Data Science and Engineering (DSE) Record, Volume 4, issue 1.

# Prediction of Employment Region of Graduates Using Machine Learning Approach

Bao Xiaohui[1] and Prompong Sugunnasil[2]

[1] Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand
[2] College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand

xiaohui_b@cmu.ac.th

**Abstract.** This study propose a comprehensive machine learning model for predicting the employment region choices of normal higher vocational college students. By undergoing meticulous data pre-processing, comparing diverse algorithms, and resampling methods, we developed an excellent predictive model whose F1-score can reach 0.991 ± 0.015. In addition, we also provide the application methods and scenarios of our model. As nations strive for educational equity, our findings offer a potent predictive framework to inform strategies for attracting capable graduates to rural teaching roles, thereby advancing educational parity and societal development.

**Keywords:** employment region, teacher selection, rural education, machine learning

## 1 INTRODUCTION

Ensuring equitable access to quality education for all students, regardless of their geographical location, is paramount in promoting equal development opportunities for all people in a country. However, the shortage of teachers, particularly in rural areas, presents a pressing challenge faced by numerous countries worldwide. China, for example, has more than 2.9 million rural teachers, with only 58.3% of them under the age of 40 1. Retirement of rural teachers will peak in the next few years, but it is always difficult to recruit teachers for rural areas. For instance, a striking shortage of 4,126 teachers is observed in rural primary and secondary schools in Baise city, with a specific deficit of 1,237 teachers in primary schools alone. These challenges are not unique to Baise, with statistics showing that the rural teacher shortage in Guangxi amounts to tens of thousands 2. In most areas of China, the special posting program①

① The special post teacher program is a special policy implemented by the Chinese central for rural teachers needs to be adjusted and recruited several times, but still remains unfilled. The insufficiency of teachers in rural regions often translates to limited educational resources and diminished educational quality. On the contrary, China has a substantial number of normal students, with relevant data indicating an annual count of over 600,000 normal student graduates. But only nearly 30% of normal students find opportunities to teach in schools 3. The recruitment of rural teachers is a critical issue that not only promotes progress and development in rural society but also bridges the gap between urban and rural areas, ultimately upholding social equity and stability, and facilitating the sustainable development and prosperity of the entire nation.

In recent years, numerous studies have focused on exploring strategies at a macro level to attract normal school students to pursue careers as rural teachers 456. However, it is essential to recognize that the decision of whether these students ultimately choose rural areas for employment is influenced not only by national policies but also by their personal considerations. Therefore, efforts to bridge the urban-rural teacher gap should extend beyond the scope of national policies and take into account the individual aspirations and motivations.

This study endeavors to contribute to the mitigation of the rural teacher shortage and the promotion of educational equity between urban and rural regions. Its expected goal is to develop a micro-level model capable of identifying whether a student has a willing for teaching in rural areas. By achieving this, both the college and the education department can effectively identify and select appropriately qualified teacher candidates for rural primary schools, consequently facilitating their admission into normal colleges.

The proposed model in this study can utilize only objective student-related information, which are more practical and different from previous research primarily relying on surveys, to predict the employment region of graduates, providing more understandable references for policy formulation. 12 machine learning algorithms and 11 resampling methods have been employed and compared in the model construction process to achieve optimal predictive performance.

## 2 RELATED WORKS

The 19th National Congress of the Communist Party of China emphasized the need to promote equality in education, promote the integrated development of compulsory education in urban and rural areas, and attach great importance to compulsory education in rural areas. Additionally, various policies and programs implemented by the Chinese

---

government for rural compulsory education in central and western regions. It aims to guide and encourage college graduates to take up compulsory education in rural areas, gradually solve the problem of insufficient and unreasonable teacher structure in rural schools, improve the overall quality of rural teachers, and promote the balanced development of urban and rural education.

government, such as the "Special Posting Program for Rural Teachers," "Rural Teacher Support Plan," and "Free Normal University Student Plan" further signify the commitment towards enhancing rural education. Furthermore, in response to the shortage of teachers in rural primary schools, the government allowed several higher vocational colleges to establish primary education departments, traditionally offered only at undergraduate universities, to train general teachers② for rural primary schools. However, despite these efforts, rural areas still face challenges due to lower incomes and limited public infrastructure compared to urban regions. Those prompts the migration of young residents to cities in pursuit of better opportunities, leaving rural areas with predominantly elderly and child populations. Similarly, most of the normal higher vocational college students still decide to work in urban areas, even though they cannot secure regular teaching positions there 7.

Scholars from various countries have explored solutions to the challenges of rural teacher recruitment using logical deduction by reasoning 891011 and statistical methods 46121314. Some studies have focused on understanding the factors influencing individuals' decisions to pursue teaching as a career 1516212223 or why rural teachers leave their positions in rural areas 2223242526. One impactful contribution is the FIT-Choice scale developed by Helen M. G. Watt and Paul W. Richardson **Error! Reference source not found.**, which has been widely utilized globally **Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.** , including in China by scholars Liu and Fang 21, to assess career choice factors in teaching. Fu Wei-dong and Fu Yi-chao 4 conducted a cross-analysis of questionnaire data from 4,500 normal education majors, revealing the complex factors influencing normal college students' willingness to teach in rural regions. While traditional methods of logical deduction by reasoning and statistical methods have provided valuable information for rural teacher recruitment, they suffer limitations, such as time-consuming conclusions and susceptibility to subjective judgment.

The widespread adoption of machine learning algorithms has revolutionized the prediction for human behavior, finding applications in areas like credit defaults 28, fraud detection 29, diabetic patients’ re-admission 30. However, there remains a lack of studies applying machine learning methods to analyze teachers' employment choices. As a result, our study draws insights from related research in areas such as graduates' employability 3132, recruitment prediction 3334, and employee retention 353637. These studies mostly rely on questionnaires as the primary data source 33343638. However, such approaches may be more suitable for purely academic research or additional assessments during recruitment processes due to the relatively small datasets and high costs associated with survey implementations. To address this issue, algorithm-focused scholars may opt to utilize publicly available datasets such as the IBM HR Dataset 323539. Nevertheless, these datasets are often outdated and have undergone multiple analyses, making it challenging to derive novel insights for

---

② A general teacher is the teacher who can teach the whole subjects.

practical use. Alternatively, some researchers opt to access data from relevant institutions, which can offer unique and relatively new information. Francisco, et. al. 31 obtained an average F1-Score of 0.62 while predicting whether the graduates would be employed based on the information of 134,129 master's degree graduates retained by the OEEU project. However, acquiring such data can be difficult, and the fixed attributes may not always align with the specific needs of researchers, impacting prediction accuracy. Table 1 provides a summary of the latest and the most relevant studies for this work.

Table 1. A Summary of the Latest Relevant Studies for This Work

| Method | Ref. | Year | data sources | data scale | Area | Performance |
|---|---|---|---|---|---|---|
| logical deduction | 8 | 2020 | — | — | Rural Teacher Retention | — |
| | 26 | 2021 | — | — | Rural Teacher Retention | — |
| Statistics | 6 | 2018 | Survey | 7834 | Rural Teacher Recruitment | — |
| | 13 | 2021 | Survey | 1328 | Rural Teacher Retention | — |
| | **Error! Reference source not found.** | 2018 | Survey | 354 | Teacher Recruitment | — |
| Machine Learning | 32 | 2020 | open sources | 1400 | Graduates' Employability | Accuracy: 78% |
| | 38 | 2021 | open sources | 1000 | Employee Retention | Accuracy: 86.39% |
| | 39 | 2023 | the database of relevant institutions | approximately 700,000 | Employee Retention | F1-Score: 99.48 |

Despite the valuable insights provided by previous studies, they have certain limitations. As a result, our research aims to contribute in the following ways: (i) Proposing a fast analysis model that continuously updates the outcomes of relevant analyses. This enables the government to make timely adjustments in response to changing preferences among young individuals. (ii) Complementing existing research by incorporating an insight from individual at the micro level. By combining the results of this study with the findings of previous research, a more comprehensive policy foundation can be established. (iii) Constructing a model based on objective and existing data that includes students' fundamental information upon entering school and their career choices after graduation, it captures the transformation of individuals from students to teachers, which is crucial for studying individual career choices.

## 3 METHODOLOGY

Through a comprehensive employment of various machine learning algorithms, data mining means and statistical resample methods, our aim is to identify the optimal model for predicting the employment region choices of normal higher vocational college students. A visual representation of our experiment's structure is provided as Fig. 1.

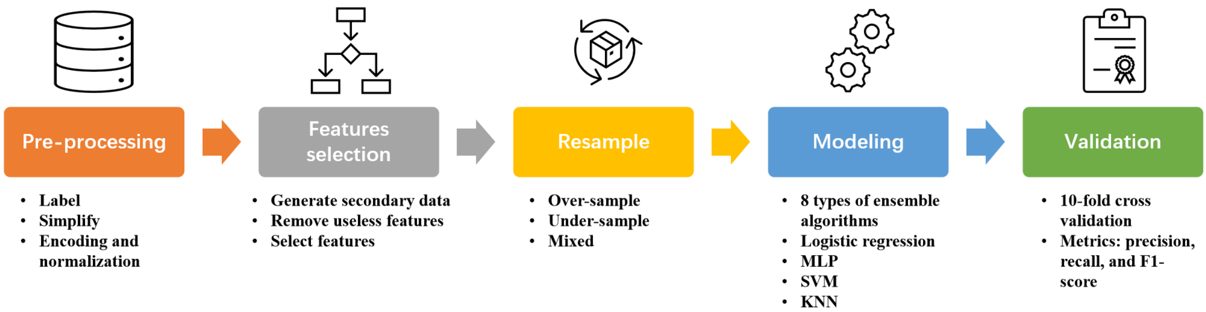


Fig. 1. The framework of the experiment

### *3.1 Data collection & Pre-processing*

The dataset utilized in this study has been sourced from a student information database maintained by the student affairs department of a higher vocational college situated in the western Chinese province of Guangxi. This region is generally acknowledged as relatively underdeveloped in China. The student information database encompasses a dataset comprised of information pertaining to 454 graduates. Each individual graduate is characterized by a distinct set of 15 features, including attributes such as their name, class, student ID, ethnicity, birthplace, and more.

To delve deeper into the data's significance, we have embarked on generating secondary data derived from metadata. The following procedures have been employed: (i) Identify the population, area, and population density of the county where each student resides. These parameters are then compared against the corresponding figures within the area of the student's workplace. (ii) Obtain data regarding the local minimum wage and per capita disposable income (PCDI) in the student's hometown. Subsequently, we compute the disparities between these amounts and their equivalents at the student's workplace. (iii) Tencent Maps is utilized to calculate both the geographical distance and driving duration between the student's hometown and their workplace.

### *3.2 Modeling*

To enhance the predictive performance and mitigate any potential bias arising from imbalanced data, 4 over-sampling, 2 under-sampling and 2 mixed resampling techniques are employed. Then, we endeavor to predict employment region choices of normal higher vocational college graduates using 12 distinct machine learning algorithms. These encompass Stochastic Gradient Descent (SGD), Decision Tree (DT), Logistic Regression (Log), AdaBoost, Multilayer Perceptron (MLP), Gradient Boost (GB), Bagging, Support Vector Machine (SVM), Random Forest (RF), Extra Trees Classifier, K-nearest neighbors (KNN), and XG Boost. This comprehensive array of algorithms ensures a comprehensive assessment of predictive capabilities.

Following the aforementioned steps, we conducted a 10-fold cross-validation process and employed precision, recall, and F1-score metrics to rigorously evaluate the predictive performance of each algorithm. These evaluations were specifically focused on predicting outcomes related to the dependent variable denoted as 'yes,' indicating a graduate's inclination to pursue a career as a rural teacher.

## 4 EXPERIMENT AND RESULTS

### *4.1 Initial model*

Initially, the original dataset was employed to assess its quality by subjecting it to the models, resulting in predictive outcomes. The left segment of Table 2 presents the computed performance metrics of the classifiers, including precision, recall, and f1-score. The findings highlight that the Decision Tree (DT) and Support Vector Machine (SVM) algorithms achieved the highest f1-scores at 0.409 and 0.508, respectively. Nevertheless, these algorithms still fell short of providing dependable predictions.

Table 2. Performance with different algorithms

| | **With original dataset** | | | **With secondary data** | | |
|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1-Score** | **Precision** | **Recall** | **F1-Score** |
| **BAG** | 0.235±0.221 | 0.123±0.139 | 0.150±0.160 | 0.978±0.028 | 1.000±0.000 | 0.989±0.014 |
| **Log** | 0.235±0.221 | 0.123±0.139 | 0.150±0.160 | 0.982±0.028 | 0.993±0.020 | 0.988±0.021 |
| **Ada** | 0.277±0.312 | 0.093±0.117 | 0.127±0.138 | 0.977±0.028 | 0.995±0.016 | 0.986±0.019 |
| **GB** | 0.070±0.155 | 0.036±0.090 | 0.047±0.113 | 0.977±0.029 | 0.993±0.021 | 0.985±0.022 |
| **XGB** | 0.421±0.109 | 0.354±0.100 | 0.376±0.083 | 0.977±0.029 | 0.988±0.024 | 0.983±0.021 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **DT** | 0.406±0.148 | 0.418±0.101 | 0.409±0.123 | 0.977±0.029 | 0.987±0.026 | 0.982±0.022 |
| **RF** | 0.388±0.164 | 0.292±0.096 | 0.326±0.112 | 0.975±0.033 | 0.988±0.024 | 0.981±0.017 |
| **MLP** | 0.480±0.315 | 0.231±0.179 | 0.266±0.154 | 0.976±0.031 | 0.975±0.060 | 0.975±0.041 |
| **SGD** | 0.257±0.180 | 0.343±0.338 | 0.271±0.220 | 0.970±0.040 | 0.978±0.036 | 0.973±0.026 |
| **SVM** | 0.435±0.087 | 0.635±0.087 | 0.508±0.087 | 0.969±0.044 | 0.967±0.035 | 0.967±0.028 |
| **ET** | 0.421±0.165 | 0.354±0.116 | 0.374±0.120 | 0.965±0.044 | 0.841±0.110 | 0.895±0.079 |
| **KNN** | 0.391±0.108 | 0.369±0.104 | 0.372±0.085 | 0.714±0.144 | 0.707±0.074 | 0.702±0.088 |

Subsequently, the processed data was fed into the same models. Notably, a significant enhancement was observed across all models in terms of precision, recall, and F1-scores due to the integration of the generated secondary data. For most models, these values surged beyond 0.9. Specifically, the Bagging (BAG), Logistic Regression (Log), and Ada Boost (Ada) models exhibited the most remarkable F1-scores, attaining 0.989, 0.988, and 0.986, respectively.

### *4.2 Enhancing Predictive Performance through Resampling Methods*

Since the model is mildly imbalanced, it may interfere with its operation. To optimize the model, we applied resampling techniques to balance the training dataset. Table 3 illustrates the predictive outcomes post-resampling. The results unveil that, for the Ada Boost and Bagging method, data pre-processing using resampling methods doesn't notably enhance performance. Nevertheless, nearly every resampling method elevates the performance of the Logistic Regression Classifier (Log). The most favorable combination yielded the F1-score of 0.991 ± 0.015.

Table 3. Model performance with various resample techniques

| | | **over-sample** | | | | **under-sample** | | **another** | |
|---|---|---|---|---|---|---|---|---|---|
| | | SO | SMOTE | BSMOTE | SVM | SU | CM | Tomek | ENN |
| **BAG** | **Precision** | 0.975 ± | 0.966 ± | 0.975 ± | 0.965 ± | 0.966 ± | 0.975 ± | 0.975 ± | 0.942 ± |
| | **Recall** | 1.000 ± | 1.000 ± | 1.000 ± | 1.000 ± | 1.000 ± | 1.000 ± | 0.995 ± | 0.972 ± |
| | **F1-Score** | 0.987 ± | 0.982 ± | 0.987 ± | 0.981 ± | 0.982 ± | 0.987 ± | 0.985 ± | 0.955 ± |
| **Log** | **Precision** | 0.976 ± | 0.981 ± | 0.982 ± | 0.982 ± | 0.982 ± | 0.976 ± | 0.981 ± | 0.853 ± |
| | **Recall** | 1.000 ± | 0.993 ± | 1.000 ± | 1.000 ± | 1.000 ± | 1.000 ± | 0.993 ± | 0.968 ± |
| | **F1-Score** | 0.987 ± | 0.987 ± | 0.991 ± | 0.991 ± | 0.991 ± | 0.987 ± | 0.987 ± | 0.905 ± |
| **Ada** | **Precision** | 0.982 ± | 0.975 ± | 0.982 ± | 0.970 ± | 0.975 ± | 0.982 ± | 0.981 ± | 0.981 ± |
| | **Recall** | 0.991 ± | 0.996 ± | 0.996 ± | 0.991 ± | 0.996 ± | 0.996 ± | 0.991 ± | 0.990 ± |
| | **F1-Score** | 0.986 ± | 0.985 ± | 0.988 ± | 0.979 ± | 0.985 ± | 0.988 ± | 0.986 ± | 0.986 ± |

# 5 DISCUSSION

The results presented in Table 2 underscores the substantial positive effect of incorporating secondary data on the performance of our model. Specifically, ensemble learning techniques, such as Bagging, AdaBoost, Gradient Boosting, and XGBoost, exhibit noteworthy enhancements in precision, recall, and F1-score metrics when enriched with supplementary data. A similar trend is observed with Logistic Regression algorithms, implying that these algorithms derive significant advantages from the inclusion of numerical data. In contrast, Support Vector Machine (SVM) algorithms and tree-based models, such as Decision Trees, Random Forest (RF), and Extra Trees (ET), demonstrate relatively strong performance in precision, recall, and F1-scores when solely applied to the original dataset, underscoring their inherent robustness. And the utilization of secondary data also substantially amplifies their performance. Nonetheless, all algorithms exhibit limitations in furnishing reliable predictions when reliant solely on the original dataset. This deficiency implies that the generated data may encompass pivotal factors influencing the selection of employment regions by graduates, warranting further exploration in subsequent phases of our research.

Our findings indicate that a combination of various resampling techniques and logistic regression classification algorithms yields highly effective results in predicting employment region choices for normal college graduates. This effectiveness becomes particularly pronounced in datasets with an imbalance. Among the resampling techniques applied, including simple under-sampling (SU), Borderline-SMOTE (BSMOTE), and SVM-SMOTE (SS), the most favorable results are produced. Conversely, the influence of resampling on Bagging and AdaBoost is relatively limited. This observation can be attributed to their status as ensemble algorithms, which gives them an inherent robustness against imbalanced data, reducing the need for resampling techniques. As such, they will be good options when we need to obtain acceptable predictive performance in a short time.

Diverging from prior research, which predominantly relied on surveys, our proposed model capitalizes solely on objective student-related information. This approach not only reduces implementation costs but also mitigates the potential for fraudulent behavior during admission tests, rendering it a more practical and reliable approach. Consequently, relevant organizations can leverage our models to efficiently identify suitable candidates for rural teaching positions based on student data. The application process can be followed as follows: (i) Target Teacher Input Regions: Identify the rural areas that require an influx of teachers. Combine this information with applicant details to create a specialized database. (ii) Model Training: Train our models, which combine the SU resampling technique with the logistic regression algorithm, using employment data from recent years in the specified regions. (iii) Prediction: Input the processed data into the trained model to obtain predictive outcomes. By following this process, educational institutions and related entities can efficiently identify and target potential candidates for rural teaching positions, aiding the great improvement of educational equity in underserved areas.

Our proposed model offers swift results and relies on data that is readily accessible through relevant government agencies and educational institutions. This simplicity enhances its potential for widespread application across diverse regions and industries for selecting high-quality candidates. Nonetheless, our study does acknowledge certain limitations. We must highlight our constrained industry expertise, which limited our capacity for data collection, secondary data generation, and processing. Hence, we recommend that further exploration and implementation of our model in diverse contexts be pursued by practitioners and educators within related fields at subsequent stages.

# 6 CONCLUSION AND OUTLOOK

We've presented a comprehensive machine learning model to predict the employment regions selected by normal higher vocational college students. This involved meticulous data preprocessing, diverse algorithms, and resampling techniques, resulting in a high-performing model achieving an impressive F1-score of $0.991 \pm 0.015$.

Our findings provide a predictive framework to inform strategies for attracting capable graduates to rural teaching roles, advancing educational equality and social development. Based on our existing research results, we can further explore and apply our findings from the following three avenues: exploring the impact of factors on graduate job region decisions, integrating the model into other management systems for real-time prediction and monitoring, and adapting the model to different countries and industries by tailoring data processing to local contexts.

## References


1. The voice of Minzu University of China during the "Two Sessions" in 2022 | alumni Xie Shangguo：We shall introduce a policy to raise the retirement age of teachers as soon as possible, [EB-OL]. (2022-03-05) [2022-12-02]. https://xiaoyou.muc.edu.cn/info/1004/1853.htm.
2. Why rural teacher recruitment is neglected, Guangming Online,[EB-OL]. (2018-09-22) [2022-12-02]. https://news.gmw.cn/2018-09/22/content_31307508.htm
3. Li Binhui, Li Shihui. (2019). Rural Teacher's Supply-side Structural Reform in Normal Universities under “New Normal University” Background. Journal of Lingnan Normal University, 40(03): 1-10+34.
4. FU Wei-dong, FU Yi-chao. (2015). An investigation of the free-tuition undergraduates and their willingness in rural teaching in 30 regional universities. Journal of Hebei Normal University(Educational Science Edition), 17(1): 114-120.
5. Zhang Xiaoying, Leng Xiaohei. (2019). On Influencing Factors of ＆ Incentive Strategies for Normal Students’ Rural-Area-working in Vitalization Process of Rural Education. Journal of Yichun University, 42(01):109-114.
6. KE Wenjin, DENG Fei, JIANG Jinqiu, et al. (2018). Factors Influencing the Willingness of Rural-Born College Students to Teach in the Countryside:A Survey of Normal Universities

in Western China. Fudan Education Forum, 16(4): 73-79.
7. Yang Wei 'an. (2019). The Changes of the Policy for Rural Primary School Teacher Supplements:Review and Prospects over the Past 70 Years. Educational Research, 40(07):16-25.
8. Li J, Shi Z, Xue E. (2020). The problems, needs and strategies of rural teacher development at deep poverty areas in China: Rural schooling stakeholder perspectives. International Journal of Educational Research, 99: 101496.
9. YU Hai-hong; WANG Dian-dong. (2017). Supply-side Reform of Training Rural Teachers at Local Normal Universities in the Era of Big Data. University Education Science(02), 62-67+126.
10. Acikgoz, Y. (2019). Employee recruitment and job search: Towards a multi-level integration. Human resource management review, 29(1), 1-13.
11. Mingren, Z., & Shiquan, F. (2018). Rural teacher identity and influencing factors in Western China. Chinese Education & Society, 51(2), 91-102.
12. Wong, A. K., Tang, S. Y., & Cheng, M. M. (2014). Teaching motivations in Hong Kong: Who will choose teaching as a fallback career in a stringent job market?. Teaching and teacher education, 41, 81-91.
13. Tan Hui-ling, WAN Xing-Rui, HUANG Jie-qiong. (2021). A Mixed-method Study on the Evaluation of the Implementation Effect of Poverty Alleviation Policies in Education——A Case Study of the Supporting Plan for Rural Teachers in J County. Journal of Educational Studies (02) : 118-128.
14. Zeng Xin, Gao Zhenyi. (2018). Empowering:Building a Need-oriented Teacher Team Construction Road for Small Rural Schools——Based on the Survey of the Implementation of The Supporting Project of Rural Teachers in 6 Provinces. Journal of Central China Normal University(Humanities and Social Sciences), 57(01):174-187.
15. Miller, L. C. (2012). Situating the rural teacher labor market in the broader context: A descriptive analysis of the market dynamics in New York State. Journal of Research in Rural Education (Online), 27(13), 1.
16. Watt, H. M., & Richardson, P. W. (2007). Motivational factors influencing teaching as a career choice: Development and validation of the FIT-Choice scale. The Journal of experimental education, 75(3), 167-202.
17. Watt, H. M., Richardson, P. W., Klusmann, U., Kunter, M., Beyer, B., Trautwein, U., & Baumert, J. (2012). Motivations for choosing teaching as a career: An international comparison using the FIT-Choice scale. Teaching and teacher education, 28(6), 791-805.
18. Fokkens-Bruinsma, M., & Canrinus, E. T. (2012). The factors influencing teaching (FIT)-choice scale in a Dutch teacher education program. Asia-Pacific Journal of Teacher Education, 40(3), 249-269.
19. Salifu, I., Alagbela, A. A., & Gyamfi Ofori, C. (2018). Factors influencing teaching as a career choice (FIT-Choice) in Ghana. Teaching Education, 29(2), 111-134.
20. Suryani, A., Watt, H. M., & Richardson, P. W. (2016). Students' motivations to become teachers: FIT-Choice findings from Indonesia. International Journal of Quantitative Research in Education, 3(3), 179-203.
21. Liu Jia, Fang Xing. (2015). An Empirical Study of the Teacher Education Students' Intention to Work in the Rural Areas and the Influencing Factors. Journal of Teacher Education, 2(5): 72-79.
22. Hammer, P. C., Hughes, G., McClure, C., Reeves, C., & Salgado, D. (2005). Rural Teacher Recruitment and Retention Practices: A Review of the Research Literature, National Survey of Rural Superintendents, and Case Studies of Programs in Virginia. Appalachia Educational Laboratory at Edvantia (NJ1).

23. Gallo, J., & Beckman, P. (2016). A global view of rural education: Teacher preparation, recruitment, and retention. Global education review, 3(1).
24. Li, J., Shi, Z., & Xue, E. (2020). The problems, needs and strategies of rural teacher development at deep poverty areas in China: Rural schooling stakeholder perspectives. International Journal of Educational Research, 99, 101496.
25. Miller, L. C. (2012). Working paper: Understanding rural teacher retention and the role of community amenities. Center for Educational Policy and Workforce Competitiveness. University of Virginia.
26. Li Caihong. (2021). Why Rural Teachers Escape? From the Formative Perspective of Informal Institution. Educational Science Research(03):78-84.
27. Chen, Z., Lin, G., & Jin, X. (2021, November). Credit Approval Prediction Based on Boruta-GBM Model. In 2021 7th International Conference on Systems and Informatics (ICSAI) (pp. 1-5). IEEE.
28. Alonso Robisco, A., & Carbo Martinez, J. M. (2022). Measuring the model risk-adjusted performance of machine learning algorithms in credit default prediction. Financial Innovation, 8(1), 70.
29. Choi, D., & Lee, K. (2017). Machine learning based approach to financial fraud detection process in mobile payment system. IT CoNvergence PRActice (INPRA), 5(4), 12-24.
30. Alloghani, M., Aljaaf, A., Hussain, A., Baker, T., Mustafina, J., Al-Jumeily, D., & Khalaf, M. (2019). Implementation of machine learning algorithms to create diabetic patient re-admission profiles. BMC medical informatics and decision making, 19, 1-16.
31. García-Peñalvo, F., Cruz-Benito, J., Martín-González, M., Vázquez-Ingelmo, A., Sánchez-Prieto, J. C., & Therón, R. (2018). Proposing a machine learning approach to analyze and predict employment and its factors.
32. Xu, L. (2020). College student employment data platform based on FPGA and machine learning. Microprocessors and Microsystems, 4, 103471.
33. Al-Radaideh, Q. A., & Al Nagi, E. (2012). Using data mining techniques to build a classification model for predicting employees performance. International Journal of Advanced Computer Science and Applications, 3(2).
34. Li, Y. M., Lai, C. Y., & Kao, C. P. (2011). Building a qualitative recruitment system via SVM with MCDM approach. Applied Intelligence, 35, 75-88.
35. Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., & Zhu, X. (2019). Employee turnover prediction with machine learning: A reliable approach. In Intelligent Systems and Applications: Proceedings of the 2018 Intelligent Systems Conference (IntelliSys) Volume 2 (pp. 737-758). Springer International Publishing.
36. Alao, D. A. B. A., & Adeyemo, A. B. (2013). Analyzing employee attrition using decision tree algorithms. Computing, Information Systems, Development Informatics and Allied Research Journal, 4(1), 17-28.
37. Chang, H. Y. (2009, March). Employee turnover: a novel prediction solution with effective feature selection. In WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering (No. 3). World Scientific and Engineering Academy and Society.
38. Qutub, A., Al-Mehmadi, A., Al-Hssan, M., Aljohani, R., & Alghamdi, H. S. (2021). Prediction of employee attrition using machine learning and ensemble methods. Int. J. Mach. Learn. Comput, 11(2), 110-114.
39. Silpa, N., Rao, V. M., Subbarao, M. V., Kurada, R. R., Reddy, S. S., & Uppalapati, P. J. (2023, May). An Enriched Employee Retention Analysis System with a Combination Strategy of Feature Selection and Machine Learning Techniques. In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 142-149). IEEE.