Random Forest," 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), Bhopal, India, 2018, pp. 1-7, doi: 10.1109/ICACAT.2018.8933580.

M. A. Hossen, R. B. Alamgir, A. U. Alam, F. Siddika, S. F. Hossain and M. S. Arman, "A Web Based Four-Tier Architecture using Reduced Feature Based Neural Network Approach for Prediction of Student Performance," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), DHAKA, Bangladesh, 2021, pp. 269-273, doi: 10.1109/ICREST51555.2021.9331003.

R. Paul and M. Rashmi, "Student Satisfaction and Churn Predicting using Machine Learning Algorithms for EdTech course," 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2022, pp. 1-6, doi: 10.1109/ICRITO56286.2022.9965116.