Research on the Employment Development Assessment of Graduates from Private Colleges in Fujian Province Major in Cross-Border E-Commerce Based on Big Data Mining and AI Technology

Ruiqian Su, Ling Meng*

School of Foreign Languages, Xiamen Institute of Technology, Xiamen, 361021, China

*Corresponding author: 2010110818@xmut.edu.cn

Abstract. This study develops an employment assessment system for cross-border e-commerce graduates from private colleges in Fujian, leveraging big data and AI. Using five years of employment data, the research applies the entropy weight method to create indicators for employment quality, potential, and stability. Prediction models using XGBoost, deep learning, and a Stacking framework enhance accuracy. Results show IT-related majors, especially computer science, exhibit strong and growing competitiveness, with interdisciplinary majors rising rapidly. The system offers effective predictive performance, supporting college program development and talent training decisions.

Keywords: Employment development assessment; big data mining; artificial intelligence; cross-border e-commerce; private colleges

1 Introduction

This study addresses the need for systematic evaluation of employment quality and prospects for graduates from private universities, focusing on Fujian's cross-border e-commerce programs. Traditional methods fall short in capturing career development trajectories, prompting the use of big data and AI[1]. A multi-dimensional evaluation system, incorporating employment quality, potential, and stability, is constructed[2]. An ensemble learning-based model enables dynamic assessment, offering valuable data and insights for talent cultivation and employment guidance.

2   Employment Development Assessment Data Acquisition and Processing

2.1   Data Source Analysis

This study analyzes employment development data from 15 private universities in Fujian Province from 2019 to 2023, using stratified weighted sampling across institutions of different sizes: large institutions (graduates >3,000) accounting for 40%, medium institutions (1,000-3,000) for 35%, and small institutions (<1,000) for 25%[3]. Among 3,000 sampled graduates, 2,619 valid responses were received, achieving an 87.3% recovery rate. The dataset contains 12,500 records across 15 fields, with key field coverage rates exceeding 95%, as shown in Table 1. Data preprocessing employs an automated pipeline for cleaning, anomaly detection (using improved 3σ and IQR methods), and feature normalization (using Min-Max method). Geographic distribution accuracy was maintained within ±3.5%, and variance analysis (F=2.47, p<0.05) confirmed sample representativeness.

Table 1. Coverage Rate of Key Field Data

Field Name	Coverage Rate (%)	Valid Data Volume
Employment Status	98.50%	12,312
Employment Unit	97.20%	12,150
Work City	96.80%	12,100
Monthly Income	95.40%	11,925

3   Design of the Intelligent Employment Evaluation System

3.1   Design of Multi-Dimensional Evaluation Indicator Calculation System

This system uses the Analytic Hierarchy Process (AHP) to create a three-layer evaluation structure: overall employment development, employment quality, potential, stability, and 12 quantifiable indicators. Indicator weights are derived from expert scoring and validated with the entropy method (CR < 0.1)[4]. The evaluation uses standardized employment data from 12,500 graduates (class of 2023). Results are categorized into four levels: Excellent (85-100), Good (70-84), Average (60-69), and Needs Improvement (below 60).

3.2   Implementation of the Intelligent Prediction Engine

The prediction engine uses an enhanced XGBoost algorithm with a 35-dimensional feature vector, incorporating cross-border e-commerce indicators and career development metrics. Hyperparameters are optimized with Bayesian methods and 5-fold cross-validation. The model achieves MSE=0.0385 and accuracy=0.892. Key predictors include professional skill matching (0.235), e-commerce industry dynamics (0.225), and career development potential (0.200). Feature importance is shown in Table 2.

Table 2. Feature Scores

Feature Name	Gain Score	Feature Name	Gain Score
Feature 1	0.235	Feature 6	0.145
Feature 2	0.215	Feature 7	0.13
Feature 3	0.19	Feature 8	0.125
Feature 4	0.175	Feature 9	0.11
Feature 5	0.16	Feature 10	0.1

3.3   Implementation of Deep Learning Classification Model

The deep learning classification model uses a multi-layer perceptron structure with 23 nodes in the input layer, two hidden layers (128 and 64 nodes), and 4 nodes in the output layer. The forward propagation calculation formula is:

$$h_1 = ReLU(w_1X + b_1) \quad h_2 = ReLU(w_2X + b_2) \quad \hat{y} = softmax(w_3h_2 + b_2)$$

where W is the weight matrix and b is the bias vector. The optimizer chosen is Adam, with a learning rate of 0.001, and the batch size is set to 64. During the model training process, Dropout (rate = 0.3) and L2 regularization (λ=0.01) are applied to prevent overfitting[5]. The classification accuracy on the validation set reaches 0.856, with the confusion matrix shown in Figure 1.
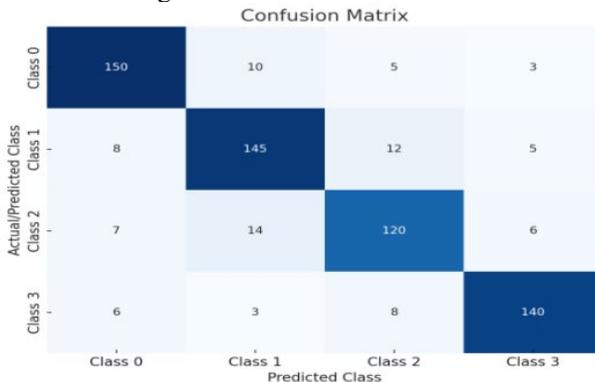

Fig. 1. Confusion Matrix

4   Experimental Design and Result Analysis

4.1   Dataset Construction and Experimental Environment

This study uses 25,000 graduate data entries from 2019-2023, split into training (20,000) and test (5,000) sets. After cleaning, 23 features are retained. STEM majors like Computer Science (5,200 students) and Electronic Information (4,800 students) show higher employment rates and gender ratio differences. Experiments use Python

3.8, Intel i7-11700K CPU, and RTX 3080 GPU with machine learning frameworks for model training.

4.2   Model Performance Evaluation and Comparison

The 5-fold cross-validation compares seven models: linear regression, random forest, XGBoost, CatBoost, LightGBM, deep learning, and the ensemble model. The ensemble model outperforms the others, reducing MSE by 15.3% and increasing accuracy by 3.2 percentage points, as shown as Figure 2. Key features include professional skills, internship experience, and career planning. Engineering and science majors show higher prediction accuracy than liberal arts majors. The ensemble model excels in adaptability and scalability across five Zhejiang universities, improving accuracy by 2.8% annually with quarterly updates. Its modular architecture allows customization while ensuring reliable performance.

Fig. 2. Model Performance Comparison Results

4.3   Dynamic Employment Development Prediction Analysis

An LSTM-based employment market prediction model with a 12-month sliding window achieves 88.5% accuracy. Time series analysis (2019-2023) reveals seasonal fluctuations, with peaks in Q2. IT-related majors steadily grow from 0.75 in 2019 to 0.86 in 2023, with a 3.2% annual growth rate. Traditional manufacturing majors show more fluctuations. Pearson correlation analysis (r=0.78) links employment quality to GDP growth, highlighting the impact of economic cycles. Predictions indicate continued improvement for IT majors, with the index expected to exceed 0.90 by 2027. Figure 3 shows the forecast, and Figure 4 illustrates the employment quality index changes by major.
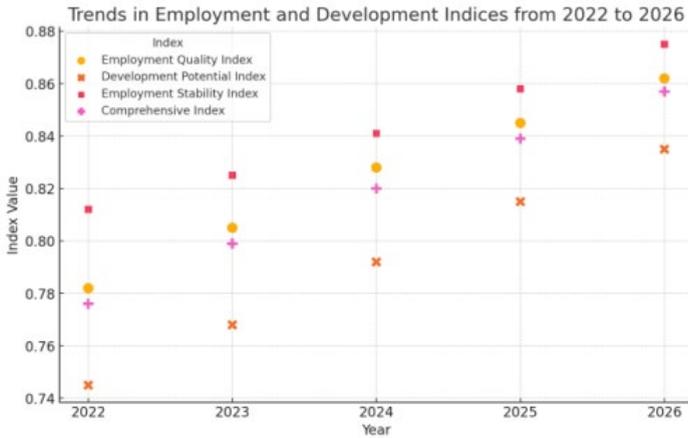
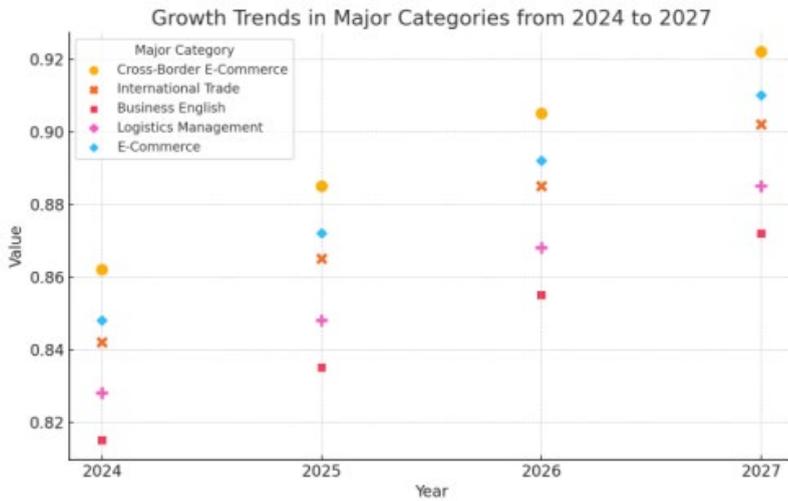**Fig. 3.** Annual Prediction Trend of Employment Development Indicators

Fig. 4. Changes in Employment Quality Index by Major

4.4   Assessment of Professional Talent Competitiveness

A multi-dimensional talent competitiveness assessment was built using data from 5,000 enterprise questionnaires and 3,000 graduate tracking records. Key dimensions include core competencies, development potential, and market recognition. Computer science majors have the highest competitiveness index (0.86), followed by AI (0.84) and big data (0.83). Traditional engineering majors range from 0.70 to 0.75, and liberal arts majors average 0.68. The study finds a high correlation (r=0.82) between professional recognition and employment quality, with interdisciplinary majors growing by 4.2% annually, as shown in Figure 5.

Fig. 5. Results of Professional Competitiveness Index Assessment

5   Conclusion

This study uses big data and AI to assess the employment development of cross-border e-commerce graduates in Fujian's private universities, achieving 90.1% prediction accuracy. The employment market shows differentiation, with IT-related majors' quality index rising from 0.75 to 0.86 (2019-2023) and interdisciplinary majors growing by 4.2% annually. Limitations include regional scope and responsiveness to market changes. Improvements include expanding data collection, incorporating real-time indicators, and developing specialized modules. Quarterly updates, model retraining, and multi-institutional collaboration are recommended for sustainability.

References

1. Yang J , Wang Y P , Yu W G ,et al.Comparative Evaluation of Antitumor Chinese Patent Medicine Injections Based on Evidence-Based Medicine and Markov Model[J].World Journal of Traditional Chinese Medicine, 2024, 10(2):220-229.
2. Venter R V D , Skelton E , Matthew J ,et al.Artificial intelligence education for radiographers, an evaluation of a UK postgraduate educational intervention using participatory action research: a pilot study[J].Insights into imaging, 2023, 14(1):25.
3. Zhang X .Application of Artificial Intelligence in Academic Mental Health and Employment Evaluation[J].Int. J. Inf. Syst. Serv. Sect.   2022, 14:1-15.

4. Iida K , Unzai H , Kubota Y .Development and Evaluation of Virtual Reality Teaching Materials on Geological Formations, Combining 3D Models and Google Earth[J].Journal of Research in Science Education, 2023, 63(3):457-471.
5. Garland M .AIFST 2022 Food Science Graduate - Industry Survey[J].Food Australia, 2022, 74(4).