Proceedings of the 4th International Conference on Literature, Language, and Culture Development
DOI: 10.54254/2753-7064/53/2025.21770

Construction of a CatBoost Classification Prediction Model for Municipal-Level Teacher Identity Based on Professional Experience

Jingyi Shan[1,a], Linan Lin[1,b,*]

[1]*College of Preschool and Primary Education, Shenyang Normal University, Shenyang, China*
a. 1420538718@qq.com, b. 18602453365@126.com
**corresponding author*

Abstract: To explore the relationship between teacher identity characteristics and professional experience, this study takes municipal-level outstanding teachers in the Beijing-Tianjin-Hebei region as the research sample and constructs a teacher identity classification prediction model based on the CatBoost machine learning algorithm. Guided by the principle of "promoting the great spirit of educators," this paper extracts 14 professional experience characteristic indicators from four dimensions: educational and teaching ability, collaboration and innovation ability, research and practice ability, among others. Leveraging the advantages of the CatBoost algorithm in handling categorical features, the study calculates the importance of various identity characteristics of outstanding teachers and applies them in the training and testing of variables. In terms of innovation, the CatBoost algorithm improves upon the traditional Gradient Boosting Decision Tree (GBDT) by refining the classification effectiveness through test data evaluation, thereby ensuring precise model assessment. The research findings indicate that machine learning has broad applicability in evaluating teacher professional development and can accurately reveal the composition of teacher professional identity. This provides strong data support and scientific evidence for advancing teacher professional development, formulating targeted training strategies, and promoting the construction of a strong education system.

Keywords: Municipal-Level Teachers, CatBoost Classification Prediction, Identity Characteristics

1. Introduction

The Education Powerhouse Development Plan (2024–2035) further emphasizes the strategic goal of "vigorously promoting the spirit of educators and cultivating a high-quality teaching workforce." As the demand for educational modernization continues to evolve, traditional teacher education models are increasingly unable to meet the requirements of building an educational powerhouse, manifesting two major shortcomings: (1) the regional imbalance in the allocation of outstanding teacher resources and (2) the lack of a scientific evaluation system for teacher professional development. To address these issues, educational administrative departments and teacher education researchers must leverage modern information technologies such as big data and artificial intelligence to identify and filter key characteristics of teacher professional development, thereby facilitating the transformation and

DOI: 10.54254/2753-7064/53/2025.21770

upgrading of teacher education paradigms. Against this backdrop, how to construct a dynamic model for evaluating teacher professional development using machine learning algorithms has become a critical focus in deepening the reform of the teacher education system.

Traditional teacher training models have long relied on theoretical coursework and the experiential transmission between mentors and trainees. However, the limitations of this approach have become increasingly apparent in practice. First, there exists a disconnect between theory and practice. The training of pre-service teachers primarily focuses on educational theory, supplemented by short-term teaching internships. Research indicates that pre-service teachers often exhibit weak awareness of professional development and insufficient job competence, resulting in prolonged adaptation periods after entering the workforce. Additionally, the homogenization of course content and the neglect of individual differences among teachers are prominent drawbacks of the traditional training model. Existing teacher education curricula fail to meet the personalized development needs of teachers across different disciplines, with the specific needs of special education teachers being particularly overlooked. Second, the evaluation criteria are overly simplistic, with the assessment of outstanding teachers overly reliant on static indicators such as academic credentials and award counts. This lack of effective measurement for dynamic capabilities such as digital literacy, research output, and teaching innovation has led to a utilitarian tendency wherein teachers develop professionally merely to achieve higher rankings. Third, there is insufficient support for lifelong teacher development. Professional development is an ongoing process that extends beyond initial teacher training to encompass all stages of a teacher's career. Schools and educational authorities should update their teacher training philosophies, establish an integrated teacher development model, and provide diverse platforms and resources to better support teachers' professional growth.

In recent years, with the rise of the artificial intelligence industry, the field of educational data mining has broken traditional assessment barriers by leveraging model construction to enhance evaluation efficiency and promote the innovative application of machine learning algorithms in teacher education. By integrating dynamic feature capture, nonlinear relationship modeling, and a dual-driven approach to prediction and attribution, intelligent and standardized teacher development assessments have become possible. Existing research has shown that machine learning-based evaluation models significantly outperform traditional methods in terms of prediction speed and accuracy. For example, the unsupervised Random Forest (RF) method, using Gini index, permutation importance, and Boruta algorithm, has been effectively employed to rank the importance of cloud computing-related variables [1] and to identify and analyze intelligent strategies for family education models [2]. Building upon this foundation, this study introduces the CatBoost machine learning classification algorithm, using the professional experience of municipal-level teachers as the basis. By extracting discriminative temporal features from categorical data, we construct an interpretable model aligned with disciplinary logic, enabling continuous monitoring and attribution analysis of teacher growth potential. This approach advances the paradigm of research on teacher professional development.

2. Literature Review

A solid disciplinary foundation and strong teaching ability are fundamental components of an outstanding teacher's professional competence. On the one hand, excellent teachers must possess profound subject knowledge. Beyond mastering fundamental educational theories, they need to have a comprehensive and in-depth understanding of the curriculum goals and core concepts of their subject areas, applying their subject knowledge system effectively in actual teaching practices [3]. On the other hand, outstanding teachers demonstrate the ability to flexibly adjust their teaching methods and strategies based on the interests and learning needs of different students. This adaptability enhances student engagement and fosters active learning. In the New Basic Education

DOI: 10.54254/2753-7064/53/2025.21770

Reform experiment, Professor Lan Ye explicitly proposed the idea of "returning the classroom to students," advocating for classrooms to become spaces where students engage in autonomous learning and active inquiry. This philosophy has been effectively implemented by outstanding teachers, who prioritize students' central role in learning, continuously innovate classroom teaching and classroom management approaches, and achieve an organic integration of teacher guidance and student autonomy.

Conviction serves as the driving force behind a teacher's continuous progress and is fundamental to their professional development. A noble educational belief reflects a teacher's passion for education and sense of social responsibility. It motivates teachers to explore new educational approaches to better cultivate students with the essential character traits and key competencies necessary for adapting to future societal developments and achieving lifelong personal growth [4]. In recent years, with the increasing emphasis on the spirit of educators, the significance of teacher ethics in the teacher education system has become more prominent. Teacher ethics represent the behavioral norms and moral standards that teachers must uphold in fulfilling their professional duties. This includes not only a commitment to the education sector but also a spirit of dedication and selflessness. The saying "A teacher must be learned to teach and upright to be a model" underscores the dual role of teachers as both knowledge disseminators and guides in students' personal growth. Teachers should lead by example with noble moral character and cultivate students with profound professional knowledge, embodying the unity of both academic teachers and moral educators.

Innovative awareness and digital competence are prerequisites for teachers to evolve into research-oriented educators [5]. The China Education Modernization 2035 policy has set higher expectations for teachers' digital literacy and educational technology capabilities. With the advent of the artificial intelligence era, the education sector has seen the emergence of a new generation of innovative and forward-thinking educators. These teachers break away from traditional teaching models, actively embrace AI technologies, and integrate them into their instructional practices, leading to the development of novel and highly effective teaching methods that significantly enhance teaching quality and efficiency. In addition to possessing a solid academic foundation, outstanding teachers adapt to technological advancements and leverage educational technology to create personalized learning platforms for students. This enables differentiated instruction to thrive in a digital and intelligent education environment. Through project-based learning and similar approaches, students are encouraged to apply AI in practice, fostering their innovative thinking and problem-solving abilities, thereby effectively advancing educational modernization and the development of a strong education system.

Continuous learning and self-improvement are indispensable driving forces in teachers' professional development. In an era of rapid advancement, where knowledge updates accelerate, and new technological fields emerge constantly, teachers must maintain a proactive learning attitude. They need to continuously update their knowledge structures, enhance their teaching proficiency, and improve their practical skills to keep pace with societal changes. Teachers who embrace lifelong learning can not only advance their own growth but also engage in continuous reflection and self-evaluation. By critically examining their teaching practices, teachers can identify and address problems in a timely manner, ultimately improving the quality and effectiveness of education.

3. Methodology Design

3.1. Data Sources

This study utilizes the professional experience of municipal-level teachers as the data source and applies the CatBoost machine learning algorithm to establish a classification prediction model. Based on the principles of the CatBoost algorithm, variables are selected from the professional experience

DOI: 10.54254/2753-7064/53/2025.21770

indicator system of municipal-level teachers. Specifically, nine indicators are set as feature variables: the presence or absence of practical experience, status as a training expert, involvement in research projects, research output, recognition of excellent teaching cases by the JiaoShiWang platform, teacher ranking, collaboration ability, educational background, and administrative experience. Correspondingly, municipal-level outstanding teachers are designated as the output variable. The constructed CatBoost model calculates the importance of teacher identity characteristics and applies these calculations in the training and testing of data. Unlike traditional models, the CatBoost algorithm evaluates model performance through classification metrics such as accuracy, recall, precision, and the F1 score, ensuring a scientifically robust classification assessment.

The CatBoost algorithm enhances the traditional Gradient Boosting Decision Tree (GBDT) by improving its efficiency in handling categorical features. In categorical feature processing, CatBoost eliminates cumbersome preprocessing steps by adopting target encoding, which directly processes categorical features through smoothed mean encoding. This approach not only accelerates training speed but also mitigates overfitting issues. Regarding sequential tree construction, CatBoost innovatively employs an ordered boosting approach, ensuring that each tree can only access the predictions of preceding trees during its construction. By flexibly utilizing various loss functions, this method effectively resolves the information leakage problem commonly found in traditional GBDT models. The CatBoost loss function, widely recognized in academic research, consists of two components: training error and a regularization term [6]. The loss function is formulated as follows:

$$\mathcal{L}(F) = \sum_{i=1}^{n} L\left(y_i, F(x_i)\right) + \sum_{k=1}^{k} \Omega(f_k)$$

where

$\Omega(f_k)$ represents the regularization term of the k-th tree, typically consisting of the number of leaf nodes and the sum of squared leaf weights. This term is used to adjust the complexity of the model and prevent overfitting. Meanwhile, the training error measures the discrepancy between the predicted values and the actual values.

To enhance prediction speed and ensure robustness against noise and outliers in the data, CatBoost employs a symmetric tree structure, where all nodes within a tree split based on the same feature and threshold. This design ensures computational efficiency and stability, making the model well-suited for educational data analysis.

4. Empirical Process and Result Analysis

Table 1: Model Parameters

Parameter Name	Parameter Value
Training Time	0.085s
Data Split Ratio	0.9
Data Shuffling	Yes
Number of Iterations	100
Learning Rate	0.1
L2 Regularization Term	1
Maximum Tree Depth	10
Overfitting Detection Threshold	0
Additional Iterations After Optimization	20

Table 1 presents the parameter configurations and training duration used for constructing the model with CatBoost. The data indicates that the training process took only 0.085 seconds, demonstrating

DOI: 10.54254/2753-7064/53/2025.21770

the high efficiency of the model training. To ensure an effective model evaluation, the dataset was split with a training-to-validation ratio of 9:1. Additionally, to enhance the model's generalization ability and prevent the order of data from affecting training outcomes, CatBoost implemented a shuffling operation. For iteration settings, the model underwent 100 iterations to better fit the data. The learning rate, a key parameter controlling model update speed, was set at 0.1 to achieve a balanced adjustment. To prevent overfitting, the L2 regularization term and the maximum tree depth were set to 1 and 10, respectively, effectively controlling the model's complexity. Furthermore, after achieving optimization, an additional 20 iterations were conducted to further refine performance. This parameter configuration effectively harnesses the advantages of the CatBoost algorithm, enabling the model to achieve high performance within a short training duration.

Table 2: Feature Importance

Feature Name	Feature Importance (%)
Practical Experience	0.00%
Training Expert	3.80%
Research Project Involvement	3.20%
Research Output	42.80%
Recognized Teaching Case (JiaoShiWang)	3.50%
Teacher Ranking	27.70%
Collaboration Ability	0.00%
Educational Background	8.20%
Administrative Experience	10.70%

Table 2 reveals that "Research Output" dominates the model's prediction, with an importance score of 42.8%, making it the most influential factor in classification outcomes. The second most critical feature is "Teacher Ranking", with an importance of 27.7%, indicating that teacher title classification significantly impacts the prediction results. Notably, "Administrative Experience" ranks third, with an importance of 10.70%, suggesting that this characteristic also plays a role in shaping the model's predictions. Meanwhile, "Educational Background" has an importance of 8.2%, signifying that academic qualifications remain a contributing factor in classification results. Three other features—"Training Expert" (3.8%), "Recognized Teaching Case" (3.5%), and "Research Project Involvement" (3.2%)—exhibit a moderate influence on prediction results but are relatively less impactful. Interestingly, "Practical Experience" and "Collaboration Ability" both have an importance score of 0.00%, implying that these characteristics do not significantly contribute to the model's predictive capability, or their effects are overshadowed by other dominant features in the dataset.

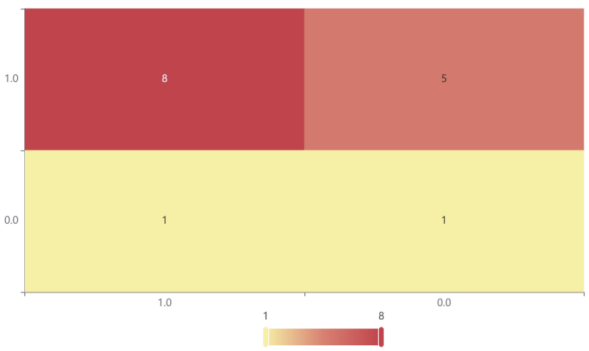

Figure 1: Confusion Matrix Heatmap

DOI: 10.54254/2753-7064/53/2025.21770

As shown in Figure 1, the data exhibits a specific distribution pattern. By analyzing the numerical values within each matrix cell, the frequency of observations under different category combinations can be determined. The top-left cell represents True Positives (TP), indicating that the model correctly predicted one sample as a positive case. The bottom-left cell represents False Positives (FP), showing that one negative sample was misclassified as positive. The top-right cell corresponds to False Negatives (FN), which has a value of zero, indicating that no positive samples were incorrectly classified as negative. The bottom-right cell (value: 8) and the adjacent cell (value: 5) collectively form the True Negatives (TN), reflecting that most negative samples were accurately predicted by the model.

Table 3: Model Evaluation Results

	Accuracy	Recall	Precision	F1 Score
Training Set	1	1	1	1
Test Set	0.6	0.6	0.793	0.664

The evaluation results indicate that the CatBoost classification model performs exceptionally well on the training dataset, achieving perfect scores of 1.000 in accuracy, recall, precision, and F1 score, implying that the model has minimal error in training data classification. However, the test set results exhibit a noticeable performance decline, highlighting the model's limitations in generalization ability. The test accuracy and recall are both 0.600, indicating that the model correctly identifies 60% of the positive class samples and retrieves 60% of actual positive cases. The precision score of 0.793 suggests that 79.3% of the samples predicted as positive were indeed correct, demonstrating relatively strong reliability in positive case classification, though some errors remain.

5. Conclusion and Educational Management Implications

This study, based on the professional experience of municipal-level teachers, introduces the CatBoost machine learning algorithm to construct a classification prediction model for outstanding teacher identity characteristics. The results indicate that the CatBoost algorithm has broad applicability in evaluating teacher professional development. The study identifies key components of outstanding teachers' professional competencies: A solid disciplinary foundation and strong teaching ability are essential to the formation of professional competence. A noble sense of conviction and professional ethics serve as the foundation for driving teacher professional development. Innovative awareness and digital competency are necessary prerequisites for teachers to evolve into research-oriented educators. Continuous learning and self-improvement are indispensable driving forces in teachers' professional growth. Based on these findings, this study provides the following insights and recommendations for teacher professional development:

(1) Enhancing Research Awareness:

Teachers should actively foster a research-oriented mindset and embrace the concept of "advancing education through research." Research should be regarded as a key tool for improving teaching quality and facilitating professional growth. Through engagement in research projects, academic writing, and scholarly seminars, teachers can continuously deepen their understanding of core educational concepts.

(2) Innovating Teaching Approaches:

Teachers should be encouraged to break free from traditional instructional models and actively explore innovative teaching methods and classroom management strategies. Approaches such as experiential learning, project-based learning, interdisciplinary thematic learning, and flipped

DOI: 10.54254/2753-7064/53/2025.21770

classrooms can significantly enhance student engagement and interest, ultimately strengthening their grasp and application of subject knowledge.

(3) Developing a Comprehensive and Multi-Level Teacher Training System:

Educational administrative departments should establish a comprehensive, multi-tiered teacher training system to support lifelong professional development. Regarding training content, the CatBoost machine learning algorithm can be utilized to analyze teachers' professional experiences, identifying personalized training needs based on different subject backgrounds and career development stages. By designing customized training curricula that align with teachers' actual needs, the quality and effectiveness of training programs can be improved, thereby contributing to the advancement of high-quality teacher development and the broader goal of building a strong education system.

References

[1] *Hui, H., Silvana, T. (2024) Analysis of cloud computing-based education platforms using unsupervised random forest. Education and Information Technologies,12,15905-15932.*

[2] *Xia, J., Zhang, S. (2023) Enhancing family education pattern recognition with a random forest algorithm. Journal of Intelligent & Fuzzy Systems,6,9803-9813.*

[3] *Imani, R., Raven, S., Dalia, S., Jamala, W., Hannah, J., René, R., Tiffany, Bergin. (2025) Integration of individuals with lived experience to improve recruitment within criminal justice research: 'experience as the best teacher'. International Journal of Social Research Methodology,1,1-14.*

[4] *Ignacio Jr. A. G., (2024) EXPLORING THE PERSPECTIVES OF PRESERVICE FILIPINO MATHEMATICS TEACHERS: BASIS FOR A PROPOSED EDUCATIONAL BELIEF MODEL. Problems of Education in the 21st Century,4,487-506.*

[5] *Sarkar, C., Mohanty, V., Balappanavar, A., Chahar, P., Rijhwani, K.(2022).Development and Validation of a Comic Tool: An Innovative Approach to Raise Awareness about Tobacco Control among School Teachers. Indian Journal of Community Medicine,4,536-542.*

[6] *Md, E.C. A.K.M., Saiful I.,Rashed,U.Z., Sharfaraj, K.(2025). A machine learning-based approach for flash flood susceptibility mapping considering rainfall extremes in the northeast region of Bangladesh. Advances in Space Research,2,1990-2017.*